Universität
Augsburg
University

Essays in physician scheduling with a special focus on breaks and flexibility

Kumulative Dissertation
der Wirtschaftswissenschaftlichen Fakultät
der Universität Augsburg
zur Erlangung des Grades eines
Doktors der Wirtschaftswissenschaften
(Dr. rer. pol.)

Vorgelegt von

Melanie Johanna Patricia Erhard
(Dipl.-Kffr.)

Erstgutachter:	Prof. Dr. Jens O. Brunner
Zweitgutachter:	Prof. Dr. Axel Tuma
Drittgutachter:	Prof. Dr. Robert Klein
Vorsitzender der mündlichen Prüfung:	Prof. Dr. Michael Krapp
Datum der mündlichen Prüfung:	05.04.2019

List of contributions

This dissertation is comprised of the following contributions submitted to or published in scientific journals. The specified categories relate to the journal ranking JOURQUAL 3 of the "Verband der Hochschullehrer für Betriebswirtschaft e.V. (VHB)".

The order of the contributions corresponds to the order of print in this dissertation:

Contribution 1

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

The final authenticated version is available online at https://doi.org/10.1016/j.ejor.2017.06.037

Status: Published in European Journal of Operational Research, category A.

Contribution 2

Erhard, M., Brunner, J.O. (2018). Physician staffing levels and absence planning under special consideration of breaks – A case study on anesthetists.

Status: Submitted to OMEGA - The international Journal of Management Science, category B.

Date of submission: August 07, 2018.

Contribution 3

Erhard, M. (2019). Flexible staffing of physicians with column generation. Flexible Services and Manufacturing Journal, in press.

The final authenticated version is available online at https://link.springer.com/article/10.1007%2Fs10696-019-09353-8

Status: Accepted for publication and published online in Flexible Services and Manufacturing Journal, category B.

Acknowledgement

First, special gratitude goes to my academic advisor Prof. Dr. Jens O. Brunner: I'm very thankful for all the support, the inspiring and reassuring discussions, and the trust you showed in me during my PhD studies. Moreover, I'm obliged for all colleagues at our chair for providing me a great time at the UNIKA-T.

I am very grateful for the support of my family and last but not least, I want to thank my wonderful husband Daniel for bringing joy and happiness in my life. Thank you for the support and all the confidence you show in me every day. And of course, beyond that, that (sometimes) special kind of understanding.

Thank you all very much.

Table of content

1. Introduction and motivation

Personnel planning and scheduling is a topic of interest since the early 20th century. As physicians constitute a subgroup of personnel having specific characteristics, the personnel planning problem is discussed first in general before bridging to the physician scheduling problem. Since employees are a crucial resource for business success, the necessity to improve efficiency in production processes to increase productivity and outcomes, e.g. by division of work and responsibilities (Scherm and Süß 2016), rises. Personnel planning mainly decomposes into three parts: Staffing decisions on a strategic level are to be made to ensure an appropriate number of employees to cover demand and provide the desired quality of service. Secondly, rosters/schedules have to be generated efficiently on an operational level to prevent possible negative consequences such as unnecessary overtime hours for personnel (Wabro et al. 2010). Thirdly, re-planning problems reschedule personnel in short term due to unforeseen absences, e.g. in case of illness. Since this dissertation focuses mainly on staffing problems combined with an operational offline evaluation, re-planning problems are not discussed in the following.

Staffing. On a strategic level, the management of an institution has to determine the required size of the workforce to be able to handle operations as expected in terms of quality and quantity. Staffing decisions therefore are comprised of two separate decisions: The determination of qualitative manpower requirements by defining a profile and the appropriate level of qualification, education, and experience for a specific employment position as well as the quantitative investigation of staff size requirements. The latter decision needs to respect quantitative and time-dependent staffing requirements (Wabro et al. 2010). In general, staffing decisions are governed by a predefined immutable framework given by organization-internal and organization-external factors (Scherm and Süß 2016). For a detailed overview, see Figure 1.

Fig. 1: Framework for staffing decisions (based on Scherm and Süß 2016, p. 8)

Organization-external factors such as governmental regulations are given by law and are therefore to be considered mandatorily. Moreover, there are various other aspects, depending on employees' individual preferences, which might influence management's decision making, such as the importance of work-life balance or the desired employment relationship. Internal factors are requirements of the institution itself affecting staffing decisions: Several facets such as organizational structure and culture as well as the production program and service range generate specific staff requirements, which are to be met to cover demand (Spengler 1999).

Rostering. When scheduling personnel, it is of major importance to investigate some key aspects to create a roster of high quality which satisfies the expectations of the workforce. Ozcan (2017) defines five different factors influencing the scheduling process: Costs, coverage of demand, roster quality, flexibility, and robustness. Especially due to the increasing relevance of generating an annual surplus and being profitable in the face of rising operating costs, the reduction of upcoming costs is a main concern for institutions (Ozcan 2017). A leverage point to decrease these costs can be seen in a better utilization of scarce resources such as staff (Bölt 2014) or expensive equipment with limited capacity. Another very important aspect in scheduling is the coverage of demand (Ozcan 2017): Even though meeting demand is a central aspect for every type of industry, it is especially important for the service sector. There are certain service-specific characteristics which vary significantly from industry and manufacturing such as the impossibility of stockpiling and the volatile provision and consumption of services (Aggarwal 1982). This makes coverage of demand in every period of each day essential. Another central issue is the quality of the roster (Ozcan 2017) which is determined by the degree of matching supply and demand and the level of accounting for fairness criteria, e.g. an equal distribution of unfavorable shifts across the workforce and individual features and preferences of staff (Gross et al. 2018). A generated schedule of a high quality leads to possible positive consequences such as an increasing quality of the delivered work and services as well as a high job satisfaction and motivation for personnel (Aiken et al. 2002). The remaining two aspects, flexibility and robustness, are positively correlated with each other. Ozcan (2017) defines flexibility in this context as the level of variability of the generated schedule, i.e. the ability to adopt changes and modifications of the initial plan derived by unforeseen events as well as short-term rescheduling due to illness. As part of this, robustness describes the reliability of the created roster (Ozcan 2017).

This illustrates the general aspects and difficulties that arise when making staffing and scheduling decisions in an institution. In this dissertation, different hierarchical planning aspects are partially combined. As will be seen in the remainder, a more detailed point of view and the tool of mathematical programming can provide more realistic and precise opportunities to improve and automate the process of scheduling personnel.

1.1. Mathematical programs and key aspects in personnel scheduling

In this subsection, mathematical modeling as a mechanism to support decision making units with staffing and scheduling personnel is presented. First, different modules of personnel planning discussed in literature are presented and second, one of the first mathematical models for personnel scheduling problems, i.e. the set covering approach based on Dantzig (1954), is introduced since this model provides the basis for the modeling approaches in this thesis.

1.1.1. Modules of personnel scheduling

In general, the required workforce size can either be approximated by a specific measure or a precise mathematical model. The former approach, approximating the number of personnel by an indicator, is not very time consuming and does not involve a multitude of key figures and information. There are several approximation mechanisms: For example, using the estimated peak in demand to determine the required staffing level is a common approach. Among others, another approximation mechanism is to determine the workforce size by the ratio of demand hours to the available working hours for each servant (Jarr 1973). However, since only few influencing factors are taken into account, these staffing decisions might be of rather low quality due to a significant underestimation of the actually required workforce size. In contrast, implementing a mathematical model may lead to a more realistic estimation of the required number of personnel since staffing and working environment can be represented: Mathematical modeling approaches can account for a higher number of operational constraints and regulations (Brunner et al. 2010) which lead to a more realistic and valid estimation of the number of required personnel. Therefore, Ernst et al. (2004) separate the staffing and rostering process into diverse modules to evolve a guideline for selecting which aspects are to be implemented in a mathematical model, depending on the purpose of the formulation.

Module 1	Module 2	Module 3	Module 4	Module 5	Module 6
Demand modeling	Days off scheduling	Shift scheduling	Line of work construction	Task assignment	Staff assignment

Tab. 1: Rostering taxonomy (based on Ernst et al. 2004, p. 5-6)

The taxonomy of Ernst et al. (2004) defines six different modules, which are presented in Table 1: Demand modeling, days off scheduling, shift scheduling, line of work construction, and staff assignment. Not each module is required in every mathematical model. Depending on the aim of the decision maker, the stated objective function for the mathematical formulation as well as the type and quantity of the considered modules vary significantly (Ernst et al. 2004).

Module 1, demand modeling, determines the quantity of required personnel for each time period within the planning horizon. Demand can either be deterministic, i.e. shift or task based, or stochastic, i.e. flexible by implementing probabilities for the occurrence of specific demand levels. An appropriate number of days off between two consecutive sequences on duty are handled in Module 2. Module 3 is the scheduling of shifts: This module determines both, the types of shifts that are available and the number of personnel that has to be assigned to each of the formerly selected shift types. Entire lines of work are generated in Module 4. In this module, individual lines of work, i.e. working schedules with an appropriate assignment of shifts, are constructed for each individual employee. It is therefore necessary to consider several different aspects and regulations, e.g. a minimum number of rest periods between two consecutive shifts, a maximum number of working hours within the planning horizon, and the structure of demand. Module 5 assigns one or more tasks to shifts while accounting for potential additional challenges determined by the type of duty and the thereby required level of education or experience. The last module, staff assignment, assigns personnel to specific lines of work constructed in Module 4 (Ernst et al. 2004). Again, personnel scheduling problems are different between business sectors and vary with respect to the considered modules and the purpose of the objective function.

1.1.2. The set covering approach as crucial mathematical modeling formulation

The personnel scheduling problem is highly constrained due to the diversity of different governmental rules, regulations, agreements, and environmental as well as workforce-specific aspects that are to be taken into account. Since mathematical modeling provides a tool to account for a high number of constraints simultaneously, decision makers and schedulers can benefit from the hereby provided support (Ernst et al. 2004). Even though there are several approaches (e.g. Blöchlinger (2004), Bechtold et al. (1991), and Aykin (2000)), Dantzig (1954) is one of the first authors who state a mathematical formulation for a generic personnel scheduling problem (based on Edie (1954)). This rather small mathematical model often builds the basic approach which is to be extended depending on specific regulations determined by a precise Rostering problem. Dantzig (1954) formulates the set covering problem as the following mixed-integer programming model (MIP):

Sets with indices

$j \in \boldsymbol{J}$	Set of schedules with index j
$t \in \boldsymbol{T}$	Set of periods with index t

Parameters

a_{tj}	1 if schedule j covers period t, 0 otherwise
b_t	Demand in period t

Integer decision variables

x_j	Number of employees being assigned to schedule j

$$\text{Minimize} \sum_{j \in \boldsymbol{J}} x_j \tag{1}$$

subject to

$$\sum_{j \in \boldsymbol{J}} a_{tj} x_j \geq b_t \qquad \forall\, t \in \boldsymbol{T} \tag{2}$$

$$x_j \geq 0 \text{ and integer} \qquad \forall\, j \in \boldsymbol{J} \tag{3}$$

As Edie (1954) determines the number of required toll booths over the year without specifying the number of required employees that are to be scheduled to run the booths, Dantzig (1954) extends this approach by proposing a set covering approach to assign personnel. The objective function therefore minimizes the number of employees which are assigned to a working sequence.

Constraints (2) ensure demand for personnel to be covered in each time interval by forcing the total number of assigned personnel to be larger or equal to the demand b_t in each period t. To ensure the assignment of feasible working patterns, the parameter a_{tj} contains various information concerning allowed working hours and sequences. The variables are defined in (3).

By applying this approach, a minimum number of personnel, starting their working sequence in different periods of the planning horizon, can be determined to cover demand in each time interval. If there are several issues to be considered, the presented modeling approach can easily be extended by implementing weights and different goals in the objective function, i.e. making use of a weighted set covering formulation (Balinski 1965). A variation of the set covering approach by Dantzig (1954) is the set partitioning model: Pursuing a similar objective function, the only difference is the formulation of the demand constraints. In a set partitioning approach, it is not allowed to assign more employees than required in any period. Instead, this approach forces demand to be met accurately in each time interval which commonly leads to a highly restricted solution space (Garfinkel and Nemhauser 1969).

The mathematical models in this thesis make use of the presented set covering approach and state modifications and extensions to apply a more precise mathematical model to tackle the physician scheduling problem.

1.2. Importance of physician scheduling

In Germany, there are three different types of hospital administrations: Private agency, public agency, and non-profit agency (Bölt and Graf 2012). As private hospitals are privately owned and non-profit hospitals are funded by associations and foundations, solely public hospitals are owned by federal government and related institutions. As a result, only public hospitals are obliged by law to provide a full spectrum of health care services to protect the population against sickness and disease whereas non-profit and privately held hospitals commonly specialize their treatments, e.g. as an orthopedic clinic (Fournier and Mitchell 1992). It is therefore not completely possible for public hospitals to select patients based on their diagnosis related group (DRG) and the associated expected revenue to generate an annual surplus. Instead, these institutions are commonly maximum care providers that ensure the overarching treatment of every disease.

However, since around one third of the hospitals in Germany generate an annual loss each year (Blum et al. 2017), the number of hospitals is continuously decreasing while the number of patients grows each year.

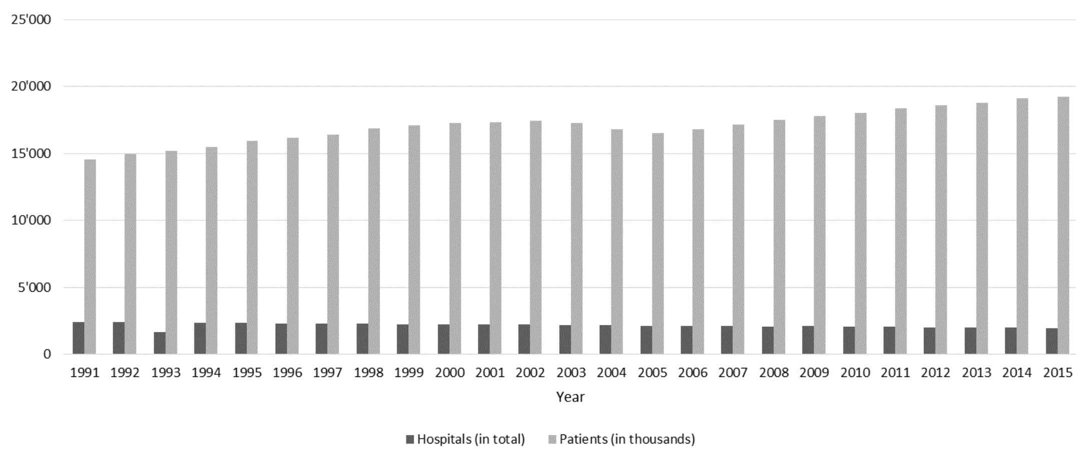

Fig. 2: Trends in number of hospitals and number of patients

As Figure 2 shows, the number of hospitals has been decreasing from 2’411 hospitals in total (Statistisches Bundesamt 2016) (996 public institutions) in 1991 (Bölt and Graf 2012) to 1’951 hospitals in total (570 public institutions) in 2015, indicating a decrease of around 10% in total (43% public hospitals) within 25 years. In contrast, the number of patients has been increasing from 14’576’000 patients in 1991 to 19’239’000 patients in 2015 (Statistisches Bundesamt 2016). As a result, hospitals are forced to treat more patients with almost similar (or less) available resources. Since supply and demand occur at the same point in time and demand cannot be backlogged due to the negative consequences on the health of patients, e.g. severe secondary failures or death, it is of major importance to cover demand in each period of every day within the year

(Aggarwal 1982). Due to the capacity limitation of resources, an efficient handling is essential: The employees, especially physicians, are considered a bottleneck resource in the care providing process, which makes it mandatory to schedule them carefully (Santos and Eriksson 2014). Planning and scheduling physicians in an efficient way leads to several benefits: Unnecessary overtime hours for physicians might decrease due to a better match of demand and supply, preventing a growth in resulting staffing costs (Villarreal and Keskinocak 2016), increasing the possibility to maintain a high quality of care while working conditions of staff are not deteriorated (for example by increasing working hours and personnel utilization each week), and maintaining job satisfaction for physicians to decrease turnover rates and associated costs (Aiken et al. 2002).

The resulting problem is referred to as the physician scheduling problem. The objective is to determine an optimal number of physicians who are subsequently scheduled efficiently subject to coverage of demand and legal working regulations. Moreover, various soft constraints can additionally (not mandatorily) be taken into account, e.g. individual preferences and fairness aspects. Analyzing the problem of physician scheduling drives the motivation for the presented essays. These are summarized and discussed in the following sections with the main aim to answer the following research questions:

Research question 1: Which characteristics and aspects of physician scheduling in hospitals are already covered in literature and what are promising areas for future research?

Research question 2: How does flexibility in the assignment of shifts and breaks affect the overall number of required physicians? How can we build a more realistic estimation for approximating the total size of the workforce?

Research question 3: What is the value of implementing flexibility in the process of physician scheduling?

1.3. Content of the thesis

The remainder of this thesis is structured as follows: Section 2 summarizes the three contributions with the full version of the appropriate papers attached in the appendix. In section 3, the contributions of this dissertation are discussed. The discussion encompasses the summary of the findings, resulting insights for hospital's management and various ideas for future research. Finally, section 4 concludes the dissertation.

2. Summary of the contributions

In this section, each of the contributions to the body of research is discussed. The individual contributions are attached in the appendix.

2.1. State of the art in physician scheduling

In practice, hospitals commonly face a great fluctuation in demand, especially for physicians. The requirement for care can vary from one hour to another and from one day to the next, which might lead to negative consequences for the scheduled staff size due to the inability to match supply and demand efficiently. As a result, overstaffing in periods of low demand or understaffing for high demand level hours can occur. Both leads to negative effects on patient care and physician's job satisfaction and utilization level. Scheduling physicians efficiently can therefore reduce upcoming personnel costs and prevent over- and understaffing. Due to the importance of the problem under consideration, there is a large body of research studying the physician scheduling problem. Erhard et al. (2018) review current literature in this field and discuss several ideas for future research. This paper provides the first literature overview on the physician scheduling problem: A framework is defined to classify literature according to topic-related features: The hierarchical level, modeling approach (problem-specific characteristics, uncertainty, and mathematical methodologies), and real life implementation. Subsequently, the transfer of research into practice is discussed and an agenda for further research is established.

In general, the physician scheduling problem is a subcategory of the field of personnel scheduling containing several problem-specific facets that differ from other areas, e.g. individual agreements between hospital and physician (Charles et al. 2013) or an increasing bargaining power of the workforce due to their central importance (Santos and Erikson 2014). The literature review discusses 60 papers, published between 1985 and 2015, which proves the relevance of the physician scheduling problem and the increasing attention in research within the last decades. Especially in European countries, research has been increasing since 2004 due to the introduction of the DRG compensation system.

The heart of the literature review is a classification of the physician scheduling problem according to the former mentioned framework. In general, there are three different hierarchical levels: Staffing, Rostering, and Re-planning problems. Staffing problems consider strategic or long-term decisions concerning the appropriate size of staff and educational aspects in residency programs, whereas Rostering problems focus on a midterm tactical or operational offline level. The latter constitute the majority of literature with the aim of creating a repeatable (cyclic) roster or detailed lines of work for personnel covering a planning horizon of several weeks. Moreover, Re-planning problems have only been considered in one publication up to now and focus on operational online rescheduling mechanisms in case of unforeseen absences.

Physicians can be classified based on two different criteria: Educational level and working hours. First, the educational level differentiates between residents and fully educated physicians as residents can perform some tasks of physicians but are still in education, leading to a restricted task set. Second, working hours define staff as full or part timers, independent of the level of education. Residents are considered in several publications, whereas part timers are mainly neglected.

To schedule physicians, there are two different types of shifts: Predefined and flexible shifts. Flexible shifts have various starting and ending periods and diverse lengths in contrast to predefined shifts. Even though flexible shifts are superior when it comes to matching supply and demand due to the uncertainty in hospital surroundings, the majority of literature solely considers predefined shifts, e.g. three eight hour shifts. Moreover, breaks are not explicitly assigned in current literature and mainly neglected.

Concerning employee-related aspects, fairness in terms of evenly distributed working hours and granting of individual preferences are considered in literature. In total, around one third of the papers considers either one of these or all staff-related features. However, this number has increased significantly within the last years due to the increasing awareness of the importance of personnel satisfaction.

Objectives in physician scheduling literature can either be financial or non-financial. Financial objectives focus on direct costs measured in monetary value, whereas non-financial goals consider patient- or employee-related aspects. In-line with personnel aspects, the number of publications implementing non-financial objectives increased significantly during past years.

The requirement for care is commonly uncertain in hospitals and governed by stochasticity. In literature, there are two ways of assuming demand: Deterministic or stochastic. Even though stochastic demand is more realistic, current literature mostly assumes demand to be deterministic to reduce complexity of the mathematical problem formulation and improve solvability. Concerning mathematical modeling and solution methodologies in use for the physician scheduling problem, the majority of research applies mathematical programming techniques, e.g. linear programming (LP), integer programming (IP), and mixed-integer programming (MIP). Other methods such as non-linear programming (NLP) and queuing models (QM) are barely in use. After modeling, the considered problem can either be solved exactly or by using a heuristic solution procedure. Despite the advantage of receiving a near optimal solution within short computation times by heuristic approaches, more than two thirds of the papers solve their mathematical model exactly. This leads to an optimal solution or at least an upper or lower bound for the objective function value as well as the optimality gap as indicator for the solution quality.

With respect to applicability, almost all papers use real life data in their experimental study to evaluate their model for real world demand. Almost one third of the approaches is implemented in a hospital department at least in a test phase. Most of these papers were published after the year 2000 and derive from Northern America which indicates a geographically higher willingness to investigate new methods in scheduling processes.

Summarizing, research interest in the physician scheduling problem has been steadily increasing for the last 30 years. Depending on the hierarchical level of the considered problem and the purpose of research, there are various aspects that are to be considered which affect the applied modeling approach and corresponding solution technique.

2.2. Physician staffing levels and absence planning under special consideration of breaks – A case study on anesthetists

The trend of constantly increasing annual costs of hospitals has been occurring for several decades and will continue into the future. As the workforce of a hospital, especially physicians, are known as a main cost driver in hospital surroundings, it is necessary to investigate the physician scheduling problem in detail. Up to now, staffing decisions are commonly based on simple approximation methods that significantly underestimate the required number of physicians due to neglecting various crucial planning aspects, e.g. the assignment of breaks or rest periods, even though these are mandatory by law. It is therefore the main goal of this paper to determine a realistic lower bound for the required workforce size in short computation time subject to coverage of the forecasted demand.

The contribution is a mixed-integer programming model that ensures a maximum amount of flexibility in terms of shift types and breaks. Shifts are allowed to have a multitude of different starting and ending periods and various lengths, whereas breaks are allowed to be assigned flexibly for each shift within a predetermined range of periods. In current literature, the assignment of a break is mostly neglected even though it is a central aspect in real life since physicians are commonly not able to take their break while being on duty. Here, we provide the first extension of the mathematical modeling approach of Bechtold and Jacobs (1990) for the assignment of breaks to periods in a way that accounts for overlapping shifts. We show the importance of accounting for breaks when scheduling or staffing physicians. Moreover, we consider additional key elements when determining the appropriate staffing level such as working regulations given by law and rest periods. Lastly, we improve absence planning by determining a weekly lower bound for the required staffing level over one entire year. By doing so, weekly variation and seasonality in the workforce size can be detected and used for vacation planning and scheduling of foreseen absences, such as conferences and further educational programs/workshops.

In our problem formulation, we estimate the required workforce size as realistic as possible. To this end, we use an aggregated formulation which determines a minimum number of physicians required, selects a number of shift types, and assigns a break. In our model, we assign shifts and breaks flexibly and ensure the coverage of demand in every period of every day and the adherence to legal and working time regulations. Moreover, since this model serves as an estimator for the workforce size, we propose a second mathematical model which uses the output of the first model as an input and creates individual lines of work, which can be assigned to the physicians.

The validation of the model and the examination of its performance in the experimental study uses two randomly generated demand scenarios as well as demand derived from a hospital. For both random profiles, 50 instances were generated and solved twice, with and without accounting for breaks, to evaluate the effect of considering breaks when determining physician staffing levels. As a result, solution times are rather low for each test instance. Moreover, neglecting breaks leads to a significant underestimation of the required workforce size, independent of the demand profile. This is the case for more than 90% of the test sets, which shows understaffing of one to five physicians when not taking breaks into account. This shows that neglecting breaks leads to a realistic estimation of the workforce size in less than 10% of the considered test settings. Moreover, this effect becomes even more severe the shorter the length of a planning day, i.e. when demand could be covered by one single shift type (without accounting for a break) and an additional physician is only required when breaks are considered.

In the second step of the experimental study, real life data from one entire year is in use. Demand is decomposed into 52 one-week planning periods and solved separately for each week. First, flexible shifts are allowed and second, to simulate shift settings found in real life, only three eight hour shifts are available. Again, each test instance is solved twice: Neglecting breaks and accounting for breaks. Even though the resulting optimal workforce size varies from one week to another, neglecting breaks leads to a significant underestimation of the required number of physicians. Especially when reducing the number of available shift types, the needed workforce size increases significantly if breaks are considered, i.e. neglecting breaks underestimates the required size of personnel by six physicians.

Summarizing, a reduced set covering approach is presented, which provides two types of flexibility when determining a lower bound for the staffing level of physicians. Flexibility in terms of shift types and the assignment of breaks enable a better match of supply and demand and ensure a rest period for each physician on duty. As our results indicate, our approach provides a superior method to determine a realistic lower bound for the workforce size compared to other approximation mechanisms in use, e.g. maximum demand and the ratio of total demand and available working hours of personnel, due to the consideration of breaks and several key aspects such as rest periods and total working hours. This way of estimating a lower bound for the number of required personnel therefore leads to the ability to cover demand in each period while creating planned idle time to handle stochasticity in demand and emergency patients.

2.3. Flexible staffing of physicians with column generation

Hospitals, especially physicians, face a great fluctuation in demand every hour of each day with high peaks (commonly) around noon. To be able to cover expected peaks in demand, hospitals schedule a large number of physicians. However, as this maximum level in demand declines after several hours, hospitals are overstaffed afterwards. As current practice commonly uses a small number of predefined shift types, e.g. three eight hour shifts, it is not possible to match supply and demand adequately. Implementing an increasing amount of flexibility in the scheduling process leads to the ability to better match supply and demand due to the large number of starting periods of shifts and various different shift lengths.

