Preference-based physician scheduling in hospitals

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

vorgelegt von
Christopher Nikolaus Gross, M. Sc.

Erstgutachter:	Prof. Dr. Jens O. Brunner
Zweitgutachter:	Prof. Dr. Axel Tuma
Drittgutachter:	Prof. Dr. Robert Klein
Vorsitzender der mündlichen Prüfung:	Prof. Dr. Michael Krapp
Tag der mündlichen Prüfung:	5. April 2019

Augsburg, Dezember 2018

List of contributions

This thesis contains the following contributions published in scientific journals. The specified categories relate to the journal ranking VHB-JOURQUAL3 of the Verband der Hochschullehrer für Betriebswirtschaft e.V. (2015). The order of the contributions corresponds to the order of print in this thesis.

Contribution 1

Gross, C. N., Fügener A., Brunner, J. O. (2018) Online rescheduling of physicians in hospitals. Flexible Services and Manufacturing Journal 30(1-2): 296-328.
`https://doi.org/10.1007/s10696-016-9274-2`
Published, category B

Contribution 2

Gross, C. N., Brunner, J. O., Blobner M. (2018) Hospital physicians can't get no long-term satisfaction — An indicator for fairness in preference fulfillment on duty schedules. Health Care Management Science.
`https://doi.org/10.1007/s10729-018-9452-8`
Published, category A

Contribution 3

Gross, C. N. (2018) Long-term workload equality on duty schedules for physicians in hospitals. PATAT 2018: Proceedings of the 12th International Conference on the Practice and Theory of Automated Timetabling, 183-197.
`https://patatconference.org/patat2018/proceedings/`
Published, unranked

Contents

1 Introduction

Personnel scheduling problems occur in a variety of domains, such as transportation, manufacturing, or the service industry. One sector where scheduling is very important is the healthcare sector. In other industries—such as transportation or manufacturing—the negative outcomes of bad scheduling range from idle employees (overstaffing) to idle vehicles or machines (understaffing). In the healthcare sector, however, bad scheduling can lead to severe illness or even death when patients are not treated due to an insufficient number of staff available. It is therefore not surprising that physician and nurse scheduling problems have seen a lot of research attention (Erhard et al., 2018; Cheang et al., 2003).

All staff scheduling problems can be categorized according to their planning horizon. We distinguish between staffing problems ("How many people should I employ?") with a long-term planning horizon of several months or years, rostering problems ("When does which employee have to work?") with a planning horizon of several weeks to a few months, and rescheduling problems ("Who can substitute for the employee who is unexpectedly absent?") with a planning horizon of a few hours or days. The contributions in this thesis focus on the rostering and rescheduling problems. The general form of the rostering problem is always the same: Given a set of employees, their availability, and a demand structure for the presence of these employees, find a schedule that ensures coverage of demand while respecting a set of constraints, such as maximum work time, labor law, fairness, and others. Rescheduling problems exhibit the same structure and additionally try to minimize deviations to an existing schedule.

While the basic problem structure of the rostering problem might fool most people into thinking that it would be easy to solve, the actual dimensions of the problem quickly become clear when illustrated with a short example. Let our illustrative example be a medium-sized hospital department with 30 physicians, out of which we need 5 to be working specific overnight duties each day. This gives us $\frac{30!}{25!} \approx 1.7 \cdot 10^7$ possibilities to assign these physicians to overnight duties on a single day. To create the roster for an entire month, we need to make this assignment decision on each day of the month. For a four-week planning horizon of 28 days, this leaves us with $(1.7 \cdot 10^7)^{28} \approx 3.3 \cdot 10^{202}$ possible schedules. To get a feeling for the magnitude of this number, contrast this with the estimated number of particles in the observable universe, which is 10^{90} (Guth, 2001). It is not hard to imagine that enumerating all the possible schedules, eliminating the invalid ones, and then finding the best schedule is not only a hard but an impossible feat for a human.

Due to the massive complexity of the problem, even before computers were readily available to every business, algorithms were developed to assist with creating valid schedules (Baker, 1974). These ensured that, by following the algorithm, schedulers would always arrive at a valid schedule that respects the constraints baked into the respective algorithm. However, as the algorithms had to be executed by hand, they were simple and inflexible by modern standards. The potential for automation of scheduling tasks was quickly apparent when the first computers were commercially available. Early computer-based systems could already create schedules automatically while respecting certain hard constraints and minimizing the violation of soft constraints, such as physician preferences or fairness (Gierl et al., 1986).

Modern scheduling approaches have access to an immensely higher amount of memory and processing power. Whereas early models were run on mainframes in data centers with a few megabytes of memory, today's consumer computers are several magnitudes more powerful. This feature, coupled with more efficient algorithms, enables today's scheduling approaches to incorporate a multitude of constraints. It is not uncommon for scheduling models, such as the ones presented in this thesis, to have complex constraints regarding a maximum amount of consecutive days of work, limits on overtime hours per week, or requirements for equal distribution of preference fulfillment or workload. This processing power has enabled schedulers and scientists to explore more sophisticated ways of how fairness constraints can be incorporated in the scheduling process. While early approaches, such as the scheduling algorithm presented by Baker (1974), were designed in such a way that they would never exceed given upper bounds on consecutive working days or consecutive assigned weekends, they could not dynamically adjust the assignments to ensure an equal satisfaction or workload among all employees. Additionally, they were not designed to create updated schedules in case of unforeseen absences and were also unable to compensate for violations incurred via rescheduling.

As physicians are a very costly resource to hospitals, managers have a high incentive to minimize turnover and physician attrition. Hospitals therefore go to great lengths to keep their physicians satisfied. This also manifests on duty rosters where schedulers incorporate various fairness measures. Physicians are usually able to submit preferences specifying when they want to or do not want to work. Schedulers then try to respect these preferences. Furthermore, they want to assign an equal workload to each physician in order to prevent physicians feeling disadvantaged in comparison with their colleagues. A multitude of further fairness measures is described in the literature (see e.g., Karsu and Morton, 2015; Bertsimas et al., 2011; Stolletz and Brunner, 2012), but in this thesis we will focus on preference fulfillment and how it interacts with other objectives, such as, e.g., equal workload distribution or stability during rescheduling.

While there is a large body of research on personnel scheduling in general (Ernst et al., 2004) and nurse (Cheang et al., 2003) and physician scheduling Erhard et al. (2018) in particular, there are still a few research gaps. The following open research questions are answered by this thesis.

1. How can existing physician schedules be adapted in case of unexpected absences?

2. How does rescheduling on physician schedules affect plan quality?

3. Does maximizing preference fulfillment on physician schedules during each planning horizon ensure fairness in the long term?

4. How can long-term fairness in preference fulfillment on physician schedules be guaranteed?

5. How can long-term equal workload distribution on physician schedules be guaranteed?

6. How does long-term equal workload distribution affect long-term fairness in preference fulfillment on physician schedules?

The remainder of this work is structured as follows. In Section 2, all contributions contained in this thesis are summarized. The contributions are included in their full version in the appendix. Section 3 provides answers to the research questions above and details how these answers are derived from the presented contributions. Finally, Section 4 concludes this work and highlights opportunities for further research.

2 Summaries of the contributions

This thesis makes several contributions to current literature, which are summarized in this section. All contributions can be found in the appendix. The order of the contributions follows the chronological order in which the contributions were submitted to scientific journals.

2.1 Online rescheduling of physicians in hospitals

Gross et al. (2018b) explore the physician rescheduling problem. They present a model to update physician rosters in the case of unexpected absences. The developed model is applied to data from a German university hospital. Results of a case study confirm that this model enables the scheduler to influence the trade-off between plan quality and plan stability. This contribution has been published in *Flexible Services and Manufacturing Journal*, which is ranked in category B in the VHB-JOURQUAL3 ranking (Verband der Hochschullehrer für Betriebswirtschaft e.V., 2015). It can be found in its entirety in Appendix A.

The creation of physician rosters is a complex task due to the multitude of constraints which need to be respected: Qualifications of physicians, individual agreements, maximum amount of sequential working days, part-time contracts, and many more. Especially the qualification constraints necessitate the development of models specific to physician scheduling and prevent the application of nurse scheduling models. Where nurses usually have at most two qualification levels (regular nurses and head nurses), physicians usually have at least five or even more. This is explained by the different specializations of physicians as well as the progress of physician education after getting their degree from university. Furthermore, physician schedules often are acyclic schedules, whereas nurses are usually scheduled cyclically.

An unsolved problem in research is the adaptation of physician schedules in the case of unexpected absences. While there are approaches for nurse rescheduling, they cannot be applied to physicians due to the different problem structure. As the updated schedule needs to respect all constraints of the original schedule and also needs to minimize the number of rescheduling operations, the rescheduling problem is even more complex than the physician scheduling problem. The discussed contribution presents the first model for physician rescheduling. When creating an updated plan, the scheduler has to make a trade-off decision between two different goals: Should the new schedule have minimal deviations from the existing plan or should it conform to plan quality goals as much as possible? This trade-off between plan *quality* and *stability* is explored in this contribution.

The presented model assigns physicians to duties (i.e., late shift or overnight work) and workstations. For duty assignments, the model respects physician qualifications, demand for duties, physician preferences for a specific duty or for not being assigned to any duty on a specific day, and fairness in terms of exceeding a maximum number of duty assignments per physician. For the workstation assignments, minimum and maximum demand for workstations and a training plan are respected. This training plan specifies which physician should be assigned to which department in order to further their education. A specialty of this model is its capability to create initial schedules as well as create an updated schedule based on an existing schedule. The approach optimizes for five different objectives, of which four are *quality* objectives and one is the *stability* objective: Physicians should be assigned to workstations according to their training plan (*training*). Duties should be assigned according to physicians' preferences (*preferences*). During each planning horizon, the duty workload should be equally distributed among all physicians (*fairness*). For all duties and workstations, the demand should be satisfied (*coverage*). During rescheduling, changes to the existing plan should be minimal (*stability*). Each objective is given an individual weight, which can be used by schedulers to rank the different objectives in importance depending on their preferences or workplace culture.

Existing approaches for physician scheduling usually consider the duty assignment and workstation assignment problems separately, i.e., use separate models for these problems, which are then run sequentially. This means that the model which is run first—usually the duty roster model—can not optimize for *quality* goals regarding the other schedule—usually the workstation roster. This inevitably leads to an inferior performance of the workstation roster in comparison with the duty roster. The model presented in this contribution integrates the workstation and duty roster, thereby allowing it to find the optimal trade-off between plan *quality* goals on the workstation and duty rosters. To enable a comparison with sequential approaches, both the integrated model and two separate (sequential) models to solve the problem are presented.

The presented model is applied to data from an anesthesiology department at a German university hospital. Two case studies are presented. In the first study, the integrated model is evaluated with different cost settings for the trade-off between plan *quality* and *stability*. The second study compares the integrated model with the sequential models to evaluate the trade-off between *quality* objectives on the duty and the workstation roster. For both studies, 50 instances of physicians becoming absent are drawn. On each of the 28 days, a randomly drawn set of physicians is marked as absent and the model is run on the respective day to create an updated schedule for this day and the future, keeping the schedule from the past intact.

The first study evaluates three different cost settings for the model objectives. In this study, violations of plan *stability* are weighted either LOW, BALANCED, or HIGH compared with plan *quality* objectives. The following process is repeated for each of the 50 instances. First, an initial plan is created using the integrated model. A vector with 28 elements defining additional absences on each day of the planning horizon is drawn randomly. Then, the integrated model is run on each day of the planning horizon, incorporating the additional absences for this day. The model is run once for each of the LOW, BALANCED, and HIGH cost settings. Results indicate that the LOW and BALANCED cost settings lead to worse performance of plan *stability* than the HIGH setting because a huge number of duty and workstation assignments are rescheduled. However, they perform better in terms of *coverage*, *training*, *fairness*, and *preferences* as they are able to fulfill more *quality* goals by reassigning more people on the schedule.

In the second study, the integrated model is compared with the sequential models. 50 instances are run and the following process is repeated for each of the instances. First, an initial schedule is created using the sequential models. The absence vector for the respective run from the first case study is used. Afterwards, the sequential models are solved on each day of the planning horizon, incorporating the additional absences on that day. For this study, only the BALANCED cost setting is used. The results from the sequential models are then compared with the results for the BALANCED cost setting from the first case study, which uses the integrated model. The results indicate that the sequential models—with the duty roster model being run first—can fulfill plan *quality* goals on the duty roster much better than the integrated model. However, plan *quality* on the workstation roster created by the sequential models is worse than on the schedule created by the integrated model. Due to the sequential models being separate models, it is not possible for the model being run first to optimize for *quality* objectives on the second roster. The integrated model, however, performs worse in terms of *stability* than the sequential models.

Summarizing, this contribution proposes the first model for physician rescheduling. It highlights the trade-off between plan *quality* and *stability*. Using the proposed model, schedulers can influence this trade-off by selecting the appropriate weight setting. Several weight settings and their outcomes on the trade-off are evaluated. For an application in practice, the BALANCED or HIGH cost setting is recommended as the LOW cost setting leads to an immense number of rescheduling operations, which are often hard to execute in practice as each rescheduled physician needs to be notified of the changed schedule. Additionally, the use of integrated models for duty and workstation rostering is recommended over the current practice of using sequential models. Using an integrated model enables schedulers to weigh *quality* on the duty roster against *quality* on the workstation roster, thereby resulting in better workstation assignments, which have a positive influence on the fulfillment of physicians' *training* requirements. The work leaves opportunities for future research. The definition of the *fairness* objective is quite limited and could be expanded. Furthermore, the model does not incorporate long-term fairness for the *preference* objective.

2.2 Hospital physicians can't get no long-term satisfaction — An indicator for fairness in preference fulfillment on duty schedules

Gross et al. (2018a) investigate the long-term impact of fairness in preference distribution. As existing work in physician scheduling only focuses on fairness within each planning horizon, they propose a satisfaction indicator which is carried between planning horizons. Different strategies are evaluated to update the satisfaction between planning horizons. Results of a computational study indicate that disregarding long-term fairness can lead to unequal treatment of physicians which accumulates over several planning horizons. This contribution has been published in *Health Care Management Science*, which is ranked in category A in the VHB-JOURQUAL3 ranking (Verband der Hochschullehrer für Betriebswirtschaft e.V., 2015). It can be found in its entirety in Appendix B.

To ensure adequate medical care around the clock, a certain number of physicians need to be present at the hospital at any time. During the day, physicians are present to perform procedures and see patients, but during the night no procedures are scheduled. Therefore, a duty roster is required to specify which physicians have to stay at the hospital overnight, i.e., to be "on duty". Due to the disruption of physicians' private lives caused by overnight work, schedulers let physicians specify preferences for when they would like to be assigned to an overnight duty and when they would prefer not to be assigned to any duty. As the creation of this roster is a complicated task for a human scheduler, mathematical models and computer assistance are often used to create the schedule. To ensure fairness among the physicians, most models try to maximize preference fulfillment among all physicians. However, due to the placement of preferences, it is often impossible to achieve completely equal preference fulfillment among all physicians, meaning some physicians will be slightly disadvantaged. As existing models do not take into account fairness data from the past, it is possible that the same physicians will be disadvantaged again and again, leading to unhappiness and physician attrition in the long run.

This contribution proposes a satisfaction indicator, a number between 0 and 1, to calculate how satisfied physicians are with a plan in terms of preference fulfillment. For each physician, the satisfaction indicator tracks how many preferences have been fulfilled per number of days in the planning horizon. This enables a comparison between physicians, highlighting how well each physician's preferences have been respected in comparison to their colleagues. Based on the satisfaction indicator, it is possible to calculate a satisfaction-based weight for each physician's preferences in the next planning horizon. This weight is selected in such a way that preferences by physicians who have had many preferences fulfilled in the previous planning horizon are now respected with a lower importance than those of physicians whose preferences have not been respected as much in the previous planning horizon.

The approach describes a general physician scheduling model which assigns physicians to overnight duties. The planning horizon for the proposed model is always an entire month, i.e., 4 or 5 weeks. Physicians are assigned to duties based on their individual qualifications. The model takes physician preferences to be assigned to a specific duty or to not be assigned to any duty as an input. The model optimizes for two main objectives: coverage of physician demand for duties and fulfillment of physician preferences according to the satisfaction-based weight. While the model is based on the requirements of an anesthesiology department, it is kept intentionally generic as to be applicable to most departments. Furthermore, specialized constraints are not incorporated into the model in order to keep the complexity of the model low and focus on the evaluation of fairness in preference fulfillment.

Three different strategies to calculate the satisfaction-based weight from physicians' individual satisfaction indicators are proposed: The constant strategy (C) disregards long-term fairness by fixing the satisfaction-based weight at a constant value and disregarding the satisfaction indicator. This is equal to not incorporating long-term fairness in the planning process and is used to simulate the approach of existing literature. The second strategy calculates the satisfaction indicator by applying exponential smoothing after the planning horizon (ESA). Initially, the satisfaction-based weight is set to a fixed value for the first planning horizon. Then, after creating the plan for a planning horizon, each physician's satisfaction indicator for the created plan is calculated and smoothed exponentially with the satisfaction-based weight from the previous planning horizon to arrive at the satisfaction-based weight for the next planning horizon. Finally, the third strategy applies exponential smoothing during plan creation for the respective planning horizon (ESD). For this strategy, the satisfaction-based weight is modeled as a decision variable which depends on the satisfaction indicator. Using this strategy, the model optimizes the satisfaction-based weight along with the fulfillment of preferences to arrive at a solution with minimal cost for preference violations.