It is therefore the major contribution of this research to test and evaluate the effect and value of flexibility in the scheduling process of physicians. In this research, a MIP model is developed which provides a maximum level of flexibility in the scheduling process with respect to the sequences of working days, starting and ending times of shifts as well as the allowed shift lengths, and the placement of the break. The objective is to generate cost-minimal schedules covering the entire planning horizon for each physician, i.e. the total salary costs are minimized in our objective function, subject to coverage of demand and several legal and labor regulations. Due to the complexity of the problem, a column generation (CG) approach is applied as solution approach, which provides at least a good lower bound to estimate the total labor costs, if it is not possible to solve the considered problem to optimality.

In the solution process, the compact formulation decomposes by physician (based on Dantzig-Wolfe 1960) to build an individual line of work for each employee that covers the whole planning period. The resulting optimization problems, restricted Master Problem (MP) and Subproblem (SP), are solved iteratively to optimality. The column with the smallest reduced costs generated by the SP is added to the linear restricted MP which is solved again. This loop continues until no column that prices out can be found. We additionally implement a second criterion for the termination of the algorithm making use of the objective function value of the linear restricted MP and the SP of the current iteration: A lower bound for the linear restricted MP is calculated in each iteration of the solution process and terminates the algorithm if the current value of the linear restricted MP is smaller than the actual value of the calculated bound (rounded off). Eventually, the restricted MP is subsequently solved as an IP, which might lead to a solution which is not optimal any longer.

To evaluate the performance of the proposed solution approach as well as to analyze the effect of flexibility in the scheduling process, an experimental study is conducted. In the first step, we account for robustness and evaluate the performance of the CG approach by testing four different lengths for the planning horizon: One, two, four, and six weeks. Additionally, we consider three different scenarios for the level of demand: Mean demand level, 75% quantile demand, and maximum (100%) demand level (based on aggregated data derived from real life). Solving the considered instances leads to an optimal solution only for a one week problem. The remaining instances cannot be solved to optimality by the compact formulation. The results from our experimental study indicate that our CG heuristic provides a good solution in appropriate computation time for each test set, which is evidence of the high quality of our solution approach. In the second step of our experimental study, a factorial analysis is conducted to evaluate the value of flexibility in the scheduling process. Day, shift, and break parameters are varied, which results in a total of 168 test sets. Each parameter setting can be solved using the CG heuristic within at most one hour of computation time, which is quite reasonable. Moreover, the study shows the central relevance of flexibility: Solution values range from 55’250€ to 198’250€ when varying day parameters and from 55’250€ to 81’250€ when varying shift and break parameters. This means, increasing flexibility in terms of sequences of working days leads to a significant reduction in the resulting total salary costs. This is also true for flexibility in shift types: The more shift types are available, the lower the required labor costs and with this, the required workforce size to cover demand. Variation in break parameters however has either a positive or no effect. The positive effect occurs especially in surrounding which provide a minimum level of flexibility in shift types.

Concluding, we formulated a MIP based on a reduced set covering approach accounting for a maximum level of flexibility in the scheduling process of physicians and apply a CG heuristic as solution approach. As the results indicate, a high level of flexibility has a significant positive effect on the total labor costs and on the required number of physicians: The less flexible the scheduling circumstances, the higher the resulting labor costs due to the inability to match supply and demand.

3. Discussion of the contributions

In this section, the basic findings and results are summarized and presented. Moreover, managerial insights derived from the contributions are provided and discussed. Eventually, approaches and areas for future research are presented.

3.1. Summary of major findings and critical evaluation of limitations

Based on the research questions proposed in an earlier subsection, the main findings are summarized in this subsection. Additionally, appropriate limitations concerning the findings are discussed.

Research question 1: Which characteristics and aspects of physician scheduling in hospitals are already covered in literature and what are promising areas for future research?

Within the last decades, the importance of the physician scheduling problem in research and practice has increased significantly. The research question addressed here is summarized und already partially answered in subsection 2.1.. Literature investigates the specific features of the physician scheduling problem in contrast to general personnel scheduling literature and the nurse scheduling problem with respect to the type of problem, problem characteristics such as employee type, experience level of physicians, shift types in use, break assignment, and fairness and individual personnel aspects as well as modeling approaches, uncertainty, solution approaches, and real life implementation. As our analysis indicates, the number of papers increases significantly and a growing number of diverse aspects is considered. However, it is necessary to define the purpose of modeling and, based on this, the relevant aspects that need to be taken into account. For example for Rostering problems, it is not sufficient to solely generate a schedule for each physician. Instead, individual lines of work, which account for personnel aspects and requirements, are to be constructed to ensure an appropriate job satisfaction and consequently a high quality of care. Thus, current literature aims to identify key concerns of the specific problem type that it focuses on to ensure taking into account all relevant aspects.

General research gaps identified in this first contribution are the consideration of an adequate length for the planning period in Rostering problems to ensure predictability for the workforce and a higher planning granularity to better match supply and demand and adopt realistic assumptions. Moreover, even though fairness aspects are gaining importance, the perception of staff is not measured. Therefore, individual perceptions and evaluation of the performance of such metrics in the scheduling process should be adopted to potentially adjust modeling criteria. A more detailed overview on ideas for future research is given in subsection 3.3..

Research question 2: How does flexibility in the assignment of shifts and breaks affect the overall number of required physicians? How can we build a more realistic estimation for approximating the total size of the workforce?

Up to now, it is common practice to estimate the required number of physicians simply by using an approximation mechanism such as the maximum level of demand or based on total demand and available working hours of an employee. Since these methods do not account for various influencing factors, the required workforce size is underestimated significantly. The second contribution in section 2.1. addresses an improved mechanism to approximate staffing levels of physicians more realistically.

The impact of accounting for the assignment of breaks in addition to several other legal regulations when estimating the required number of physicians is modeled using a MIP model. As confirmed by our results, neglecting the assignment of breaks leads to a significant underestimation of the required number of personnel. Consequently, hospitals are often unable to cover demand within regular working hours of staff, which leads to an increasing probability of overtime hours for the workforce and growing staffing costs for the hospital. Especially in real word settings which provide less flexibility, the situation is even more severe: Neglecting breaks might underestimate the workforce size by around one third. This results in additional negative consequences for the quality of patient care and the satisfaction and job motivation of staff. Additionally, the explicit consideration of a break is crucial for physicians since most of them are usually unable to take their breaks even though it is mandatory by law. To answer the first part of our research question: The assignment of breaks in general leads to a more realistic estimation of the workforce size by (majorly) increasing the number of required physicians whereas a flexible placement does either have a positive or no effect. Moreover, implementing flexibility in shift types increases improvement further due to additional precision because of an increasing ability to better match supply and demand.

In our approach, shifts and breaks are assigned in an aggregated way: Our model selects shifts to cover demand subject to several workforce-related legal regulations, but not on an individual level. To generate feasible lines of work covering the entire planning horizon for each physician, an additional mathematical model is required which accounts for physician-specific individual regulations. Similarly for the assignment of breaks: Since our model solely assigns a number of breaks to a specific period, a post-processing procedure is necessary to assign breaks individually to each physician on duty, e.g. by a heuristic procedure. Moreover, we assume personnel to be available 365 days a year as we do not account for vacation and national holidays or absences due to illness. For a more detailed approach, these aspects are to be taken into account or at least additional personnel has to be employed to smooth the resulting effect. This means for the second part of our research question: A more detailed mathematical model investigating several workforce- and physicians-related working regulations improves the approximation of the estimated workforce size.

Research question 3: What is the value of implementing flexibility in the process of physician scheduling?

In general, practice as well as research do not allow for a high level of flexibility in the physician scheduling process. This rigid scheduling environment prohibits an adequate matching of supply and demand. The third contribution therefore addresses the implementation of day, shift, and break flexibility in the physician scheduling problem.

The answer for this research question depends on the degree of flexibility in the scheduling process: The higher the level of flexibility, the lower the occurring total labor costs to cover demand and the required size of the workforce. Accounting for flexibility in patterns of working days leads to a significant decrease in total labor costs. In the extremes, total salary costs decrease by around 350% when increasing working day flexibility to a maximum level. When implementing flexibility in available shift types, results also show a positive effect on the occurring costs. A maximum level of flexibility in shift types leads to a decrease of total salary costs of around 32%. Inserting additional flexibility in the scheduling process by enlarging the size of the break window either has a positive or no effect on the objective. For almost 50% of the different parameter settings, the maximum size of the break window decreases the total labor costs. Flexibility in break assignment therefore particularly affects settings that do not provide a high level of flexibility in working days and available shift types, i.e. especially for real life assumptions.

In the experimental study of this contribution, real life data from the central operating theater is used. Even though this data is derived from a large teaching hospital, it might be biased due to the planning policies of the hospital. Reported demand is mainly based on several hospital-internal planning assumptions such as the predefined Master Surgery Schedule (MSS) for the operating theater as well as the proposed Case Mix (CM) of the hospital, which is defined at the beginning of each year. This might affect demand and change if general long-term assumptions of the hospital are modified. Moreover, the proposed MIP mainly focuses on flexibility aspects in the scheduling process rather than considering personnel aspects of the workforce which might be of central relevance on a practical experimental level.

3.2. Managerial insights improving hospital's administrative and organizational system

As hospitals are confronted with an increasing cost pressure within past years, efforts to become profitable have risen significantly within the last decades. Since the workforce of a hospital generates a major part of the operating costs, improving the scheduling of employees promises large potential cost savings. In this dissertation, we identified four different managerial insights which are discussed in the following that might provide potential to reduce costs and improve the process of physician scheduling.

I) *Realistic estimation of the workforce size.* Hospital's management commonly uses simple approximation mechanisms to estimate the required workforce size since these methods are neither cost nor time intensive. Applying a single formula or basic statistics does not require much time and there is no need to collect many different types of information. However, this leads to a rather bad estimation. Instead, hospital's management should consider using several additional information to create a more detailed mathematical model which (for example) implements breaks to estimate the required number of personnel. This leads to a more precise estimation of the number of required physicians and to a reduction in overtime hours and related costs.

II) *Improvement in absence planning.* In general, hospitals use a mean level for the forecasted demand as basis to determine the workforce size. As a result, no conclusions can be made for vacation policies and absence planning throughout the year. Instead, when the weekly number of required physicians is determined for one entire year, seasonality and trends in demand can be identified and beneficially used for absence planning, e.g. by defining a range for vacation or improving planning of known absences, such as conferences and medical education.

III) *Importance of implementing flexibility in the scheduling process of physicians.* In current literature and practice, it is still quite common to assume a rigid scheduling environment. This results in a decreasing ability to match supply and demand and therefore in an increasing number of required personnel and related costs. Implementing a higher degree of flexibility in terms of patterns of working days, available shift types, and the placement of breaks results in a significant decrease in total labor costs. Hospital's management should therefore consider investigating a higher amount of flexibility in the scheduling process. Moreover, if the implementation of flexibility is not possible due to some reasons and hospitals need to stick to a small number of shifts, shifts having a short duration are superior to shifts with a long duration and should therefore be preferred when flexibility is not adoptable.

IV) *Considering break assignments is crucial even on high hierarchical levels.* Even though the assignment of a break is a central aspect when generating lines of work to ensure an appropriate rest period for each physician while being on duty, the consideration of breaks is also a key aspect on higher hierarchical planning levels. Even on a strategic and on a tactical level, when the size of the required workforce is estimated and seasonality in demand is analyzed to plan predictable absences of personnel, the assignment of breaks is to be taken into account. By integrating several key elements such as minimum rest periods, maximum weekly working hours, and the assignment of breaks, the required size of the workforce can be estimated more realistically. As a result, an appropriate number of physicians to cover demand is determined and employed which leads to a decrease in potential overtime hours due to inadequate planning.

3.3. Areas for future research

The contributions that are presented in this dissertation close several existing research gaps in current literature. However, as the former discussion indicates, there are still diverse limitations that raise a variety of topics and ideas for further research. In this subsection, some ideas are discussed in more detail.

Further aspects for modeling the physician scheduling problem to improve applicability in a practical environment

As various aspects are to be considered in literature handling the physician scheduling problem, there are several gaps in current research leading to ideas for further research. First, further research might promote the importance of considering breaks in the scheduling process since physicians are commonly not able to take their breaks in practice, even though these are mandatory by law. Investigating an efficient way to model the assignment of breaks mathematically might lead to acceptance of the increasing complexity because of the positive effect on physicians' utilization and working hours.

Second, operational online planning provides the most potential for future research: Since only one publication handles Re-planning problems, further research might improve the rescheduling process and investigate several aspects and methods to tackle unforeseen absences of personnel.

Third, demand patterns used in research are often not entirely realistic, even though these are derived from historical data. Commonly, former requirements of care are aggregated over a specific time interval and subsequently used to estimate future demand. By doing so, seasonality and trends in demand are neglected. Since demand in hospitals fluctuates heavily, future research should investigate how to determine more realistic demand patterns, which represent real life demand more precisely.

Fourth, practice-oriented research might develop user interfaces for hospital surroundings to improve performance and acceptance of mathematical modeling approaches derived from research into practice. Commonly, research proposes a well-performing mathematical model and solution approach to tackle real life problems but does not account for an appealing user interface to increase hospital's willingness to implement approaches from literature.

Flexibility in the physician scheduling problem

Since physicians face a great fluctuation in demand, the employment of part time workers might be sufficient. In our approaches, a homogenous group of full timers is considered which all have similar minimum and maximum working hours. Flexibility can be implemented by employing part time physicians to better match supply and demand especially for peaks due to the diversity in available working hours, i.e. part timers might be on duty for only four hours, whereas full timers are to be on duty for at least seven hours. Moreover, future research should investigate flexible sequences of working days and increase the number of available shifts by applying flexible shifts. Even though complexity rises due to the increasing level of flexibility, it is possible to better match supply and demand and with this, to prevent possible negative consequences, such as excessive overtime hours for personnel.

Increasing importance of accounting for work-life balance of staff

Especially when considering residents, additional ergonomic rules ensuring their wellbeing and job satisfaction are important. Up to now, residents' schedules are created for a long-term perspective focusing on educational concerns but neglecting individual requests. Moreover, the schedule of teaching physicians should also be taken into account when scheduling residents to ensure an efficient education and medical progress.

Furthermore, the wellbeing of the workforce also needs to be taken into account, especially on an operational level. Our approach assumes that each physician is willing to start a shift whenever it is necessary, but this might not be applicable in real life surroundings. Therefore, additional ergonomic regulations such as shift starting time windows, individual preferences, and the assignment of multiple breaks for long shifts are to be considered in future research. Moreover, future research might develop a guideline for the trade-off decision of hospitals: Reducing costs by implementing a maximum level of flexibility vs. satisfaction and job motivation of personnel. As a result, a balanced policy between cost- and employee-satisfaction is to be determined.

4. Conclusion

This dissertation on the physician scheduling problem is comprised of three parts. First, the physician scheduling problem is introduced and the content and focus of this dissertation is motivated. Second, the major part of this dissertation, the three contributions of this thesis are summarized. Note, the full version of each contribution is attached in the appendix. And third, each contribution is presented in detail and critically discussed with respect to a more comprehensive point of view. Additionally, resulting managerial insights derived from the contributions are presented and an agenda for further research is provided.

The central contributions of this dissertation are as follows: The first literature overview containing a framework to classify research focusing on the physician scheduling problem is provided in contribution 1. The second contribution presents a mechanism to create an approximation for the required workforce size which is more realistic than approximations created using well-known techniques whereas the value of a maximum level of flexibility in the scheduling process is evaluated in contribution 3.

Since it is the main aim of this dissertation to promote the importance and complexity of the physician scheduling problem, this thesis extends current literature by three aspects: First, contribution 2 implements flexibility and various labor regulations and legal rules in the physician scheduling problem to build a realistic estimation of the required number of physicians. Moreover, the assignment of breaks is considered to ensure a break for each physician and prevent excessive overtime hours. A major advantage of the presented approach compared to other approximation mechanism is the consideration of several key aspects, such as breaks and maximum weekly working hours, which ensures a more precise estimation. A realistic estimation is crucial to prevent unnecessary overtime hours and decreasing job satisfaction. In the third contribution, a maximum level of flexibility, i.e. flexibility in working days, shift types, and break assignment, is provided to evaluate the effect of flexibility. The results from the experimental study show the high value of flexibility for the physician scheduling problem: Especially flexibility in working days and shift types results in a significant decrease in total labor costs due to the increasing ability to better match supply and demand especially for peak periods.

Concluding, applying the presented framework developed in our literature review, key elements of the physician scheduling problem can be investigated. Since hospitals are facing an increasing pressure to become profitable and physicians are still considered as a crucial and scarce resource, hospitals are forced to establish policies to decrease upcoming costs while ensuring personnel satisfaction and maintaining the desired quality of care. The physician scheduling problem is therefore of major importance and expected to gain further relevance in research and practice in near future.

References

Aggarwal, S.C. (1982). A focussed review of scheduling in services. European Journal of Operational Research, 9(2), 114-121.

Aiken, L.H., Clarke, S.P., Sloane, D.M., Sochalski, J., Silber, J.H. (2002). Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction. Jama, 288(16), 1987-1993.

Aykin, T. (2000). A comparative evaluation of modeling approaches to the labor shift scheduling problem. European Journal of Operational Research, 125(1), 381-397.

Balinski, M.L. (1965). Integer Programming: Methods, Uses, Computations. Management Science, 12(3), 253-313.

Bechtold, S.E., Brusco, M.J., Showalter, M.J. (1991). A comparative evaluation of labor tour scheduling methods. Decision Sciences, 22(4), 683-699.

Bechtold, S.E., Jacobs, L.W. (1990). Implicit modeling of flexible break assignments in optimal shift scheduling. Management Science, 36(11), 1339-1350.

Blöchlinger, I. (2004). Modeling staff scheduling problems. A tutorial. European Journal of Operational Research, 158(1), 533-542.

Bölt, U. (2014). Statistische Krankenhausdaten: Grund- und Kostendaten der Krankenhäuser 2011. In J Klauber, M Geraedts, J Friedrich, J Wasem (Ed), Krankenhaus-Report 2014: Schwerpunkt: Patientensicherheit. Schattauer: Stuttgart, 293-328.

Bölt, U., Graf, T. (2012). Gesundheit-20 Jahre Krankenhausstatistik. Wirtschaft und Statistik, 2, 112-139.

Blum, K., Löffert, S., Offermanns, M., Perner, P. (2017). Krankenhaus Barometer - Umfrage 2017. Technical Report, Deutsches Krankenhaus Institut e.V.: Düsseldorf.

Brunner, J.O., Bard, J.F., Kolisch, R. (2010). Midterm scheduling of physicians with flexible shifts using branch and price. IIE Transactions, 43(2), 84-109.

Brunner, J.O., Edenharter, G.M. (2011). Long term staff scheduling of physicians with different experience levels in hospitals using column generation. Health Care Management Science, 14(2), 189-202.

Charles, A.G., Ortiz-Pujols, S., Ricketts, T., Fraher, E., Neuwahl, S., Cairns, B., Sheldon, G.F. (2013). The Employed Surgeon: A Changing Professional Paradigm. JAMA Surgery, 148(4), 323-328.

Dantzig, G.B. (1954). Letter to the Editor - A Comment on Edie's "Traffic Delays at Toll Booths". Operations Research Society of America, 2(3), 339-341.

Dantzig, G.B., Wolfe, P. (1960). Decomposition Principle for Linear Programs. Operations Research, 8(1), 101-111.

Edie, L.C. (1954). Traffic delays at toll booths. Journal of the Operations Research Society of America, 2(2), 107-138.

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

Ernst, A.T., Jiang, H., Krishnamoorthy, M., Sier, D. (2004). Staff scheduling and rostering: A review of applications, methods and models. European Journal of Operational Research, 153(1), 3-27.

Fournier, G.M., Mitchell, J.M. (1992). Hospital costs and competition for services: A multiproduct analysis. The Review of Economics and Statistics, 74(4), 627-634.

Garfinkel, R.S., Nemhauser, G.L. (1969). The Set-Partitioning Problem: Set Covering with Equality Constraints. Operations Research, 17(5), 848-856.

Gross, C.N., Fügener, A., Brunner, J.O. (2017). Online rescheduling of physicians in hospitals. Flexible Services and Manufacturing Journal, 30(1-2), 296-328.

Jarr, K. (1973). Probleme, Grundlagen und Methoden der Personalplanung. Wirtschftsdienst, 53(8), 431-438.

Ozcan, Y.A. (2017). Analytics and Decision Support in Health Care Operations Management. 3rd edition, Jossey-Bass: New York, NY.

Santos, M., Eriksson, H. (2014). Insights into physician scheduling: A case study of public hospital departments in Sweden. International Journal of Health Care Quality Assurance, 27(2), 76-90.

Scherm, E., Süß, S. (2016). Personalmanagement. 3. Auflage, Vahlen: München.

Spengler, T. (1999). Grundlagen und Ansätze der strategischen Personalplanung mit vagen Informationen. Rainer Hampp Verlag: München und Mering.

Statistisches Bundesamt (2016). Gesundheit. Grunddaten der Krankenhäuser 2016, 12(6), 1-156.

Villarreal, M.C., Keskinocak, P. (2016). Staff planning for operating rooms with different surgical service lines. Health Care Management Science, 19(2), 144-169.

Wabro, M., Matousek, P., Aistleither, R. (2010). Handbuch für die Personalplanung. Bundesgesundheitsagentur: Wien.

A. Appendix

A1. State of the art in physician scheduling

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

The final authenticated version is online available at https://doi.org/10.1016/j.ejor.2017.06.037

Status: Published in European Journal of Operational Research, category A.

A2. Physician staffing levels and absence planning under special consideration of breaks – A case study on anesthetists

Erhard, M., Brunner, J.O. (2018). Physician staffing levels and absence planning under special consideration of breaks – A case study on anesthetists.

Status: Submitted to OMEGA - The international Journal of Management Science, category B.

Date of submission: August 07, 2018.

Physician staffing levels and absence planning under special consideration of breaks – A case study on anesthetists

Melanie Erhard (corresponding author)[a,b,*]

Jens O. Brunner[a,b]

[a] Chair of Health Care Operations/Health Information Management

Faculty of Business and Economics

University of Augsburg

Universitätsstraße 16, 86159 Augsburg, Germany

[b] University Center of Health Sciences at Klinikum Augsburg (UNIKA-T)

Neusässer Straße 47, 86156 Augsburg, Germany

*melanie.erhard@unikat.uni-augsburg.de

Submission: August 07, 2018

Physician staffing levels and absence planning under special consideration of breaks – A case study on anesthetists

Abstract

In hospitals, personnel generate the biggest and most important cost. This research handles the physician staffing problem in hospitals on an integrated tactical level by focusing on the investigation of break assignments in staffing levels as major objective. In particular, we consider modeling the placement of breaks within shifts to ensure an appropriate approximation for the size of the workforce for demand coverage. Current literature mainly neglects the consideration of breaks whereas practice uses manual planning approaches that are time and cost intensive. Focusing on a staffing level with tactical applicability, we minimize the number of assigned physicians. To determine a more realistic lower bound for the required staffing level several essential rules such as demand coverage and labor regulations are thereby taken into account. Moreover, trends in personnel utilization throughout the year that provide a valuable contribution for absence time planning (e.g. due to vacation and conferences) are identified. We formulate a mixed-integer program and solve it with standard software (like CPLEX). In our experimental study, real world data from a large hospital in Germany is used. Computational results show that the consideration of breaks is highly relevant especially when approximating required personnel capacity in staffing level decisions: Neglecting breaks, even on a high hierarchical level, leads to a significant underestimation of the number of required physicians particularly under rigid real life assumptions when hardly any flexibility in shift types is provided. Accordingly, understaffing and an increase in unplanned working hours in terms of overtime might be possible consequences. Moreover, legal regulated break periods for physicians cannot always be ensured. Therefore, we recommend the consideration of break assignments even when approximating staffing policies to create leeway in planned absence decision-making.

Keywords: Physicians, break assignment, staffing levels, hospitals, mixed-integer programming

1. Introduction

Within the last twenty years, the number of patients for hospitals is growing significantly due to the increasing age of society and the additional reduction in patient's length of stay (e.g. Schelhase 2014). This leads to an increase in annual costs for hospitals, e.g. in Germany from 50 billion euro to 82 billion euro in 2011. The budget which is consumed by hospitals each year is continuously ascending and this trend will not stop in near future. Since staff generates more than half of these costs, hospitals' managers are forced to schedule their staff efficiently to reduce personnel expenses (Bölt 2014). Assigning physicians in an appropriate way enables hospitals to provide the same high quality of service while reducing the upcoming labor costs, e.g. for overtime hours. In many cases, this results in a significant growth of the number of patients which can be treated while increasing the time patients spent with a physician for treatment. Additionally, a decreasing number of working hours for physicians can be noticed (Rising et al. 1973).

In this context, a lot of work has been done with respect to nurse rostering (e.g. see Burke et al. 2004 for a comprehensive overview or Van den Bergh et al. 2013 for a more recent overview on personnel scheduling including nurse scheduling). According to physician scheduling, there is still a lack in research (e.g. see Erhard et al. 2018). A reason for this deficit is the complexity of the physician scheduling problem since hospitals are facing a great fluctuation in demand. The level of required care changes from hour to hour and from one day to the next which makes it difficult to cover demand in every period (Brunner et al. 2009).

The purpose of this paper is to determine a more realistic estimation of physician staffing levels to facilitate and support absence planning. The contribution of our work is manifold. We develop a mathematical model which provides full flexibility during the process of approximating physician staffing levels with respect to starting and ending times of shifts, the length of shifts, and the placement of a break for each shift. Accounting for breaks when deciding about staffing policies is especially important since physicians are commonly not able to take their breaks in practice, even though it is mandatory by law. This is partially reasoned by environmental uncertainty, e.g. due to stochasticity in surgery durations (Fügener et al. 2017) as well as emergency patients (Venkat et al. 2015), but also due to understaffing as a result of neglecting breaks and unrealistic approximation techniques for the required workforce size, e.g. a simple workload bound. Therefore, implementing breaks is necessary in staffing level decisions on a tactical basis to ensure an appropriate size of the overall workforce on duty. This effect already occurs in flexible shift settings but as an insight of our study we can show that this problem gets even worse in more rigid real life settings, where less flexibility is provided.

Our main objective is to approximate the minimum number of physicians required as realistic as possible, subject to coverage of anticipated demand. At the same time, flexible breaks for each physician as well as general labor regulations, such as rest periods and working hours, are taken into account. The mathematical model selects specific shift types including break assignments. Since we are focusing on a staffing level decision, we do not generate individual schedules which can be applied on an operational level when estimating the required workforce size. In lieu thereof, the selected shifts solely ensure the practicability of our decision by matching supply and demand, assigning breaks, and restraining major scheduling regulations over all personnel, when determining the required staffing level. But, we additionally present a second mathematical model which can be used in a downstream step to create appropriate, feasible individual schedules that are applicable on an operational level, if necessary. Being almost neglected by current literature up to now, one major contribution is the first methodological extension of the modeling approach presented by Bechtold and Jacobs (1990) in a way that accounts for an individual break assignment even in case of overlapping shifts. In our computational study, we randomly generated two different demand patterns that correspond to real world settings (see Brunner et al. 2011) to test for the effect of including breaks on the optimal workforce size. Based on our case study with real world demand, we finally derive substantial insights for research and practice concerning the resulting utilization of personnel as well as the necessity and the effect of accounting for breaks in the staffing process.

The paper is structured in the following way: In Section 2, we review the relevant literature concerning this topic of research. Afterwards, we provide a precise problem definition followed by the mathematical models (see Section 3). In Section 4, we evaluate and test our models by using randomly generated and real life data (in our case study) to analyze the effect of implementing breaks in the scheduling process on the overall size of the workforce when approximating staffing levels. We show the high quality of the resulting staffing policy, guaranteeing a break for every assigned physician. Eventually, we summarize our main findings and give insights before concluding the paper by giving ideas for further research.

2. Literature review

Unlike other branches, business in the service sector is denoted by specific characteristics, e.g. high variation in demand. Since demand cannot be backlogged without generating indirect costs, there is need for excessive (over-) capacity because institutions, especially in the health care sector, have to be prepared to handle peaks in demand. As a consequence, this leads to a large amount of idle time during periods of low demand (Aggarwal 1982). For hospitals, the situation is even more serious since there is need for care 24/7. The negative effect of unmet demand has already been mentioned in the early 1980s (Aggarwal 1982). In the following, we provide a review of the most relevant research for our work. A detailed bibliography for physician scheduling can be found in Erhard et al. (2018).

Beaulieu et al. (2000) tackle the problem with a mathematical programming approach to schedule up to 20 physicians in a hospital in Montreal. Using predefined shifts, the corresponding mixed-integer programming (MIP) model is solved by a Branch and Bound (B&B) approach. The resulting schedule provides increased quality with respect to the number of working hours and rule violations. Similar work scheduling emergency room physicians is presented in Carter and Lapierre (2001). The authors analyze scheduling procedures in six different hospitals in Montreal and build feasible schedules by using tabu search. White and White (2003) apply a constraint programming model to schedule physicians, residents, and medical students in Ottawa hospital in short term. The initial schedule is created using a bin packing procedure. Rousseau et al. (2002) use a similar approach to schedule physicians in the emergency department of a hospital in Montreal. A solution approach combining a local search and a genetic algorithm is introduced, leading to a generic method. A genetic algorithm is also applied in Puente et al. (2009) to solve their goal programming model for short term physician scheduling in a Spanish hospital. Using predefined shifts, the created roster violates less soft constraints than the one currently in use. Additionally, shifts are distributed fairly among the workforce. More recently, Ferrand et al. (2011) present a goal programming approach whose main concern is building a cyclic roster. Their research results in well-balanced working patterns and the opportunity to take care of physicians' interests. Gierl et al. (1993) apply queuing theory solved by a knowledge-based system to consider fairness in the distribution of shifts and working hours. The system selects physicians out of a particular group according to several specific conditions to design a duty roster. The resulting schedule increases job motivation while decreasing costs and absenteeism due to the increase in fairness. Carrasco (2010) also focusses on an equal distribution of workload: In a pediatric department in a hospital in Spain, night and weekend shifts are to be assigned fairly. The problem is modeled as a MIP and solved by using a greedy algorithm that identifies feasible solutions while satisfying fairness constraints. Ganguly et al. (2014) take different skill levels of physicians into account and develop a staff assignment algorithm to balance staffing costs and service levels.

Focusing on more flexibility, Brunner et al. (2009) formulate their flexible shift scheduling problem of physicians as a MIP. They solve their model in two different ways: Exact as a MIP and by using a decomposition heuristic. The heuristic decomposes the planning horizon in separate weeks and uses the preceding week as input for the following one. The model as well as the decomposition heuristic are tested by using real world data derived from the anesthesia department of a large teaching hospital in Germany. Solving the problem as a MIP generates high quality solutions at the cost of huge runtimes. In comparison, the heuristic is much faster and provides almost similar results. The computational study shows that final schedules do not require any overtime. In a subsequent work, Brunner et al. (2010) introduce part time physicians and solve the problem by using an exact Branch and Price (B&P) approach. From a methodologic point of view, the authors introduce and investigate two different branching strategies in the B&P framework. Computational experiments analyze the performance of the algorithm for various lengths of the planning horizon, i.e. two, four, and six weeks. Overall, schedules of high quality are generated. Another extension deals with different skill/experience levels (Brunner and Edenharter 2011): Over a 1-year planning horizon, senior and junior physicians of a teaching hospital in Germany are scheduled. The problem is formulated as a MIP and solved by using a column generation heuristic. The approach provides near optimal solutions which allow for flexibility in terms of assigned shifts and, at the same time, an appropriate level of service and coverage of demand. In contrast, Stolletz and Brunner (2011) develop a reduced set covering problem formulation which uses a shift matrix as input, in comparison to an implicit approach. Their main objective is to minimize the paid out hours by using flexible shifts. Additionally, breaks are assigned to shifts in a pre-processing step, i.e. breaks are considered in the shift matrix. As a result, the reduced set covering formulation outperforms the implicit modeling approach with respect to computation time and solution quality. Furthermore, the authors investigate different fairness measures and conclude that minimal cost schedules can be fair. Implementing breaks in the scheduling process of physicians is relevant to ensure an appropriate staffing level and with this a high quality roster that guarantees appropriate rest periods, increases job satisfaction, and therefore the quality of care. Despite the considerable effect on overtime hours and coverage of demand, breaks are typically not integrated in the physician scheduling process. If the number of assigned physicians is not reduced by the number of personnel on break, the coverage of demand cannot be guaranteed for every period of each day of the planning horizon.

Even in general personnel scheduling literature, only few authors consider the assignment of breaks: Gaballa and Pearce (1979) assign shifts and breaks to an airline call center's personnel. Focusing on a discontinuous planning problem, breaks are assigned implicitly for each shift and every period. Extending this approach by multiple breaks with various lengths, Aykin (1996) develops a model that further increases the flexibility in the scheduling process. A more aggregated implicit modeling approach is used in Bechtold and Jacobs (1990) who introduce a variable that determines the number of personnel on break for a specific period.

In current research, there exists a lack with respect to the assignment of breaks for physicians, as it has only been addressed in few articles to date. Our research contributes to filling this gap and puts the break assignment in the center of investigation. In particular, we show the importance of taking account of breaks on a high hierarchical level of decision making, i.e. when approximating the minimum workforce size required to cover demand in staffing level policies. Therefore, detailed and more precise models are needed to prevent for undercoverage of demand.

3. Problem description

The problem under consideration covers a planning horizon consisting of a set of $\boldsymbol{D}$ days spanning several weeks which all have $\boldsymbol{P}$ periods (e.g. 1-hour periods). We consider a number of physicians which are to be scheduled having specific characteristics such as the maximum (minimum) regular amount of working hours $\overline{R}$ ($\underline{R}$) per time period which are either stated in the labor contract or based on special agreements between the concerned parties. We do not permit a weekly number of overtime hours since our tactical planning problem determines a good lower bound for the size of the workforce to cover the forecasted demand and facilitates planning predictable absences throughout the year, i.e. by revealing required staffing levels for each week in advance. In the following, we consider a homogeneous group of physicians.

We introduce two kinds of flexibility in our model. First, a set of flexible shifts $s \in \boldsymbol{S}$. Each shift s is defined by a flexible starting period as well as a flexible shift length which is between a minimum length S^{min} and a maximum length S^{max}, i.e. seven and 13 hours. To provide access to the resulting available shift types, a shift matrix is generated with the before mentioned parameter values. This matrix defines the binary parameter A_{sp} which indicates a working period p for a specific shift s by being equal to one and zero otherwise. The appropriate assignment of a shift type s to a number of physicians on a specific day d is ensured by the integer decision variable z_{sd}. Considering a discontinuous problem, a shift cannot span over two planning days. In other words, each shift starting on a (planning) day ends on the same (planning) day. Note that the number of planning periods per (planning) day must not span 24 hours and can overlap two real days. For instance, a planning day starts at 5 am and ends at 1 am the next real day, i.e. $|\boldsymbol{P}| = 20$ 1-hour periods. Between two consecutive shift assignments, a minimum rest time P^{rest} must be ensured. Second, we consider flexibility in setting breaks within shifts. Each shift includes a single break that can be assigned flexibly within a predefined break window that is determined by a minimum number of B^{pre} working periods before the break placement and the minimum number of B^{post} working periods before a shift ends. To facilitate our study, we assume each break to have a duration of exactly one period, i.e. 1 hour, per shift. In case of investigating period lengths other than one hour, additional constraints have to be implemented to ensure an appropriate length for the assigned break, e.g. two periods in case of a period duration of half an hour. To guarantee a certain quality and quantity of service and care, the generated schedule has to ensure that (forecasted) demand N_{dp} is covered in each period p of every day d of the planning horizon. By using different assumptions on future demand, i.e. mean or 75th percentile of past demand, staffing level decisions account for the uncertainty in demand realizations. Again, these exemplary schedules are generated and assigned to ensure feasibility of our decision and do not claim to be in use on an operational level, but the provided solution guarantees a valid lower bound for the workforce size of superior quality compared to other estimation techniques, e.g. maximum demand. We formulate the problem as a MIP and assign breaks by the mathematical formulation to periods (labeled as Model 1 in the remainder of the paper).

Model 1 – Breaks to periods. We introduce the essential notation and present the identified superior formulation for the break assignment. Since breaks are modeled implicitly, these have to be assigned to each physician in a post-processing step which is computationally easy (Bechtold and Jacobs 1990). Moreover, since we use an aggregated model to schedule physicians instead of creating individual schedules and lines of work for each employee, personalized rosters have to be generated in a subsequent step, e.g. by a downstream MIP. We therefore state an additional mathematical model (Model 2) which can be used for this purpose. Our main objective is to estimate the minimum workforce size more realistic and determine a lower bound for the number of required physicians subject to providing an appropriate level of care, i.e. forecasted demand coverage. To formulate our model, we need an integer variable Y indicating the number of employed physicians and an integer z_{sd}, indicating the number of physicians working a specific shift on a particular day.

Sets with indices

$d \in \boldsymbol{D}$	Set of days with index d
$p \in \boldsymbol{P}$	Set of day-periods with index p
$s \in \boldsymbol{S}$	Set of shifts with index s
$b^{start} \in \boldsymbol{B^{start}}$	Set of all possible starting periods for a break within a planning day
$b^{end} \in \boldsymbol{B^{end}}$	Set of all possible ending periods for a break within a planning day

Parameters

S^{min}	Minimum shift length
S^{max}	Maximum shift length
A_{sp}	1 if shift s covers period p, 0 otherwise
F_s	First working period in shift s
L_s	Last working period in shift s
W_s	Number of working periods in shift s
B^{pre}	Minimum amount of working periods before the break is allowed to start
B^{post}	Minimum amount of working periods after the break has ended
P^{rest}	Minimum number of rest periods between two consecutive shifts
$\overline{R}$	Maximum amount of regular working periods for physician i per week

$\underline{R}$	Minimum amount of regular working periods for physician i per week
N_{dp}	Demand in period p of day d physician i
B^{first}	Earliest possible period for a break for each shift type per planning day
B^{last}	Latest possible start for the beginning of a break for each shift per planning day

Integer decision variables

Y	Number of employed physicians
z_{sd}	Number of physicians assigned to shift s on day d
o_d	Number of off shifts on day d
b_{dp}	Number of break assignments in period p on day d

$$\text{Minimize } Y \tag{1}$$

subject to

$$\sum_{s\in\boldsymbol{S}} z_{sd} + o_d = \sum_{s\in\boldsymbol{S}} z_{s(d+1)} + o_{(d+1)} \qquad d \in \boldsymbol{D}\backslash|\boldsymbol{D}| \tag{2}$$

$$z_{sd} \leq \sum_{k\in\boldsymbol{S}:|P|-L_s+F_k-1\geq P^{rest}} z_{k(d+1)} + o_{(d+1)} \qquad \forall\, s \in \boldsymbol{S}, d \in \boldsymbol{D}\backslash\{|\boldsymbol{D}|\} \tag{3}$$

$$z_{sd} \leq \sum_{k\in\boldsymbol{S}:|P|-L_k+F_s-1\geq P^{rest}} z_{k(d-1)} + o_{(d-1)} \qquad \forall\, s \in \boldsymbol{S}, d \in \boldsymbol{D}\backslash\{1\} \tag{4}$$

$$\sum_{s\in\boldsymbol{S}} \sum_{d\in\boldsymbol{D}} W_s z_{sd} \leq \overline{R} \cdot Y \tag{5}$$

$$\max_{d\in\boldsymbol{D},p\in\boldsymbol{P}} N_{dp} \leq Y \tag{6}$$

$$\sum_{d\in\boldsymbol{D},p\in\boldsymbol{P}} N_{dp} \leq \overline{R} \cdot Y \tag{7}$$

$$\sum_{s \in \boldsymbol{S}} A_{sp} z_{sd} - b_{dp} \geq N_{dp} \qquad \forall d \in \boldsymbol{D}, p \in \boldsymbol{P} \tag{8}$$

$$\sum_{s \in \boldsymbol{S}} z_{sd} = \sum_{p \in \boldsymbol{P}} b_{dp} \qquad \forall\, d \in \boldsymbol{D} \tag{9}$$

$$\sum_{t=B^{first}}^{b^{end}} b_{dt} \geq \sum_{\substack{s \in \boldsymbol{S} \\ L_s - B^{post} \leq b^{end}}} z_{sd} \qquad \forall\, d \in \boldsymbol{D}, b^{end} \in \boldsymbol{B^{end}} \backslash \{B^{last}\} \tag{10}$$

$$\sum_{t=b^{start}}^{B^{last}} b_{dt} \geq \sum_{\substack{s \in \boldsymbol{S} \\ b^{start} \leq F_s + B^{pre}}} z_{sd} \qquad \forall\, d \in \boldsymbol{D}, b^{start} \in \boldsymbol{B^{start}} \backslash \{B^{first}\} \tag{11}$$

$$\sum_{t=q}^{\tau} b_{dt} - \sum_{\substack{s \in \boldsymbol{S} \\ F_s + B^{pre} \geq q \\ \tau \geq L_s - B^{post}}} z_{sd} \geq 0 \qquad \begin{array}{l} \forall\, d \in \boldsymbol{D}, q \in \left\{B^{first}, \dots, B^{last}\right\}, \\ \tau \in \{q, \dots, \min(q + (S^{max} - B^{pre} - B^{post}) - 1, B^{last})\} \end{array} \tag{12}$$

$$Y, z_{sd}, b_{pd} \geq 0 \text{ and integer} \qquad \forall i \in \boldsymbol{I}, s \in \boldsymbol{S}, d \in \boldsymbol{D} \tag{13}$$

The objective function (1) minimizes the total number of (identical) physicians required over the planning horizon to cover the forecasted demand in order to determine a good lower bound on the workforce size. The first block of constraints (2) to (4) models flexible shift assignments. For this, flow balance constraints, i.e. constraints (2), guarantee the assignment of an appropriate number of shift types for each day within the planning horizon. Additionally, for each day, an appropriate number of off shifts is to be assigned to ensure a feasible combination of shifts over the entire workforce. This combination of the number of assigned on and off shifts on any day d has to correspond to the selected number of assignments on the subsequent day. Not any two shift assignments on consecutive working days are allowed due to working regulations. Particularly, the minimum rest time P^{rest} between two shift assignments on consecutive working days d and $d+1$, as well as on working days d and $d-1$ respectively, has to be enforced. Constraints (3) and (4) are in use for this purpose. Therefore, constraints (3) ensure an appropriate assignment of shifts and off days that are allowed to be assigned on consecutive working day and force the two consecutive shift assignments to account for enough rest periods P^{rest}. Constraints (4) handle the opposite direction by selecting off days and affected shifts for a working day d and

$d-1$, for which the amount of periods between the first working period for one shift on any day d and the last working period for another shift on the preceding day is more than P^{rest}. As a result, a shift on day d can only be assigned if there is a minimum amount of off periods between the first working period of the shift assigned on day d and the last working period of the shift assigned on day $d-1$. It is therefore for example not feasible to assign a shift s, starting at 7 am, on day d when the assigned shift of the previous working day $d-1$ ends later than 7 pm. The parameter P^{rest} must be updated if the total amount of planning periods per day do not span 24 hours. The second set of constraints (5) handles the working time for the employed workforce according to the regulations in their labor contract, i.e. not exceeding $\overline{R} \cdot Y$. The summation counts the weekly working time based on the shift assignments where W_s determines the number of working periods for shift type s that counts towards regular working time. Note, the thereby selected number of shift assignments can create an infeasible solution when determining individual schedules for each physician in a subsequent step. The third block of constraints (6) and (7) set well-known lower bounds on the workforce size: Constraint (6) defines a lower bound determined by the peak in demand within the planning horizon, i.e. the required number of personnel is larger or equal to the maximum demand in the planning period (LB^{Peak}). Constraint (7) defines a lower bound for the workforce based on the weekly workload of each employed physician (LB^{Work}) to determine the size of the necessary workforce. Constraints (8) take care about the demand being in covered every period p on every day d even though, some physicians are assigned to a break in a specific period. Therefore, the shift matrix A_{sp} is predetermined by legally adjusted parameter values and serves as an input. Here, the number of personnel available to cover demand has to be reduced by the number of physicians having their break in a period. In other words, a minimum number of physicians to be on duty is enforced.

The additional set of constraints (9) to (12) determines the flexible break patterns. We fundamentally extend the modeling idea presented by Bechtold and Jacobs (1990) since the approach does not account for overlapping shifts. In particular, constraints (9) assure that the number of shift assignments on any day d is equal to the number of break assignments on the same day. In addition, the appropriate location of the break according to the labor regulations must be assured. Constraints (10) are called forward passing constraints (Bechtold and Jacobs 1990): They force the number of flexible breaks assigned from the first possible break period up to each possible ending period b^{end} to be bigger than or equal to the minimum number of break assignments in $t \in \{B^{first}, \dots, b^{end}\}$. The right hand side counts the number of shift assignments with break window ending by no later than b^{end} and hence determines the number of break assignments needed. In contrast, constraints (11) are called backward passing constraints and work in the opposite direction. They force the number of break assignments after period b^{start} to be bigger than or equal to the number of shift assignments with break window after b^{start}. Again, the latter determines the number of break assignments in periods $t \in \{b^{start}, \dots, B^{last}\}$. The previously introduced constraints are sufficient to assign the appropriate number of breaks if the break window of any shift type is not a real subset of another shift type (Bechtold and Jacobs 1990). However, our modeling of flexible shifts allows several different shift types to overlap each other that may subsequently result in a break window of one specific shift becoming a real subset of the break window of another. This raises the necessity of an extension of the underlying mathematical formulation to account for such situations and consequently, some further constraints are required: Constraints (12) ensure that every shift s gets a break within its break window assigned, even though this break window is part of the break window of another shift. In the following, we give an intuitive example why the constraints are necessary. Consider two shifts s_1, having its break window from period two to period seven, and s_2, having its break window from period three to five. A feasible assignment without constraints (12) could assign a break in period two and another break in period six. The newly introduced constraints forbid this assignment where shift s_1 gets two breaks while shift s_2 gets no break. For this, constraints (12) consider each break window explicitly. Finally, variable domain definitions are given in (13).

Model 2 – Individual schedule creation. Since Model 1 determines a superior lower bound on the workforce size due to the more realistic estimation of the number of required physicians, no individual shift schedules for personnel are created. Model 1 ensures the adherence of the sum of legal working hours over the entire workforce, not for each physician individually. To generate individual lines of work for each employee, an additional mathematical model is necessary which accounts for satisfying legal rest periods and labor working hour regulations (for the mathematical formulation and description see Appendix I). But, this might lead to infeasibility: Model 1 solely determines how often a shift is to be assigned despite accounting for a certain variety in shift types and with this, starting and ending times and lengths. As a result, it might be not possible to ensure individual working regulations for each physician. For example, if especially long shifts are selected to be assigned, ensuring $\underline{R}$ minimum working hours might not be insurable for each personnel due to running out of shifts, i.e. not having enough shifts (especially shorter ones) to meet staff-specific weekly working hours.

To counteract infeasibility, another approach might be to integrate Model 2 in Model 1. In this case, the minimum number of employed physicians is determined by accounting for labor regulations of each employee individually. This might result in an increasing number of required personnel due to considering physicians individually. Consequently, a prime lower bound for the number of required physicians might result. But, at the cost of computational effort and significantly ascending solution times.

But again, this way of assigning breaks requires an additional post-processing mechanism to create final schedules with appropriate break allocations. However, the post-processing has almost no computational burden since a greedy procedure is all what is needed. For each day, the greedy heuristic considers each period (in ascending order) and assigns the breaks to the physician first whose break window has the smallest number of periods left (cf. Bechtold and Jacobs 1990). Remember the tactical focus of our work which is the realistic estimation of the workforce size to build a good lower bound and the improvement in planning foreseen absences robustly, the individual shift schedules including appropriate break assignments are a downstream aim.

4. Experimental study

In this section, we examine the performance of our proposed modeling approach by using randomly generated as well as real world demand patterns. For this, we have implemented Model 1 and Model 2 in IBM ILOG OPL Studio 6.3 and CPLEX 12.0. All computations are performed on a 3.30GHz PC (Intel(R) Core(TM) i5-4590QM CPU) with 8 GB RAM to analyze the effect of considering breaks on the approximation of a good lower bound on the optimal staff size.

In the second part of our experimental study, we consider the demand pattern for anesthetists of the year 2010 delivered by the central operating theatre of a large teaching hospital in Munich. To clarify the general structure, we build an exemplary standard week. Therefore, we aggregated demand over one entire year to build up the mean demand level. For a five day working week, the first day of the week is Monday (Mon through Fri). On weekend days, there does not occur any regular demand since we focus on treatments for elective patients that are commonly known in advance. Demand in off hours, i.e. periods of non-elective demand such as night or early morning hours and weekend days, are handled by additional staff being assigned to on-call service: The appropriate number of physicians just adds to the workforce size. The distribution of mean demand is shown in Figure 1.

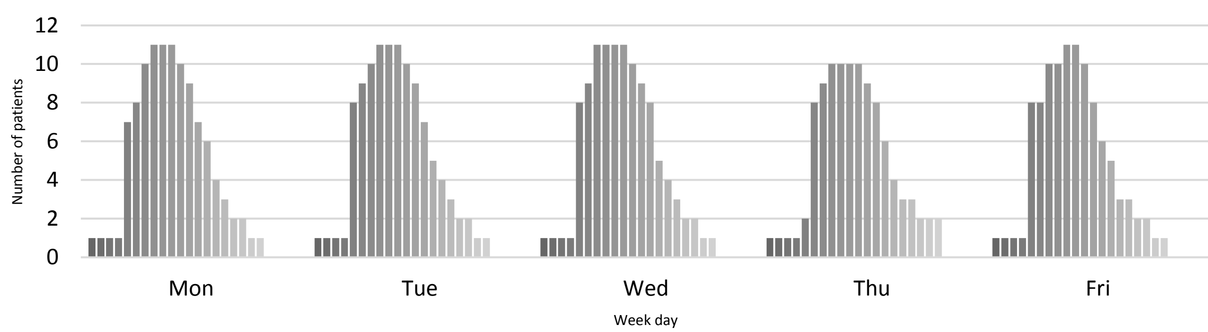

Figure 1: Mean (50% quantile) demand scenario of care for the standard week

Every day consists of 20 periods, each of them having a duration of one hour. This corresponds to a time window which lasts from 5 am to 12 am for the occurring demand. We do not take the remaining hours into account, as there do not show up any elective patients. Emergency patients in off hours (night and weekend) are treated by physicians on duty. In our standard week, the maximum demand is eleven patients (around noon), whereas the minimum number of elective surgeries supposes to be equal to zero. This is not only for weekend days but also for a few other periods the case. Generally, the required level of care is low during morning hours. In period five, which corresponds to 9 am, demand surges from one to at most eight patients. Period eleven is a reversal point: At this time, demand decreases slowly and reaches its minimum around period 18.

To assure feasibility in our model, each physician has the same set of skills and experience level. In general, the total number of physicians employed bases either on the volume of working hours for personnel $LB^{Work} = \frac{\sum_{p \in \boldsymbol{P}} \sum_{d \in \boldsymbol{D}} N_{pd}}{\overline{R}}$ or the maximum demand $LB^{Peak} = \max N_{pd}$ in the planning horizon as lower bound (for more detail, we refer to Brunner and Edenharter 2011). In particular, the first bound (workload bound) is strongly influenced by the total number of working periods for each physician and the sum of demand within the planning days whereas the second bound (maximum demand) is determined by peaks in demand. Former results in an increasing complexity in the solution process due to various resulting possibilities for the design of the shift schedule whereas the latter only occurs in a few periods but might affect the size of the workforce over the whole planning horizon since the required level of care is artificially high. As a result, the larger the considered planning period, the higher the number of required physicians to cover demand. Since we focus on a tactical problem, we use this information to define a superior approximation for the number of required physicians. To keep the study more simple, we suppose the maximum number of regular working hours for each physician to be equal to 40 hours a week, i.e. $\overline{R} = 40$ and permit weekly only 5 undertime hours for each physician, i.e. $\underline{R} = 35$. The shift lengths are between seven and 13 periods (hours) including a flexible break of one period. After a shift ends, the corresponding physician is off for at least twelve periods (hours). This value is reduced by the implicitly assigned off hours within two consecutive planning days since the total amount of the planning periods is less than 24 hours. Considering 20 planning periods results in four implicitly assigned off hours. Additionally, we define the break window to start at least after three working periods and at most three periods before the end of the appropriate shift. This leads to $B^{pre} = B^{post} = 3$. These values are constant over the entire planning horizon and throughout the experimental study. A detailed overview of the parameter setting is given in Table 1.