For the C and ESA strategies, the proposed model is linear as the satisfaction-based weight is modeled as a parameter. For the ESD strategy, however, the model becomes quadratic as the satisfaction-based weight is modeled as a decision variable and is multiplied with the sum of preference violations—also a decision variable—in the objective function. In order to efficiently solve the quadratic model, an equivalent linear formulation is described. There can only be one preference per day, so the amount of preference violations is bounded by the number of days in the planning horizon. As the value of the satisfaction indicator depends on the amount of preference violations, it can only take on discrete values and only as many as the maximum amount of preference violations. The amount of possible values for the satisfaction-based weight is therefore bounded by the number of days in the planning horizon. Consequently, it is possible to pre-calculate all possible values of the satisfaction-based weight for each physician, derive the costs induced by assigning each amount of preference violations to each physician, and supply these cost values as parameters. A new decision variable can then be added to select the appropriate cost value and replace the quadratic term in the objective function. The model thereby becomes a linear decision model and can be solved as efficiently for the ESD strategy as it can be for the C and ESA strategies.

In a case study, the three strategies are compared. To enable a comparison, two performance indicators are defined. The APS indicator measures the variance of the average of the satisfaction indicator per physician over all planning horizons. In other words, when this indicator is smaller, satisfaction is more equally distributed among physicians after several planning horizons. Small values of this indicator translate to long-term equal preference fulfillment over all physicians. The ASV performance indicator measures the average of the satisfaction indicator variance per physician between planning horizons. When this indicator is smaller, the satisfaction for each individual physician does not change as much from one planning horizon to the next. Small values of this performance indicator indicate a stable level of satisfaction for each physician over several planning horizons.

The models for each strategy are applied to data from an anesthesiology department at a German university hospital. The data spans 24 months and the model is run for each month. Afterwards, the satisfaction indicator is calculated and the satisfaction-based weight is updated according to the respective strategy. This real-life data does not contain conflicting preferences—i.e., several physicians requesting the same duty on the same date—as workplace culture at the department dictates to avoid making conflicting requests. Consequently, the models can rarely make a trade-off between several physicians with the same request based on their satisfaction-based weights. This is also visible in the performance indicators, which hardly differ between all strategies. It follows that the strategies for fairness in preference fulfillment do not have a high impact when there are no conflicting preferences which would require trade-off decisions based on physician satisfaction.

In order to evaluate the performance for data with conflicting preferences, additional data is generated based on the real-life data from the hospital. The data generation algorithm takes the target conflict rate, i.e., the percentage of preferences which are in conflict with at least one other preference, as an input. These are preferences where at least one other physician has specified a preference for the same duty on the same date. The algorithm is run 11 times to generate data sets with the target conflict rate ranging from 0 % to 100 % in increments of 10 percentage points. For each of these data sets, the model is run for each of the three strategies for each of the 24 months. Results indicate that the ESD strategy performs best in terms of the APS and ASV indicator. This means that the ESD strategy is able to provide both a more equal distribution of satisfaction among all physicians as well as a more stable level of satisfaction over several planning horizons for each individual physician. Furthermore, results show that the C strategy—no tracking of long-term fairness of preference fulfillment—leads to a skewed distribution of satisfaction after several planning horizons. For this strategy, the distribution of preference fulfillment depends on the underlying solver implementation.

In summary, this contribution proposes an indicator to measure physician satisfaction with a plan in terms of preference fulfillment. It illustrates how ignoring long-term satisfaction tracking can lead to unequal treatment of physicians, which accumulates over several planning horizons. A model is presented which assigns physicians to duties while respecting their preferences according to three different strategies: no satisfaction tracking (C), satisfaction tracking after each planning horizon (ESA), and satisfaction tracking during model solving (ESD). As the last strategy results in a quadratic decision model, an equivalent linear formulation is derived. The proposed model is evaluated using data from a German university hospital. Results indicate that satisfaction tracking is not effective when there are no conflicting preferences in the data and therefore no trade-off decisions between physicians with conflicting preferences need to be made. Additional data is generated based on the real-life data to evaluate the impact of long-term satisfaction tracking when the data contains preferences. The model is applied to the generated data and results show that the ESD strategy performs best for all performance indicators. It is therefore recommended to implement satisfaction tracking over several planning horizons and to do so using the ESD strategy. The work opens up possibilities for further research. The definition of "satisfaction" is limited to the amount of satisfied preferences for duties. However, there might be additional factors influencing physicians' satisfaction with a plan, such as the equal distribution of workload.

Furthermore, the general idea of satisfaction tracking could be applied to scheduling in other industries or non-medical personnel in the health care sector.

2.3 Long-term workload equality on duty schedules for physicians in hospitals

Gross (2018) evaluates the trade-off between long-term fairness in preference fulfillment and long-term equal workload distribution. He proposes a workload indicator to measure physicians' workload during each planning horizon and explores different strategies to update the workload indicator between planning horizons. In a computational study, he finds that disregarding long-term equal workload distribution accumulates to unequal treatment of physicians over several planning horizons. The proposed model enables schedulers to manage the trade-off between long-term satisfaction in preference fulfillment and long-term equal workload distribution. This contribution has been published in *PATAT2018: Proceedings of the 12th International Conference on the Practice and Theory of Automated Timetabling*, which is not ranked in the VHB-JOURQUAL3 ranking (Verband der Hochschullehrer für Betriebswirtschaft e.V., 2015). Additionally, this contribution has been submitted to a PATAT special issue of *Annals of Operations Research*, which is ranked in category B in the VHB-JOURQUAL3 ranking. It can be found in its entirety in Appendix C.

In order to provide medical care in emergencies, physicians are assigned to overnight duties via a schedule respecting their preferences. In a previous contribution (Gross et al., 2018a, see Section 2.2), a satisfaction indicator is introduced which ensures fairness in preference fulfillment among physicians over several planning horizons. This ensures that no physician is disadvantaged in the long term with regards to preference fulfillment. However, some duties are not requested by any physician. These duties still need to be assigned in order to satisfy physician demand during the night. A policy usually employed for the assignment of these duties is equal distribution of workload. However, as many physician scheduling models operate on a planning horizon of a single month, it is often impossible to achieve a completely equal distribution of workload within the planning horizon. Therefore, in each planning horizon some physicians will be disadvantaged in comparison to other physicians. If this disadvantage is not compensated in the following planning horizons, it is possible that over a longer time span the workload of some physicians accumulates to a much higher workload than that of other physicians.

In this contribution, a workload indicator to track the workload of an individual physician in a single planning horizon is introduced. This workload indicator is defined as the amount of overnight duties assigned to the physician divided by the total number of days in the planning horizon. This is an approximation of the number of overnight duties a physician is assigned per day. The workload indicator is then incorporated in the physician scheduling model proposed by Gross et al. (2018a). Based on the workload indicator, a workload-based weight is calculated to guide the model towards assigning duties to certain physicians. Simultaneously, the model takes into account physician satisfaction in terms of preference fulfillment by using a satisfaction indicator and a satisfaction-based weight. This enables the model to respect long-term equal preference fulfillment as well as long-term equal workload distribution.

The model defines two types of strategies, one for each of these objectives. For long-term fairness in preference fulfillment, the model offers the unfair (U) strategy, which disables long-term fairness considerations and only maximizes preference fulfillment in each planning horizon. The second long-term fairness strategy for preference fulfillment is the fair (F) strategy, which is equal to the ESD strategy proposed by Gross et al. (2018a). This strategy calculates a satisfaction-based weight for each physician based on the satisfaction indicator for the current plan and optimizes it during model solving. For long-term equal workload distribution, there are also two available strategies. The first strategy uses a constant workload-based weight (CL). This eliminates all considerations of equal workload distribution. The second strategy applies exponential smoothing to the workload-based weight from the previous planning horizon and the workload indicator from the current planning horizon to calculate the workload-based weight for the current planning horizon during model solving.

When running the model, one of each type of strategy—for fairness in long-term equal preference fulfillment as well as for long-term equal workload distribution—has to be selected. The F and ESL strategies both result in a quadratic decision model. As the F strategy is equivalent to the ESD strategy described by Gross et al. (2018a), the linearization described in that work can be applied to this strategy. For the ESL strategy, a linearization is described in the contribution as follows. The amount of assigned duties per physician is bounded by half the number of days in the planning horizon as duties can at most be assigned any other day. As the workload-based weight depends only on parameters and the amount of duties assigned in the current planning horizon, it is possible to pre-calculate all possible values of the workload-based weight for each physician. Based on the workload-based weights, the cost values can be calculated which will be incurred when assigning a certain amount of duties to a given physician. These cost values can then be added to the model as parameters. An additional decision variable which selects the correct cost value can be added and the quadratic term in the objective function can be substituted with a linear term which adds the selected cost value. The model for the ESL strategy can then be solved as a linear decision model.

To evaluate the model's performance in terms of long-term equal preference fulfillment, the APS and ASV indicators by Gross et al. (2018a) are used. Additionally, the APL and ALV performance indicators are introduced to evaluate the performance in terms of long-term equal workload distribution. The APL indicator is defined as the variance over all physicians of the average workload indicator per physician over all planning horizons. Lower values indicate a more equal distribution of workload among all physicians after several planning horizons. The APS indicator is defined as the average over all physicians of the per-physician variance of workload between planning horizons. Lower values of this indicator signify a more stable level of workload for each physician between planning horizons. These indicators are used to evaluate model performance in two computational studies.

The first study uses data presented by Gross et al. (2018a) for the evaluation of their model. This data is generated based on real-life data from a German university hospital. The study uses only the data set with a 0 % conflict rate for preferences and the unfair (U) strategy is selected for preference fulfillment. For workload distribution, the model is run for all 24 months of data with the CL and then the ESL strategy. This effectively disables any optimizations for long-term fairness in regards to preference assignment and only compares the strategy without any workload considerations (CL) with the proposed strategy for long-term equal workload distribution (ESL). As the roster does not consider fairness in preference fulfillment, the APS and ASV indicators are not evaluated because they only evaluate plan quality in terms of preference fulfillment. Results indicate that the ESL strategy is able to improve the values of the APL and ALV indicators distinctly in comparison with the CL strategy. It follows that the ESL strategy can achieve a more equal distribution of workload among all physicians over all planning horizons as well as a more stable level of workload for each individual physician between different planning horizons. Next, all 11 data sets with varying conflict rate between 0 % and 100 % are evaluated with the F-CL and F-ESL strategy combination. However, none of the performance indicators change notably between the different conflict rates. It can therefore be concluded that the conflict rate of preferences and improvements by the workload-based weight are not directly correlated.

For the second study, data with varying preference probability, i.e. the probability with which a physician submits a preference on any given day, is generated. In total, 11 data sets with a preference probability ranging from 0 % to 100 % with increments of 10 percentage points are generated. As before, the models for the F-CL and the F-ESL strategy combinations are run for all 24 months of generated data for each preference probability and all performance indicators are compared. Results indicate that the biggest improvements of the F-ESL over the F-CL strategy for the APL and ALV indicators can be achieved on the data set with a 0 % preference probability. Improvements then decrease with increasing preference probability and converge to a small but still significant improvement for preference probabilities of 50 % and higher. For the APS and ASV indicators, no improvements are possible for the data set with 0 % preference probability as it does not contain any preferences, therefore optimization of preference-based assignments can not affect the performance indicators. For data sets with higher preference probabilities, the improvement in the APS and ASV indicators converges to a small but noticeable improvement and remains at that value for data with preference probabilities of 50 % and more. The fact that no stronger improvement in any performance indicators can be seen for preference probabilities over 50 % is explained by the implicit limit on the number of preferences defined by model constraints: As duties can only be assigned on non-sequential days, preferences can also only be submitted on non-sequential days. Therefore, even if the probability to submit a preference on a day is at 100 %, preferences can only be submitted at most on 50 % of days.

In conclusion, this work introduces an indicator to measure workload distribution among physicians on duty rosters. The presented model incorporates this workload indicator into a physician scheduling model and integrates it with the satisfaction indicator for long-term equal preference fulfillment. Two different strategies for the calculation of a workload-based weight are evaluated: constant workload-based weight (CL), i.e., no tracking of workload, and exponential smoothing of the workload based-weight during the current planning horizon (ESL). As the model for the ESL strategy is quadratic, a linearization for the model is provided. In a computational study, the model is applied to generated data based on data from a German university hospital. Results indicate that the workload indicator and the workload-based weight derived from it are effective in ensuring an equal distribution of workload over several planning horizons. The effectiveness of the workload-based weight does not correlate with the amount of conflicting preferences in the data. However, it does correlate with the amount of preferences and is more effective for data with less preferences.

3 Discussion of the contributions

This thesis comprises the three contributions presented above. Each contribution fills a research gaps in existing literature. This section illustrates how the contributions presented above can answer the following research questions.

1. How can existing physician schedules be adapted in case of unexpected absences?

2. How does rescheduling on physician schedules affect plan quality?

3. Does maximizing preference fulfillment on physician schedules during each planning horizon ensure fairness in the long term?

4. How can long-term fairness in preference fulfillment on physician schedules be guaranteed?

5. How can long-term equal workload distribution on physician schedules be guaranteed?

6. How does long-term equal workload distribution affect long-term fairness in preference fulfillment on physician schedules?

Each research question is answered in a separate section below.

3.1 How can existing physician schedules be adapted in case of unexpected absences?

The problem of rescheduling physicians is a daily nuisance for many schedulers. Due to the lack of literature, physicians are rescheduled manually, resulting in scheduling overhead for the manual scheduler and suboptimal plans for the affected physicians. While there is a huge body on literature on creating physician schedules, none of it considers the problem of rescheduling physicians. Approaches for the rescheduling of nurses exist, but cannot be used to create physician schedules, as physician schedules are usually acyclic and require more diverse qualifications than nurse schedules. The first contribution in this thesis (Gross et al., 2018b, see Section 2.1) explores the physician rescheduling problem.

When rescheduling physicians, the resulting schedule should be of high quality. The new schedule should therefore perform well with regards to the same objective measurements which were applied when creating the initial schedule. Additionally, another factor comes into play: Rescheduling physicians on a plan is simple, but informing the rescheduled physicians of their new assignments is time-consuming. It is therefore imperative to keep the number of rescheduling operations as low as possible. Any rescheduling model must therefore optimize for those two objectives—plan quality and stability. The situation is further complicated by scheduling processes which include the solving of several models sequentially, e.g., to first create a duty roster and then a workstation roster. Not only does this process lead to suboptimal schedules as the model run first can never optimize for objectives from the model run second, it also further constrains the rescheduling problem: Rescheduling on the roster which is created second might be facilitated by being able to reschedule on the roster created first, resulting in better plan quality and lower number of rescheduling operations overall.

The model presented in the first contribution allows the creation of an initial plan as well as the rescheduling of this plan in case of unexpected absences. It enables the assignment of duties and workstations to physicians via an integrated model, thereby guaranteeing to find a schedule which is optimal for both types of assignments. Plan quality goals as well as plan stability are incorporated in the model. A case study illustrates its suitability for practical applications. The contribution also provides a comparison between the integrated model and sequential models and thereby illustrates how interdependencies between separate rosters can not be respected when creating separate models and running them sequentially.

Summarizing, when creating an updated physician schedule, schedule changes should be minimized. Simultaneously, the updated schedule should be optimized for the same plan quality goals as the initial schedule. The contribution provides a model which can be used to adapt physician schedules in case of unexpected absences while optimizing for plan quality as well as plan stability.

3.2 How does rescheduling on physician schedules affect plan quality?

In an ideal world, the schedule would be at least as good in terms of plan quality after rescheduling as it was before. Realistically, this is impossible unless the initial schedule had massive overstaffing which is then reduced by unexpected absences. Most of the time, however, the initial schedule will not have a lot of overstaffing and instead just cover the demand. In this case—assuming the initial schedule is the optimal schedule—the adapted schedule will inevitably be of inferior quality in comparison with the initial schedule. It has less resources to cover the same demand and fulfill plan quality objectives as some physicians are unexpectedly absent. Therefore, the schedule is more highly constrained than the initial schedule and no model can ever find a schedule with higher plan quality than the initial one.

However, the plan quality objectives from the initial plan are just a part of what makes a good adapted schedule. As each rescheduling operation on a plan needs to be followed by communication with the rescheduled physician to inform them of their updated schedule, rescheduling operations are costly for schedulers. A good rescheduled plan should therefore make the minimum number of assignment changes possible. Rescheduling therefore should always optimize for plan quality in terms of objectives defined for the initial schedule as well as stability in terms of a low number of rescheduling operations. Unfortunately, a better fulfillment of plan quality objectives often comes at the cost of a high number of rescheduling operations. It follows that rescheduling models need to make a trade-off between plan quality and plan stability.