$\underline{R} = 35$	$S^{min} = 7$	$B^{pre} = 3$
$\overline{R} = 40$	$S^{max} = 13$	$B^{post} = 3$
	$P^{rest} = 12$	

Table 1: Parameter setting Base Case

This parameter setting is called the Base Case and serves as reference point for the analysis of the outcome. To compare for the effect of the consideration of breaks, we additionally solve different demand scenarios with a reference model (denoted by Model 0) which neglects the consideration of flexible breaks at all to provide information about the effect of the break assignment itself. Since no breaks are considered in Model 0, we use the mathematical formulation of Model 1 without constraints (9) – (12) and reformulate constraints (8) by neglecting the number of physicians having their break in a specific period when covering demand. Additionally, we set the parameters $S^{min} = 6$ and $S^{max} = 12$ and update W_s, which counts the regular working hours of a shift s, accordingly.

Solving Models 0 and 1 for our mean demand scenario using the fixed parameter setting of our Base Case to optimality results in an optimal workforce size of 14 physicians for Model 0 and 15 physicians for Model 1 respectively. If no breaks are assigned, the minimum number of physicians required for this specific demand scenario is equal to the analytical lower bound LB^{Work} for the workforce size based on Brunner and Edenharter (2011). Therefore, for this test instance, neglecting breaks does influence the total size of the workforce. This might not be the case for some specific demand patterns in combination with large scheduling flexibility for constructing shifts. First, we evaluate the effect of taking breaks into account on the objective value by comparing Model 1 with our reference Model 0 on two different generated demand patterns with random realizations. Eventually, we present a case study for the whole year 2010 under two different input assumptions, i.e. maximum flexibility in shift types and a more realistic three eight hour shift system. The latter analysis shows the serious effect of neglecting breaks in real world settings.

1.1. Effects of break assignment from practical point of view

In this section, we evaluate the performance of our model using two different generated demand profiles based on real life experience (for more details, see Brunner et al. (2010)) to gain insight into the effect of including breaks in the staffing process. By doing so, we would like to mirror current staffing levels and scheduling practice and show the significant impact of break assignments. Based on the following assumptions, we generated 50 different test instances for each demand structure. Each test instance is solved to optimality using Model 0 and Model 1. Runtimes ranged from ten to 120 seconds and can be neglected for tactical decision making. For the general configuration, see Figure 2.

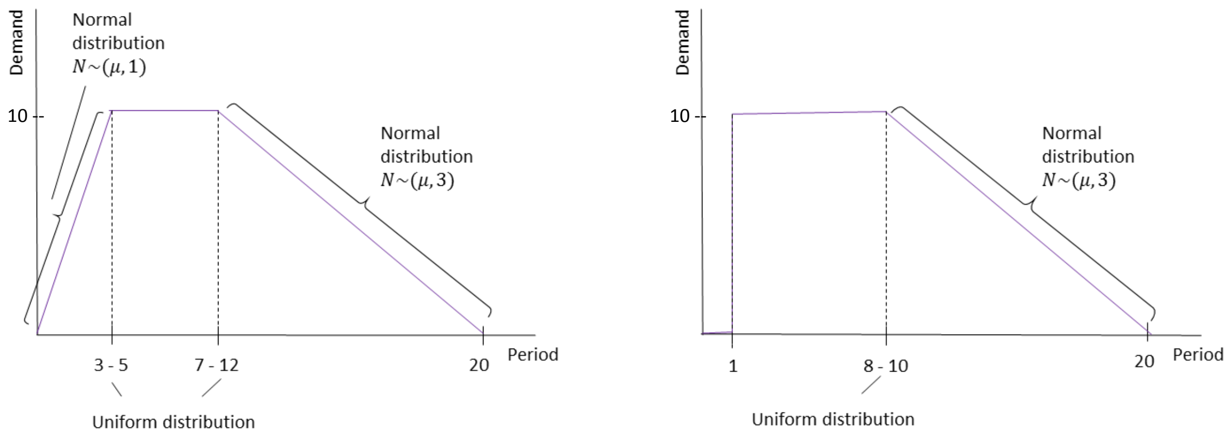

Figure 2: Randomly generated demand patterns

First, we generate the requirements for care according to the left hand side of Figure 2 randomly. In this case, demand increases stochastically with an increasing μ up on period one (which corresponds to 5 am) to a maximum of ten patients. This peak in demand is constant in its altitude but duration varies: It is constant from period five to period seven. In period three to five, respectively seven to twelve, requirement for care can still adopt the maximum value due to the underlying uniform distribution in these hours of the (planning) day. This means, the resulting peak in demand ranges at least from period five to seven but at most from period three to twelve. Subsequently, the required level of care declines stochastically following the underlying Normal distribution with a decreasing μ to zero until the end of the day, i.e. in period 20. As a result, the optimal size of the workforce ranges from eleven to 13 in case of excluding breaks and from twelve to 15 if breaks are considered (for a more detailed overview see Appendix II). The minimum number of physicians required is therefore underestimated when using the reference Model 0 in all 50 instances even with high flexibility. This means, the required size of the workforce is underestimated in each executed test, i.e. the optimal objective value is one or two employees (at most) less. Not considering breaks therefore leads to an unrealistic and too low estimation of the number of required physicians and with this the inability to cover demand in all test instances.

An underestimation of two physicians seems to be the case especially for instances, where the optimal solution provided by Model 0 is disparate from the predetermined analytical working hour bound LB^{Work}. A reason for this might be the specific structure of these demand profiles:

The underlying patterns are commonly denoted by a heavily fluctuating demand containing a wide range for the peaks' lengths. Due to the emerging instability and inconsistency in the required level of care over the considered days and hours, matching supply and demand is rather complicated and makes the implementation of breaks particularly important.

Second, we use the structure shown on the right hand side of Figure 2. This profile for demand corresponds significantly to operating rooms in real life surroundings which are determined by the opening hours of the operating theater. In this setting, demand starts occurring in period one with its maximum value, i.e. number of parallel open operating rooms where in each one an anesthetist has to be available. Again, the length of the peak in demand is stochastic and is therefore at least until period eight constant, but at most until period ten. This is determined by the underlying uniform distribution for the length of the occurring peak. Subsequently, the required level of care decreases slowly to its minimum (equal to 0) in period 20, i.e. at the end of the planning day. Results provide similar insight: The optimal number of employed physicians ranges from 15 to 20 without the assignment of breaks and from 16 to 23 in case of applying Model 1 (see Appendix III for detailed results). Albeit underestimation of personnel does not occur that often for this structure of demand, it is still the case for 86% of the tests. This means, in 43 out of 50 instances, hospital would not be able to cover demand if the number of employed physicians is determined without considering breaks since the optimal workforce size is one to five physicians less. This means, the resulting gap between the workforce size considering breaks and neglecting breaks is much larger for demand patterns following this specific, more realistic, profile. Note that the optimal size increases compared to the first profile. At a first glance, this result is a kind of counterintuitive but can be explained by the wide spread of demand throughout the day which results in more physicians needed even for no break assignments (Model 0), i.e. z^{Flex}. This increased availability gives much more flexibility for assigning appropriate breaks. Therefore, decreasing the number of possible demand periods within a day (as it is the case for many service organizations) will strongly increase the underestimation by neglecting breaks in the staffing process. In other words, the underestimation of the workforce size increases when the planning day gets shorter, i.e. less than 20 periods, and demand can be covered by single shifts, i.e. no two shifts are necessary to cover early and late demand.

Concluding, neglecting breaks leads to an inappropriate management decision and increases probability of understaffing which might possibly result in unplanned overtime hours for personnel. Depending on the structure of the demand pattern, this effect shows up for some profiles more often and in a more aggravating way than for others.

1.2. Case study

In further studies, we saw a significant reduction in the workforce size by the implementation of flexibility in terms of shift types and the size of the break window in the staffing process. Even though the computational effort rises marginally due to the increasing assignment opportunities, the generated benefit is valuable for hospitals to reduce their expenses for personnel while ensuring the requested level and quality of care. In particular, there is a positive effect on the workforce size by enlarging the break window, i.e. enlarging the options of the placement of a break has a positive effect on the workforce size especially in situations of minor flexibility in shift types. Furthermore, since we are focusing on approximation mechanisms for staffing level decisions, the scope of solution time should be of minor relevance.

Therefore, this section composes of two parts: First, Model 1 with full flexibility is used to solve the real world problem instances for the entire year 2010. We use the parameter values of our Base Case and decompose the annual demand into 52 seven day planning horizons. Second, assumptions derived from current literature and real life in terms of available shift types are integrated, i.e. reducing flexibility by using three different eight hour shifts only (see Erhard et al. (2018)). For the first share of the study, aggregated statistics such as the maximum demand in each week are presented with 15 as the overall maximum requirement for care in a few periods. Throughout the year, the required level of care increases in the second and third month and decreases afterwards. Over the summer months, demand seems to be almost stable. Around fall, demand is growing again for around eight weeks and decreases in December.

Based on this demand patterns, optimal schedules are generated to determine the minimum size of the workforce. This requires Model 1 with 1'339 constraints and 827 variables. We present the minimum workforce size (z^{Flex_Break}) per week as well as LB^{Peak} and LB^{Work} in Figure 3. Focusing on the computational effort, the required solution time is solely a few seconds for all test instances.

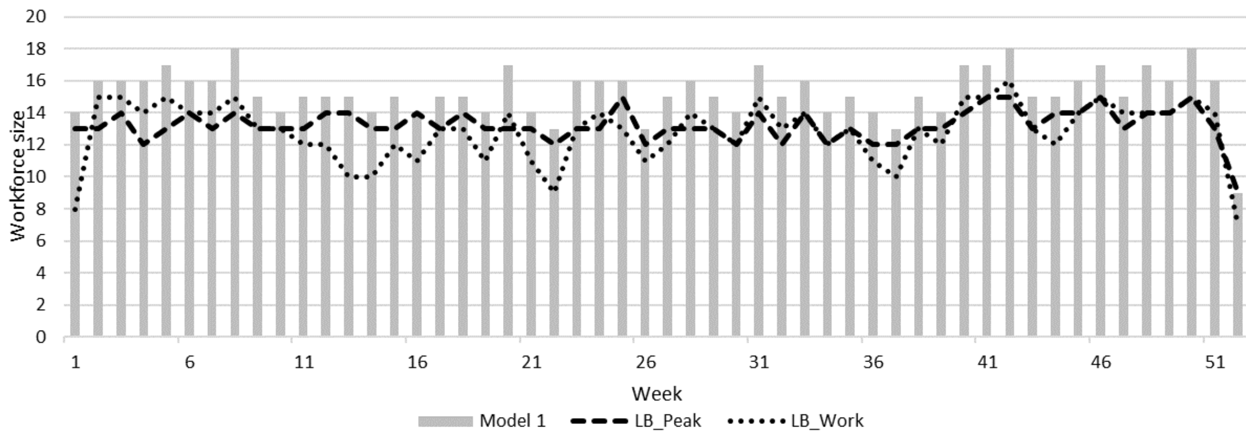

Figure 3: Annual size of the workforce per week

Due to the variation in demand, the optimal size of the workforce varies over the different weeks significantly. In the first week of January 2010, at least 14 physicians are required to cover demand. Since we consider the number of elective patients, this rather small number at the beginning of the year seems to be representative for the occurring demand. There are several national holidays at the beginning of a new year where no planned surgeries take place. Moreover, patients might be on vacation and do not want their surgeries afterwards. Subsequently, the minimum size of the workforce decreases for the following months. Even though, demand is quite stable in the summer months, the minimum number of personnel ranges between nine and 17 physicians during that time. The minimum number of physicians is strictly higher than both common lower bounds for the workforce size LB^{Peak} and LB^{Work}. This is not surprising since the available working hours of the workforce commonly serve as lower bound for the number of personnel which significantly underestimates the required number of physicians. Our approach therefore provides a superior lower bound for the approximation of staff size in short time leading to a more realistic estimation of personnel required. Since our Model 1 still underestimates the required workforce size (compared to the integration of Model 2 in Model 1), the maximum objective value of all weeks should be in use to determine the overall size of the workforce. This leads to the conclusion that at least 18 physicians are required to cover the annual demand from year 2010. Comparing the optimal size of the workforce for each week and the minimum number of personnel required with mean demand (where 15 physicians are assigned), need for care could not be covered in 23 weeks of the year. As a result, if mean demand is used to determine the optimal number of personnel, around 44% of the year is understaffed. Of course, employing the number physicians that is required to cover maximum demand, only overstaffing occurs.

Based on the thereby determined optimal number of physicians required for each week, hospital's management can use this valuable information to handle predictable absences throughout the year on a more operational level: Vacation, medical conferences, health workshops as well as postgraduate training can be scheduled well-conceived. Due to the knowledge of the weekly required workforce size, weeks of low demand (and with this high idle capacity) can especially be considered to schedule vacation of staff and other plannable absences. As a result, demand is still covered without the necessity of overtime hours of personnel and physicians can take their holiday and attend medical conferences while maintaining an appropriate level of medical staff and quality of care at the hospital.

To evaluate the quality of the provided solution, additional performance indicators such as idle time and the average weekly utilization of personnel are determined by the following formula:

$$Utilization = \frac{\sum_{d \in \boldsymbol{D}} \sum_{p \in \boldsymbol{P}} N_{dp}}{\overline{R} \cdot Y}$$

Since we assume each employee to work 40 hours each week, a total of 720 hours of care is available each week for a workforce size of 18 physicians. These are confronted with a weekly demand of 494 hours on average. This results in an average of 226 hours idle time weekly which is decomposed over the total number of staff. Therefore, assuming the size of the workforce to be 18, the average utilization is 68.57% which seems to be appropriate, especially for hospitals.

In contrast, staffing each planning period with its individual minimum number of personnel (see Figure 3) leads to an increase in the utilization of staff as well as a reduction of idle time. As an illustrative example, we determine these performance indicators for the second week of the year 2010: The minimum number of required physicians is 16 opposed to a demand level of 575. The resulting idle time of 65 hours decompose over all physicians. Considering the whole year, average weekly utilization is 80.35%. The resulting weekly utilization of staff for varying size (flexible number of physicians) as well as 18 physicians (max. number physicians) over the whole year are shown in Figure 4. Note, not all physicians are available 52 weeks of the year due to absences and vacation which are not taken into account by our model.

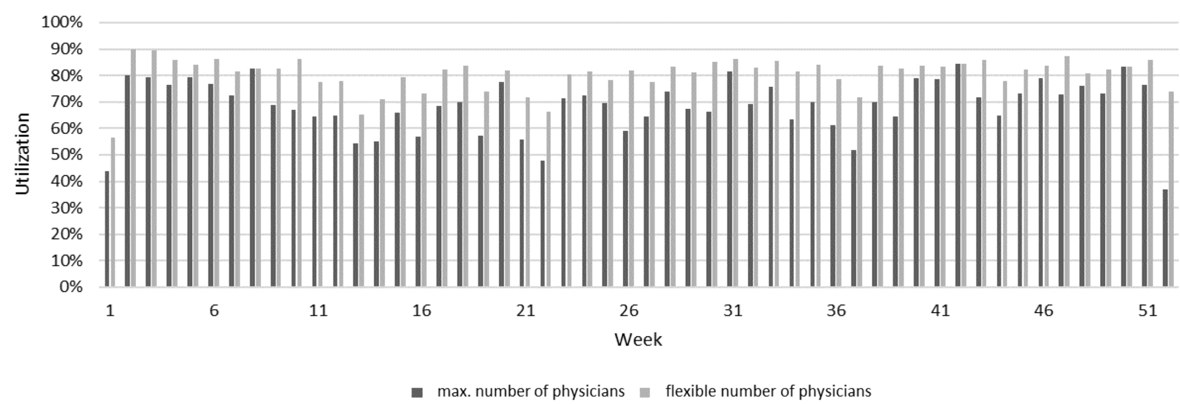

Figure 4: Utilization of physicians in 2010

Since a maximum level of flexibility in shift types is quite uncommon in practice and hardly viable due to potential negative effects for personnel and with this patient care, the number of available shift types is reduced in the second part of the case study. We therefore consider solely three overlapping eight hour shifts which are uniformly distributed over the planning day, i.e. starting in period one, seven and twelve respectively when breaks are considered. Again, the former presented 52 one-week instances are solved to optimality. The Base Case parameter setting is used as input data, except the shift-related ones. Each instance is solved using Model 1 and Model 0 to test for the effect of break assignments in real life surroundings.

Solving both models to optimality for each week results in a disparate range for the objective function value: Applying Model 0, the minimum workforce size ranges from 17 to 33 ($z^{3shifts}$). In contrast, the range provided by Model 1 has a minimum value of 23 and a maximum of 44 physicians, i.e. $z^{3shifts\_Break}$ (see Figure 5).

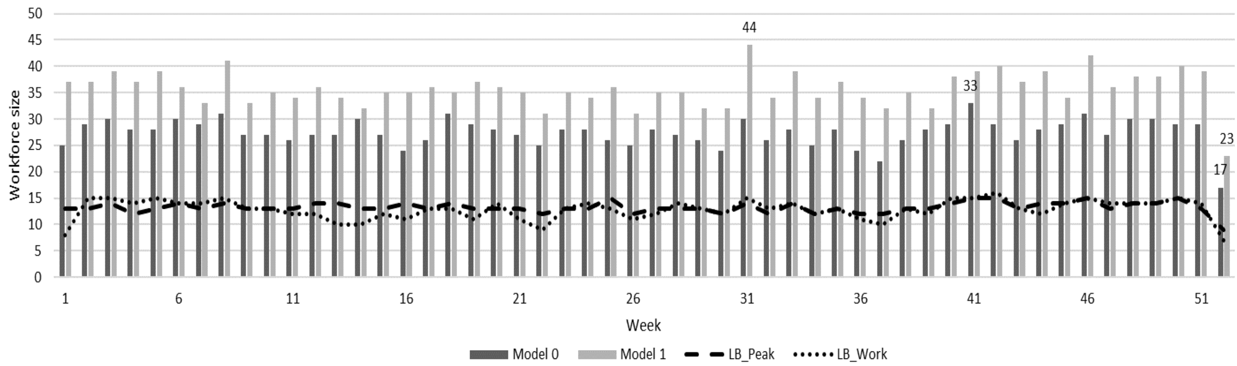

Figure 5: Optimal objective value in 2010, breaks in- and excluded

Applying either minimum or maximum objective value provided by Model 0 to determine the optimal number of physicians leads to a significant underestimation of the number of required personnel. Referring to the maximum objective of Model 0 for the minimum workforce size employed leads to occupying far too few staff, i.e. six physicians less than for Model 1. The results provided by our experimental study lead to the insight that is it important for hospital's management in particular to consider more detailed mathematical models assigning shifts aggregated to a number of physicians and accounting for breaks when deciding about the required staff size rather than using simple approximation schemes such as LB^{Peak} or LB^{Work}. A precise scheduling model leads to a more realistic estimation of the required number of personnel and prevents for being significantly understaffed, i.e. not being able to cover annual demand during physicians' regular working hours. A simple approximation leads to a greater probability of a higher underestimation of the required number of physicians, as can be seen by the following example using demand of the fourth week of the year 2010: $LB^{Peak} = 12$, $LB^{Work} = 14$, $z^{Flex} = 15$, $z^{Flex\_Break} = 16$, $z^{3shifts} = 28$, and $z^{3shifts\_Break} = 37$. This leads to $LB^{Peak} \leq LB^{Work} \leq z^{Flex} \leq z^{Flex\_Break} \leq z^{3shifts} \leq z^{3shifts\_Break}$.

Since no legal nor workforce constraints are taken into account when approximating the workforce size by LB^{Peak} and LB^{Work}, the determined staff size provides the smallest values, i.e. leading to the highest underestimation of personnel. A more realistic estimation is achievable, if a more detailed modeling approach considering many additional labor characteristics, i.e. Model 1, is in use. In this case, a maximum level of flexibility results in an objective function value which is at least equal to the determined bound. But, the investigation of breaks leads to a further increase in the workforce size due to the additionally integrated accuracy. Reducing the solution space by decreasing the level of flexibility in terms of shift types, the optimal objective function value of a 3-shift system is at most as good as the optimal objective function value when including maximum flexibility in shift types, i.e. $z^{3shifts} \geq z^{Flex}$. Moreover, including breaks in the scheduling process might result in an additional growth in the objective function value, i.e. $z^{3shifts\_Break} = 37$. As a result, the number of overtime hours for personnel might decrease due to using a precise modeling approach which accounts for breaks when determining the minimum staffing level, e.g. compared to simple approximation approaches. As already mentioned, one method to integrate further accuracy in our model is the combination of Model 1 and Model 2 to assign shift schedules and breaks individually to each employee. But, this leads to a significant increase in solution time. Our model is therefore a middle course between solution quality, precision, and computational effort since Model 1 provides a high quality approximation of staff size within short computation times. So, additional accuracy is available, e.g. by individual shift and break assignment, but at the expense of higher solution times.

Additionally, employing the appropriate number of physicians leads to a potential reduction in staff utilization. The resulting unplanned working hours per day are not specifically scheduled and can therefore be used for administrative and organizational tasks as well as to handle stochasticity in demand, e.g. smooth variation in the required level for care and handle unforeseen events.

2. Summary and conclusion

In this study, a reduced set covering approach is used to schedule flexible shifts and flexible breaks for physicians in hospitals. In particular, we have put a focus on the assignment of breaks because these are commonly neglected when approximating a lower bound for the required workforce size. Since physicians are often not able to take their break while being on duty even though it is mandatory by law, the relevance is conspicuous. We have developed a new MIP formulation to model the assignment of breaks. Based on real life experience, two demand patterns with 50 randomly generated instances each show the importance of implementing breaks in staffing level decisions due to the resulting disparate optimal workforce size. Additionally, for the considered tactical planning problem, we decompose real life data of 2010 into weekly planning periods in the second part of our experimental study. Model 1 which takes breaks into account provides an optimal solution for all instances in a computation time which is less than five minutes for all test sets. Considering breaks therefore does not harm in terms of computational effort, since solution times are in any case rather low. Comparing the minimum number of personnel for one entire year when applying simple approximation mechanism, e.g. LB^{Peak} or LB^{Work}, Model 0 (without breaks), and Model 1, neglecting breaks and less modeling accuracy leads to a significant under-estimation in the optimal number of physicians needed. Considering labor regulations, personnel needs, and breaks in the physician staffing process therefore results in a more realistic estimation of the required workforce size and could result in a potential for reducing overtime hours and related costs. Moreover, accounting for trends and seasonality in the number of required staff each week leads to a significant improvement when planning predictable absences such as vacation, medical conferences, and workshops: Since the levels of weekly demand is forecasted, weeks of low utilization can be selected for the placement of vacation and occupational trips. This leads to a higher utilization during this time of the year and a decrease in the probability of understaffing for the remaining weeks of the year. Concerning the level of flexibility, the lower the degree of flexibility in terms of shift types, the higher the number of required physicians to cover demand and with this the upcoming labor costs. Especially under such kind of circumstances, additional flexibility in break assignment is rather important. By enlarging the size of the break window, the number of options for the placement of the break increases which leads to a reduction in the number of required physicians.

For future research, the employment of part time physicians in addition to full timers can generate further flexibility and improvement in the staffing process. Daily working hours for part time physicians can be reduced to a greater extent compared to full time employees. Assigning part timers can improve matching supply and demand during times of peaks within the day. Another approach to implement supplementary flexibility is a flexible sequence of working days instead of standard working days from Monday to Friday. Even though solution times might increase partially with the degree of flexibility, hospital's management should seek to approve this additional effort in order to achieve a solution of higher quality. Albeit a high level of flexibility in shift and break assignment leads to decreasing personnel costs, it is also necessary to take the wellbeing of personnel into account. Inserting too much flexibility in staff's schedule might have negative consequences on the job motivation as well as absenteeism and can cause a high probability for turnovers. For current practice, hospitals have to find a middle course to match hospitals and personnel's interest. By determining an optimal size of the workforce, regular demand can be covered while ensuring an appropriate amount of planned slack time to care for emergency patients and unforeseen peaks in demand. Moreover, taking care of social and psychological aspects of flexible schedules as well as the implementation of ergonomic rules into the roster generation process is important to ensure personnel's wellbeing. On an operational level, further research can contain aspects such as to consider individual preferences for the placement of the break. Additionally, assigning multiple breaks efficiently for long shifts is still an open issue.

Appendix I

To build a shift schedule for each physician, additional notation and a second mathematical model, i.e. a feasibility model which does not have an objective function, is presented:

Sets with indices

$i \in \boldsymbol{I}$	Set of physicians with index i

Parameters

Z_{sd}	Number of physicians assigned to shift s on day d (determined by Model 1)

Binary decision variables

x_{isd}	1 if physician i is assigned to shift s on day d, 0 otherwise

$$\sum_{s \in \boldsymbol{S}} x_{isd} \leq 1 \qquad \forall\, i \in \boldsymbol{I}, d \in \boldsymbol{D} \tag{14}$$

$$x_{is_1d} + x_{is_2(d+1)} \leq 1 \qquad \forall\, i \in \boldsymbol{I},\ s_1, s_2 \in \boldsymbol{S}\colon |P| - L_{s_1} + F_{s_2} - 1 < P^{rest}, d \in \boldsymbol{D} \setminus \{|\boldsymbol{D}|\} \tag{15}$$

$$\underline{R} \leq \sum_{s \in \boldsymbol{S}} \sum_{d \in \boldsymbol{D}} W_s x_{isd} \leq \overline{R} \qquad \forall\, i \in \boldsymbol{I} \tag{16}$$

$$\sum_{i \in \boldsymbol{I}} x_{isd} = Z_{sd} \qquad \forall\, s \in \boldsymbol{S}, d \in \boldsymbol{D} \tag{17}$$

$$x_{isd} \in \{0,1\} \qquad \forall i \in \boldsymbol{I}, s \in \boldsymbol{S}, d \in \boldsymbol{D} \tag{18}$$

Constraints (14) and (15) handle the individual assignment of flexible shifts. For this, constraints (14) guarantee that every physician i is assigned to at most one shift out of many flexible types per day. A minimum number of rest periods between two consecutive shift assignments for each physician is ensured by constraints (15). The next set of constraints enforces the adherence of the minimum and maximum weekly working hours for each physician i according to the regulations in his labor contract to be within the allowed range from $\underline{R}$ to $\overline{R}$. Especially maintaining the minimum number of weekly working hours might result in an infeasible solution if there are not enough appropriate shifts selected by Model 1 that can be assigned to the workforce. Constraints (17) render the optimal decision variable values determined by Model 1 for the number of physicians assigned to shift s on day d as input parameter to Model 2 by converting Z_{sd} to x_{isd}. Eventually, variables are defined in constraints (18).

When solving Model 2, possibly no feasible solution can be found since Model 1 solely approximates the size of the workforce due to the aggregated formulation which does not generate individual schedules and does not account for legal working hour and rest period regulations individually. As a result, using Z_{sd}, determined by Model 1, as input for Model 2 might provide no feasible solution.

Appendix II

Test set	Bound		Objective function		Underestimation
	LB^{Peak}	LB^{Work}	Model 0	Model 1	Model 0 vs. Model 1
1	10	13	13	15	-2
2	10	11	12	13	-1
3	10	12	12	14	-2
4	10	11	11	13	-2
5	10	11	12	13	-1
6	10	12	12	14	-2
7	10	11	12	13	-1
8	10	11	12	13	-1
9	10	12	12	14	-2
10	10	12	13	14	-1
11	10	13	13	15	-2
12	10	12	12	14	-2
13	10	11	12	13	-1
14	10	11	12	14	-2
15	10	12	12	13	-1
16	10	11	12	13	-1
17	10	13	13	14	-1
18	10	12	12	14	-2
19	10	10	11	13	-2
20	10	13	13	14	-1
21	10	11	12	13	-1
22	10	13	13	15	-2
23	10	12	12	13	-1
24	10	12	12	14	-2
25	10	11	11	13	-2
26	10	11	12	13	-1
27	10	11	12	13	-1
28	10	12	12	14	-2
29	10	10	11	12	-1
30	10	11	12	14	-2
31	10	11	12	13	-1
32	10	12	12	14	-2
33	10	12	13	14	-1
34	10	11	12	13	-1
35	10	11	12	14	-2
36	10	11	12	13	-1
37	10	12	12	13	-1
38	10	11	12	13	-1
39	10	11	12	13	-1
40	10	11	12	13	-1
41	10	12	13	14	-1
42	10	10	11	13	-2
43	10	11	12	13	-1
44	10	11	11	13	-2
45	10	12	12	14	-2
46	10	11	11	13	-2
47	10	12	12	14	-2
48	10	12	12	14	-2
49	10	11	11	13	-2
50	10	10	11	13	-2

Table 2: Detailed results Figure 2, first analysis (right hand side of Figure 2)

Appendix III

Test set	Bound LB^{Peak}	Bound LB^{Work}	Objective function Model 0	Objective function Model 1	Underestimation Model 0 vs. Model 1
1	10	15	15	20	-5
2	10	16	16	18	-2
3	10	16	17	18	-1
4	10	17	17	19	-2
5	10	16	17	18	-1
6	10	15	15	17	-2
7	10	16	17	18	-1
8	10	15	15	17	-2
9	10	16	16	17	-1
10	10	16	16	17	-1
11	10	16	17	18	-1
12	10	15	15	17	-2
13	10	16	16	18	-2
14	10	16	16	17	-1
15	10	17	17	18	-1
16	10	15	15	17	-2
17	10	16	17	18	-1
18	10	16	16	17	-1
19	10	16	16	18	-2
20	10	16	17	18	-1
21	10	15	15	17	-2
22	10	16	17	18	-1
23	10	16	16	17	-1
24	10	16	17	18	-1
25	10	16	17	17	0
26	10	16	16	17	-1
27	10	16	16	17	-1
28	10	16	16	17	-1
29	10	16	16	18	-2
30	10	16	16	18	-2
31	10	15	15	16	-1
32	10	17	17	19	-2
33	10	16	16	17	-1
34	10	15	15	16	-1
35	10	16	16	17	-1
36	10	15	16	16	0
37	10	16	17	18	-1
38	10	14	15	16	-1
39	10	16	17	18	-1
40	10	17	17	18	-1
41	10	20	20	23	-3
42	10	18	20	21	-1
43	10	18	20	20	0
44	10	18	20	20	0
45	10	16	17	19	-2
46	10	18	20	21	-1
47	10	17	20	20	0
48	10	18	19	20	-1
49	10	16	18	18	0
50	10	18	20	20	0

Table 3: Detailed results Figure 2, second analysis (left hand side of Figure 2)

References

Aggarwal, S.C. (1982). A focussed review of scheduling in services. European Journal of Operational Research, 9(2), 114-121.

Aykin, T. (1996). Optimal shift scheduling with multiple break windows. Management Science, 42(4), 591-602.

Beaulieu, H., Ferland, J.A., Gendron, B., Michelon, P. (2000). A mathematical programming approach for scheduling physicians in the emergency room. Health Care Management Science, 3(3). 193-200.

Bechtold, S.E., Jacobs, L.W. (1990). Implicit modeling of flexible break assignments in optimal shift scheduling. Management Science, 36(11), 1339-1350.

Bölt, U. (2014). Statistische Krankenhausdaten: Grund- und Kostendaten der Krankenhäuser 2011. In J Klauber, M Geraedts, J Friedrich, J Wasem (Ed), Krankenhaus-Report 2014: Schwerpunkt: Patientensicherheit. Schattauer: Stuttgart, 293-328.

Burke, E.K., Causemaecker, P., Berghe, G., Landeghem, H. (2004). The state of the art of nurse rostering. Journal of Scheduling, 7(6), 441-499.

Brunner, J.O., Bard, J.F., Kolisch, R. (2009). Flexible shift scheduling of physicians. Health Care Management Science, 13(3), 285-305.

Brunner, J.O., Bard, J.F., Kolisch, R. (2010). Midterm scheduling of physicians with flexible shifts using branch and price. IIE Transactions, 43(2), 84-109.

Brunner, J.O., Edenharter, G.M. (2011). Long term staff scheduling of physicians with different experience levels in hospitals using column generation. Health Care Management Science, 14(2), 189-202.

Carrasco, R.C. (2010). Long-term staff scheduling with regular temporal distribution. Computer Methods and Programs in Biomedicine, 100(2), 191-199.

Carter, M.W., Lapierre, S.D. (2001). Scheduling emergency room physicians. Health Care Management Science, 4(4), 347-360.

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

Ferrand, Y., Magazine, M., Rao, U.S., Glass, T.F. (2011). Building cyclic schedules for emergency department physicians. Interfaces, 41(6), 521-533.

Fügener, A., Schiffels, S., Kolisch, R. (2017). Overutilization and underutilization of operating rooms - insights from behavioral health care operations management. Health Care Management Science, 20(1), 115-128.

Gaballa, A., Pearce, W. (1979). Telephone sales manpower planning at quantas. Interfaces, 9(3), 1-9.

Ganguly, S., Lawrence, S., Prather, M. (2014). Emergency department staff planning to improve patient care and reduce costs. Decision Sciences, 45(1), 115-145.

Gierl, L., Pollwein, B., Heyde, G., Kurt, H. (1993). Knowledge-based scheduling of duty rosters for physicians. Informatics for Health and Social Care, 18(4), 355-366.

Puente, J., Gómez, A., Fernández, I., Priore, P. (2009). Medical doctor rostering problem in a hospital emergency department by means of genetic algorithms. Computers & Industrial Engineering, 56(4), 1232-1242.

Schelhase, T. (2014). Statistische Krankenhausdaten: Diagnosedaten der Krankenhäuser 2011. In J Klauber, M Geraedts, J Friedrich, J Wasem (Ed), Krankenhaus-Report 2014: Schwerpunkt: Patientensicherheit. Schattauer: Stuttgart, 329-356.

Rising, E.J., Baron, R., Averill, B. (1973). A system analysis of a University-Health-Service Outpatient clinic. Operations Research, 21(5), 1030-1047.

Rousseau, L., Gendreau, M., Pesant, G. (2002). A general approach to the physician rostering problems. Annals of Operations Research, 155(1-4), 193-205.

Stolletz, R., Brunner, J.O. (2011). Fair optimization of fortnightly physician schedules with flexible shifts. European Journal of Operational Research, 219(3), 622-629.

Van den Bergh, J., Beliën, J., De Bruecker, P., Demeulemeester, E., De Boeck, L. (2013). Personnel scheduling: A literature review. European Journal of Operational Research, 226(3), 367-385.

Venkat, A., Kekre, S., Hegde, G.G., Shang, J., Campbell, T.P. (2015). Strategic management of operations in the emergency department. Production and Operations Management, 24(11), 1706-1723.

White, C.A., White, G.M. (2003). Scheduling Doctors for Clinical Training Unit Rounds Using Tabu Optimization. In EK Burke, H Rudová (Ed), Practice and Theory of Automated Timetabling IV. Springer: Heidelberg, 120-128.

A3. Flexible staffing of physicians with column generation

Erhard, M. (2019). Flexible staffing of physicians with column generation. Flexible Services and Manufacturing Journal, in press.

The final authenticated version is available online at https://link.springer.com/article/10.1007%2Fs10696-019-09353-8

Status: Accepted for publication and published online in Flexible Services and Manufacturing Journal, category B.

Flexible staffing of physicians with column generation

Melanie Erhard (corresponding author)[a,b, *]

[a] Chair of Health Care Operations/Health Information Management

Faculty of Business and Economics

University of Augsburg

Universitätsstraße 16, 86159 Augsburg, Germany

[b] University Center of Health Sciences at Klinikum Augsburg (UNIKA-T)

Neusässer Straße 47, 86156 Augsburg, Germany

*melanie.erhard@unikat.uni-augsburg.de

Submission: July 23, 2018

Flexible staffing of physicians with column generation

Abstract

In Germany, around 40% of the hospitals do not generate an annual surplus. This leads to an increasing pressure on hospital's management to reorganize and restructure their processes and resources to decrease the upcoming costs and become profitable. Since personnel, especially physicians, generates a major part of the arising costs, assigning staff efficiently provides an opportunity to decrease associated expenses. Up to now, experienced physicians create rosters manually which is cost and time intense due to the problem's complexity and especially the fluctuation in demand. To circumvent this difficulty, it is our main aim to create a new mathematical modeling approach to implement additional flexibility in the rostering process to better match supply and demand.

Therefore, we formulate the problem as mixed-integer programming model with the objective to minimize occurring labor costs of physicians over the considered planning horizon subject to coverage of demand to make flexibility monetarily evaluable. In our approach, full flexibility in terms of patterns of working days, shift types, and the placement of the break is provided. To solve the problem under consideration, a column generation heuristic is presented. In our experimental study, the performance of the provided solution approach as well as the effect of additional flexibility in the rostering process are evaluated using real life data. Results indicate the significant impact of implementing flexibility in the scheduling process on the salary costs of the number of required physicians and evidence the superior quality of our solution approach.

Keywords: Physicians, Column Generation, flexibility, hospital, mixed-integer programming

1. Introduction

Within the last decades, hospitals face an increasing pressure to become profitable. Almost one third of Germany's hospitals generated a significant loss in 2016. Solely 61% of the hospitals negotiated the year with a surplus. The remaining 10% account a more or less balanced budget (Blum et al. 2017). Since this proportion did only change marginally over the last ten years and this ongoing negative trend will not stop in near future, the situation is severe (Blum et al. 2007). As a result, hospital's management is forced to counteract the cost pressure it is confronted with, e.g. by reducing upcoming costs. Since more than half of the operating costs are generated by the workforce (especially physicians), this seems to be a leverage point (Statistisches Bundesamt 2016). But, as the number of patients for hospitals is increasing at the same time (Bölt 2014), it is not sufficient to just reduce the number of personnel to save money since this might raise negative consequences for patient care and the quality of the provided service.

Therefore, it is necessary to reduce personnel costs by planning and scheduling staff efficiently to potentially decrease unplanned overtime hours and corresponding expenses (Thungjaroenkul et al. 2007). Up to now, an experienced physician creates schedules commonly manually which is a quite complex task: There are various rules that have to be taken into account, e.g. governmental workforce regulations (Ernst et al. 2004), individual agreements between physicians and hospitals (Brunner et al. 2011), and the length of the schedule as well as personnel requests (Damci-Kurt et al. 2017). Apart from this, it is the main aim to ensure an appropriate number of physicians scheduled to cover demand in each period of every day. In hospitals, demand is liable to heavy fluctuations occurring every day and every hour of the year, which makes it hard to predict in advance. Especially due to peaks in the required level of care, a large number of physicians has to be on duty to handle those (Aggarwal 1982). Since such peaks commonly have a duration of only a few hours, hospitals are overstaffed after its declination. This leads to a lot of idle time during the remaining hours of the day and raises the necessity to construct lines of work for employees which better match supply and demand (Brunner and Edenharter 2011). A possible approach to implement more flexibility in the scheduling process is to apply a greater number of shift types having different starting periods and variable lengths instead of solely using predefined eight hour shifts having fixed starting times (Isken 2004). Moreover, additional flexibility can be investigated by employing part time physicians having a reduced number of weekly working hours and can specifically be assigned to cover (expected) peaks in demand (Van den Bergh 2013).

The purpose of this paper is to evaluate flexibility in the physician scheduling process. Our contribution is manifold: We develop a mathematical model that provides full flexibility when scheduling physicians with respect to the patterns of sequences of working days, starting and ending times of shifts, and the placement of the break for each physician. Our main objective is to create a (near) optimal schedule that minimizes the weekly labor costs of physicians assigned to different lines of work within the planning horizon. At the same time, the forecasted demand is to be covered and breaks are to be assigned. The problem is formulated as mixed-integer programming model (MIP) with the objective to minimize weekly salary costs for personnel and solved applying a column generation (CG) heuristic approach to generate specific lines of work for each individual physician. The resulting solution provides a cost minimum staffing policy (when being solved to optimality), or at least a lower bound for estimating the resulting total labor costs. Furthermore, we examine the effect of additional flexibility in the scheduling process on the overall salary expenses.

The paper is structured as follows: In the subsequent section, current literature and research in the field of physician scheduling is reviewed. In section 3, the underlying problem structure is discussed briefly and the stated mathematical model is presented. Afterwards, we present a CG heuristic as appropriate solution approach due to the problem size and complexity (see section 4). Our experimental study is presented in section 5, where the effect of implementing a high degree of flexibility and the assignment of breaks is analyzed. Moreover, we evaluate the performance of our solution approach. We show the positive effect of flexibility in the scheduling process as well as the high quality of the resulting planning decisions. Eventually, the main findings and insights are summarized as the paper concludes by identifying some ideas for further research.

2. Overview on current literature

In contrast to the nurse scheduling problem, physician scheduling is still backward in research due to the complexity of the problem reasoned by the multitude of regulations and individual agreements that have to be taken into account. This makes the physician scheduling problem less generalizable and highly constraint. Even though, research has amplified significantly from 1985 up to now, i.e. from one publication in 1985 to nine papers in 2016. In the following, the relevant literature and developments in this field of research are reviewed. For a more detailed bibliography, see Erhard et al. (2018).

In current literature, flexibility in terms of days on and days off duty are mostly considered when scheduling residents, night and weekend shifts, and on-call shifts for residents and/or physicians. Beliën and Demeulemeester (2006) focus on scheduling trainees in a university hospital in Belgium. In their approach, hard constraints, such as labor rules, and soft constraints, e.g. duty preferences in terms of working days, are taken into account. Solving the problem using a Branch and Price (B&P) approach results in an optimal schedule of high quality with respect to fulfilled preferences, computational effort, and the length of the planning horizon. A one-year on-call schedule for residents using five different types of shifts is created in Cohn et al. (2009). Combining professional's expertise with a heuristic solution algorithm leads to an improved schedule which is generated in significantly reduced amount of computation time. A long term planning horizon is also considered in Brunner and Edenharter (2011): To determine the optimal number of personnel required over one entire year, a mixed-integer programming model is stated and solved applying a column generation heuristic. Since the number of available shifts is generated implicitly by the mathematical formulation, almost a maximum level of flexibility in shift types is provided.

Implicitly generated flexible shifts are defined by their starting and ending period having various different allowed shift lengths. These are also in use in Brunner et al. (2009): To minimize labor costs, a mixed-integer program is formulated and solved using a decomposition heuristic dividing the planning horizon in weekly subproblems. As a result, schedules of high quality reducing staffing costs are created in short amount of time. This research is developed further by Brunner et al. (2010) to analyze the effect of applying different branching strategies within the proposed B&P approach. Focusing on a similar objective function, it is the aim of Stolletz and Brunner

(2012) to minimize paid out costs and maximize fairness for physicians. In their approach, flexible shifts are also in use. In contrast to Brunner et al. (2009), shift types were generated by a shift matrix in a preprocessing step and serve as input for the stated reduced set covering formulation. Applying a heuristic which decomposes the considered problem by week leads to a high quality schedule violating significantly less soft restrictions in an appropriate solution time. Investigating uncertainty in terms of patient arrivals in their problem formulation, Ganguly et al. (2014) consider three different emergency departments. Scheduling physicians with different skill levels, exact results of their MIP formulation indicate potential for balanced staffing costs and levels of patients to improve staffing policies. Also considering an emergency department, it is the major aim of El-Rifai et al. (2014) to minimize the waiting time of patients while providing a fair distribution of shifts for physicians at the same time. Their stochastic mixed-integer programming formulation is solved exact and the results are subsequently evaluated using a simulation software. Results indicate potential improvement and differences for the considered staffing strategies.

In contrast, the assignment of breaks is almost neglected in current literature. Even though the positive effect of an appropriate number of break periods within a working day on the performance of personnel is well-known (Janaro and Bechtold 1985), recent research hardly deals with it. In general staff scheduling literature, Bechtold and Jacobs (1990) were one of the first to investigate the assignment of breaks to increase utilization of personnel. Therefore, a linear program (LP) is formulated to implicitly assign rest periods flexibly. Comparing their approach to an explicit set covering formulation results in improved performance with respect to solution time and the number of generated integer solutions, even for large test instances. Thompson (1995) also assigns shifts and breaks implicitly. Extending the formulation of Bechtold and Jacobs (1990) by a larger number of available shift types and the consideration of overtime hours, optimal solutions are created in a shorter amount of time without violating operational constraints on the placement of the break. Adopting the set covering formulation developed by Dantzig (1954), Aykin (1996) stated an integer programming (IP) model to assign multiple rest and lunch breaks to break windows. Considering a predetermined small number of shifts, every feasible combination for a break to shift assignment is defined. Evaluating their modeling approach for five different demand scenarios leads to the insight that this is a suitable approach to solve larger instances to optimality. Further extension is conducted in Rekik et al. (2010) by considering fractional breaks. When focusing on physician scheduling literature, the situation is even severe: Up to now, solely four papers take breaks into account. Brunner et al. (2009, 2010) propose an implicit approach for the assignment of breaks to physicians whereas Stolletz and Brunner (2012) assign breaks explicitly by a predetermined shift matrix. Comparing implicit and explicit modeling formulation leads to the insight, that an explicit reduced set covering formulation outperforms the implicit modeling approach with respect to solution times and the number of required constraints and variables. A combined approach is in use in Erhard and Brunner (2018): Considering predefined flexible overlapping shifts in a reduced set covering problem formulation, breaks are assigned implicitly by an extended formulation of Bechtold and Jacobs (1990). Evaluating the effect of taking breaks in the scheduling process into account leads to the insight, that those are especially important even on a high hierarchical level. Not considering an appropriate number of breaks in staffing level decisions therefore leads to a significant underestimation of the size of the workforce which is required to cover demand. This may lead to a lot of overtime hours and dissatisfaction of physicians, especially under real world circumstances where rather low flexibility in shift types is provided. This effect decreases the more flexibility is implemented in the scheduling process.

To the best of our knowledge, there exists no approach in literature that considers all features of flexibility with respect to the patterns of sequences of working days, starting and ending times of shifts, and the placement of flexible breaks. With our research we close the gap.

3. Problem description and statement of the mathematical model

The problem under consideration is to minimize weekly salary costs of the required workforce to cover demand for a given planning horizon consisting of $|\boldsymbol{W}|$ weeks, each of them comprising of a set of $\boldsymbol{D}$ days. Each of these days d spans $|\boldsymbol{P}|$ periods, e.g. stated in 1-hour increments. To create individual lines of work, a set of physicians $\boldsymbol{I}$ is available to be assigned to various shift types $s \in \boldsymbol{S}$. Each physician i can be modeled with his specific characteristics and individual restrictions that are either stated in the labor contract or base on special agreements between the physicians and the hospital they work for. In the following, we consider a homogeneous group of physicians, i.e. all parameters are introduced without index i. However, we could also model each physician by individual characteristics, e.g. part and full time physicians.

Flexibility in terms of working patterns is implemented by the minimum number of consecutive days on duty $\underline{D}^{on}$ as well as the maximum number of consecutive days on duty $\overline{D}^{on}$. Between each consecutive sequence of days on duty for each physician i, a minimum number of days off duty $\underline{D}^{off}$ has to be ensured.

Flexibility in shifts is provided by the various available shift types $s \in \boldsymbol{S}$. These are determined by specific characteristics, such as the minimum and maximum allowed shift length, i.e. S^{min} and S^{max}. In our case, this results in a shift stretch of seven to twelve hours as a maximum value since we define the length of a period to correspond to one hour. The available set of shifts $\boldsymbol{S}$ is predetermined by a shift matrix which serves as input for the mathematical model by the binary parameter A_{sp}. This parameter indicates if a period p is a working period for a specific shift type s by being equal to 1. Otherwise, the parameter takes 0 as its value for the remaining periods. The binary decision variable x_{isd} ensures the assignment of a specific shift to an individual physician for each day on duty. The resulting shift schedule for each physician i is created in a way that does not exceed a maximum number of weekly working hours $\overline{R}$ which is determined by labor regulations. Moreover, not any two shifts are allowed to be assigned consecutively for each physician i since a minimum number of rest periods P^{rest} between two shift assignments has to be ensured. Note, we are considering a discontinuous planning problem since shifts are not allowed to spill over from one to the subsequent planning day.

Flexibility in the placement of breaks within shifts is provided by the integrated break window for the individual assignment of a break in a period p for each physician i working shift s on day d. The size of the break window is determined by the length of the assigned shift and the minimum number of working periods after a shift has started and before the shift ends, i.e. B^{pre} and B^{post}. The resulting lines of work are assigned in a way that ensures demand N_{dp} to be covered in every period p of every day d in the planning horizon.

An example of a line of work for one single physician is provided in Figure 1. We consider a planning horizon of five working weeks Monday to Sunday. Each row denotes a planning period within a (planning) day, i.e. we consider 20 planning periods p. Columns represent the flexible shift assignments for the physician on any day. A 1 denotes a working period. A 0 within a sequence of 1s indicates a break assignment. For instance, the physician working this line of work has a break in period 4 on Monday in the first week. The schedule shows that sequences of working days vary in the first day on duty and the number of consecutive working days, i.e. ranging from four to seven consecutive days on duty. Moreover, shifts start at various planning periods and have different lengths. Each shift has an appropriate break which is flexibly assigned, e.g. in period four on Monday in the first week and in period fourteen on Thursday in the subsequent week. The first positive demand is normalized to period 1.

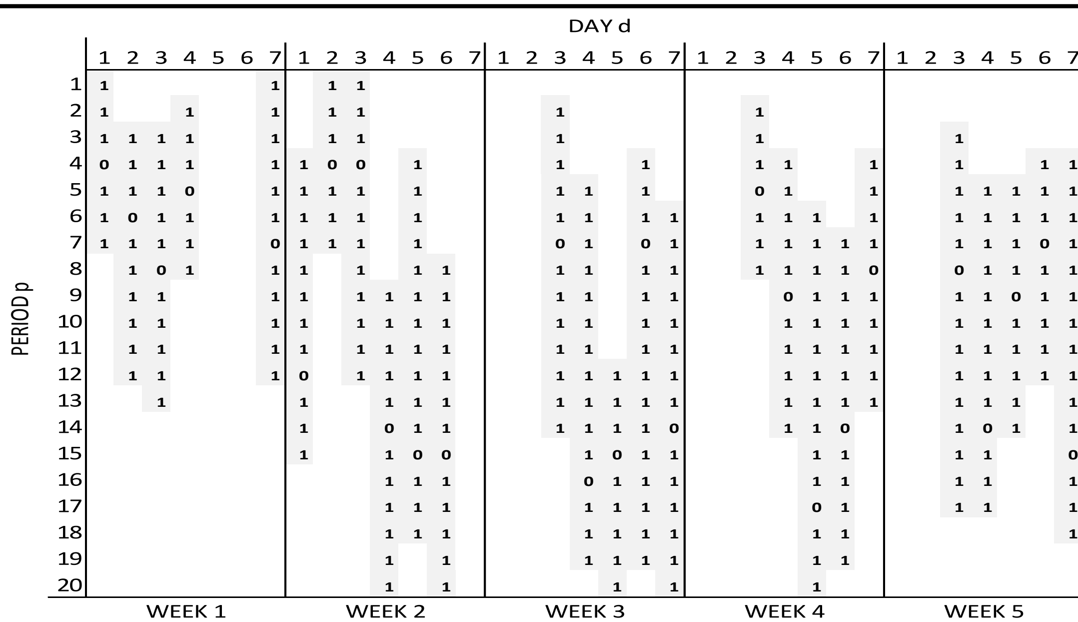

Figure 1: Example for possible line of work

In the following, we formulate the problem as mixed-integer programming model. In our formulation, we provide a maximum level of flexibility in terms of working days, shift types, and the assignment of breaks. Since we focus on the evaluation and comparison of our mathematical modeling and solution approach, our objective is the minimization of the total weekly salary costs subject to coverage of demand. Appraising the required size of the workforce by occurring labor costs within the planning horizon assigns a monetary value to scheduling flexibility. This makes our results more comparable.