This trade-off is modeled in the approach presented in the first contribution. The presented model allows the scheduler to set cost values for plan quality and plan stability objectives. According to these cost values, the resulting plan will perform better in terms of quality or have less rescheduling operations. The contribution explores several different cost settings in a case study. Results indicate that the proposed model effectively allows schedulers to influence the trade-off between plan quality and plan stability. The evaluation of the different cost settings shows that rescheduling costs should be kept balanced with plan quality costs or even higher to keep the number of rescheduling operations low and avoid communication overhead. However, very high rescheduling costs also severely impact plan quality objectives and might even lead to plans with insufficient demand coverage.

Summarizing, rescheduled plans are always at most equally good—but usually worse—in terms of plan quality when compared with the initial plan. Rescheduling is always a trade-off between keeping plan quality high and rescheduling operations low.

3.3 Does maximizing preference fulfillment on physician schedules during each planning horizon ensure fairness in the long term?

Physicians have to perform overnight duties, i.e., stay at the hospital over night, which obviously interferes with their private lives. As physicians are a costly resource at the hospital and high turnover should be avoided, schedulers allow physicians to submit preferences for when they would like to work overnight and when they would prefer to be home at night. To keep physician satisfaction high, scheduling models usually optimize the fulfillment of these preferences. This is often done by optimizing some fairness measure, such as maximizing the amount of fulfilled preferences or minimizing the amount of violated preferences. However, this optimization often only tries to optimize the solution for the current planning horizon. It does not take into account data from previous planning horizons regarding how many preferences have been satisfied for each individual physician.

As planning horizons in physician scheduling are usually a single month or 4 to 5 weeks and due to the placement of preferences, it is often impossible to achieve a completely equal distribution of preference fulfillment among all physicians within one planning horizon. Some physicians might be a bit better off while a few are a bit worse off, even though the model might try to maximize preference fulfillment. One might argue that these random variations should cancel out over time and that after several planning horizons it can be expected that all physicians have been treated equally. However, this is not the case. The second contribution (Gross et al., 2018a, see Section 2.2) investigates this problem. It proposes a physician scheduling model which is evaluated in a case study over 24 planning horizons. The model tries to maximize the number of fulfilled preferences in each planning horizon. Results of the case study indicate that some physicians are consistently treated more favorably than others when there is no long-term fairness guidance in the model. The selection of physicians who are treated better seems to be based on underlying heuristics in the solver implementation.

Summarizing, simply optimizing for maximum preference fulfillment within each planning horizon will inevitably lead to small unfairness within each planning horizon. This unfairness can accumulate over several planning horizons and heavily skew the distribution of preference fulfillment towards a few physicians. To ensure long-term fairness, fairness measurements need to be used which take into account data from previous planning horizons to inform decisions on preference fulfillment in the current planning horizon.

3.4 How can long-term fairness in preference fulfillment on physician schedules be guaranteed?

As discussed above, optimizing for maximum preference fulfillment within each planning horizon does not guarantee long-term fairness over several planning horizons. Instead, small inequalities in treatment within individual planning horizons stack up and lead to heavily unequal treatment in the long term. To counter this effect, physician scheduling models need to take into account data from previous planning horizons when creating a schedule for the current planning horizon. This data needs to reflect the history of how well each individual physician was treated in the past and must be used to guide decisions on whose preferences should be fulfilled in the current planning horizon. Over several planning horizons, this leads to an equal distribution of preferences among all physicians.

The second contribution (Gross et al., 2018a) proposes a physician scheduling model and introduces a satisfaction indicator which quantifies how well an individual physician's preferences have been respected on a schedule. The proposed satisfaction indicator is always a rational number between 0 and 1 and approximates the fulfilled preferences per day in the planning horizon. This indicator can be calculated for each physician and each schedule. There are several strategies how this indicator can be incorporated into the scheduling model: not at all (C strategy)—thereby creating schedules which are unfair in the long term—by updating it after each planning horizon (ESA strategy), or by updating it during the solving of the model for the current planning horizon (ESD strategy). In a case study, all three strategies are evaluated to see which performs best in terms of long-term fairness. Unsurprisingly, the C strategy performs worst as it disregards long-term fairness and creates schedules which are highly unequal in terms of preference fulfillment after several planning horizons. The ESA strategy manages to create schedules which have a more equal distribution of preference fulfillment after several planning horizons, but with low stability of satisfaction for each physician between several planning horizons. The most equal distribution of preference fulfillment after several planning horizons as well as the most stable level of satisfaction between planning horizons is achieved by the ESD strategy.

Summarizing, a measure of how well each physician has been treated in terms of preference fulfillment in the past must be incorporated into scheduling models in order to ensure long-term fairness. This measure must then be used to guide the model towards granting preferences for physicians who have had less fulfilled preferences in the past. The second contribution introduces a satisfaction indicator which exhibits these features. In a case study, results indicate that the proposed satisfaction indicator combined with the ESD strategy can achieve fairness in preference fulfillment after several planning horizons as well as maintain a stable level of satisfaction between planning horizons for each physician.

3.5 How can long-term equal workload distribution on physician schedules be guaranteed?

Another important objective when creating physician schedules is equal workload distribution among all physicians. As discussed above, physicians can submit preferences for when they do or do not want to work. However, not all overnight duties are requested by at least one physician, meaning there are several duties left to assign without any requests for them. Not assigning the duties is not an option as their coverage is required to ensure adequate medical care around the clock. Therefore, a good policy for the assignment of these unrequested duties is assigning them in such a way that the workload is distributed equally among all physicians.

A completely equal distribution of workload among all physicians is often impossible within one planning horizon. This leads to a slightly unequal distribution of workload within each planning horizon, which may accumulate to a heavily unequal distribution of workload in the long term. To avoid this behavior, a measure of how much workload has been assigned to each physician on past schedules must be incorporated in the scheduling model. This workload measure must then be used to ensure that physicians who have been assigned a high workload in the past are assigned a lower workload in the current planning horizon. The third contribution (Gross, 2018, see Section 2.3) proposes a workload indicator to track the workload of each individual physician on a schedule. The workload indicator is defined as the amount of overnight duties a physician performs per day of the planning horizon. Two different strategies are defined for the evaluation of the workload indicator: No long-term equal workload distribution or incorporating the workload indicator into the model by smoothing the workload-based weight derived from it during the solution process. The latter strategy leads to a quadratic decision model, for which a linearization is described. In a case study, the effectiveness of the workload indicator in regards to equal workload distribution after several planning horizons is proven. Results show that using the workload indicator leads to a more equal distribution of workload among all physicians over several planning horizons as well as a more stable level of workload for each individual physician between several planning horizons.

Summarizing, optimizing for equal workload distribution within each planning horizon is not sufficient to ensure equal workload distribution over several planning horizons. A workload indicator needs to be incorporated into the scheduling model to guide the assignment of duties in such a way that physicians with a high workload in the past are assigned less work and vice versa. The third contribution introduces a workload indicator with these properties. A case study confirms that the use of the workload indicator in the model ensure an equal distribution of workload over several planning horizons.

3.6 How does long-term equal workload distribution affect long-term fairness in preference fulfillment on physician schedules?

As we have seen in the above sections, long-term equal preference distribution and long-term equal workload distribution should both be respected when creating physician schedules. However, a schedule that might be optimized for one of these objectives does not necessarily have to perform well in terms of the other objective. It is easy to see how these objectives can be contrary to each other: Physician preferences are often not equally distributed, meaning some physicians will usually request more duties than others. In this case, optimizing for preference fulfillment would obviously be detrimental to the equal workload distribution objective. On the other hand, when there are no or only very few preferences to take into account, equal workload distribution could be achieved more easily.

The model proposed in the third contribution (Gross, 2018) explores this trade-off. It incorporates both a satisfaction indicator for long-term equal preference fulfillment as well as a workload indicator for long-term equal workload distribution. The scheduler is then able to assign weights to these objectives in order to decide whether preference fulfillment or workload distribution is more important when creating schedules. As physicians usually care more about being assigned according to their preferences than they care about being assigned to exactly as many duties as their colleagues, the contribution's case study evaluates a cost setting where the cost for violating equal preference fulfillment is higher than the cost for violating equal workload distribution. The model is then evaluated using data with a varying amount of preferences. Results indicate that the performance regarding equal workload distribution is better for a low amount of preferences. With an increasing amount of preferences, performance in term of equal preference fulfillment improves while performance of equal workload distribution decreases.

Summarizing, long-term equal workload distribution and long-term equal preference fulfillment are incongruent objectives. Depending on how the preferences are selected, optimizing for one will harm the other. It is therefore important for schedulers to decide how they want to make the trade-off between these two objectives. The model proposed in the third contribution allows schedulers to influence this trade-off using cost settings. A case study confirms that the long-term equal preference fulfillment and long-term equal workload distribution objectives optimize for different performance measures and that this manifests more clearly the higher the amount of preferences in the data.

4 Conclusion

This thesis examines several aspects of the physician scheduling problem with consideration of physician preferences. It motivates the need for research on the rescheduling problem as well as the problems of long-term fairness in preference fulfillment and long-term equal workload distribution. The main part of this thesis consists of three contributions to literature. A summary of each of the contributions is provided and the contributions are applied to answer open research questions introduced in this thesis.

The first contribution introduces a rescheduling model for physicians. This is the first model to support rescheduling of physicians in current literature. A case study underlines the effectiveness of this model in making a trade-off decision between plan quality and plan stability when creating updated plans in case of unexpected absences. In the second contribution, a satisfaction indicator for physician preference fulfillment over several planning horizons is proposed. Results of a case study indicate that the use of a satisfaction indicator is mandatory to achieve long-term fairness in preference fulfillment. The contribution proposes a strategy how this satisfaction indicator can be included in linear physician scheduling models. Afterwards, the third contribution examines the long-term equal distribution of workload on physician schedules and introduces a workload indicator. This indicator is effective in measuring the workload and the contribution details how it can be incorporated into physician scheduling models, together with the satisfaction indicator. A case study explores how the trade-off between long-term fairness in terms of equal preference fulfillment and long-term equal workload distribution can be managed depending on the amount of preferences in the input data.

The thesis poses several research questions which have not been answered by existing literature. The first contribution shows how rescheduling can be implemented on physician schedules. It further explores how rescheduling affects plan quality and how to manage the trade-off between plan quality and plan stability. The second contribution illustrates that optimizing for maximum preference fulfillment during each planning horizon does not guarantee equal preference fulfillment over several planning horizons. The proposed satisfaction indicator can be used to ensure that created plans exhibit fairness in preference fulfillment in the long term. In the third contribution, a workload indicator is introduced which is able to distribute workload equally over several planning horizons. It examines the trade-off between long-term fairness in preference fulfillment and long-term equal workload distribution.

The following opportunities for further research are opened up by this thesis. While all contributions propose linear programming models, those usually take several seconds up to several minutes to solve or are not solved to optimality. Depending on the practical application, this time might be too long and heuristic approaches should be considered, which are not investigated in this thesis. Furthermore, the definition of preferences and workload could be expanded. Currently, each preference has the same weight when considering fairness, just as each duty has the same weight when considering workload distribution. In reality, physicians might want to be able to assign different weights to preferences depending on how happy or unhappy they would be with the preference being satisfied or ignored. Similarly, hospitals usually have different kinds of shifts and duties to assign, which might be assigned different weights for workload depending on when the duty needs to be performed and how many hours it comprises. Physician education is only considered in the first contribution, but no long-term considerations are included. Further work should examine how physician schedules can ensure equal educational progress among all physicians over several planning horizons. The proposed indicators in the second and third contribution could be adapted for this task.

Bibliography

Baker, K. R. (1974). Scheduling full-time and part-time staff to meet cyclic requirements. *Journal of the Operational Research Society*, 25(1):65–76.

Bertsimas, D., Farias, V. F., and Trichakis, N. (2011). The price of fairness. *Operations Research*, 59(1):17–31.

Cheang, B., Li, H., Lim, A., and Rodrigues, B. (2003). Nurse rostering problems — a bibliographic survey. *European Journal of Operational Research*, 151(3):447–460.

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

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

Gierl, L., Pollwein, B., Heidenreich, D., and Heyde, G. (1986). Personaleinsatzplanung für die Anästhesiologie mit Hilfe eines Expertensystems (PEP). In Horbach, L., editor, *Medizinische Informatik in der Schweiz*, pages 211–217, Basel u.a. Schwabe.

Gross, C. N. (2018). Long-term workload equality on duty schedules for physicians in hospitals. In Burke, E. K., Di Gaspero, L., McCollum, B., Musliu, N., and Özcan, E., editors, *PATAT 2018: Proceedings of the 12th International Conference on the Practice and Theory of Automated Timetabling*, pages 183–197.

Gross, C. N., Brunner, J. O., and Blobner, M. (2018a). Hospital physicians can't get no long-term satisfaction — an indicator for fairness in preference fulfillment on duty schedules. *Health Care Management Science*.

Gross, C. N., Fügener, A., and Brunner, J. O. (2018b). Online rescheduling of physicians in hospitals. *Flexible Services and Manufacturing Journal*, 30(1):296–328.

Guth, A. H. (2001). Time since the beginning. *Astrophysical Ages and Time Scales, ASP Conference Series*, 245:3–17.

Karsu, Ö. and Morton, A. (2015). Inequity averse optimization in operational research. *European Journal of Operational Research*, 245(2):343–359.

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

Verband der Hochschullehrer für Betriebswirtschaft e.V. (2015). Complete list of the journals in VHB-JOURQUAL3 in alphabetical order. https://vhbonline.org/en/service/jourqual/vhb-jourqual-3/complete-list-of-the-journals/, accessed on November 16, 2018.

Appendix A

Online rescheduling of physicians in hospitals

The following contribution (Gross et al., 2018b) has been published in "Flexible Services and Manufacturing Journal", which is ranked in category B in the VHB-JOURQUAL3 ranking (Verband der Hochschullehrer für Betriebswirtschaft e.V., 2015). The submitted version reproduced below in its entirety.

Online rescheduling of physicians in hospitals

Christopher N. Gross (corresponding author)

Andreas Fügener

Jens O. Brunner

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

Department of Health Care Operations / Health Information Management

Faculty of Business and Economics

University of Augsburg

Universitätsstraße 16, 86159 Augsburg, Germany

e-mail: christopher.gross@unikat.uni-augsburg.de

Submission date: February 25, 2016

Revision date: October 27, 2016

The final publication is available at

https://link.springer.com/article/10.1007/s10696-016-9274-2

Online rescheduling of physicians in hospitals

Abstract

Scheduling physicians is a complex task. Legal requirements, different levels of qualification, and preferences for different working hours increase the difficulty of determining a solution that simultaneously fulfills all requirements. Unplanned absences, e.g., due to illness, additionally drive the complexity. In this study, we discuss an approach to deal with the following trade-off. Changes to the existing plan should be kept as small as possible. However, an updated plan should still meet the requirements regarding work regulation, qualifications needed, and physician preferences. We present a mixed-integer linear programming model to create updated duty and workstation rosters simultaneously following absences of scheduled personnel. To enable a comparison with previous sequential approaches, we separate our model into two models for the duty and workstation roster which generate plans sequentially. In a case study, we apply our integrated and sequential models to real-life data from a German university hospital with 133 physicians, 17 duties, and 20 workstations. We consider a planning horizon of 4 weeks and reschedule physicians on each day for three different cost settings for the trade-off between plan quality (in terms of preferences, fairness, coverage and training) and plan stability, resulting in a total of 4,201 model runs. We demonstrate that our integrated model can achieve near-optimal results with reasonable computational efforts. In each of these runs our model reschedules physicians within 1 to 21 seconds. We run the sequential models on the same data, but for only one cost setting, resulting in 1,401 runs. The results indicate that our integrated model manages to respect interdependencies between duty and workstation roster whereas the sequential models will always optimize for the plan which is created first. Overall, results indicate that our integrated model parameters allow managing the trade-off between plan quality goals and plan stability.

Keywords: OR in health services; Physician Rescheduling; Online planning; Mixed-integer linear program

Submission: February 25, 2016

Revised: October 27, 2016

1 Introduction

In recent years competition in the health care market has increased. This forces hospitals to reduce their operating costs in order to remain competitive. Since personnel costs make up almost 60% of a hospital's total costs (Bölt, 2015), an easy way of cutting costs is to improve personnel usage by scheduling staff more efficiently. This can increase staff efficiency and reduce the number of required personnel. Therefore the problem of shift scheduling for physicians has gained increasing attention in the last 30 years (Erhard et al., 2016).

The task of physician scheduling is a complex one. This is due to the multitude of different qualifications a physician can have and the need to service a hospital 24 hours a day. Different assignments may require physicians with different qualifications. Additionally, physicians' working times are bound by legal regulations, e.g., concerning maximum working time and minimum staffing levels. Finding a schedule that meets the demand for physicians of a certain qualification while simultaneously complying with said legal requirements is a complex problem (Brunner et al., 2009).