Therefore, the following binary decision variable is defined, i.e.:

$$Y_i = \begin{cases} 1, if\ physician\ i\ is\ on\ duty\ within\ the\ planning\ horizon \\ 0, otherwise. \end{cases}$$

On a daily basis, we have to decide if a physician works or is off duty. We use a binary variable y_{id} defined as:

$$y_{id} = \begin{cases} 1, if\ physician\ i\ works\ on\ day\ d \\ 0, otherwise. \end{cases}$$

To assure feasible schedules, we define another binary decision variable z_{id} that determines the start of a sequence of working days, i.e.

$$z_{id} = \begin{cases} 1, if\ physician\ i\ starts\ a\ sequence\ of\ working\ days\ on\ day\ d \\ 0, otherwise. \end{cases}$$

On each working day, the physician is assigned to a shift. We model this by the following binary variables:

$$x_{isd} = \begin{cases} 1, if\ physician\ i\ is\ assigned\ to\ shift\ s\ on\ day\ d \\ 0, otherwise. \end{cases}$$

Moreover, each physician being assigned to a shift also gets a break assigned which is modeled by:

$$b_{ispd} = \begin{cases} 1, if\ physician\ i\ is\ assigned\ to\ a\ break\ in\ period\ p\ when\ working\ shift\ s\ on\ day\ d \\ 0, otherwise. \end{cases}$$

Now we introduce the essential notation and present the basic model.

Sets with indices

$i \in \boldsymbol{I}$	Set of physicians with index i
$w \in \boldsymbol{W}$	Set of weeks with index w
$d \in \boldsymbol{D}$	Set of days with index d
$d \in \boldsymbol{D}_w$	Subset of days, i.e. Monday to Sunday, within week w with D^{week} as the number of days per week, i.e. 7 days, and $\boldsymbol{D}_w = \{(w-1) \cdot D^{week} + 1, \dots, w \cdot D^{week}\}$
$p \in \boldsymbol{P}$	Set of day-periods with index p
$s \in \boldsymbol{S}$	Set of shifts with index s

Parameters

c^{plan}	Salary costs of a physician (depending on the length of the planning horizon)
S^{min}	Minimum shift length
S^{max}	Maximum shift length
A_{sp}	1 if shift s covers period p, 0 otherwise
F_s	First working period in shift s
L_s	Last working period in shift s
W_s	Number of working periods in shift s
B^{pre}	Amount of working periods before the break is allowed to start
B^{post}	Amount of working periods after the break has ended
$\underline{D}^{on}$	Minimum number of consecutive days on
$\overline{D}^{on}$	Maximum number of consecutive days on duty
$\underline{D}^{off}$	Minimum number of days off between consecutive sequences of working days on duty
P^{rest}	Minimum number of rest periods between two consecutive shifts
$\overline{R}$	Maximum amount of regular working periods for physician i per week
N_{dp}	Demand in period p of day d physician i

Binary decision variables

Y_i	1 if physician i works in the planning horizon, 0 otherwise
y_{id}	1 if physician i is on duty on day d, 0 otherwise
z_{id}	1 if the sequence of working days of physician i begins on day d, 0 otherwise
x_{isd}	1 if physician i is assigned to shift s on day d, 0 otherwise
b_{isdp}	1 if physician i working shift s in period p of day d is assigned a break, 0 otherwise

$$\text{Minimize} \sum_{i \in I} c^{plan} \cdot Y_i \tag{1.1}$$

subject to

$$\sum_{s \in S} \sum_{i \in I} (A_{sp} x_{isd} - b_{isdp}) \geq N_{dp} \qquad \forall d \in \boldsymbol{D}, p \in \boldsymbol{P} \tag{1.2}$$

$$y_{id} \leq Y_i \qquad \forall i \in \boldsymbol{I}, d \in \boldsymbol{D} \tag{1.3}$$

$$z_{id} = y_{id}(1 - y_{i(d-1)}) \qquad \forall i \in \boldsymbol{I}, d \in \boldsymbol{D} \tag{1.4}$$

$$\sum_{t=d}^{d+\overline{D}^{on}} y_{it} \leq \overline{D}^{on} \qquad \forall i \in \boldsymbol{I}, d \in \boldsymbol{D} \tag{1.5}$$

$$z_{id} \leq y_{it} \qquad \forall i \in \boldsymbol{I}, d \in \left\{1, \dots, |\boldsymbol{D}| - \underline{D}^{on}\right\}, t \in \left\{d, \dots, d + \underline{D}^{on} - 1\right\} \tag{1.6}$$

$$z_{id} \leq 1 - y_{it} \qquad \forall i \in \boldsymbol{I}, d \in \left\{1 + \underline{D}^{off}, \dots, |\boldsymbol{D}| - \underline{D}^{on} + 1\right\}, t \in \left\{d - \underline{D}^{off}, \dots, d - 1\right\} \tag{1.7}$$

$$\sum_{s \in \boldsymbol{S}} x_{isd} = y_{id} \qquad \forall i \in \boldsymbol{I}, d \in \boldsymbol{D} \tag{1.8}$$

$$x_{is_1 d} + x_{is_2(d+1)} \leq 1 \qquad \forall i \in \boldsymbol{I},\ s_1, s_2 \in \boldsymbol{S}: |P| - L_{s_1} + F_{s_2} - 1 < P^{rest}, d \in \boldsymbol{D} \backslash |\boldsymbol{D}| \tag{1.9}$$

$$\sum_{s \in S} \sum_{d \in D_w} W_s x_{isd} \leq \overline{R} \qquad \forall i \in \boldsymbol{I}, w \in \boldsymbol{W} \tag{1.10}$$

$$\sum_{p=F_s+B^{pre}}^{L_s-B^{post}} b_{isdp} = x_{isd} \qquad \forall i \in \boldsymbol{I}, s \in \boldsymbol{S}, d \in \boldsymbol{D} \tag{1.11}$$

$$Y_i, y_{id}, z_{id}, x_{isd}, b_{isdp} \in \{0,1\} \qquad \forall i \in \boldsymbol{I}, s \in \boldsymbol{S}, d \in \boldsymbol{D}, p \in \boldsymbol{P} \tag{1.12}$$

The objective function (1.1) minimizes the total weekly salary costs of the number of employed physicians.

Flexible days. The first block of constraints (1.2) – (1.7) models flexible working days. Constraints (1.2) take care about demand being covered in every period p of every day d within the planning horizon. It is therefore necessary to reduce the number of assigned physicians in each period p by the number of employees having their break in this specific period of the day. As a result, the number of assigned physicians is forced to be greater than or equal to the level of demand in each period of every day, even though some physicians are assigned to a break in this period. Moreover, constraints (1.3) indicate if physician i has at least one day on duty during the whole planning horizon, i.e. $\forall d \in \boldsymbol{D}, y_{id} = 1 \rightarrow Y_i = 1$. Constraints (1.4) determine the start of a sequence of consecutive working days (i.e. $z_{id} = 1$) by linking variables y_{id} with z_{id}. In other words, if there is a switch from 0 to 1 in the y-variables then the corresponding z-variable is set to 1. Those constraints can be easily linearized (see Appendix I). The next constraints (1.5) force a maximum number of consecutive working days $\overline{D}^{on}$. In other words, in any sequence of $\left(\overline{D}^{on} + 1\right)$ consecutive y-variables at least one of those must be 0; otherwise the constraints are not fulfilled. Next, we enforce a minimum number of $\underline{D}^{on}$ consecutive working days for each valid schedule. If a sequence of working days starts on day d, i.e. $z_{id} = 1$, then constraints (1.6) assure that the y-variables for the days from d to $d + \underline{D}^{on}$ are bound from below by 1. Mathematically speaking, the constraints ensure the following: $\forall i \in \boldsymbol{I}, d \in \boldsymbol{D}, z_{id} = 1 \rightarrow y_{it} = 1$ for $t \in \left\{d, \dots, d + \underline{D}^{on} - 1\right\}$. Ensuring a minimum number of days off after each sequence of days on, we use a similar modeling idea (see constraints 1.6). To be able to utilize the z-variables, we enforce the appropriate number of off days before a sequence of days on starts; otherwise we would need additional variables indicating the end of a working sequence. In particular, constraints (1.7) assure a minimum number of $\underline{D}^{off}$ days off between two consecutive sequences of working days. As before, if a sequence of working days starts on day d, i.e. $z_{id} = 1$, then constraints (1.7) model that the y-variables for the days from $d - \underline{D}^{off}$ to $d - 1$ are forced to 0, otherwise the constraints would not be valid, i.e. $\forall i \in \boldsymbol{I}, d \in \boldsymbol{D}, z_{id} = 1 \rightarrow y_{it} = 0$ for $t \in \left\{d - \underline{D}^{off}, \dots, d - 1\right\}$. If one of the y_{it} variables before a working sequence starts is equal to 1, the constraints (1.7) are not fulfilled since the right hand side would be 0 as well but the left hand side would be 1 due to $z_{id} = 1$. This is a contradiction.

Flexible shifts. After the conditions of the various sequences of working days have been regulated, the allowed distribution of shift types to each individual physician i is to be considered. The second set of constraints (1.8) – (1.9) handles the flexible assignment of shifts: Constraints (1.8) ensure that each physician being on duty on a specific day d gets exactly one shift s assigned. If a day is not a working day for a physician, i.e. $y_{id} = 0 \rightarrow x_{ids} = 0$, $\forall i \in \boldsymbol{I}, d \in \boldsymbol{D}, s \in \boldsymbol{S}$. To ensure a minimum number of rest periods P^{rest} between two consecutive shift assignments, constraints (1.9) do not allow some specific combinations of shift assignments on day d and $d + 1$ where off periods are less than P^{rest}. Not any number of working hours are permitted within a week for each physician due to labor regulations and individual agreements of hospitals and physicians. Therefore, constraints (1.10) ensure a maximum number of working hours within a week for each physician. Depending on the number of assigned shifts and corresponding working hours for each shift W_s, constraints (1.10) sum up all duties within one week w for each physician i and force those not to exceed a maximum value $\overline{R}$.

Flexible breaks. Constraints (1.11) ensure exactly one break assignment for each physician on duty. In particular, the break is to be placed within a predetermined break window defined by the minimum number of working hours before a break is allowed to be placed B^{pre} and the minimum number of working periods after a break B^{post} individually for each assigned shift type s. Eventually, variable definitions are given in (1.12).

4. Solution approach

Since standard software is not able to find a feasible solution for the majority of parameter settings, planning horizons of more than one week, and demand scenarios except for the mean demand level, a CG heuristic is applied to find good integer solutions in reasonable runtime (Desaulniers et al. 2005). As can be seen easily, our model decomposes by physician. Hence, an individual line of work covering the entire planning horizon is constructed, i.e. specifying sequences of working days, days off, the assigned shift type for each day on duty, and the period of the break within the assigned shift.

Due to the specific block structure of our problem, Dantzig-Wolfe decomposition is applied (Dantzig and Wolfe 1960; Desrosiers and Lübbecke 2005). The general procedure of the solution algorithm is shown in Figure 2.

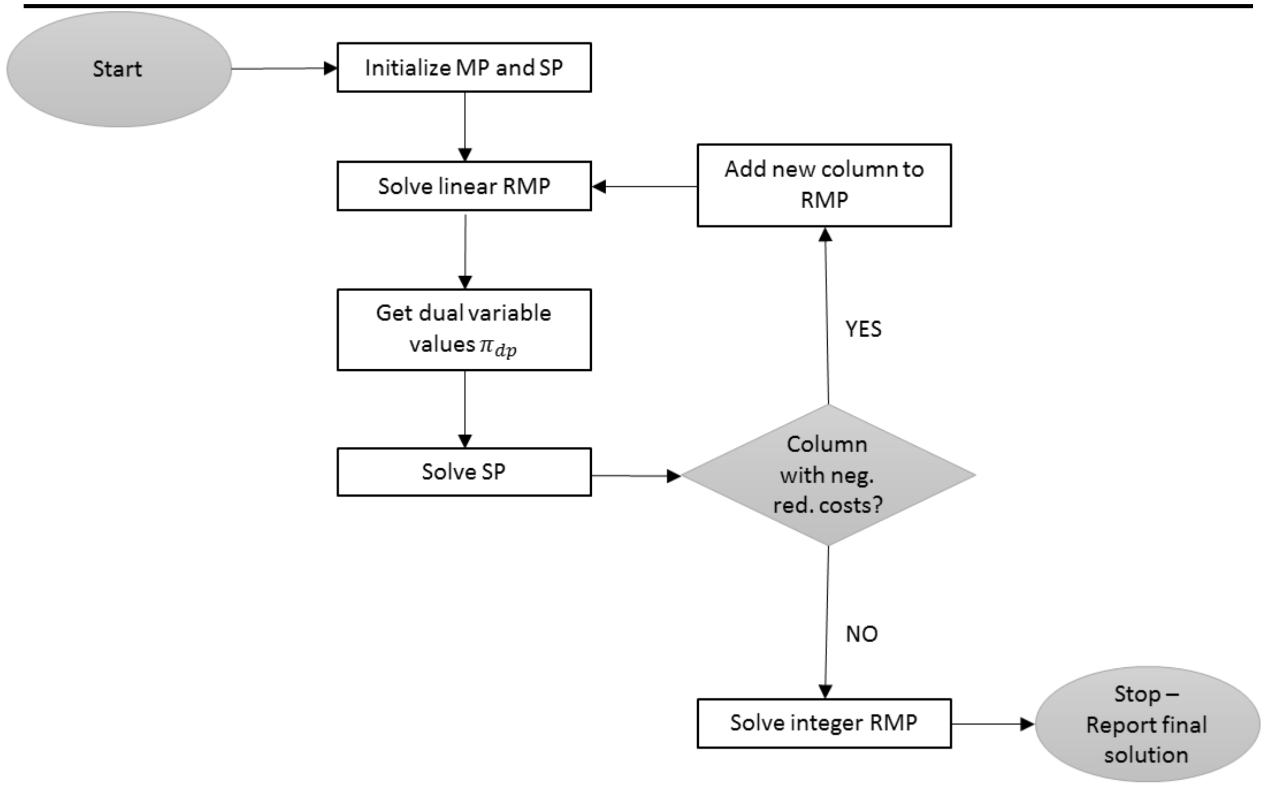

Figure 2: Flowchart column generation heuristic

Therefore, we decompose our initial model and formulate a Master Problem (MP) and a Subproblem (SP) for each physician (Dantzig and Wolfe 1960). Since we solve the extended version using

CG, we have to relax the integrality constraints for our linear MP (MP-LP). Considering only a subset of feasible columns, the resulting restricted MP (RMP) does not contain the integrity of all feasible solutions, solely the specific ones which are generated by solving the SP. This leads to a significant reduction in solution times (Lübbecke and Desrosiers 2005). Consequently, two separate optimization problems are formulated and solved iteratively during the process of CG.

The SP in this coherence serves as a generator of new columns (variables) to enter the basis (Barnhart et al. 1998). Here, the physicians are aggregated as the SP creates feasible lines of work that are to be assigned to a number of physicians which results in an additional reduction in symmetry. After solving the linear relaxation of the RMP, the resulting dual variable values (π_{dp}) of the demand constraints (2.2) are consigned to the SP which is subsequently solved to optimality as integer program (IP) to determine a new promising column (Desrosiers and Lübbecke 2005). Generated columns in this context are evaluated according to their reduced cost which are to be negative in our minimization problem to generate a positive effect on the objective function value of the RMP. Since our SP is solved to optimality as well, we are able to identify the one column that generated the largest improving (in our case: minimizing) effect on the objective function (Brunner and Edenharter 2011). Afterwards, the new detected column is added the RMP which is solved again. The algorithm terminates if no column that prices out can be found.

As a result, the optimal solution for the linear RMP is simultaneously the optimal solution to the linear MP. But, up to now, the provided optimal solution might be continuous and provides therefore a lower bound. Hence, as last step of our CG heuristic, we solve the RMP as IP to provide a feasible integer solution, i.e. an upper bound. The thereby provided solution must not be optimal any longer. The resulting gap between the solution provided by CG which serves as lower bound (LB) and the objective function value provided by solving the RMP as IP (serving as upper bound (UB)) is small as will be shown soon.

Additional notation for MP and SP:

Sets with indices

$r \in \boldsymbol{R}$	Set of working rosters with index r

Parameters

X_{dp}^{r}	1 if roster r covers period p on day d, 0 otherwise

Integer decision variables

λ_r	Number of physicians working roster r in the planning horizon

Binary decision variables

y_d	1 if day d is a day on duty (working day), 0 otherwise
z_d	1 if the sequence of working days begins on day d, 0 otherwise
x_{sd}	1 if shift s is assigned on day d, 0 otherwise
b_{sdp}	1 if shift s gets a break assigned in period p on day d, 0 otherwise

Mathematical Modell – Master Problem

$$\text{Minimize} \sum_{r \in R} c^{plan} \lambda_r \tag{2.1}$$

subject to

$$\sum_{r \in R} X_{dp}^{r} \lambda_r \geq N_{dp} \qquad \forall d \in \boldsymbol{D}, p \in \boldsymbol{P} \qquad (\pi_{dp}) \tag{2.2}$$

$$\lambda_r \geq 0 \text{ and integer} \qquad \forall r \in \boldsymbol{R} \tag{2.3}$$

The objective function (2.1) minimizes the total weekly labor costs of personnel being assigned to a roster over the planning horizon.

Constraints (2.2) ensure that the number of assigned physicians in every period p is as least as high as the required level of care (or higher) such that the arising demand is covered throughout the entire planning horizon. The last constraints (2.3) define the newly introduced decision variable λ_r which determines the number of physicians working a specific roster $r \in \boldsymbol{R}$**.** The dual variables $\pi_{dp} \geq 0$ derived from contraints (2.2) determine the generic reduced cost of a MP column as given in (3).

$$1 - \sum_{d \in \boldsymbol{D}} \sum_{p \in \boldsymbol{P}} \pi_{dp} \; X_{dp}^{r} \tag{3}$$

In SP notation, the reduced costs are stated in (4).

$$1 - \sum_{s \in \boldsymbol{S}} \sum_{d \in \boldsymbol{D}} \sum_{p \in \boldsymbol{P}} \pi_{dp} \cdot (A_{sp} x_{sd} - b_{sdp}) \tag{4}$$

In the following, we state our generic SP.

Mathematical Model – Subproblem

$$\text{Minimize } c^{plan} - \sum_{s \in \boldsymbol{S}} \sum_{d \in \boldsymbol{D}} \sum_{p \in \boldsymbol{P}} \pi_{dp} \cdot (A_{sp} x_{sd} - b_{sdp}) \tag{3.1}$$

subject to

$$z_d = y_d(1 - y_{(d-1)}) \qquad \forall d \in \boldsymbol{D} \tag{3.2}$$

$$\sum_{t=d}^{d+\overline{D}^{on}} y_t \leq \overline{D}^{on} \qquad \forall d \in \boldsymbol{D} \tag{3.3}$$

$$z_d \leq y_t \qquad \forall d \in \{1, \dots, |\boldsymbol{D}| - \underline{D}^{on}\}, t \in \{d, \dots, d + \underline{D}^{on} - 1\} \tag{3.4}$$

$$z_d \leq 1 - y_t \qquad \forall\, d \in \{1 + \underline{D}^{off}, \dots, |\boldsymbol{D}| - \underline{D}^{on} + 1\}, t \in \{d - \underline{D}^{off}, \dots, d - 1\} \tag{3.5}$$

$$\sum_{s\in \boldsymbol{S}} x_{sd} = y_d \qquad \forall d \in \boldsymbol{D} \tag{3.6}$$

$$x_{s_1 d} + x_{s_2 (d+1)} \leq 1 \qquad \forall\, s_1, s_2 \in \boldsymbol{S}\colon |P| - L_{s_1} + F_{s_2} - 1 < P^{rest}, d \in \boldsymbol{D} \backslash |\boldsymbol{D}| \tag{3.7}$$

$$\sum_{s\in S} \sum_{d \in D_w} W_s x_{sd} \leq \overline{R} \qquad \forall w \in \boldsymbol{W} \tag{3.8}$$

$$\sum_{p=F_s+B^{pre}}^{L_s-B^{post}} b_{sdp} = x_{sd} \qquad \forall s \in \boldsymbol{S}, d \in \boldsymbol{D} \tag{3.9}$$

$$z_d, y_d, x_{sd}, b_{sdp} \in \{0,1\} \qquad \forall s \in \boldsymbol{S}, d \in \boldsymbol{D}, p \in \boldsymbol{P} \tag{3.10}$$

Searching for a new promising column in each iteration, dual values π_{dp} derived from the demand constraints (2.2) are required. Therefore, the objective function of the corresponding SP (3.1) is to minimize the reduced cost of a new column.

As a result, the subproblem generates specific lines of work for generic physicians, including an individual sequence of working days, the assignment of a shift for each day on duty, and the flexible assignment of a break during this shift. Based on the concept of column generation, only columns having negative reduced costs are added to RMP to improve the objective function value, i.e. revealing an objective function value of the SP smaller than 0. The remaining constraints correspond to constraints (1.4) to (1.11) which are already discussed in detail in former sections. Therefore, these are not explained any further. Note, we have dropped index i. Constraints (3.10) define the decision variables.

5. Experimental study

In this section, the performance of our heuristic solution approach in comparison to an exact solution (where disposable) is evaluated by using real life data, i.e. aggregated demand patterns. Therefore, the compact formulation of our model as well as the solution approach are implemented in IBM ILOG OPL Studio 7.0 and CPLEX 12.0. All computations were executed on a 2.60GHz Intel(R) Xeon(R) CPU E5-2650 v2 Machine with 8 GB RAM running under the Windows 10 Enterprise operating system. The SP is solved to optimality for each test instance while at the end of the algorithm a time limit is set to 600 seconds when solving the RMP as IP.

2.1. Underlying demand pattern

In our experimental study, we use data derived from a large teaching hospital in Germany. In more detail, the occurring demand of an operating theater over one entire year is analyzed. We aggregate the demand (from Monday to Sunday) over one year and build three different demand scenarios, i.e. mean demand pattern (50% quantile), 75% quantile, and maximum demand (100% quantile). Figure 3, 4, and 5 exhibit two different bounds for approximating the required workforce size in addition to the structure of the considered demand scenarios. The workload bound (LB^{work}) is determined by the maximum amount of allowed working periods per physician per week and the maximum demand bound (LB^{max}) which is defined by the overall peak in demand within the planning horizon (for more detail we refer to Brunner and Edenharter 2011). Note, for each test instance, the required workforce size provided by the workload bound is larger than provided by the maximum demand bound which might lead to a significant increase in solution times. According to Brunner and Edenharter (2011), a personnel scheduling problem becomes NP-hard, if $LB^{work} \geq LB^{peak}$. This means, LB^{work} is the binding bound for the overall workforce size due to the rather low level of the peak in demand (for a detailed proof see Brunner and Edenharter 2011). We include requirements for care on weekend days in our data to be able to evaluate our theoretical approach especially with respect to flexibility in sequences of working days, even though, this specific demand is covered by on-call shifts in practice.

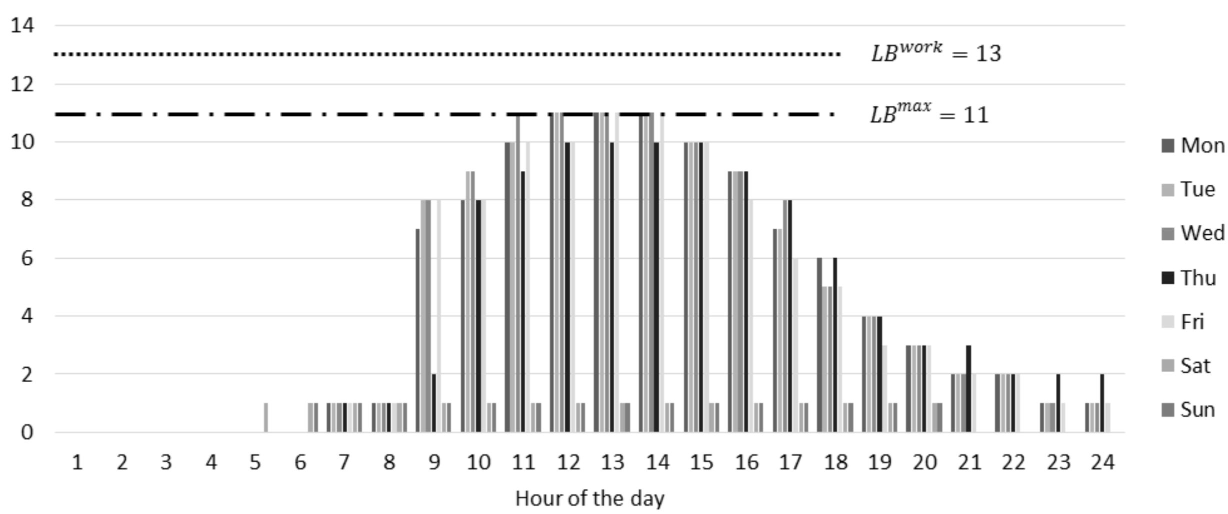

Figure 3: Real life demand – Mean (50% quantile) demand scenario

Decomposing annual demand in one week planning periods (Monday to Sunday), demand of each hour of every day is aggregated. As can be seen easily, each day consists of 24 1-hour periods. In general, demand starts occurring at 5 am and lasts until 12 am, as we consider a stretch of 20 periods for the upcoming demand. At the beginning of the day, the requirement for care is rather low, but increases significantly at 9 am to its maximum value around noon. During this time of the day, demand is fairly constant, at least for four hours. Subsequently, the required level of care decreases slowly to its minimum value at 12 am. For the specific case of our mean demand scenario, there is barely no requirement for care in the first two hours, i.e. 5 am and 6 am. This is not the case for the 75% (quantile) and the maximum demand level (100% quantile). Additionally, the maximum demand varies between the considered scenarios, i.e. eleven for the mean demand scenario, twelve for 75% quantile demand scenario, and 15 for the maximum demand scenario (100% quantile) respectively.

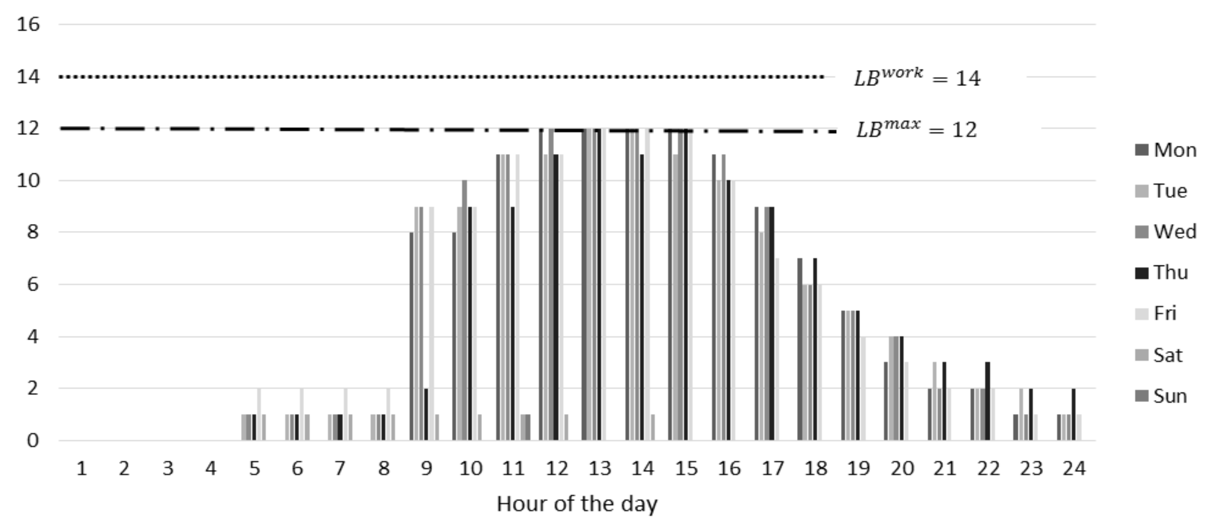

Figure 4: Real life demand –75% (quantile) demand scenario

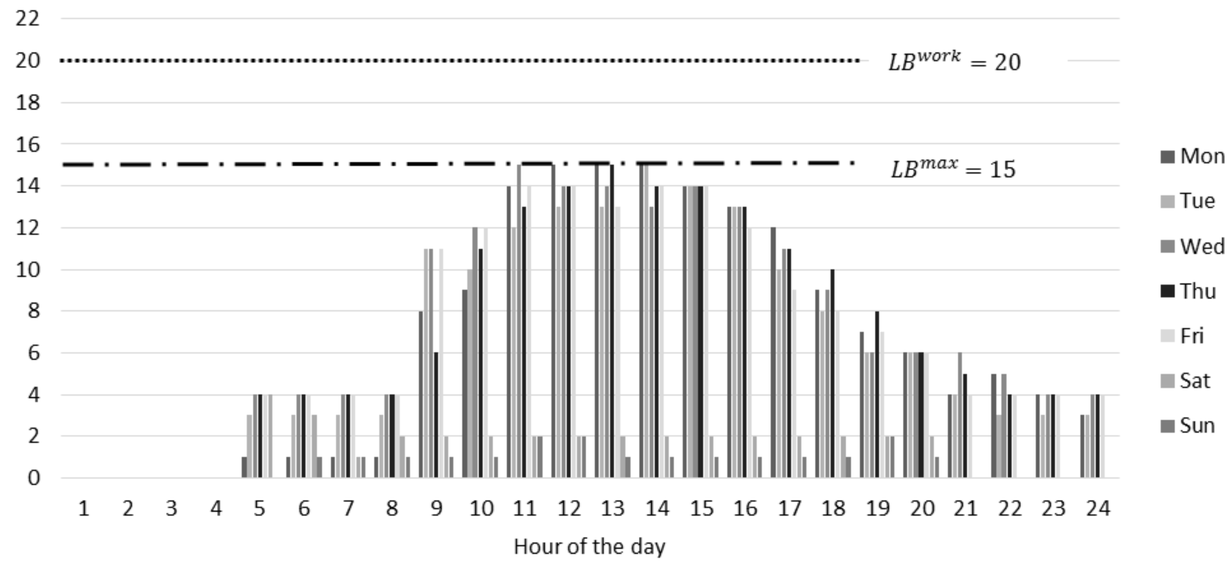

Figure 5: Real life demand – Maximum (100% quantile) demand scenario

Remaining hours were not taken into account when creating our mean demand level since demand during these hours is commonly handled by the assigned on-call physicians. As we are focusing on a discontinuous planning problem, shifts do not spill over to the next planning day. Our planning day is defined from 5 am to midnight.

To make our results comparable, we additionally define a standard parameter setting which is in use throughout the entire experimental study, except in subsection 5.3. where we conduct a factorial analysis to determine the effect of an increasing respectively decreasing level of flexibility in the scheduling process. For more detail, see Table 1.

Day parameter	$\underline{D}^{on} = 3$	$\overline{D}^{on} = 6$	$\underline{D}^{off} = 2$
Shift parameter	$S^{min} = 7$		$S^{max} = 12$
Break parameter	$B^{pre} = 3$		$B^{post} = 3$
Ergonomic/Legal regulated parameter	$\overline{R} = 45$	$P^{rest} = 11$	$c^{plan} = \|\boldsymbol{W}\| \cdot 1'625$

Table 1: Standard parameter setting

As a result, created lines of work have to assign at least three but at most six consecutive working days to a physician. Moreover, between two subsequent sequences of working days, a minimum of two days has to be off. Since the assignment of a break is not mandatory until a shift length of six hours, we define the minimum length for our shifts to be seven periods, i.e. seven hours. The maximum allowed shift length is twelve hours. The remaining shift dependent parameter values are determined implicitly by the defined values for S^{min} and S^{max}, such as the binary values for the shift matrix A_{sp}, the first and last working period for each shift (F_s and L_s respectively), and the number of on duty periods for each shift type W_s.

In our approach, only one type of break with a length of one hour is considered. But, this formulation can easily be modified and extended, e.g. by varying the length of the assigned break. The placement is determined according to the break parameters. In our standard parameter setting, a physician being assigned to a shift has to work at least three periods ($B^{pre} = 3$) until a break is assigned but no later than three hours before the shift ends ($B^{post} = 3$).

The subsequent set of parameters ensures a minimum number of rest periods between two consecutive shifts by defining $P^{rest} = 11$, as regulated by law as well as a maximum amount of weekly working hours for each physician, i.e. $\overline{R} = 45$. Moreover, the weekly average salary is assumed to be 1'625€ (Marburger Bund 2016) which are to be multiplied by the number of weeks in the planning horizon to determine the labor cost of one entire roster. Note, in our experimental study, we suppose a homogeneous group of physicians.

In general, when running the CG heuristic, a large computational effort is investigated to prove for optimality when solving MP-LP. As the so-called tailing-off effect leads to a significant growth in solution times (Gilmore and Gomory, 1963), a feasible lower bound for the MP-LP (LB-MP) is implemented to improve the algorithm's performance (Volland et al. 2017). To state LB-MP, we use information derived in each iteration of the solution process, i.e. the current objective function values of the MP-LP (z_i^{MP-LP}) and the SP (z_i^{SP}). Based on Lübbecke and Desrosiers (2005), the bound is stated in (5).

$$z_i^{LB-MP} = max\{z_{i-1}^{LB-MP};\ z_i^{LB-MP}\} \tag{5}$$

Calculating the lower bound in each iteration of the solution process, we initialize $z_0^{LB-MP} = -\infty$ and define $z_i^{LB-MP} = \left\lceil \frac{z_i^{MP-LP}}{1-z_i^{SP}} \right\rceil$. The algorithm terminates due to LB-MP if $z_i^{LB-MP} \geq z_i^{MP-LP}$, i.e. once the calculated bound is larger than or equal to the current MP objective function value, before finding the optimal MP-LP solution. This is in around one third of our test instances the case.

2.2. Evaluation of solution approach

In this subsection, the performance of our solution approach in comparison to the exact solution of our compact formulation is evaluated. To do so, the standard parameter setting as well as the three predefined demand scenarios are in use. First, various lengths for the planning horizon are analyzed with respect to solvability, solution times, and the quality of the provided solution. Second, a labor cost minimal schedule is determined for the different demand patterns by solving the compact formulation to optimality (if possible) and by applying the suggested CG heuristic.

The performance of the proposed solution approach is tested by implementing various lengths for the planning horizon, i.e. one, two, four, and six weeks. Here, the standard parameter setting as well as the mean demand pattern serve as input. For more detail, see Table 2.

	CG					
	Objektive MP-LP	Objective MP-IP	Optimality gap	# columns	Solution time: MP-LP	Solution time: MP-IP [sec.]
1 week roster	22'750	26'000	3'250	548	4.68 min.	600
2 week roster	45'500	55'250	9'750	1'730	49.76 min.	600
4 week roster	97'500	110'500	13'000	6'060	6.02 h	600
6 week roster	146'250	165'750	19'500	7'109	10.82 h	600

Table 2: Objective value three demand scenarios

Table 2 presents the solution values and algorithm performance measures for the considered CG approach. The first column represents the length of the planning period of the underlying demand pattern. Column two displays the optimal solution value for MP-LP whereas column three contains the integer solution for the CG heuristic (MP-IP). Column four evaluates the quality of the provided solution by calculating the absolute optimality gap, i.e. $opt.\,gap\ (absolute) = (MP - IP) - (MP - LP)$. The last three columns provide information about the number of required columns for the CG heuristic and solution times for MP-LP and MP-IP.

An exact solution derived by the compact formulation is only provided for an one week roster. The optimal solution of 24’375€ is provided in 2 hours and 15 minutes solution time. Planning periods of more than one week cannot be solved to optimality, i.e. even a feasible solution cannot be found within a time limit of three hours. Moreover, an optimal solution for the LP relaxation requires more than three hours for some test instances. With respect to solution times, run times providing an exact solution are more than ten times higher for the MIP compared to our CG heuristic (for the one week roster).

In contrast, the CG heuristic is able to find a good solution in acceptable amount of time, i.e. 13 minutes, 1 hour, 6 hours, and 10 hours for a six week planning period respectively. Note, solution times might be significantly lower for test instances with a high peak due to the dominance of LB^{peak} over LB^{work} for these specific test instances. Considering the objective function values, a salary cost minimal roster provided by the CG heuristic ranges from 26’000€ to 165’750€, i.e. corresponding to two respectively three additionally required physicians. But, the smaller value of 26’000€ is only appropriate when focusing on a planning horizon of one week which is rather short. Since the resulting minimum salary costs depend on the length of the planning period, the resulting objective value increases the longer the length of the roster.

Figure 6 displays the development of the implemented LB (stated in (5)) exemplary for the CG heuristic when generating a two-week roster:

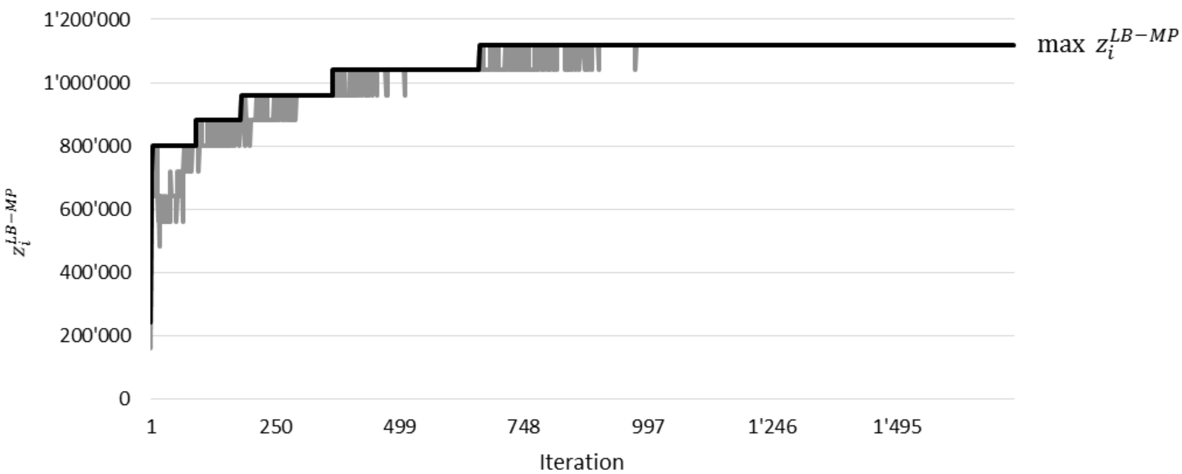

Figure 6: Real life demand – Maximum (100% quantile) demand scenario

The figure shows the lower bound z_i^{LB-MP} for each iteration when running the CG heuristic as well as the maximum value which is in use as early termination criterion during processing the algorithm. The value of the LB is rather small in the first iteration but increases significantly when generating the next column. Afterwards, the LB increases successively.

When solving the different demand scenarios (i.e. 50%, 75%, and 100% quantile) in the second step of the algorithmic evaluation, a planning horizon of two weeks is considered for each instance (see Table 3). Table 3 gives the solution and computational values for the CG heuristic. The first column represents the underlying scenario for the demand pattern. Again, column two to seven provide the objective function values, absolute optimality gap, number of iterations and solution times.

	CG					
	Objektive MP-LP	Objective MP-IP	Optimality gap	# columns	Solution time: MP-LP	Solution time: MP-IP [sec.]
50 % quantile	45'500	55'250	9'750	1'730	49.76 min.	600
75 % quantile	52'000	58'500	6'500	1'012	22.45 min.	600
100 % quantile	74'750	84'500	9'750	1'633	50.03 min.	600

Table 3: Objective value three demand scenarios for two weeks

Solving the formulated problem using our CG heuristic for mean demand, an objective function value of 45’500€ is proposed as MP-LP solution, whereas total labor costs of 55’250€ is proposed when forcing integer values (MP-IP).

Also for the remaining two demand patterns (75% and 100% quantile), a total salary for physicians of 58’500€ respectively 84’500€ are required. Both values are provided in quite an acceptable amount of time, even though computation time is much higher for the maximum demand level, i.e. 22.45 minutes compared to 50.03 minutes respectively.

Due to the complexity derived by the three different types of implemented flexibility, the compact formulation of our mathematical program is not solvable for the majority of parameter settings, a planning horizon of more than one week, and other demand pattern than mean level. Since our column generation approach performs quite satisfying with respect to solution quality and computational effort, it is further in use for the following steps of our experimental study.

2.3. Factorial analysis of flexibility parameters

In this subsection, we conduct a factorial analysis to test for the effect of the different types and degrees of flexibility in the staffing process on the total salary costs of the workforce. In our analysis, we subdivide our flexibility parameters into two discrete subsets according to the effect we focus on: day parameters and shift parameters.

In the following, each of these values is varied successively while keeping the remaining parameters constant. For the minimum number of consecutive working days, the minimum number of consecutive days off as well as the minimum shift length, the value of the specific parameter is increased iteratively. For the maximum number of consecutive days on duty and the maximum shift length, we vary the appropriate value in the opposite direction. This results in a total of 168 test instances whose coherence is stated in Table 4. Note, some of the tested parameter-combinations might not be applicable in real life to build a working roster, but for completeness of our factorial design, these are also conducted.

Parameter	$\underline{D}^{on}$	$\overline{D}^{on}$	$\underline{D}^{off}$	S^{min}	S^{max}	B^{pre}	B^{post}	
Standard parameter setting	3	6	2	7	12	3	3	168 test instances
Variation in day parameters	1 to 6	6 to 1	1 to 5	7	12	3	3	
Variation in shift parameters	3	6	2	7 to 12	12 to 7	3 to 1	3 to 1	

Table 4: Factorial Design

We solve each variation of the stated parameters applying our CG heuristic whereas the former stated mean demand scenario serves as input for the underlying requirement for care having a two-week planning horizon.

2.3.1. Variation of day parameters

For each parameter setting of $\underline{D}^{on}$ and $\overline{D}^{on}$, the minimum number of days off is additionally increased iteratively from one to five to investigate the effect additional flexibility in terms of daily working patterns on total workforce size. In each step, the maximum number of consecutive days on duty is decreased from the initial value of six to the current value of $\underline{D}^{on}$, until $\overline{D}^{on} = \underline{D}^{on}$. This results in a total of 105 different settings. A detailed overview on the objective function values, absolute optimality gap, number of columns, and computational effort is provided in Table 5 in Appendix II. Again, $\underline{D}^{on}$ is increased iteratively whereas the parameter value of $\overline{D}^{on}$ is decreased.

In general, solution times are rather low for each parameter setting, i.e. ranging from 37 seconds to 59.17 minutes. Solely around 30% of the considered test sets cannot be solved within the defined time limit when solving MP-IP. In these cases, the solution process is aborted after a computation time of 600 seconds. The remaining 72 instances are solvable in less than ten minutes. It is therefore possible, to find a good feasible solution with low computational effort, at least when solving a two-week problem, independent of the input data. This again serves as evidence for the satisfying performance of our provided solution approach in contrast to the compact MIP formulation.

Concerning the objective function value, the minimum labor costs for physicians range from 55'250€ to 198'250€ at its maximum. The resulting salary costs vary depending on the provided level of flexibility in patterns of working days: Full flexibility results in an objective value of 55'250€

whereas a parameter setting with a lower level of flexibility provides 198’250€ as solution, i.e. for $\overline{D}^{on} = \underline{D}^{on} = 1$ and $\underline{D}^{off} = 5$. A more realistic parameter setting, defining $\underline{D}^{off} = 2$, results in personnel costs ranging from 55’250€ to 110’500€ when varying the remaining day parameter $\overline{D}^{on}$ and $\underline{D}^{on}$, i.e. three additional physicians are required in less flexible scheduling settings. Due to the reduction in flexibility as the minimum and the maximum number of allowed consecutive working days converge, the available combination of assignable days on and off duty decrease. As a result, labor costs increase significantly. Even though, the spread of the minimum and the maximum total wage costs is not that significant for each set of test instances, similar results can be ascertained for the remaining parameter combinations, e.g. $\overline{D}^{on} = 6, \underline{D}^{on} = 2$ and $\underline{D}^{off} = 1$ in comparison to $\overline{D}^{on} = 2, \underline{D}^{on} = 2$ and $\underline{D}^{off} = 1$. As a result, if the appropriate workforce size derived by the personnel costs provided by the CG heuristic for a specific parameter setting is employed, the average weekly utilization of personnel ranges from 50.85% to 72.03% at its maximum (see column ten in Table 5 in Appendix II). Utilization levels are derived by the ratio of available hours of supply and hours of occurring demand (in each week), given in (6):

$$Weekly\ utilization = \frac{\sum_{d \in D} \sum_{p \in P} N_{dp}}{\overline{R} \cdot \sum_{r \in R} \lambda_r} \tag{6}$$

Since we use our 50% quantile demand scenario as input data, the requirement for care is 551 hours each week. These are confronted with the working hours of the employed personnel. As an example, weekly utilization for $\overline{D}^{on} = 6, \underline{D}^{on} = 1$ and $\underline{D}^{off} = 1$ is calculated: The bi-weekly labor costs provided by our CG heuristic is 55’250€ which leads to a required number of 17 physicians. This results in a total of 765 hours of available working hours each week. Proportioning these two measures devotes an average utilization of staff each week of $\frac{551}{765} = 0.7203$, i.e. 72.03%.

Contrary to the increasing labor costs when decreasing the level of flexibility, the average utilization of staff each week drops the lower the degree of flexibility. This is reasoned by the growing number of required personnel and with this, an increasing number of available working hours weekly (in case of less flexibility) to cover demand. But, opposed to the increasing available working hours, demand for care is constant in each test instance. Therefore, average utilization decreases. Although utilization fluctuates, the achieved levels seem to be quite reasonable especially for hospitals, i.e. in a health care providing context. Since surgeries commonly have a duration of several hours and require a great amount of concentration, utilization should not be too high to create hours of idle time (around the scheduling of surgeries) which can be used to handle administrative tasks as well as emergency patients. Note, weekly utilization of staff is strongly influenced by the predefined underlying scheduling rules: The higher the degree of flexibility, the higher the utilization of physicians due to the increasing capability to better match supply and demand. The more rigid the rostering regulation, the less options for the assignment of days on duty and different shifts occur.

With respect to the quality of the provided solutions, our column generation approach seems to work quite well: Optimality gap is 3’750€ on average, i.e. approximately one physician. The absolute optimality gap is at most 9’750€ for all 105 test instances. Moreover, almost half of the considered test instances provide an absolute optimality gap of 1’625€, around one fourth of our test sets have an absolute optimality gap of 3’750€, additional 4’875€ labor expenses are solely required in around 2% of the cases, and even 26.7% are solved to optimality by the CG heuristic, i.e. optimality gap equals 0€.

An exemplary depiction of the development of the resulting objective function values for MP-LP as well as the solution of MP-IP for a minimum number of one day on duty, i.e. $\underline{D}^{on} = 1$, is shown in Figure 7.

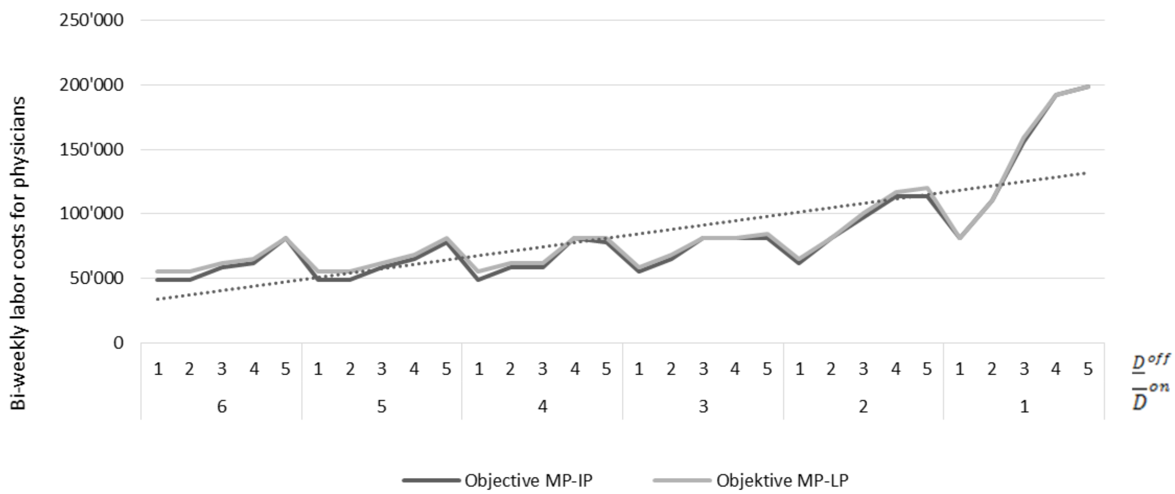

Figure 7: Trend in in day parameter flexibility (for $\underline{D}^{on} = 1$)

The vertical axis shows the minimum number of days off (duty) ranging from one to 5 for each value of $\overline{D}^{on} = 1, \dots, 6$. As the figure displays, the objective function value for MP-LP as well as for MP-IP increases for each value of $\overline{D}^{on}$ when reducing flexibility by increasing $\underline{D}^{off}$ from one to five. Moreover, there is a trend discernible: Despite the increasing objective function value for each manifestation of $\overline{D}^{on}$ itself, the total weekly labor costs of physicians grow linearly in the reduction of flexibility in the maximum number of days on duty. As a result, the lower the degree in flexibility in $\overline{D}^{on}$, the higher the total salary expenses.

Once employees have to have at least some consecutive days off after being on duty, the size of the workforce and with this the total salary costs increase dramatically. Generally speaking, the higher the convergence between the minimum and the maximum number of allowed consecutive working days, the lower the computational effort due to the decreasing number of feasible assignments. Additionally, the total number of required physicians and salary costs increase. Less flexibility in patterns of working days results in a significant increase in the value of the objective. But, even though a higher number of physicians is needed to cover demand for less flexible working patterns, especially the number of consecutive days off might be of major importance for the satisfaction of the workforce. Having only one single day off between two consecutive sequences of working days leads to discontent and a higher probability for employee's turnover. Therefore, it is sufficient to find the balance between the desired level of flexibility and an acceptable burden for staff to ensure their wellbeing.

2.3.2. Variation of shift parameters

In this section, the effect of flexibility in the allowed length of shifts is decreased and analyzed while the flexibility in starting times of shifts remains. Similar to the former subsection, the values of the different shift parameters (concerning its length) are varied iteratively. In the first step, S^{min} is increased from seven to twelve hours whereas the value of S^{max} is decreased from an initial maximum length of twelve to the current value of S^{min}. Due to the effect of the shift length on the placement of the break and with this, coverage of demand, the size of the break window is additionally considered. Therefore, B^{pre} and B^{post} are varied simultaneously, i.e. $B^{pre} = B^{post} = 3$ in the initial step. Subsequently, the values are decreased iteratively to enlarge the size of the break window. Increasing the number of possibilities for the placement of the break provides additional flexibility in the scheduling process. This results in a total of 63 different combinations for the setting of the shift parameters (see Table 6 in the Appendix III for a detailed overview on the solution and evaluation values, computational effort, and the average weekly utilization as performance indicator).

Similar to the variation in day parameters, the computational effort to solve the different shift parameter settings is rather low, respectively even lower, i.e. ranging from nine seconds to 56.03 minutes. Solving MP-IP in less than ten minutes is only achievable for less than one third of the test instances. For the remaining ones, the predetermined time limit of 600 seconds is effective. Generally, the higher the convergence between minimum and maximum allowed shift length, the smaller the solution time due to the reduced number of available shift types. For a maximum level of flexibility, the computational effort is still rather low which is reasonable. Increasing the flexibility for the assignment of the break by enlarging the size of the break window does not lead to a growth in solution times.

Moreover, there is a significant effect of the degree of shift flexibility on workforce costs. A larger number of shift types results in lower expenses for physicians. For the maximum level of flexibility with $S^{min} = 7$ and $S^{max} = 12$, solely 55’250€ labor costs are required to employ enough personnel to cover demand. Since the number of allowed shift types is reduced due to the convergence of S^{min} and S^{max}, total salary costs increase to a maximum of 81’250€. In this extreme, the fluctuation in demand can only be encountered by assigning one single type of shift having various starting periods, e.g. with a duration of twelve periods. Considering the case that a hospital does solely consider flexible starting periods for shifts rather than flexible shift lengths to schedule their personnel for whatever reason, the results provided by the table indicate that shorter shifts might be more suitable to cover standard demand compared to shifts with a long duration. In general, implementing a higher amount of flexibility in the length of available shifts results in a smaller objective function value and with this a decreasing number of physicians required. Concerning the average weekly utilization of staff, a minimum value of 66.43% and a maximum of 72.03% is provided which in turn seems to be quite appropriate for hospital settings.