As physicians are a scarce resource, hospitals have a high incentive to keep them satisfied. An important factor in employee satisfaction is fairness (Stolletz and Brunner, 2012). One example to include fairness in physician scheduling is to enforce an even distribution of unpopular shifts among all physicians. Another possibility of including fairness in the schedule is providing physicians with the ability to influence the schedule by expressing preferences for certain shifts or tasks. The consideration of preferences and fairness further complicates the creation of a valid duty roster.

Existing work focuses on the creation of new schedules. However, already created schedules often need to be adjusted. This frequently happens when employees assigned to certain shifts become unavailable due to various reasons, e.g., illness. In this case rescheduling needs to fill the gaps in the plan to make sure that staff demand is still met. Poor rescheduling can lead to reduced quality of care, increased operating cost and sinking staff morale (Clark et al., 2015). Consequently, Petrovic and Vanden Berghe (2012) find that rescheduling capability of a staff scheduling model is an important factor in determining the model's applicability in practice and Bai et al. (2016) conclude that rescheduling of physicians is a promising future research area. Changes to the plan induce significant management effort (Clark et al., 2014), e.g., because other employees who need to be present in the unavailable person's stead must be contacted and notified of the changed schedule. To reduce this overhead, changes to the plan should be minimal.

Similar to Fügener et al. (2015), we differentiate between duties (i.e., assignments where physicians have to work longer hours than regular) and workstations (i.e., concrete workplaces where physicians work during their regular working hours). We propose an approach to generate an updated version of both a duty and a workstation roster based on a previous plan and incorporating additional, previously unknown and unplanned absences. Our approach models the following trade-off. First, we want to obtain a high plan quality for our roster as measured by the following goals. For duties we wish to incorporate physicians' *preferences* (either for a specific duty or for not being on duty on specific days) and *fairness* (an even distribution of duties among all physicians). For workstations we wish to consider *training* goals (to assign physicians to workstations where they should work according to their training plan) and a good *coverage* of all workstations (no understaffing according to minimum demand levels). Second, we aim for plan *stability*, i.e., we try to minimize deviations from the previously published plan as much as possible. We solve this problem using a mixed-integer program with hard and soft constraints. Note that while we focus on absences, our approach could also cope with changed demand constraints.

Our contribution is threefold. We believe that each part of our contribution closes a different research gap. First, we propose an integrated model to simultaneously tackle the problems of duty rostering (when does which physician have to work longer) and workstation rostering (when does which physician have to work at which workstation). This avoids local optima, i.e., each assignment's consequences are always considered for both rosters. To our knowledge, we are the first to propose an integrated model for simultaneous duty and workstation rostering. We also create a sequential formulation of our model which creates a duty roster first and a workstation roster afterwards for comparison with the integrated model. Second, we propose the first approach for rescheduling physicians in hospitals. Some approaches for nurse rescheduling exist (Clark et al., 2015). However, they do not incorporate the trade-off between plan quality and plan stability we described above. Third, we apply our model with data from a German hospital. We explore different parameter values for rescheduling penalties where rescheduling (plan stability) costs are LOW, BALANCED, or HIGH in relation to other plan quality cost values and draw managerial insights. We believe we are the first to explore the trade-off between plan quality and plan stability during rescheduling of physicians and during rescheduling in the health care sector in general. Our data set is obtained from a German university hospital with 133 physicians, 17 duties and 20 workstations. We consider a 4-week planning horizon which contains 243 duty requests and 140 no-duty requests. We first use our model to generate an initial plan and then randomly draw absences for physicians on each day. Our model is then used to reschedule physicians on each day for each cost setting. We run 50 iterations of this study, resulting in a total of 4,201 runs. In this study we achieve good rescheduling results within 1 to 21 seconds. We compare our integrated model with a sequential version of itself. The sequential models are run on the same data set, resulting in 1,401 runs. The data shows that the sequential models will always produce better solutions for the plan which is created first, whereas the integrated model allows fine-tuning the trade-off between different plans and between plan quality and plan stability using cost parameter values. In particular, training aspects are considered by far less in the sequential models even though they are of high relevance for physicians and the attraction of the hospital. We suggest using the integrated model with sufficiently high cost values for rescheduling in order to reduce rescheduling-induced management overhead.

The remainder of this paper is structured as follows. Section 2 gives an overview of existing literature regarding physician scheduling. As no literature currently tackles the problem of physician rescheduling, we discuss literature concerning the related problem of nurse rescheduling. Section 3 introduces our mixed-integer linear programming model for physician rescheduling and its sequential formulation. The performance of the models is tested by a case study in Section 4. Finally, Section 5 summarizes our findings and discusses future research opportunities.

2 Related Literature

In this section we provide an overview of existing literature on scheduling and rescheduling of nurses and physicians. Personnel scheduling has been a focus of research for several decades. A recent literature review by van den Bergh et al. (2013) summarizes and categorizes existing literature on personnel scheduling. Most literature on health care personnel scheduling focuses on nurses. Burke et al. (2004) provide a detailed review of existing nurse scheduling approaches. A smaller body of literature exists on physician scheduling. For a more extensive overview we refer to Erhard et al. (2016). As to our knowledge there are no studies in the OR/MS literature on physician rescheduling, we focus on two directions in this section. First, we discuss relevant papers on incorporating preferences, fairness, and training requirements in physician scheduling. Second, as no previous work on physician rescheduling exists, we review the related, yet less complicated, setting of nurse rescheduling.

2.1 Physician Scheduling

All research on physician scheduling has been published in the last 30 years, most of it within the last 10 years (Erhard et al., 2016). In this section we focus on a few papers that discuss either consideration of fairness and preferences or training requirements in physician scheduling.

There are few papers that discuss fairness and preferences (for a discussion of fairness on a larger scope we refer to Bertsimas et al. (2011)). One possible interpretation of fairness is equality of physicians, e.g., each physician is assigned to the same number of overnight duties. Stolletz and Brunner (2012) discuss different measures of fairness, such as equal distribution of on-call duties and working hours, in physician scheduling. They compare the classical set covering approach with pre-defined shifts to an implicit formulation, where shift generation is modeled as constraints. Shifts are defined as assignments of physicians to demand periods. Baum et al. (2014) discuss the assignment of physicians in a radiology division under revenue and fairness aspects. They incorporate fairness regarding revenue with a maxmin approach, where the value of the least well-off is maximized. Besides, desirable and burdensome duties should be evenly assigned among the physicians. The authors apply a mixed-integer program and demonstrate that their approach allows significant revenue increases under strict fairness constraints. Gunawan and Lau (2013) apply an approach where physicians need to give preferences for each of their shifts. Using two different models, they assign duties to physicians so that the number of unscheduled duties is minimized and physicians' preferences are respected or so that economic constraints are obeyed. Running their model on random instances, they find that the model can find schedules with less than 3% of duties being unscheduled. Fügener et al. (2015) differentiate between duty and workstation rostering. The authors discuss two separate models: The duty roster combines fairness (in terms of even assignment of overnight duties) and preferences (of certain duty or no-duty requests), while the workstation roster focuses on the connection to the duty roster (assignment of all physicians on duty) and training requirements (assignments based on a long-term training plan).

Some papers discuss training aspects. There is a subset of the physician scheduling literature focusing on resident scheduling. Resident scheduling problems may be differentiated into the long-term rotation assignment problem, the mid-term shift scheduling problem, and the short-term task assignment problem. We introduce three papers that discuss resident shift scheduling problems which are similar to our problem setting. Different seniority levels, demand, and preferences for work-nights are discussed in Sherali et al. (2002). The authors formulate a mixed-integer program and solve it exactly using CPLEX and heuristically exploiting network structures of the problem. Three scenarios representing different demand structures are analyzed. The authors conclude that the best approach depends on the scenario. Topaloglu and Ozkarahan (2011) integrate quality of training, resident well-being, and quality of care in a mixed-integer programming model. They define hard constraints, such as regulations regarding working hours, and soft constraints, such as hospital demand for residents. They propose a column generation-based solution approach, where the feasible schedules are detected using constraint programming. On-call shifts during training periods and regular shifts of Oncology and Hematology are assigned to residents in Elomri et al. (2015). As in the previous paper, hard constraints, such as compliance with legislation, and soft constraints, such as equal assignment of workload, are considered. The mixed-integer goal programming approach can be solved with standard software within few seconds.

2.2 Rescheduling approaches in health care

While there is no literature on physician rescheduling, the problem of nurse rescheduling has been addressed in a few publications. We summarize the most relevant articles in this section, a review on nurse rescheduling is provided in Clark et al. (2015).

Most research in this area has been done in the past 15 years. As one of the first researchers in this field, Moz and Pato (2003) tackle the nurse rescheduling problem with an integer multicommodity flow model. In their succeeding research, they improve the flow model (Moz and Pato, 2004) and later develop a genetic algorithm (Moz and Pato, 2007) to solve the problem. The models try to keep changes to the plan minimal but do not consider the quality of the generated plan for the nurses affected by rescheduling.

Clark and Walker (2011) find that the acceptance of a schedule is tied to the realization of nurses' preferences for shifts. They therefore propose models that incorporate nurse preferences to switch certain shifts. Depending on each individual nurse's weight-ing factor, the model takes into account nurse preferences, nurse surpluses and shortages, while also trying to minimize the deviation from the existing plan. Kitada et al. (2010) tackle the nurse rescheduling problem with a recursive heuristic algorithm. When a nurse becomes absent, they try to assign the next available nurse with the same skill level and then repeatedly reschedule until the plan is free of conflicts. Instead of running the rescheduling process when it is required, we could also run it cyclically. Bard and Purnomo (2005) propose a model that is run every 8 hours. This decouples the model execution start time from changes in the personnel structure, i.e., the model is not necessarily run separately for each newly reported absence. The schedule generated by the model spans up to three shifts, depending on how much the data changed.

We conclude that our approach is to the best of our knowledge the first paper discussing physician rescheduling, the first paper integrating duty and workstation rostering, and the first paper considering training plans in workstation rostering. Even though some articles consider fairness and preferences, none of the discussed studies include the preferences, fairness, and training requirements simultaneously. Furthermore, we could not find any literature discussing the trade-off between plan quality and plan stability as described in our approach.

3 Model

We present a model to reschedule physicians on duty and workstation rosters. First, we introduce our model and then explore similarities and differences to an approach by Fügener et al. (2015).

Our model updates a schedule concerning a set of physicians $j \in \mathcal{J}$ during a set of sequential weeks $w \in \mathcal{W}$, with each week consisting of seven days $t \in \mathcal{T}$. The number of weeks to schedule is adjustable. Physicians are assigned to duties $i \in \mathcal{I}$ and workstations $h \in \mathcal{H}$. Next, we describe the logic of creating and updating rosters.

Our model defines a duty roster (decision variables x_{jiwt}) which assigns physicians to duties on a given day. The model is usually run for an entire month, spanning 4 to 6 weeks. On each day, each duty has a certain demand $\bar{d}_{it}^{\text{duty}}$ for physicians. Each physician holds a set of qualifications, i.e., binary parameter E_{je}^{phy} is 1 if physician j holds qualification e. Each duty requires a set of qualifications of which at least one must be held by the physicians to be assigned to this duty, i.e., binary parameter E_{ie}^{duty} is 1 if duty i requires qualification e. Physicians are considered qualified for a duty when they hold at least one of the qualifications which are required by the duty. Our model assigns physicians to duties according to their qualifications until the demand is met. For example, let physician 1 hold qualifications 1 and 2 ($E_{1,1}^{\text{phy}} = 1$ and $E_{1,2}^{\text{phy}} = 1$) and let physician 2 hold qualification 1 ($E_{2,1}^{\text{phy}} = 1$). Let duty 1 require qualification 2 or 3 ($E_{1,2}^{\text{duty}} = 1$ and $E_{1,3}^{\text{duty}} = 1$). Physician 1 is then qualified for duty 1, physician 2 is not. Additionally, preferences for specific duties (parameters $g_{jiwt}^{\text{req-on}}$) or for not being assigned to any duty (parameters $g_{jwt}^{\text{req-off}}$) and fairness (in terms of exceeding a maximum number g^{24h} of duty assignments per physician) are considered during duty roster creation.

Simultaneously, physicians who are present are assigned to a workstation on each day via a workstation roster (decision variables y_{jhwt}). To advance their education, physicians are required to perform a specified amount of certain procedures. As the likelihood of performing a procedure differs between workstations, it is in physicians' interests to be assigned to a workstation where they are likely to perform procedures which they still require for their education. For example, to become an anesthesia specialist, physicians need to perform, among other procedures, a certain number of narcoses on children. As many children visit the hospital to have their tonsils taken out, the likelihood of performing a narcosis on a child is higher in the operating room where otorhinolaryngology (ear, nose and throat, ENT) surgeries are performed. A provided training plan (parameters g_{jhp}^{station}) gives priorities to which workstation each physician should ideally be assigned. The model assigns physicians preferably to the workstations with the highest priority in the training plan (with the highest priority being 1), then to the next-lowest priority and so on and afterwards to workstations which are not in the physician's training plan. Continuing our example from above, physicians who still need to perform child narcoses will be assigned to the otorhinolaryngology operating room in the training plan with priority 1. This assignment in the training plan will then make the model prefer assigning this physician to the otorhinolaryngology operating room during workstation roster creation. Furthermore, each workstation has a minimum (parameters $\underline{d}_{ht}^{\text{station}}$) and a maximum demand (parameters $\bar{d}_{ht}^{\text{station}}$) for physicians on each day. The dependencies between the different entities and plans are visualized in Figure 1.

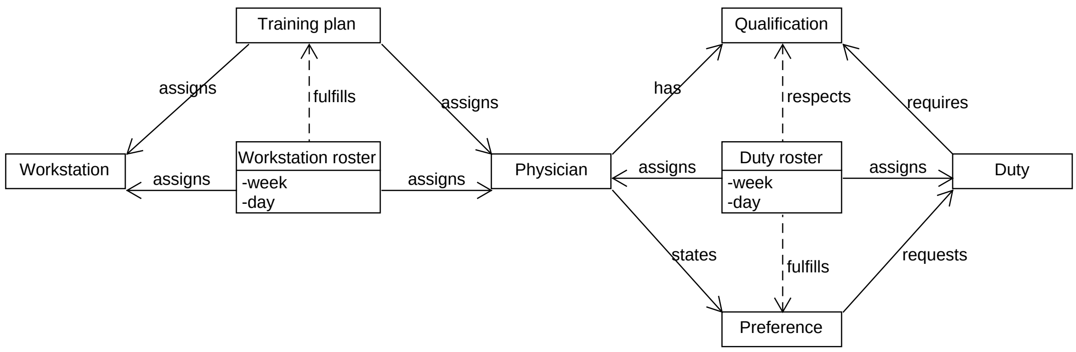


Figure 1: Dependencies between physicians, workstations, duties and the respective plans/rosters

To avoid local optima by considering the models sequentially, our model creates and updates the duty and workstation rosters simultaneously. This avoids situations where the sequential execution of separate duty and workstation models create problems for our goals in the model which is run last. It is, for example, possible that the created duty roster respects physician preferences and thereby creates duty assignments in such a way that it is impossible for the workstation roster model to create a good solution respecting the training plan. If several physicians which are assigned to the same workstation with priority 1 in the training plan are granted a request for an overnight duty on the same day, they will all be given a day off on the following day. This means there will not be enough physicians with that workstation in their training plan present on the following day, meaning the workstation roster model will have to assign physicians to that workstation who are assigned to a different workstation in their training plan, thereby violating those physicians' training plan. While we elaborate this argument only for preferences and overnight duties, it is equally valid for 4-day late duties (see definition in Section 3.1) and fairness considerations regarding duty distribution among physicians. Evidence of these problems with local optima can be found in the results of our case study in Section 4.2.

The proposed model is loosely based on an approach by Fügener et al. (2015). We highlight the major similarities and differences. Their approach also assigns physicians to duties and workstations. However, they propose two separate models which leads to problems with local optima as described above. The main additional contributions of our model are its rescheduling capability and the integration of both models into one. To enable a comparison of our model with sequential approaches, we also present a sequential version of our integrated model in Section 3.2. The models discussed in Fügener et al. (2015) only support the initial creation of plans. Should these plans need to be changed, the planner was left with the choice of running the model again and creating a new plan (which might be completely different from the previous plan) or fixing undercoverage problems manually. Both options are undesirable. A large amount of changes in the plan induces a significant management overhead and staff dissatisfaction. Manual changes are not trivial and often generate a suboptimal solution, eventually also leading to staff dissatisfaction because of violation of preferences or the training plan. Our model enables the planner to reschedule physicians based on an existing plan. The trade-off between plan quality and plan stability can be influenced by setting our model parameters. However, our model still supports the initial generation of plans. We provide the required parameters for initial plan creation in Section 3.4. Furthermore, the formulation for the training plan is less flexible than in our model.

3.1 Model formulation

We now introduce our model. First, we explain duty categories and notation, then the mathematical formulation of our model follows. Lastly, we discuss the terms of the objective function and all constraints. Our model groups duties into the following categories:

4-day late duties ($\mathcal{I}^{\textbf{late}}$) These duties cover tasks after regular operating hours. They are always assigned for 4 days a week with the fifth working day in that week being a day off. This day off pattern is chosen because the late duty leads to a working time of 10 hours a day. After 4 days of late duty, physicians have already worked their regular weekly working time of 40 hours, so they are given the remaining day in the week off to prevent overtime buildup.

Overnight duties ($\mathcal{I}^{\textbf{24h}}$) Overnight duties require the employee to be present at the hospital overnight, resulting in a total working time of 24 sequential hours (i.e., eight hours of regular work followed by an on-call period of 16 hours). These duties require that the employee is given a day off after their shift.

We use the following notation in our model.

Sets and indices	
$i \in \mathcal{I}$	Duties
$i \in \mathcal{I}^{\text{late}} \subseteq \mathcal{I}$	4-day late duties
$i \in \mathcal{I}^{\text{24h}} \subseteq \mathcal{I}$	Overnight duties
$h \in \mathcal{H}$	Workstations
$j \in \mathcal{J}$	Physicians
$e \in \mathcal{E}$	Qualifications/Experiences
$p \in \mathcal{P}$	Priorities
$w \in \mathcal{W}$	Weeks in the planning horizon
$t \in \mathcal{T} = \{1, ..., 7\}$	Days in a week
$t \in \mathcal{T}^{\text{work}} = \{1, ..., 5\}$	Working days in a week
Parameters	
$T_i^{\text{fix}} \in \mathcal{T}^{\text{work}}$	day off for duty $i \in \mathcal{I}^{\text{late}}$
T_{jwt}^{off}	1 if physician j is absent on day t in week w, 0 otherwise

E_{ie}^{duty}	1 if duty i requires qualification e, 0 otherwise
E_{je}^{phy}	1 if physician j has qualification e, 0 otherwise
X_{jiwt}	1 if physician j is assigned to duty i on day t of week w in the existing roster, 0 otherwise
Y_{jhwt}	1 if physician j is assigned to workstation h on day t of week w in the existing roster, 0 otherwise
$t_0 \in \mathcal{T}$	current day
$w_0 \in \mathcal{W}$	current week
$g_{jiwt}^{\text{req-on}}$	1 if physician j requests duty i on day t in week w, 0 otherwise
$g_{jwt}^{\text{req-off}}$	1 if physician j requests to be assigned no duty on day t in week w, 0 otherwise
g^{24h}	Maximum number of overnight duties to be assigned to a single physician in a single week
g_{jhp}^{station}	1 if physician j is assigned to workstation h with priority p in the training plan, 0 otherwise
$\bar{d}_{it}^{\text{duty}}$	Maximum demand of physicians for duty i on day t
$\bar{d}_{ht}^{\text{station}}$	Maximum demand of physicians for workstation h on day t
$\underline{d}_{ht}^{\text{station}}$	Minimum demand of physicians for workstation h on day t
$c^{\text{req-on}}$	Cost for violating a request for a duty
$c^{\text{req-off}}$	Cost for violating a request for no duty
c^{24h}	Cost for assigning too many overnight duties to a single physician
c^{late}	Cost per day for not assigning late duty to a physician who is assigned this late duty for the week
$c^{\text{station-dem}}$	Cost per physician missing to satisfy minimum demand on a workstation
$c^{\text{re-duty}}$	Cost for rescheduling assignment to a duty
$c^{\text{re-station}}$	Cost for rescheduling assignment to a workstation

r^{duty}	Reward for assigning a physician to a duty
r^{station}	Reward for assigning a physician to a workstation
$r_p^{\text{station-plan}}$	Reward for assigning a physician to a workstation as planned in the training plan with priority p
Decision variables	
x_{jiwt}	1 if physician j is assigned duty i on day t of week w, 0 otherwise
y_{jhwt}	1 if physician j is assigned to workstation h on day t of week w, 0 otherwise
z_{jiw}	1 if physician j is assigned duty $i \in \mathcal{I}^{\text{late}}$ in week w, 0 otherwise
$\Delta_{jiwt}^{\text{req-on}}$	1 if request by physician j for duty i on day t in week w is violated, 0 otherwise
$\Delta_{jwt}^{\text{req-off}}$	1 if request by physician j for no duty on day t in week w is violated, 0 otherwise
Δ_j^{24h}	Number of overnight duties assigned to physician j exceeding g^{24h}
$\Delta_{jiw}^{\text{late+}}$	Number of late duties i assigned to physician j even though physician is not supposed to do late duty in week w ($z_{jiw} = 0$)
$\Delta_{jiw}^{\text{late}-}$	Number of late duties i not assigned to physician j even though physician is supposed to do late duty in week w ($z_{jiw} = 1$)
$\Delta_{hwt}^{\text{station-dem}}$	Number of physicians missing to satisfy minimum demand for workstation h on day t in week w
$\Delta_{jiwt}^{\text{duty+}}$	1 if physician j is assigned to duty i on day t of week w but was not assigned to this duty in the existing roster ($X_{jiwt} = 0$), 0 otherwise
$\Delta_{jiwt}^{\text{duty}-}$	1 if physician j was assigned to duty i on day t of week w in the existing roster ($X_{jiwt} = 1$) but is no longer assigned to this duty, 0 otherwise

$\Delta^{\text{station+}}_{jhwt}$	1 if physician j is assigned to workstation h on day t of week w but was not assigned to this workstation in the existing roster ($Y_{jhwt} = 0$), 0 otherwise
$\Delta^{\text{station}-}_{jhwt}$	1 if physician j was assigned to workstation h on day t of week w in the existing roster ($Y_{jhwt} = 1$) but is no longer assigned to this workstation, 0 otherwise

Maximize

$$\sum_{j\in\mathcal{J}}\sum_{i\in\mathcal{I}}\sum_{w\in\mathcal{W}}\sum_{t\in\mathcal{T}}\left(r^{\text{duty}}\cdot x_{jiwt}\right)+\sum_{j\in\mathcal{J}}\sum_{h\in\mathcal{H}}\sum_{w\in\mathcal{W}}\sum_{t\in\mathcal{T}}\left(r^{\text{station}}\cdot y_{jhwt}\right)+ \tag{1a}$$

$$\sum_{j\in\mathcal{J}}\sum_{h\in\mathcal{H}}\sum_{w\in\mathcal{W}}\sum_{t\in\mathcal{T}}\sum_{p\in\mathcal{P}}\left(r^{\text{station-plan}}_{p}\cdot g^{\text{station}}_{jhp}\cdot y_{jhwt}\right)- \tag{1b}$$

$$\sum_{j\in\mathcal{J}}\sum_{i\in\mathcal{I}}\sum_{w\in\mathcal{W}}\sum_{t\in\mathcal{T}}\left(c^{\text{req-on}}\cdot\Delta^{\text{req-on}}_{jiwt}\right)-\sum_{j\in\mathcal{J}}\sum_{w\in\mathcal{W}}\sum_{t\in\mathcal{T}}\left(c^{\text{req-off}}\cdot\Delta^{\text{req-off}}_{jwt}\right)- \tag{1c}$$

$$\sum_{j\in\mathcal{J}}\left(c^{\text{24h}}\cdot\Delta^{\text{24h}}_{j}\right)-\sum_{j\in\mathcal{J}}\sum_{i\in\mathcal{I}^{\text{late}}}\sum_{w\in\mathcal{W}}\left(c^{\text{late}}\cdot\left(\Delta^{\text{late+}}_{jiw}+\Delta^{\text{late}-}_{jiw}\right)\right)- \tag{1d}$$

$$\sum_{h\in\mathcal{H}}\sum_{w\in\mathcal{W}}\sum_{t\in\mathcal{T}}\left(c^{\text{station-dem}}\cdot\Delta^{\text{station-dem}}_{hwt}\right)- \tag{1e}$$

$$\sum_{j\in\mathcal{J}}\sum_{i\in\mathcal{I}}\sum_{w\in\mathcal{W}}\sum_{t\in\mathcal{T}}\left(c^{\text{re-duty}}\cdot\left(\Delta^{\text{duty+}}_{jiwt}+\Delta^{\text{duty}-}_{jiwt}\right)\right)- \tag{1f}$$

$$\sum_{j\in\mathcal{J}}\sum_{h\in\mathcal{H}}\sum_{w\in\mathcal{W}}\sum_{t\in\mathcal{T}}\left(c^{\text{re-station}}\cdot\left(\Delta^{\text{station+}}_{jhwt}+\Delta^{\text{station}-}_{jhwt}\right)\right) \tag{1g}$$

subject to

$$\sum_{j\in\mathcal{J}}x_{jiwt}\le\bar{d}^{\text{duty}}_{it}\quad\forall i\in\mathcal{I},w\in\mathcal{W},t\in\mathcal{T} \tag{2}$$

$$x_{jiwt}\le\sum_{e\in\mathcal{E}}\left(E^{\text{phy}}_{je}\cdot E^{\text{duty}}_{ie}\right)\quad\forall j\in\mathcal{J},i\in\mathcal{I},w\in\mathcal{W},t\in\mathcal{T} \tag{3}$$

$$x_{jiwt}+\Delta^{\text{req-on}}_{jiwt}\ge1\quad\forall j\in\mathcal{J},i\in\mathcal{I},w\in\mathcal{W},t\in\mathcal{T},g^{\text{req-on}}_{jiwt}=1 \tag{4}$$

$$\sum_{i\in\mathcal{I}}x_{jiwt}-\Delta^{\text{req-off}}_{jwt}\le0\quad\forall j\in\mathcal{J},w\in\mathcal{W},t\in\mathcal{T},g^{\text{req-off}}_{jwt}=1 \tag{5}$$

$$\sum_{i \in \mathcal{I}} x_{jiwt} \leq 1 - T^{\text{off}}_{jwt} \quad \forall j \in \mathcal{J}, w \in \mathcal{W}, t \in \mathcal{T}, w > w_0 \vee (w = w_0 \wedge t \geq t_0) \tag{6}$$

$$\begin{aligned} \sum_{i \in \mathcal{I}^{24\text{h}}} \left(3 \cdot x_{jiw(t-1)}\right) \leq 3 - \sum_{i \in \mathcal{I}} x_{jiwt} - \sum_{h \in \mathcal{H}} y_{jhwt} - T^{\text{off}}_{jwt} \\ \forall j \in \mathcal{J}, w \in \mathcal{W}, t \in \mathcal{T}, t-1 > 0, w > w_0 \vee (w = w_0 \wedge t > t_0) \end{aligned} \tag{7a}$$

$$\sum_{i \in \mathcal{I}} (2 \cdot x_{jiw_0t_0}) \leq 2 - T^{\text{off}}_{jw_0t_0} - \sum_{i \in \mathcal{I}^{24\text{h}}} x_{jiw_0(t_0-1)} \quad \forall j \in \mathcal{J}, t_0 > 1 \tag{7b}$$

$$\sum_{h \in \mathcal{H}} (2 \cdot y_{jhw_0t_0}) \leq 2 - T^{\text{off}}_{jw_0t_0} - \sum_{i \in \mathcal{I}^{24\text{h}}} x_{jiw_0(t_0-1)} \quad \forall j \in \mathcal{J}, t_0 > 1 \tag{7c}$$

$$\begin{aligned} \sum_{i \in \mathcal{I}^{24\text{h}}} \left(3 \cdot x_{ji(w-1)7}\right) \leq 3 - \sum_{i \in \mathcal{I}} x_{jiw,1} - \sum_{h \in \mathcal{H}} y_{jhw,1} - T^{\text{off}}_{jw,1} \\ \forall j \in \mathcal{J}, w \in \mathcal{W}, w > w_0 \end{aligned} \tag{7d}$$

$$\sum_{i \in \mathcal{I}} (2 \cdot x_{jiw_0,1}) \leq 2 - T^{\text{off}}_{jw_0,1} - \sum_{i \in \mathcal{I}^{24\text{h}}} x_{ji(w_0-1)7} \quad \forall j \in \mathcal{J}, w_0 > 1 \tag{7e}$$

$$\sum_{h \in \mathcal{H}} (2 \cdot y_{jhw_0,1}) \leq 2 - T^{\text{off}}_{jw_0,1} - \sum_{i \in \mathcal{I}^{24\text{h}}} x_{ji(w_0-1)7} \quad \forall j \in \mathcal{J}, w_0 > 1 \tag{7f}$$

$$\sum_{j \in \mathcal{J}} z_{jiw} = 1 \quad \forall i \in \mathcal{I}^{\text{late}}, w \in \mathcal{W} \tag{8}$$

$$\sum_{i \in \mathcal{I}^{24\text{h}}} x_{jiwt} + \sum_{i \in \mathcal{I}^{\text{late}}} z_{jiw} \leq 1 \quad \forall j \in \mathcal{J}, w \in \mathcal{W}, t \in \mathcal{T} \tag{9}$$

$$\sum_{i \in \mathcal{I}^{24\text{h}}} x_{ji(w-1)7} + \sum_{i \in \mathcal{I}^{\text{late}}} z_{jiw} \leq 1 \quad \forall j \in \mathcal{J}, w \in \mathcal{W}, w > 1 \tag{10}$$

$$\sum_{t \in \mathcal{T}^{\text{work}} \setminus \left\{T^{\text{fix}}_i\right\}} x_{jiwt} - \Delta^{\text{late+}}_{jiw} + \Delta^{\text{late}-}_{jiw} = 4 \cdot z_{jiw} \quad \forall j \in \mathcal{J}, i \in \mathcal{I}^{\text{late}}, w \in \mathcal{W} \tag{11}$$

$$\begin{aligned} 3 \cdot z_{jiw} \leq 3 - \sum_{i' \in \mathcal{I}} x_{ji'wT^{\text{fix}}_i} - \sum_{h \in \mathcal{H}} y_{jhwT^{\text{fix}}_i} - T^{\text{off}}_{jwT^{\text{fix}}_i} \\ \forall j \in \mathcal{J}, i \in \mathcal{I}^{\text{late}}, w \in \mathcal{W}, w > w_0 \vee \left(w = w_0 \wedge T^{\text{fix}}_i \geq t_0\right) \end{aligned} \tag{12}$$

$$\sum_{i \in \mathcal{I}^{24\text{h}}} \sum_{t \in \mathcal{T}} x_{jiwt} - \Delta^{24\text{h}}_j \leq g^{24\text{h}} \quad \forall j \in \mathcal{J}, w \in \mathcal{W} \tag{13}$$

$$\sum_{j \in \mathcal{J}} y_{jhwt} \leq \bar{d}^{\text{station}}_{ht} \quad \forall h \in \mathcal{H}, w \in \mathcal{W}, t \in \mathcal{T} \tag{14}$$

$$\sum_{j \in \mathcal{J}} y_{jhwt} + \Delta^{\text{station-dem}}_{hwt} \geq \underline{d}^{\text{station}}_{ht} \quad \forall h \in \mathcal{H}, w \in \mathcal{W}, t \in \mathcal{T} \tag{15}$$

$$\sum_{h\in\mathcal{H}} y_{jhwt} \leq 1 - T^{\text{off}}_{jwt} \quad \forall j \in \mathcal{J}, w \in \mathcal{W}, t \in \mathcal{T}, w > w_0 \vee (w = w_0 \wedge t \geq t_0) \tag{16}$$

$$X_{jiwt} - \Delta^{\text{duty}-}_{jiwt} + \Delta^{\text{duty}+}_{jiwt} - x_{jiwt} = 0 \quad \forall j \in \mathcal{J}, i \in \mathcal{I}, w \in \mathcal{W}, t \in \mathcal{T} \tag{17}$$

$$Y_{jhwt} - \Delta^{\text{station}-}_{jhwt} + \Delta^{\text{station}+}_{jhwt} - y_{jhwt} = 0 \quad \forall j \in \mathcal{J}, h \in \mathcal{H}, w \in \mathcal{W}, t \in \mathcal{T} \tag{18}$$

$$x_{jiwt} = X_{jiwt} \quad \forall j \in \mathcal{J}, i \in \mathcal{I}, w \in \mathcal{W}, t \in \mathcal{T}, w < w_0 \vee (w = w_0 \wedge t < t_0) \tag{19}$$

$$y_{jhwt} = Y_{jhwt} \quad \forall j \in \mathcal{J}, h \in \mathcal{H}, w \in \mathcal{W}, t \in \mathcal{T}, w < w_0 \vee (w = w_0 \wedge t < t_0) \tag{20}$$

$$\begin{aligned} x_{jiwt}, y_{jhwt}, z_{jiw}, \Delta^{\text{req-on}}_{jiwt}, \Delta^{\text{req-off}}_{jwt}, \Delta^{\text{duty}+}_{jiwt}, \Delta^{\text{duty}-}_{jiwt}, \Delta^{\text{station}+}_{jhwt}, \Delta^{\text{station}-}_{jhwt} \in \{0, 1\} \\ \forall j \in \mathcal{J}, h \in \mathcal{H}, i \in \mathcal{I}, w \in \mathcal{W}, t \in \mathcal{T} \end{aligned} \tag{21}$$

$$\Delta^{24\text{h}}_{j}, \Delta^{\text{late}+}_{jiw}, \Delta^{\text{late}-}_{jiw}, \Delta^{\text{station-dem}}_{hwt} \in \mathbb{N}_0 \quad \forall j \in \mathcal{J}, h \in \mathcal{H}, i \in \mathcal{I}, w \in \mathcal{W}, t \in \mathcal{T} \tag{22}$$

Our objective function can be decomposed into five subobjectives which can be grouped into our two main objectives of plan quality and stability. Plan quality is described by the following aspects.