Inserting additional flexibility in the scheduling process by enlarging the size of the break window does either have a positive or no effect on the objective function value. For almost 50% of the different parameter settings, the maximum size of the break window decreases the total labor costs. Especially for long shift durations providing a minimum amount of flexibility, e.g. $S^{min} = 11$ and $S^{max} = 12$, enlarging the feasible combinations for the placement of the break does not affect the resulting objective. As a result, flexibility in break assignment particularly affects settings that do solely provide short shifts (less than ten hours) with a low level of flexibility, i.e. especially for real life assumptions. For instances, allowing a high degree of flexibility in the number of available shift types, additional flexibility by enlarging the size of the break window does not have a significant effect. In contrast to the factorial analysis of day parameters, the solution quality of the shift parameter analysis is not of similar performance, but still appropriate. On average, optimality gap is 5'829€, i.e. corresponding to an additional employment of at most two physicians. This disparity is rather small and serves additionally as proof for the high performance of our solution approach. Around 17% of the considered test sets provide a total optimality gap of 9'750€, 6'500€ additional salary costs are required in more than 50% and 3'250€ higher wage costs are required in 20% of the test instances. Actually, five parameter settings are solved to optimality, i.e. with a MP-LP solution equal to the MP-IP solution. An example for the development of the objective function values of the CG heuristic and the MP-LP in case of $S^{min} = 7$ is provided in Figure 9.

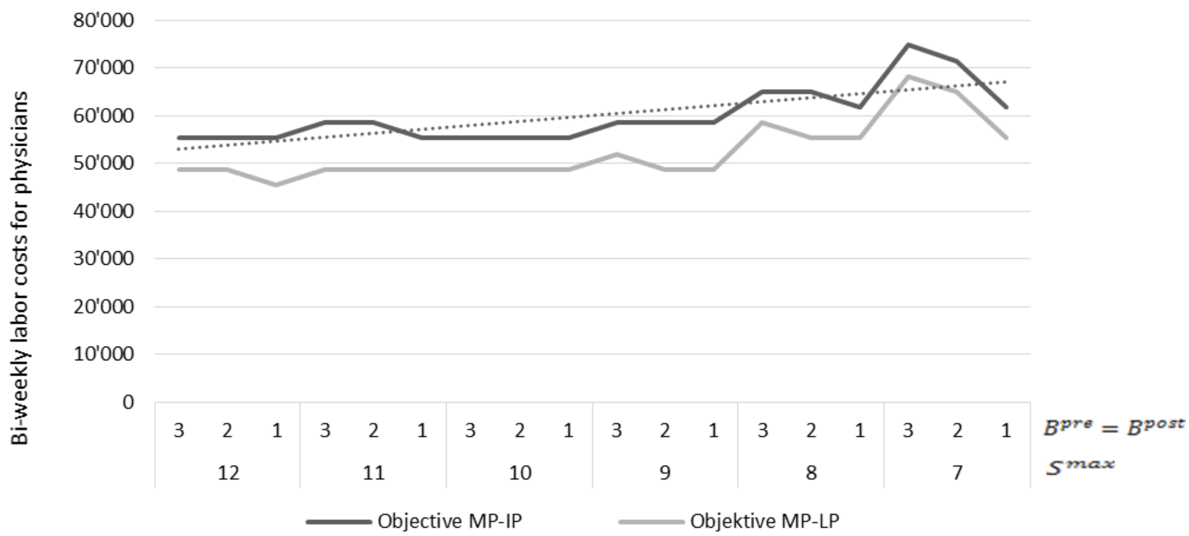

Figure 9: Trend in in shift parameter flexibility (for $S^{min} = 7$)

Again, the total weekly labor costs and with this, the number of required personnel either stays constant or decreases due to the implementation of additional flexibility by enlarging the break window. But this effect is rather small. But, similar to the variation in day parameters, the higher the convergence between S^{min} and S^{max}, the higher the required number of physicians to cover demand.

Generally, in the scheduling process, the considered amount of flexibility in terms of shift types and the size of the break window provide a significant improvement concerning the total labor costs and therefore the overall number of physicians that is required to cover mean demand. Even though the computational effort rises due to the increasing level of assignment opportunities, the generated benefit is valuable for hospitals to reduce their expenses for personnel while ensuring the requested level of service.

3. Summary and Conclusion

In this research, we formulated a MIP based on a reduced set covering approach to consider the implementation of flexible sequences of working days, flexible shifts, and flexible break assignments when scheduling physicians in hospitals. Due to the complexity of the stated model, a column generation heuristic is proposed to provide solutions of high quality in reasonable amount of solution time. To evaluate the performance of the solution approach as well as the effect of implementing a high degree of flexibility in the scheduling process (i.e. in terms of total salary costs as monetary value for the required workforce size), real life data from a large teaching hospital in Germany of the year 2010 is used. In our experimental study, demand of each period of every day over one entire year is aggregated and three different demand scenarios are defined: 50% quantile, 75% quantile, and maximum (100% quantile) demand. Moreover, the performance of the proposed CG heuristic is evaluated by varying the length of the planning horizon from one to six weeks. Finally, to test for the effect of additional flexibility in the scheduling process, a factorial analysis varying day, shift, and break parameters is conducted.

In general, our CG approach performs quite well: Due to the complexity of our mathematical model, it is not possible to solve the MIP to optimality for at almost all test instances. Albeit only a heuristic solution is provided by the CG approach, each of our test sets is solved in an appropriate amount of time. Additionally, heuristic solutions of high quality are provided which is proven by the optimality gap between MP-LP and MP-IP solution. Furthermore, as an outcome of our sensitivity analysis, the lower the degree of flexibility in terms of sequences of working days and shift types, the higher the upcoming labor costs due to the increasing number of required physicians to cover demand. Focusing on the variation of shift parameters, an increasing level of flexibility leads to a reduction in salary costs. Besides, additional reduction is achievable by enlarging the size of the break window. But, this effect is only of reduced significance compared to the level of flexibility in day and shift parameters. As a result, the optimal scheduling strategy depends on the policy of hospital's management and the hospital's ambition in terms of quality of care, desired service level, and employee satisfaction. Especially in case of employing a workforce size below the maximum level, it is barely impossible to stick to a predefined inflexible eight hour shift system to prevent for huge wage expenses. Due to the reduced number of physicians, flexible shifts have to be implemented to create an appropriate flexible shift system adopting the fluctuation in demand.

For future research, a B&P approach can be implemented to generate optimal integer solutions for the considered problem. Moreover, concerning the underlying demand pattern, further research might be investigated to build a more realistic estimation of demand profiles. Up to now, research commonly uses randomly generated demand patterns based on predefined assumptions or real life data derived from a hospital. But, even if real life demand patterns are in use, these may be biased due to specific regulations and other schedules in hospital: In general, especially when scheduling elective patients, available operating room allocations based on specialty defined by the Master Surgery Schedule (MSS) and the number of assignable available shift types determine the schedule of surgeries, i.e. demand. Therefore, demand is partially driven by the shift design of a hospital and might change, if number and type of available shifts change.

Despite the large potential to decrease hospital's operating and labor costs, on an operational level, additional aspects concerning the wellbeing of employees should be taken into account to prevent from turnover of physicians. Moreover, a time window constraint regulating the starting times of consecutive shifts can be implemented to ensure more balanced starting periods for each physician within one sequence of working days. It is also important to integrate the patients' well-being and therefore the quality of the provided service. Implementing a lot of flexibility especially in terms of sequences of working days might result in its extreme to a roster where physicians are one or two days on duty and are off for the subsequent day(s). Therefore, an appropriate level of flexibility is to be determined for the trade-off between monetary value of flexibility and quality of care. As a result, it might make sense to provide a behavioral template/scheme for hospital's management how to operate depending on the institution's goal.

Appendix I

Linearization of constraint 1.3

$$z_{id} = y_{id}(1 - y_{i(d-1)}) \qquad \forall i \in \boldsymbol{I}, d \in \boldsymbol{D} \tag{1.3}$$

$$z_{id} - y_{id} \leq 0 \qquad \forall i \in \boldsymbol{I}, d \in \boldsymbol{D} \tag{1.3.1}$$

$$z_{id} + y_{i(d-1)} \leq 1 \qquad \forall i \in \boldsymbol{I}, d \in \boldsymbol{D} \tag{1.3.2}$$

$$y_{id} - y_{i(d-1)} - z_{id} \leq 0 \qquad \forall i \in \boldsymbol{I}, d \in \boldsymbol{D} \tag{1.3.3}$$

Appendix II

$\underline{D}^{on}$	$\overline{D}^{on}$	$\underline{D}^{off}$	Objektive MP-LP	Objective MP-IP	Opt. gap (alsolut)	# columns	Solution time: MP-LP [min.]	Solution time: MP-IP [sec.]	Weekly utilization (average)
1	6	1	48'750	55'250	6'500	1'608	49.07	600	72.03%
		2	48'750	55'250	6'500	1'640	51.09	600	72.03%
		3	58'500	61'750	3'250	163	2.20	1	64.44%
		4	61'750	65'000	3'250	291	4.48	600	61.22%
		5	81'250	81'250	0	98	1.27	0	48.98%
	5	1	48'750	55'250	6'500	1'535	44.57	600	72.03%
		2	48'750	55'250	6'500	1'752	57.66	600	72.03%
		3	58'500	61'750	3'250	275	3.97	4	64.44%
		4	65'000	68'250	3'250	283	4.21	2	58.31%
		5	78'000	81'250	3'250	117	1.40	0	48.98%
	4	1	48'750	55'250	6'500	552	9.94	600	72.03%
		2	58'500	61'750	3'250	232	3.25	1	64.44%
		3	58'500	61'750	3'250	383	5.97	24	64.44%
		4	78'000	81'250	3'250	85	1.07	0	48.98%
		5	81'250	81'250	0	225	2.96	4	48.98%
	3	1	55'250	58'500	3'250	291	3.99	600	68.02%
		2	65'000	68'250	3'250	276	4.08	35	58.31%
		3	81'250	81'250	0	108	1.40	1	48.98%
		4	81'250	81'250	0	127	1.72	0	48.98%
		5	81'250	84'500	3'250	508	8.01	1	47.09%
	2	1	61'750	65'000	3'250	240	3.18	600	61.22%
		2	81'250	81'250	0	172	2.35	33	48.98%
		3	97'500	100'750	3'250	196	2.50	3	39.50%
		4	113'750	117'000	3'250	106	1.24	0	34.01%
		5	113'750	120'250	6'500	168	2.10	3	33.09%
	1	1	81'250	81'250	0	178	2.24	55	48.98%
		2	110'500	110'500	0	150	2.62	38	36.01%
		3	156'000	159'250	3'250	162	1.74	0	24.99%
		4	191'750	191'750	0	114	1.18	0	20.75%
		5	198'250	198'250	0	176	1.88	0	20.07%
2	6	1	48'750	55'250	6'500	1'520	44.29	600	72.03%
		2	48'750	55'250	6'500	1'720	55.63	600	72.03%
		3	58'500	61'750	3'250	175	2.41	17	64.44%
		4	61'750	65'000	3'250	301	4.67	600	61.22%
		5	81'250	81'250	0	73	0.93	0	48.98%
	5	1	48'750	55'250	6'500	1'696	53.28	600	72.03%
		2	48'750	55'250	6'500	1'670	54.76	600	72.03%
		3	58'500	61'750	3'250	246	3.64	26	64.44%
		4	65'000	68'250	3'250	290	4.41	3	58.31%
		5	81'250	81'250	0	144	1.97	1	48.98%
	4	1	48'750	55'250	6'500	567	9.86	600	72.03%
		2	58'500	61'750	3'250	223	3.15	1	64.44%
		3	58'500	61'750	3'250	401	6.51	173	64.44%
		4	81'250	81'250	0	98	1.31	0	48.98%
		5	78'000	81'250	3'250	212	2.69	15	48.98%
	3	1	55'250	58'500	3'250	338	5.08	600	68.02%
		2	65'000	68'250	3'250	295	4.49	31	58.31%
		3	81'250	81'250	0	109	1.49	2	48.98%
		4	81'250	81'250	0	123	1.66	1	48.98%
		5	81'250	84'500	3'250	401	5.83	1	47.09%
	2	1	61'750	65'000	3'250	250	3.43	554	61.22%
		2	81'250	81'250	0	181	2.52	1	48.98%
		3	113'750	113'750	0	75	0.99	0	34.98%
		4	113'750	113'750	0	99	1.17	0	34.98%
		5	113'750	117'000	3'250	155	1.87	0	34.01%
3	6	1	48'750	55'250	6'500	1'714	53.73	600	72.03%
		2	48'750	55'250	6'500	1'798	59.16	600	72.03%
		3	58'500	61'750	3'250	161	2.30	1	64.44%
		4	61'750	65'000	3'250	284	4.45	600	61.22%
		5	81'250	81'250	0	71	0.94	0	48.98%
	5	1	48'750	55'250	6'500	1'600	49.16	600	72.03%
		2	45'500	55'250	9'750	1'646	46.74	600	72.03%
		3	58'500	61'750	3'250	279	4.32	24	64.44%
		4	65'000	68'250	3'250	309	4.90	13	58.31%
		5	81'250	81'250	0	143	2.05	2	48.98%
	4	1	48'750	55'250	6'500	617	10.17	600	72.03%
		2	58'500	61'750	3'250	254	3.35	33	64.44%
		3	58'500	61'750	3'250	355	5.70	64	64.44%
		4	81'250	81'250	0	82	1.13	0	48.98%
		5	120'250	120'250	0	52	0.67	0	33.09%
	3	1	55'250	58'500	3'250	404	5.87	44	68.02%
		2	65'000	68'250	3'250	339	4.69	28	58.31%
		3	81'250	81'250	0	89	1.22	1	48.98%
		4	81'250	81'250	0	122	1.71	0	48.98%
		5	81'250	84'500	3'250	418	6.13	1	47.09%
4	6	1	48'750	55'250	6'500	1'735	56.94	600	72.03%
		2	48'750	55'250	6'500	1'628	51.65	600	72.03%
		3	58'500	61'750	3'250	161	2.39	58	64.44%
		4	61'750	65'000	3'250	213	3.25	225	61.22%
		5	81'250	81'250	0	57	0.77	0	48.98%
	5	1	48'750	55'250	6'500	1'725	57.37	600	72.03%
		2	48'750	55'250	6'500	1'631	54.67	600	72.03%
		3	58'500	61'750	3'250	259	4.22	4	64.44%
		4	65'000	68'250	3'250	224	3.64	1	58.31%
		5	81'250	81'250	0	121	1.86	1	48.98%
	4	1	52'000	58'500	6'500	452	6.74	600	68.02%
		2	58'500	61'750	3'250	193	2.67	21	64.44%
		3	58'500	61'750	3'250	415	6.58	527	64.44%
		4	81'250	81'250	0	79	0.99	0	48.98%
		5	120'250	120'250	0	50	0.61	0	33.09%
5	6	1	48'750	55'250	6'500	1'771	57.41	600	72.03%
		2	48'750	55'250	6'500	1'715	55.46	600	72.03%
		3	58'500	61'750	3'250	141	1.97	14	64.44%
		4	61'750	65'000	3'250	197	2.97	66	61.22%
		5	61'750	65'000	3'250	303	4.86	600	61.22%
	5	1	45'500	55'250	9'750	1'867	59.17	600	72.03%
		2	48'750	55'250	6'500	1'768	57.93	600	72.03%
		3	58'500	61'750	3'250	216	3.04	22	64.44%
		4	65'000	68'250	3'250	198	2.77	19	58.31%
		5	81'250	81'250	0	98	1.27	0	48.98%
6	6	1	48'750	55'250	6'500	845	18.17	600	72.03%
		2	48'750	55'250	6'500	780	16.04	600	72.03%
		3	58'500	61'750	3'250	154	2.11	14	64.44%
		4	58'500	61'750	3'250	321	5.09	174	64.44%
		5	61'750	65'000	3'250	183	2.63	68	61.22%

Table 5: Factorial analysis – Flexibility in day parameter

Appendix III

S^{min}	S^{max}	$B^{pre} = B^{post}$	Objektive MP-LP	Objective MP-IP	Opt. gap (alsolut)	# columns	Solution time: MP-LP [min.]	Solution time: MP-IP [sec.]	Weekly utilization (average)
7	12	3	48'750	55'250	6'500	1'608	49.07	600	72.03%
		2	48'750	55'250	6'500	770	16.96	600	72.03%
		1	45'500	55'250	9'750	664	11.97	600	72.03%
	11	3	48'750	58'500	9'750	1'805	56.03	600	68.02%
		2	48'750	58'500	9'750	1'113	26.02	600	68.02%
		1	48'750	55'250	6'500	831	17.02	600	72.03%
	10	3	48'750	55'250	6'500	1'280	27.49	600	72.03%
		2	48'750	55'250	6'500	1'518	38.05	600	72.03%
		1	48'750	55'250	6'500	1'068	22.30	600	72.03%
	9	3	52'000	58'500	6'500	422	4.27	600	68.02%
		2	48'750	58'500	9'750	739	8.83	600	68.02%
		1	48'750	58'500	9'750	1'089	17.27	600	68.02%
	8	3	58'500	65'000	6'500	330	2.36	600	61.22%
		2	55'250	65'000	9'750	556	4.81	600	61.22%
		1	55'250	61'750	6'500	683	6.91	600	64.44%
	7	3	68'250	74'750	6'500	135	0.43	215	53.24%
		2	65'000	71'500	6'500	295	1.55	600	55.66%
		1	55'250	61'750	6'500	683	6.91	600	64.44%
8	12	3	48'750	55'250	6'500	1'804	53.53	600	72.03%
		2	48'750	55'250	6'500	662	11.53	600	72.03%
		1	45'500	55'250	9'750	484	6.32	600	72.03%
	11	3	48'750	55'250	6'500	1'410	33.67	600	72.03%
		2	48'750	55'250	6'500	1'056	20.93	600	72.03%
		1	48'750	55'250	6'500	589	8.68	600	72.03%
	10	3	48'750	55'250	6'500	1'028	16.57	600	72.03%
		2	48'750	55'250	6'500	1'610	37.63	600	72.03%
		1	48'750	55'250	6'500	774	11.24	600	72.03%
	9	3	52'000	58'500	6'500	521	4.63	600	68.02%
		2	48'750	58'500	9'750	666	6.69	600	68.02%
		1	48'750	55'250	6'500	849	10.57	600	72.03%
	8	3	58'500	65'000	6'500	497	3.63	600	61.22%
		2	55'250	61'750	6'500	884	17.78	600	64.44%
		1	55'250	61'750	6'500	576	4.56	600	64.44%
9	12	3	48'750	55'250	6'500	884	17.78	600	72.03%
		2	48'750	55'250	6'500	1'526	43.69	600	72.03%
		1	45'500	55'250	9'750	665	9.12	600	72.03%
	11	3	48'750	58'500	9'750	542	7.32	600	68.02%
		2	48'750	55'250	6'500	1'095	21.67	600	72.03%
		1	48'750	55'250	6'500	1'430	30.45	600	72.03%
	10	3	52'000	58'500	6'500	565	6.24	600	68.02%
		2	48'750	55'250	6'500	683	7.62	600	72.03%
		1	48'750	55'250	6'500	881	12.08	600	72.03%
	9	3	52'000	58'500	6'500	362	2.11	600	68.02%
		2	48'750	55'250	6'500	448	3.03	600	72.03%
		1	48'750	55'250	6'500	611	4.92	600	72.03%
10	12	3	58'500	61'750	3'250	220	1.60	19	64.44%
		2	58'500	61'750	3'250	192	1.39	2	64.44%
		1	58'500	61'750	3'250	175	1.24	1	64.44%
	11	3	61'750	65'000	3'250	170	0.87	1	61.22%
		2	61'750	65'000	3'250	142	0.66	3	61.22%
		1	61'750	65'000	3'250	132	0.62	1	61.22%
	10	3	65'000	68'250	3'250	175	0.66	4	58.31%
		2	65'000	65'000	0	194	0.82	8	61.22%
		1	65'000	65'000	0	184	0.73	91	61.22%
11	12	3	58'500	61'750	3'250	196	1.10	21	64.44%
		2	58'500	61'750	3'250	160	0.80	2	64.44%
		1	58'500	61'750	3'250	146	0.72	1	64.44%
	11	3	61'750	65'000	3'250	132	0.45	1	61.22%
		2	61'750	65'000	3'250	123	0.41	1	61.22%
		1	61'750	65'000	3'250	89	0.26	1	61.22%
12	12	3	81'250	81'250	0	97	0.31	0	48.98%
		2	81'250	81'250	0	80	0.23	0	48.98%
		1	81'250	81'250	0	59	0.14	0	48.98%

Table 6: Factorial analysis – Flexibility in shift parameter

References

Aggarwal, S.C. (1982). A focused review of scheduling in services. European Journal of Operational Research, 9(2), 114-121.

Aykin, T. (1996). Optimal Shift Scheduling with Multiple Break Windows. Management Science, 42(4), 591-602.

Barnhart, C., Johnson, E.L., Nemhauser, G.L., Martin, W.P., Savelsbergh, V.P.H. (1998). Branch-and-Price: Column Generation for Solving Huge Integer Programs. Operations Research, 46(3), 316-329.

Bechtold, S.E., Jacobs, L.W. (1990). Implicit Modeling of Flexible Break Assignments in Optimal Shift Scheduling. Management Science, 36(11), 1339-1351.

Beliën, J., Demeulemeester, E. (2006). Scheduling trainees at a hospital department using a branch-and-price approach. European Journal of Operational Research, 175(1), 258-278.

Blum, K., Löffert, S., Offermanns, M., Perner, P. (2007). Krankenhaus Barometer – Umfrage 2017. Technical Report, Deutsches Krankenhaus Institut e.V.: Düsseldorf.

Blum, K., Löffert, S., Offermanns, M., Perner, P. (2017). Krankenhaus Barometer – Umfrage 2017. Technical Report, Deutsches Krankenhaus Institut e.V.: Düsseldorf.

Bölt, U. (2014). Statistische Krankenhausdaten: Grund- und Kostendaten der Krankenhäuser 2011. In J Klauber, M Geraedts, J Friedrich, J Wasem (Ed), Krankenhaus-Report 2014: Schwerpunkt: Patientensicherheit. Schattauer: Stuttgart, 293-328

Brunner, J.O., Bard, J.F., Kolisch, R. (2009). Flexible shift scheduling of physicians. Health Care Management Science, 12(3), 285-305.

Brunner, J.O., Bard, J.F., Kolisch, R. (2010). Midterm scheduling of physicians with flexible shifts using branch and price. IIE Transactions, 43(2), 84-109.

Brunner, J.O., Edenharter, G.M. (2011). Long term staff scheduling of physicians with different experience levels in hospitals using column generation. Health Care Management Science, 14(2), 189-202.

Cohn, A., Root, S., Kymissis, C., Esses, J., Westmoreland, N. (2009). Scheduling medical residents at Boston university school of medicine. Interfaces, 39(3), 186-195.

Damci-Kurt, P., Zhang, M., Marentay, B. (2017). Improving Physician Schedules by Leveraging Equalization: Cases from Hospitals in US. Working Paper, Kennesaw State University, Kennesaw, Georgia.

Dantzig, G.B. (1954). Letter to the editor - A comment on Edie's "Traffic delays at toll booths". Journal of the Operations Research Society of America, 2(3), 339-341.

Dantzig, G.B., Wolfe, P. (1960). Decomposition Principle for Linear Programs. Operations Research, 8(1), 101–111.

Desrosiers, J., Lübbecke, M.E. (2005). A Primer in Column Generation. In G Desaulniers (Ed), Column generation, GERAD 25th anniversary series, Vol 5. Springer: New York, NY, 1-32.

Desaulniers, G., Desrosiers, J., Solomon, M.M. (2005). Column generation. Springer: New York, NY.

El-Rifai, O., Garaix, T., Augusto, V., Xie, X. (2015). A stochastic optimization model for shift scheduling in emergency departments. Health Care Management Science, 18(3), 289-302.

Erhard, M., Brunner, J.O. (2018). Physician staffing levels and absence planning under special consideration of breaks – A case study on anesthetists. Working Paper, University of Augsburg, Augsburg, Germany.

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

Ernst, A.T., Jiang, H., Krishnamoorthy, M., Sier, D. (2004). Staff scheduling and rostering: A review of applications, methods and models. European journal of operational research, 153(1), 3-27.

Ganguly, S., Lawrence, S., Prather, M. (2014). Emergency department staff planning to improve patient care and reduce costs. Decision Sciences, 45(1), 115-145.

Gilmore, P.C., Gomory, R.E. (1963). A linear programming approach to the cutting stock problem—Part II. Operations research, 11(6), 863-888.

Isken, M.W. (2004). An implicit tour scheduling model with applications in Health-care. Annals of Operations Research, 128(1), 91-109.

Janaro, R.E.; Bechtold, S.E. (1985). A study of the reduction of fatigue impact on productivity through optimal rest break scheduling. Human factors, 27(4), 459-466.

Lübbecke, M.E., Desrosiers, J. (2005). Selected topics in column generation. Operations Research, 53(6), 1007-1023.

Marburger Bund (2016). Entgelttabelle zum TV-Ärzte/VKA. https://www.marburger-bund.de/sites/default/files/tarifvertraege/2017/tv-aerzte-vka/entgelttabelle-zum-tv-aerztevka.pdf, Marburger Bund, Berlin.

Rekik, M., Cordeau, J.F., Soumis, F. (2010). Implicit shift scheduling with multiple breaks and work stretch duration restrictions. Journal of Scheduling, 13(1), 49-75.

Statistisches Bundesamt (2016). Gesundheit - Kostennachweis der Krankenhäuser 2015. Fachserie 12 Reihe 6.3, Statistisches Bundesamt: Wiesbaden, 1-65.

Stolletz, R., Brunner, J.O. (2012). Fair optimization of fortnightly physician schedules with flexible shifts. European Journal of Operational Research, 219(3), 622-629.

Thompson, G.M. (1995). Improved Implicit Optimal Modeling of the Labor Shift Scheduling Problem. Management Science, 41(4), 595-607.

Thungjaroenkul, P., Cummings, G.G., Embleton, A. (2007). The impact of nurse staffing on hospital costs and patient length of stay: A systematic review. Nursing Economics, 25(5), 255-265.

Van den Bergh, J., Beliën, J., De Bruecker, P., Demeulemeester, E., De Boeck, L. (2013). Personnel scheduling: A literature review. European Journal of Operational Research, 226(3), 367-385.

Volland, J., Fügener, A., Brunner, J.O. (2017). A column generation approach for the integrated shift and task scheduling problem of logistics assistants in hospitals. European Journal of Operational Research, 260(1), 316-334.

Wolsey, L.A. (1998). Integer Programming. Wiley: New York, NY.