Training Physicians should be assigned to workstations according to priorities given to workstations in their training plan.

Preferences Physicians' preferences for duty/no duty should be respected.

Fairness Workload should be equally distributed among physicians.

Coverage Demand for duties/workstations should be satisfied.

Plan stability is described by the following aspect.

Stability Changes due to rescheduling should be minimal.

In our objective function, we reward the assignment of physicians to duties and workstations in subobjective (1a) (*coverage*). Additional rewards are associated with the assignment of physicians to workstations corresponding to the training plan (*training*) in subobjective (1b). As the reward is specific to the respective training plan entry's priority, this favors training plan entries with a higher priority over those with a lower priority. Penalties are given for violation of requests for a duty and of requests for no duty (*preferences*) in subobjective (1c). Subobjective (1d) penalizes assigning too many overnight duties to a single physician in a week (*fairness*) and not assigning all days of a late duty to the same physician in a week. The latter is a necessary modeling device as absences may break the 4-day week assignment of a specific physician. Penalties are given for the violation of minimum demand for a workstation (*coverage*) in subobjective (1e). Our model also penalizes rescheduling of assignments (*stability*) to duties in subobjective (1f) and workstations in subobjective (1g).

As we reward assignment of duties to physicians, constraints (2) make sure that we cannot assign more physicians to a duty than this duty requires. Duties should also not be assigned to physicians who do not have the required qualifications. This is enforced by constraints (3). Note that with this formulation the physician only needs one of the required qualifications. A formulation where all required qualifications need to be met is straightforward.

Physicians can request a certain duty or request that they are not assigned to a duty on a given date. To track violations, constraints (4) and (5) make sure that the respective decision variables are set when a request for a duty (or no duty) has not been met.

There is a maximum of one duty per day for each physician. Also, physicians who are not present cannot be assigned to any duties. Both of these conditions are enforced by constraints (6). These constraints are limited to the current and future days. We add this restriction to prevent the model from becoming infeasible in case data in the past is entered incorrectly, e.g., a physician in the past is absent and assigned to a duty. As we also fix the assignments in the past (see below, constraints (19) and (20)), this would otherwise introduce an infeasibility into our model.

Overnight duties require that the physician be given a day off on the following day. Consequently, we cannot assign this physician to a duty or a workstation on the following day. Also, we want to make sure that the day off does not coincide with an existing absence as this would lead to overtime buildup. This is guaranteed by constraints (7a) (for all days but Sunday) and (7d) (for Sundays). As it is possible that a physician was on overnight duty and becomes sick today, we need to limit these constraints to the future only. For the current day, we have to allow that the day off after an overnight duty and an absence (due to sickness) are on the same day. To ensure that the physician is still not assigned to a duty or workstation on the current day, we use constraints (7b) and (7c) for all days but Sunday and constraints (7e) and (7f) for Sundays.

Late duties should be performed by the same physician 4 days a week. Constraints (8) ensure that exactly one physician is given the late duty in each week. This prevents a situation where multiple physicians are given the free day associated with the late duty and also a situation where no physician is assigned the late duty and the single days of the late duty are instead distributed among several different physicians. Note that we do not limit these constraints to the future only. This is done to ensure that the variable z_{jiw} is also set for each week in the past, because this variable is required for the correct computation of the $\Delta_{jiw}^{\text{late}+}$ and $\Delta_{jiw}^{\text{late}-}$ variables.

During each week of late duty, physicians should dedicate all of their time to the late duty and not be given additional duties. To ensure this, constraints (9) make sure that physicians on a late duty are not assigned overnight duties in the same week. Similarly, constraints (10) ensure that physicians on a late duty are not assigned an overnight duty on the Sunday before their late duty starts, as that would mean that their first day of late duty coincides with their day off after the overnight duty.

As only one physician is assigned to the late duty in a week, there are two possible deviations: not assigning all days to the physician on late duty and assigning days to physicians not on late duty. Constraints (11) track both these deviations. We model this using a soft constraint to cover cases where a physician who is assigned a late duty becomes absent. In this case it should be possible to assign the remaining days of the late duty in this week to another physician. The physician on late duty in a week needs to be given a day off in this week. Constraints (12) ensure that the physician is not assigned to a duty or a workstation on their day off and that this day off does not coincide with an existing absence.

Overnight duties are unpopular with physicians as they require physicians to be present for 24 consecutive hours. In order to distribute overnight duties in a fair way among all available physicians, we use constraints (13) to discourage the assignment of too many overnight duties in a week to a single physician. This fairness measure was chosen in discussions with scheduling experts at our partner hospital.

Similar to duties, we reward the assignment of physicians to workstations. However, it might happen that we have more physicians available than are required to cover demand of all workstations. We therefore use constraints (14) to make sure that the maximum number of physicians to be assigned to each workstation is not exceeded. As we want to guarantee that a minimum number of physicians is assigned to each workstation, constraints (15) calculate the number of missing physicians on each workstation to satisfy our minimum demand.

Physicians cannot be assigned to a workstation if they are not present. Constraints (16) prevent these physicians from being assigned to a workstation. Similar to constraints (6), we limit these constraints to the current day and future as the plan in the past cannot be changed anymore. If past data is entered incorrectly, e.g., an absent physician is assigned to a workstation, this would otherwise introduce infeasibilities.

Constraints (17) and (18) calculate the deviation from the existing plan due to rescheduling. We use separate variables for positive and negative deviation (i.e., added or removed assignments) from the previous plan. While our constraints do not require the positive and negative deviation variables to not both be 1 simultaneously, this will be taken care of during optimization as both variables induce costs when they are 1. These variables are used in the objective function to penalize changes to the existing plan.

Obviously, rescheduling should not change plans from the past. Constraints (19) enforce that the duty roster for all weeks before the current week remains unchanged, whereas constraints (20) provide the same guarantee for the workstation roster.

Constraints (21) and (22) restrict the value range for our binary and integer decision variables, respectively. Note that it would be sufficient to keep these constraints for our decision variables x, y, z and drop them for all Δ. The integrality conditions for all Δ are implicitly provided as the Δ are constrained by integer variables and parameters.

3.2 Sequential models

We propose a sequential version of our integrated model to enable a comparison between existing approaches which use sequential models and our integrated model. This sequential model is similar to the approach by Fügener et al. (2015). Here, we separate constraints into two separate models for the duty and workstation roster. The duty roster model is always run first, the workstation roster model is always run afterwards and has to respect assignments made by the duty roster model. Unlike the integrated model, the combination of these two models makes it impossible for the duty roster model to take into account constraints and objectives of the workstation roster, such as the training plan. We add new parameters Z_{jiw} to the workstation roster to respect the free days of late duties. In order to prevent infeasibilities in the workstation roster model, we add the following constraints to the duty roster model:

$$
\begin{gathered}
z_{jiw} \leq 0 \quad \forall j \in \mathcal{J}, i \in \mathcal{I}^{\text{late}}, w \in \mathcal{W}, \\
(w < w_0 \vee (w = w_0 \wedge T_i^{\text{fix}} < t_0)) \wedge \sum_{h \in \mathcal{H}} Y_{jhwT_i^{\text{fix}}} > 0
\end{gathered} \tag{23}
$$

These constraints prevent the duty roster from assigning a 4-day late duty week to a physician where the free day is in the past and the physician had a workstation assignment (and was therefore present) on that day. Without these constraints, the workstation roster could become infeasible because we disallow changing assignments in the past but also force the model to not assign physicians to a workstation on a free day of a late duty week. Aside from these changes, the sequential models only contain constraints and variables from the integrated model.

3.3 Model limitations

Our models have limitations which we explore in this section. These limitations arise from our desire to keep the models lean in order to make them easier to understand for the reader. We provide ideas for alternative or extended formulations which could mitigate these limitations. When applying these models in practice, most or all of these limitations will have to be mitigated, together with customization of the models to the specific application. As these mitigations and customization further obscure the models with complicated and non-generic formulations which distract from the core idea, we have chosen to not apply them to our model in this work. Note that the limitations are mostly of a technical nature. Applying the proposed mitigations does not change the key results we derive from the case study of our model. Some of these mitigations, however, introduce new dimensions for parameters and might increase processing time.

In its current form, the models do not support week-specific demands for duties and workstations. If required, this can easily be rectified by adding a week index to the demand parameters $\bar{d}_{it}^{\text{duty}}$, $\bar{d}_{ht}^{\text{station}}$, and $\underline{d}_{ht}^{\text{station}}$. The input parameter for the training plan g_{jhp}^{station} is also missing the time dimension. In practice, physicians' assignments in the training plan change over time which may require adding a week and day index. However, since these changes are infrequent (about 2 or 3 times a year) and we are discussing a short-term problem, we omit it in our models.

Our models do not consider data outside of the planning horizon. This means that overnight duties on the Sunday immediately preceding our planning horizon are not considered when assigning duties on the first day of the planning horizon. It is therefore possible that a physician is assigned two consecutive overnight duties. To mitigate this in practice, data from the day before the planning time span needs to be provided as an input parameter. Constraints (7d) need to be modified to take this data into account or need to be supplemented with new constraints.

Fairness considerations in the models are currently limited to the maximum amount of overnight duties to which a physician is assigned per week. This could be refined by adding a week index to $\Delta_j^{24\mathrm{h}}$ to get individual values for each week. Furthermore, this is a week-only measurement and does not consider equal distribution of duties over the entirety of the planning horizon. As interpretations of fairness vary and are not the main focus of this paper, we choose to not include additional fairness constraints. However, it is trivial to add them to our models.

We limit ourselves to assigning only one physician per 4-day late duty. This is sufficient for our application as the demand for late duties is always one in our case. If a late duty with higher demand needs to be scheduled, our models support this by adding several identical virtual late duties until the actual demand is met. For example, to schedule a late duty with a demand of 3 physicians, add three virtual late duties to the models.

3.4 Creation of initial plans

Our models do not only support rescheduling but also allows for the initial generation of plans by simultaneous consideration of duty and workstation rostering. They are purposely formulated in a way that allows switching between initial plan creation and rescheduling by changing parameters only, without the need to add or remove constraints or objective function terms. To enable initial plan creation, the following parameters need to be set:

$$X_{jiwt} = 0 \quad \forall j \in \mathcal{J}, i \in \mathcal{I}, w \in \mathcal{W}, t \in \mathcal{T} \tag{24}$$

$$Y_{jhwt} = 0 \quad \forall j \in \mathcal{J}, h \in \mathcal{H}, w \in \mathcal{W}, t \in \mathcal{T} \tag{25}$$

$$c^{\text{re-duty}} = c^{\text{re-station}} = 0 \tag{26}$$

$$t_0 = w_0 = 1 \tag{27}$$

This will make the models assume that the existing duty (24) and workstation roster (25) are empty. Term (26) sets the cost of rescheduling to zero while term (27) sets the current day to the first day in the planning time span. This means we start with an empty plan, allow changes during our entire planning horizon, and do not punish deviations from the existing empty plan.

4 Case Study

In this section we describe the application of our model at a German university hospital with 1 200 beds. The focus of the case study is to analyze the trade-off decision between plan stability and plan quality. Plan quality will be evaluated in terms of *fairness* and *preference* considerations for the duty roster, and in terms of demand *coverage* and fulfillment of *training* requirements for the workstation roster. Plan *stability* will be evaluated in terms of duty and workstation reschedules during the planning horizon. We consider a department of anesthesiology with $|\mathcal{I}| = 17$ duties ($|\mathcal{I}^{\text{late}}| = 11$ 4-day late duties and $|\mathcal{I}^{\text{24h}}| = 6$ overnight duties), $|\mathcal{H}| = 20$ workstations, and $|\mathcal{J}| = 133$ physicians. Unless otherwise noted, data used in this case study is real-life data from this department. The planning horizon consists of four weeks, i.e., 28 days. In this horizon there are $\sum_{j\in\mathcal{J},i\in\mathcal{I},w\in\mathcal{W},t\in\mathcal{T}} g^{\text{req-on}}_{jiwt} = 243$ duty requests, $\sum_{j\in\mathcal{J},w\in\mathcal{W},t\in\mathcal{T}} g^{\text{req-off}}_{jwt} = 140$ requests for no duty, and $\sum_{j\in\mathcal{J},w\in\mathcal{W},t\in\mathcal{T}} T^{\text{off}}_{jwt} = 1{,}177$ preexisting absent days. The number of absences is comparatively high because our data set contains physicians which are absent during the entire planning horizon due to parental leave, maternity leave, or vacation.

Each overnight duty has a demand of 1 for all days, while each 4-day late duty has a demand of 1 for all weekdays. Detailed demand values for all duties ($\bar{d}^{\text{duty}}_{it}$) are given in Table 3 and Table 4 (Appendix). We display minimum ($\underline{d}^{\text{station}}_{ht}$) and maximum ($\bar{d}^{\text{station}}_{ht}$) demand of all workstations in Table 5. Maximum demand is the real demand from the anesthesiology department. We automatically generated a slightly lower minimum demand to ensure equal coverage of workstations in case of personnel shortage.

We automatically generate assignments with different priority on a training plan as explained in Section 3 for all physicians. Each physician is assigned to three different workstations in the training plan: one workstation with priority 1 and two workstations with priority 2 or 3.

We run two different case studies. First, we evaluate the trade-off between plan stability and quality for the integrated model with different cost settings in Section 4.1. Afterwards, we compare the sequential models as specified in Section 3.2 with our integrated model in Section 4.2. In both studies, we consider 50 instances where we randomly draw a vector of absent physicians for each day of the planning horizon. Based on historical observations, we assume the probability of a physician becoming absent on any given day to be 2.5%. Thus, the duty and workstation roster is updated on every day of the planning horizon, corresponding to solving the model 28 times. For the study of our integrated model, each instance is solved with three different parameter settings:

LOW Plan stability costs are relatively low compared to plan quality costs

BALANCED Plan stability costs and plan quality costs are balanced

HIGH Plan stability costs clearly exceed those of plan quality

A summary of cost parameters is found in Table 2. All cost values are not absolute but relative costs in relation to other cost values. Non-rescheduling costs are obtained by testing real data and expert interviews with planners at our partner hospital. We keep these costs constant for all different cost settings in order to evaluate the effects of varying rescheduling costs only. With our cost settings, the cost of rescheduling a duty is always higher than the cost of rescheduling a workstation. This is because a rescheduled duty has a more substantial impact on the rescheduled physician's life due to changed working times, whereas a rescheduled workstation assignment only requires the physician to work the same hours at a different workstation. The different rescheduling cost settings were chosen in relation to the existing cost values. For the LOW setting we keep duty rescheduling costs below duty/no-duty request violation costs in order to enable the model to reschedule physicians to fulfill their duty requests. Similarly, workstation rescheduling costs are below the workstation assignment reward. The BALANCED setting defines higher duty rescheduling costs which are above the duty preference and no-duty preference violation costs, meaning this cost setting will not reschedule physicians because of preferences only. However, workstation rescheduling costs are still below the workstation assignment reward, making workstation reschedules possible. The HIGH cost setting defines a very high duty rescheduling cost and a very high workstation rescheduling cost. This makes workstation reschedules unprofitable for the model unless the reschedule fulfills a first-priority training plan assignment.

The non-rescheduling cost values have been fine-tuned for application at our partner hospital via an iterative cycle of expert interviews. The rescheduling costs for our three different cost settings are a recommendation on our part which we found by experimenting with cost values and evaluating the generated plans. It is to be expected that all costs will have to be adjusted for each application in practice. This is mainly because each department has its own culture, specifying, e.g., how physician preferences and number of overnight duties are weighed against each other. The cost values we propose should only serve as a rough guidance when implementing this model at another hospital.

All computations are performed on an openSUSE 13.1 64bit system, kernel 3.12.59-47 with 10 GB of RAM and four virtual cores of an Intel Core i7-4700MQ processor with 2.40 GHz. The system is run as a VirtualBox 5.1.4 machine. All models are solved with IBM ILOG CPLEX 12.6.2.0.

We report the average values of a set of key performance indicators over all 50 instances. Regarding plan *stability*/rescheduling, we report the number of changed duty and workstation assignments. Regarding plan quality for duties, we report the number of assigned duties (*coverage*), the number of declined duty/no-duty *preferences*, and the number of deviations from the maximum number of overnight duties per physician (*fairness*). Regarding plan quality for workstations, we report the number of assigned workstations, the number of negative deviations from minimum demand requirements (both *coverage*), and the number of deviations from the *training* plan assignments.

Parameter	LOW	BALANCED	HIGH
$c^{\text{re-duty}}$	4	20	90
$c^{\text{re-station}}$	3	10	30
$c^{\text{req-on}}$	5	5	5
$c^{\text{req-off}}$	9	9	9
c^{24h}	5	5	5
$c^{\text{station-dem}}$	10	10	10
r^{duty}	100	100	100
r^{station}	11	11	11
$r_p^{\text{station-plan}}$ $(p = 1/p = 2/p = 3)$	30/20/10	30/20/10	30/20/10

Table 2: Summary of cost parameters for all three settings

4.1 Integrated model

We apply the integrated model as specified in Section 3.1 and discuss the results of our case study. First, we create an initial plan using our model as described in Section 3.4. Then we solve each of the 50 instances for all 28 days and all three cost parameter settings resulting in a total of $1 + 50 \cdot 28 \cdot 3 = 4{,}201$ model runs. Each model for updating a roster takes between 1 and 21 seconds to be solved to a gap of at most 0.2% between the solution's objective value and the lowest known upper bound on the objective value for all feasible solutions (optimality gap). In the following we report average values based on the 50 iterations for all three cost parameter settings. The horizontal axis of all figures represents the time in days and weeks starting at the beginning of our 4-week planning horizon. The cost parameter setting LOW is displayed with a light grey line, BALANCED with a dark grey line, and HIGH with a black line.

We present the values for the number of assigned duties and workstations in Figure 2. These values represent the sum over all assignments in the entire planning horizon after execution of the model on the respective day. We can clearly see that the number of assigned duties and workstations decreases as time progresses. This happens because with each passing day more people are absent due to our generated additional absences, meaning physicians who were previously assigned to a duty are now absent. The same logic applies to workstation assignments. However, the decrease in workstation assignments is more pronounced than the decrease in duty assignments. This can be explained by the fact that present physicians are always assigned to a workstation and only sometimes assigned to a duty. When a physician now becomes absent, they will be removed from their workstation, thereby decreasing the total number of workstation assignments. As all other physicians are also already assigned to a workstation, it is not possible to assign a different physician instead of the absent physician without also removing this replacement physician from another workstation. It follows that, because all physicians are assigned to a workstation, we will always encounter a decreasing number of workstation assignments when a physician becomes absent. For duties, the situation is somewhat different. Most physicians are not assigned to a duty, so their absence does not pose a problem for duty coverage. In case a physician who is assigned to a duty becomes absent, there are still many other surplus physicians who are not assigned to a duty who can then replace the now absent physician. The number of duty assignments is highest in the LOW cost setting, followed by BALANCED and HIGH. This is a direct result of the different rescheduling weights. All cost settings allow the model to reschedule physicians whenever an uncovered duty can be assigned because the reward for assigning a duty is bigger than the cost of rescheduling a duty ($r^{\text{duty}} > c^{\text{re-duty}}$). However, the cost of rescheduling a duty is higher in the HIGH and BALANCED setting than in the LOW setting. It follows that the amount of additional violations incurred by rescheduling a duty, such as too many overnight duties in a week or violations of duty preferences, allowed by the model during rescheduling is smaller for the BALANCED and HIGH settings than for the LOW setting. Specifically: The HIGH setting will only allow additional violation costs of at most 10 ($r^{\text{duty}} - c^{\text{re-duty}}$) during duty rescheduling, translating to, e.g., two additional violations of duty requests ($c^{\text{req-on}} = 5$) or one additional violation of a no-duty request ($c^{\text{req-off}} = 9$). Comparing this with the LOW setting, we find that the additional violation cost incurred by duty rescheduling can be 96, translating to, e.g., 19 additional violations of duty requests or 10 additional violations of no-duty requests. The argument is analogous for the BALANCED setting. Thus, it follows that the LOW setting will reschedule duties accepting more violations of plan quality goals, thereby generating a better coverage of duties. The HIGH setting, on the other hand, is more restrictive regarding violations of other plan quality goals during rescheduling, leading to a smaller number of duty reschedules and thereby leaving more duties uncovered. However, the amount of uncovered duties at the end of the month is similarly small for all cost settings. Between 1 and 3 duties could not be assigned at the end of the month in comparison to the number of duties assigned at the beginning of the month. The assignment of workstations does not depend on the cost parameter setting. This is a result of the fact that all physicians are always assigned to a workstation as explained above. It is not possible to increase total coverage of workstations via workstation rescheduling as maximum possible coverage is already provided at the beginning of the planning horizon. As explained above, it is then impossible to replace an absent physician on a workstation without removing the replacement physician from a different workstation. This leads to the decrease in workstation assignments over all cost settings, because the total coverage will go down regardless of rescheduling. For workstation assignments, rescheduling can only influence coverage of workstation-specific minimum demand and training plan fulfillment, which we will explore later in the discussion of Figure 4.

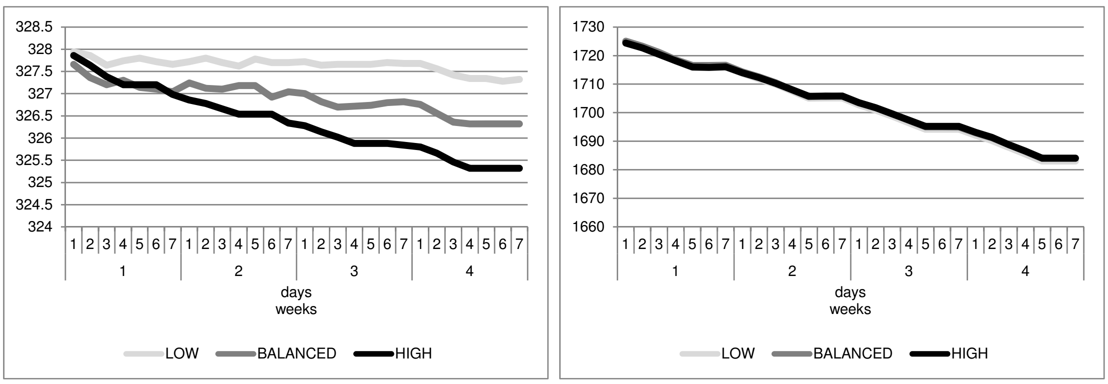


Figure 2: Average of assigned duties (x, left) and workstations (y, right)

In Figure 3 and Figure 4 we present the additional quality key performance indicators of the duty roster and the workstation roster, respectively. As before, these values represent the sum over the entire planning horizon, sampled after the rescheduling model for the respective day has run. The average of declined duty requests displayed on the left of Figure 3 shows greater values for the BALANCED setting compared to the LOW setting. This can be explained by the fact that lower rescheduling costs enable the model to make more reschedules and thereby fulfill more requests while keeping the same number of duties assigned. The number of declined duty requests is lowest for the HIGH scenario. All scenarios start from the same initial plan and preference fulfillment only gets harder as the amount of absences increases. The HIGH setting only allows a comparatively small number of additional preference violations during rescheduling, thereby keeping the number of preference violations approximately on the same level as in the initial plan. The LOW and BALANCED scenarios, however, allow a higher amount of additional violations during rescheduling, which then deviate from the initial plan. A similar effect can be observed in the right graph, which displays the number of overnight duties exceeding the maximum threshold for physicians. This number increases considerably for the LOW and BALANCED settings, where additional absences in combination with existing duty preferences and rescheduling shift the load of the overnight duties to a small number of physicians. Smaller costs for rescheduling allow the LOW setting to reschedule more physicians than the BALANCED setting, thereby achieving a slightly more balanced plan. The rather constant values for the HIGH setting can be explained by the lower number of reschedules compared to the other settings, leading to a distribution that remains close to the initial plan through all rescheduling operations.

Both graphs of Figure 4, as before, show average of sums of the respective values over the entire planning horizon, sampled after the rescheduling run on the respective day. In the initial workstation roster, all physicians are assigned to a workstation which respects at least one priority in their training plan. As time passes and the amount of absent physicians increases, so does the undercoverage of workstations (left graph) and the deviation from the training plan (right graph). Regarding the undercoverage of workstations, there is only a slight difference between the cost settings, most notably between the HIGH setting on the one hand and the LOW and BALANCED settings

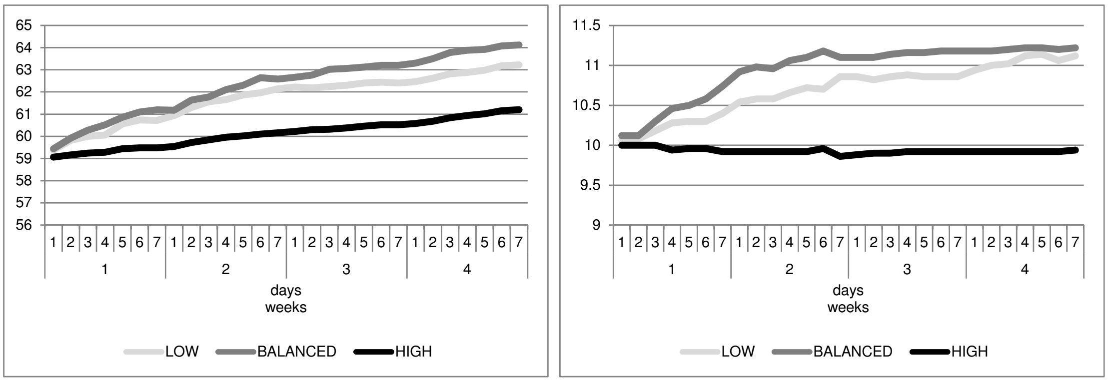


Figure 3: Average of declined duty requests ($\Delta^{\text{req-on}} + \Delta^{\text{req-off}}$, left) and exceeding the maximum number of overnight duties ($\Delta^{24\text{h}}$, right)

on the other hand. This can be explained by the fact that the HIGH setting does not reschedule physicians just to satisfy minimum demand of a workstation, thereby leaving a few more workstations below minimum demand. For the same reason, the proportion of workstation assignments not according to the training plan remains constant through all rescheduling operations for the HIGH setting: It simply does not reschedule physicians from a different workstation to the workstation with undercoverage but instead just leaves the workstation below minimum demand. The LOW and BALANCED settings, on the other hand, reschedule physicians to keep undercoverage in control. During this process, it is unavoidable to assign some physicians to workstations which are not in their training plan, simply because there are still open slots which need to be filled. Due to lower rescheduling costs, the LOW setting manages to assign more physicians according to their training plan than the BALANCED setting.

The graphs in Figure 5 follow a similar logic as previous graphs: Here, we show the number of rescheduling operations which were made during the rescheduling run on the respective day. This includes rescheduling operations which do not affect the respective day but were made during the rescheduling run on that day. Note that workstations do not have a demand on weekends, explaining the lower number of rescheduling operations for workstations on days 6 and 7. These graphs clearly show us that the LOW setting reschedules most physicians, while the HIGH setting reschedules the least. It is also visible

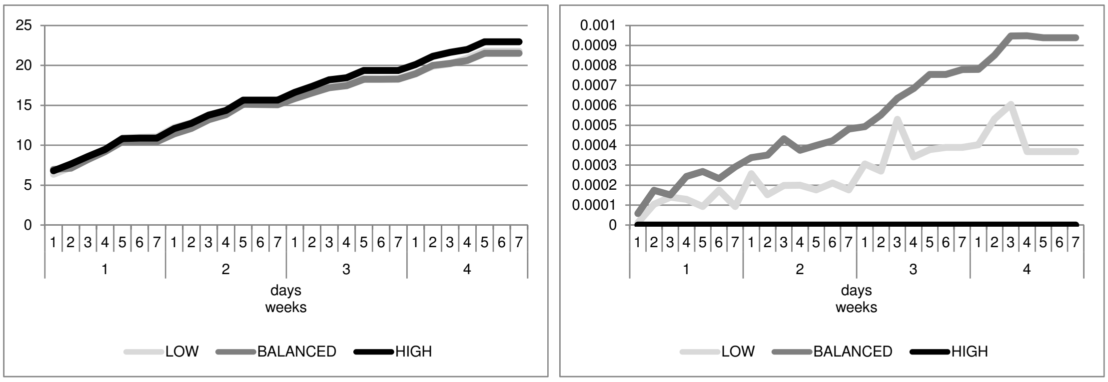


Figure 4: Average of undercoverage of workstations ($\Delta^{\text{station-dem}}$, left) and the proportion of workstation assignments not conforming to training plan ($\sum_{j\in\mathcal{J},h\in\mathcal{H},w\in\mathcal{W},t\in\mathcal{T},\sum_{p\in\mathcal{P}}(g^{\text{station}}_{jhp})=0} y_{jhwt} / \sum_{j\in\mathcal{J},h\in\mathcal{H},w\in\mathcal{W},t\in\mathcal{T}} y_{jhwt}$, right)

that the number of reschedules declines as time passes. This is a logical consequence of constraints (19) and (20) which disallow changing assignments in the past: As time passes, more and more assignments are fixed, resulting in a decreasing number of possible rescheduling operations.

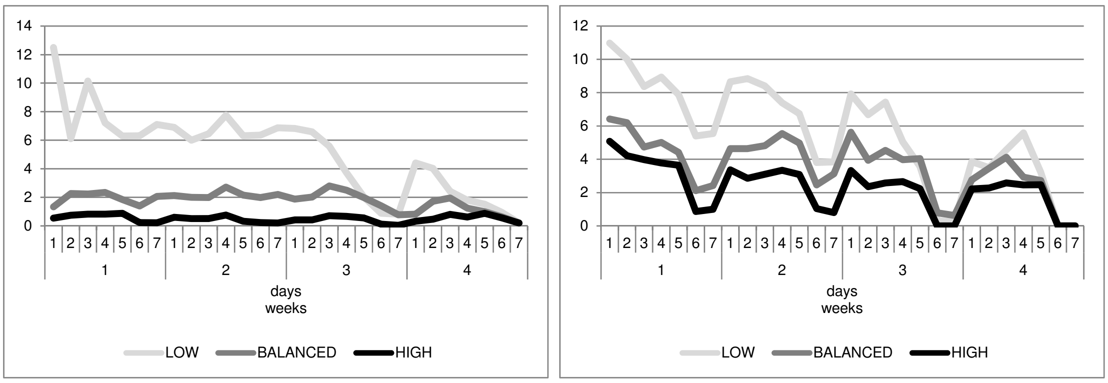


Figure 5: Average of replanned duty ($\Delta^{\text{duty}+} + \Delta^{\text{duty}-}$) and workstation ($\Delta^{\text{station}+} + \Delta^{\text{station}-}$) assignments

4.2 Sequential models

For this part of our case study, we test the sequential version of the integrated model as specified in Section 3.2 and compare it with the integrated model. As in the study for the integrated model, we run 50 iterations of the sequential models. The same base data (physicians, duties, etc.) and the same absences are used. For this study, we use only the BALANCED cost setting. As before, we first create an initial plan using the sequential models and then solve the sequential models on each day of the 28-day planning horizon to incorporate new absences for the respective day. This gives us a total of $1+50\cdot 28 = 1{,}401$ instances of the sequential models. To enable a fair comparison, we solve both duty and workstation rosters to the same 0.2% optimality gap we already used for the integrated model study. The model to reschedule the duty roster is always solved within 1 second, whereas the workstation roster rescheduling model takes up to 3 seconds. Data from the previous case study is used for the comparison between the integrated and sequential models. We discuss the key performance indicators of the sequential models in relation to the integrated model in the following. The results from the sequential models are displayed using a grey line, results from the integrated model are displayed with a black line.

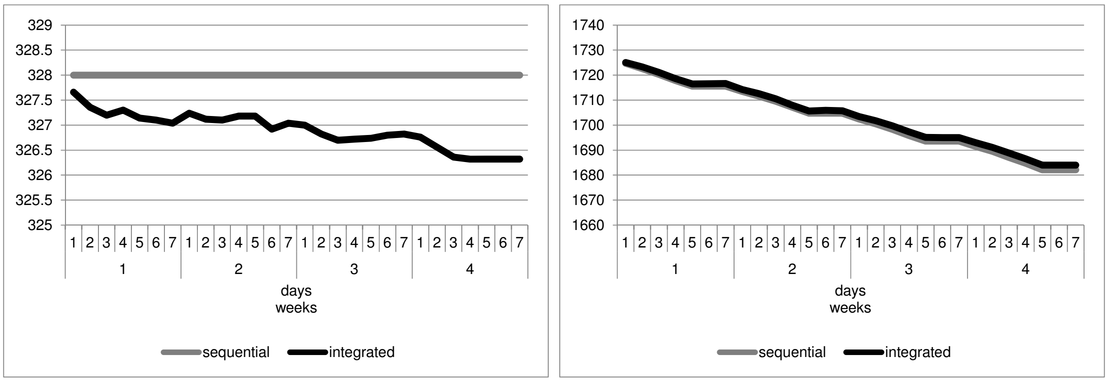


Figure 6: Average of assigned duties (x, left) and workstations (y, right) for the sequential and integrated models

Figure 6 shows the number of assigned duties and workstations summed over the entire planning horizon after rescheduling on the respective day has run. It is clearly visible and not surprising that the sequential models perform better in assigning duties. During the entire planning horizon, all duties are always assigned. This result is to be expected because the sequential models create the duty roster first, neglecting any negative influence which the assignment of duties might have on the workstation roster.

As the assignment of a duty also mandates the assignment of a day off and the sequential models assign more duties than the integrated model, the sequential models will have to assign more days off than the integrated model. Consequently, the integrated model performs marginally better for the number of workstation assignments because more physicians are available due to less days off. Considering the amount of workstation assignments over the entire planning horizon after the last day of rescheduling has run, the integrated model manages to make about 2 workstation assignments more than the sequential models.

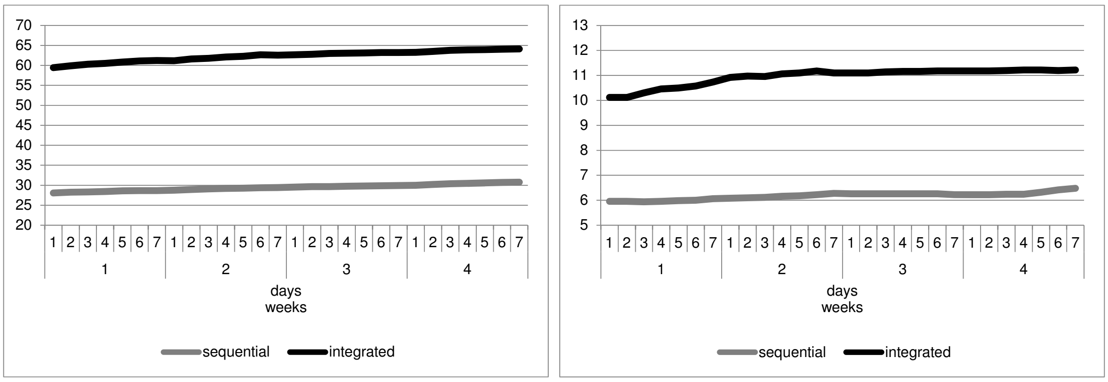
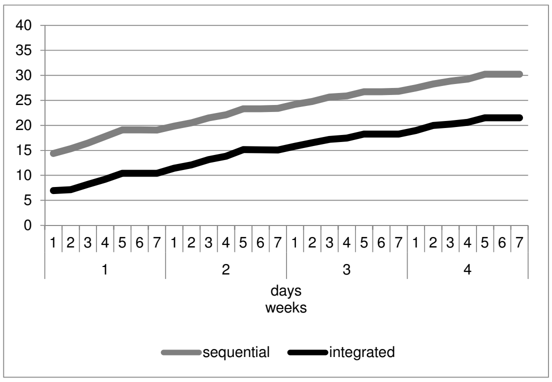


Figure 7: Average of declined duty requests ($\Delta^{\text{req-on}} + \Delta^{\text{req-off}}$, left) and exceeding the maximum number of overnight duties ($\Delta^{24\text{h}}$, right) for the sequential and integrated models

As the sequential models always create the duty plan first, they are able to achieve better values for the key performance indicators regarding duty plan quality. This is visible in both graphs of Figure 7. Throughout the entire planning horizon, the number of violated duty preferences in the integrated model is at about 200% of the value for the sequential models. We find a similar picture for the violation of fairness: The number of unfairly assigned overnight duties in the integrated model is at about 170% of the value for the sequential models.

However, the sequential models' focus on good fulfillment of quality goals for the duty plan comes at the expense of the workstation roster. Figure 8 shows that the quality of the workstation assignments, both in terms of physician training and in terms of equal



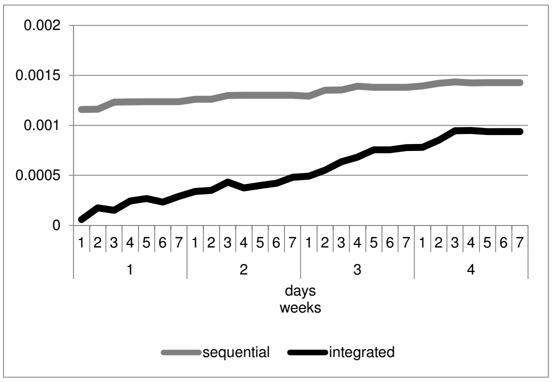


Figure 8: Average of undercoverage of workstations ($\Delta^{\text{station-dem}}$, left), the proportion of workstation assignments not conforming to training plan ($\sum_{j\in\mathcal{J},h\in\mathcal{H},w\in\mathcal{W},t\in\mathcal{T},\sum_{p\in\mathcal{P}}(g^{\text{station}}_{jhp})=0} y_{jhwt}/\sum_{j\in\mathcal{J},h\in\mathcal{H},w\in\mathcal{W},t\in\mathcal{T}} y_{jhwt}$, right) for the sequential and integrated models

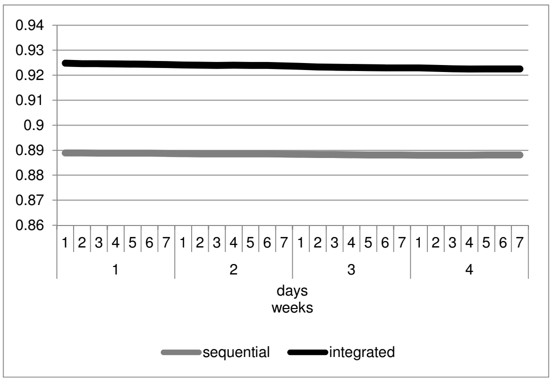


Figure 9: Proportion of workstation assignments conforming with the first priority in the training plan ($\sum_{j\in\mathcal{J},h\in\mathcal{H},w\in\mathcal{W},t\in\mathcal{T},g^{\text{station}}_{jh,1}=1} y_{jhwt}/\sum_{j\in\mathcal{J},h\in\mathcal{H},w\in\mathcal{W},t\in\mathcal{T}} y_{jhwt}$) for the sequential and integrated models

coverage of workstations, is considerably worse with the sequential models. On the last day of the planning horizon, both performance indicators are about 50% worse for the sequential models when compared to the integrated model: The number of uncovered workstations to satisfy minimum demand coverage rises from about 22 to about 30 while the percentage of workstation assignments not according to training plan rises from about 0.1% to about 0.15%. Pictured in Figure 9 is the difference in workstation assignments conforming to the first priority in the training plan: Whereas the integrated model can make over 92% of workstation assignments on a workstation with priority 1 in the training plan, the sequential models can achieve this only for less than 89% of all workstation assignments. In other words, the proportion of physicians who are not assigned to a workstation according to their first priority in the training plan rises from 8% to nearly 12%. Again, this can be explained by the fact that the duty roster model is run first. Consider that there are several physicians who are qualified to do the same overnight duty. However, they have different first-priority assignments in the training plan. The integrated model will try to assign the physicians to the duty in such a way that the day off after the duty coincides with a day where there are no open slots on the physician's first-priority workstation from the training plan, thereby maximizing the physician's assignments to this workstation. The sequential models, however, will assign the duties without taking this into account. The result is that the days off for some physicians will be on days where there would have been open slots on their first-priority workstation. Analogously, on the days where they then need to work, it is possible that there are no open slots for their first-priority workstation, so they are assigned to another workstation instead. Note that while this argument only considers overnight duties, it also holds for 4-day late duties.

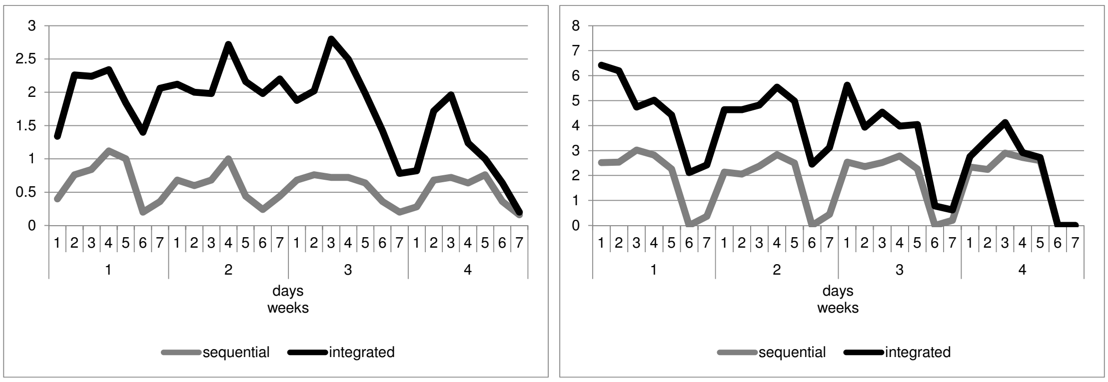


Figure 10: Average of replanned duty ($\Delta^{\text{duty}+} + \Delta^{\text{duty}-}$) and workstation ($\Delta^{\text{station}+} + \Delta^{\text{station}-}$) assignments for the sequential and integrated models

Lastly, we examine the rescheduling behavior of both models in Figure 10. The number of reschedules for duties and workstations is always higher for the integrated model. This difference can be explained by the fact that the sequential models do not respect interdependencies: Both duty and workstation model optimize for their own objectives only. The duty roster model can influence the workstation roster by assigning duties which require free days and thereby restrain the solution space of the workstation roster model. The workstation roster model can not influence the outcome of the duty roster model in any way. This lack of knowledge about the objectives and requirements of the respective other model constrains the solution space to solutions which are local optima, first in the duty and then in the workstation roster model. For this reason, the sequential models perform less rescheduling operations because many constellations where rescheduling would be beneficial to plan quality concern both workstation and duty roster. These dependencies can not be respected by the sequential models. It is therefore not surprising that the sequential models exhibit faster computation times as they are two much smaller models which together form a decomposition of the integrated model. The solution quality of the sequential models can not exceed the solution quality of the integrated model as each optimal solution in the sequential models is also a feasible solution in the integrated model.

4.3 Managerial insights

Some managerial insights can be drawn from the results of this case study. It should be noted that the different rescheduling cost settings need to be considered in relation to the other cost/reward values. Depending on the costs/rewards for different plan quality goals, rescheduling will or will not accept violations of the respective goals. Using the sequential models leads to a slightly better *coverage* of duties, but slightly worse *coverage* of workstations compared to the integrated model. The sequential models better fulfill physician *preferences*, but at the same time perform worse regarding physician *training*. This is explained by the fact that the sequential models cannot respect interdependencies between the duty and workstation rosters. They always create the duty plan first, leading to a better fulfillment of duty plan-related goals. If cost weights were set in such a way that duty plan-related goals are weighted much higher than workstation plan-related goals, the results of the sequential models and the integrated model would be much more similar. While fulfillment of physician *preferences* is also important to physicians, they usually value their *training* more. Additionally, hospitals have a legal obligation to enable physicians to reach *training* milestones within a certain time frame. It is therefore also in hospitals' interests to ensure a swift advancement of physicians' *training*. The integrated model allows managers to influence how physician *training* should be weighted against fulfillment of *preferences*, whereas the sequential model will always give a higher priority to *preferences*. More generally, the integrated model allows weighing all different plan quality goals (*training*, *preferences*, *fairness*, *coverage*) and plan *stability* against each other. The sequential models, on the other hand, split these goals into two different models and will always put more emphasis on the goals respected in the duty plan, most noticeably *preferences* and *fairness*, and will inevitably neglect the *training* goal, which only concerns itself with the workstation plan. We therefore recommend using the integrated model.

In our study, we explore different rescheduling cost values for use with the integrated model. Using too low cost values for rescheduling, as in the LOW setting, leads to a large number of reschedules of up to 400% of the BALANCED results (see Figure 5). While this does result in a few more (about 0.6%, see Figure 3) respected preferences and a few more (0.2%, see Figure 2) assigned duties, the induced management overhead of heavy rescheduling is prohibitive. The HIGH setting performs similar to the BALANCED setting in terms of reschedules and even leads to up to 50% less reschedules than the BALANCED setting. However, it also leaves a few more duty (about 0.5%) requests unfulfilled. In line with these results we suggest using the BALANCED or HIGH cost settings. The decision over which setting is chosen should be made depending on how costly reschedules and preference/training/coverage/fairness violations are to management. The rescheduling cost values we tested are based on the other cost values we use. These other cost values have been fine-tuned in cooperation with our partner hospital. It is possible and likely that these cost values will need to be adjusted based on the actual application of our model. Management should carefully monitor the current requirements of their workforce. If, e.g., physicians complain a lot about regular rescheduling (and attrition rises), management should focus more on stable plans by increasing the cost values for *stability*. Analogously, managers should change cost values for other subobjectives when required.

5 Conclusions

In the paper at hand we present a mixed-integer programming model that generates high-quality updates of duty and workstation rosters. Unlike previous approaches, it considers both problems simultaneously. Additionally, to enable a comparison with existing approaches, we present a sequential formulation of this model. We show that there exists a trade-off between having plans with high *stability* (i.e., minimizing the number of reschedules of duty or workstation assignments) and having plans with high quality (i.e., respecting *fairness* of duty distribution, fulfillment of *preferences* and *training* requirements, and *coverage* of duties and workstations). Thus, management should carefully monitor the current requirements of their workforce to understand which quality or *stability* goals are central to physician satisfaction. Our integrated model allows managers to control this trade-off between plan quality and plan *stability* of physician rosters by changing the cost values. It further allows managers to control the trade-off between quality goals: Besides the number of assignments (*coverage*), plan quality is measured in terms of *preferences* and *fairness* for the duty roster, and in terms of demand *coverage* and *training* requirements for the workstation roster. In a case study with 133 physicians, 17 duties, 20 workstations and 383 duty/no-duty requests we simulate one month based on real data and compare three parameter settings where rescheduling costs are LOW, BALANCED, or HIGH compared to the remaining cost parameters. This gives us a total of 4 200 rescheduling runs which can be solved within 1 to 21 seconds each. The results demonstrate that a rescheduling logic that maintains relatively high plan quality with relatively few changes, i.e., good plan stability, is possible. We also demonstrate the effect of cost parameters on the trade-off: in all cost settings, the number of assigned duties only differs negligibly. The number of duty reschedules per day, however, is up to more than 95% lower with the HIGH setting compared to the LOW setting.

Additionally, we evaluate sequential models which are based on our integrated model. We run these models on the same case study data, resulting in another 1 400 rescheduling runs. Using the sequential model separates our quality and stability goals into two separated models, making it impossible to influence the performance in relation to our goals by adjusting weight parameters only. The sequential models always put more emphasis on the goals which concern the model which is run first, i.e., the duty plan model. This means that sequential models will always perform well for physician *preferences* and *fairness*, because those are goals respected by the duty plan only, while at the same time performing worse for the goal of physician *training*, which concerns the workstation plan. Our integrated model allows influencing the trade-off between goals by adjusting cost values and is therefore an improvement over sequential approaches, such as the one proposed by Fügener et al. (2015).

Our work will enable hospitals to get timely decision support for rescheduling physicians in daily operations. The case study analyzes a real-life setting, an implementation in the staffing system at the partner hospital is currently in preparation. We believe that approaches such as proposed in this paper may enable hospitals to improve physician scheduling while maintaining good performance regarding *preferences*, *fairness*, *coverage*, *training* and limiting rescheduling of physicians (*stability*).

Our work has some limitations and offers room for future research. First, the definitions of *fairness* and *preferences* for duty rosters are quite limited. Especially *fairness* might differ a lot between hospitals, and the perception of *fairness* by employees and the planner might differ as well. Our paper defines *fairness* as equality - each physician should not get more than a certain number of duty assignments. However, it might be considered fair that a specific group of physicians (e.g., young physicians with stronger physics and at an earlier point in their career) are assigned to more duties than others. Besides, *fairness* could be considered for longer time horizons. E.g., a physician that had to do more duties in the previous planning horizon might get a bonus in the current one. *Preferences* are currently unlimited, and each request for a specific duty has the same value for the plan. Future research should analyze corrections to consider how many requests physicians give, and may compute different values for different physicians (e.g., physicians who got declined requests in the past might have a higher probability for being granted future requests). The workstation roster considers *coverage* and *training*. Regarding *coverage*, additional priorities over all workstations might result in better rosters. For *training* purposes, all physicians have a number of priorities which in our work remain static during the planning horizon. A dynamic view could incorporate different combinations of workstations for the *training* plan.

While our rescheduling model manages to generate updated duty and workstation rosters within 21 seconds in our computational study, this might still be too slow for some applications in practice. Rescheduling is a daily task for many managers. When using a computer system, managers should be presented a solution within 10 seconds if they are to maintain thought continuity during resolution of the current rescheduling problem (Miller, 1968). As practice applications will likely necessitate the addition of more constraints to our model, solution time will probably increase beyond our current measurements and heuristic approaches might be required to find good solutions in acceptable time.

Currently most hospitals do not use any mathematical or software-based decision support when rescheduling physicians. We are convinced that such approaches may help hospitals to find acceptable solutions, especially in situations with a lack of resources and where timely decisions are vital.

References

Bai, J., Fügener, A., Schoenfelder, J., and Brunner, J. O. (2016). Operations research in intensive care unit management: a literature review. *Health Care Management Science*, pages 1–24.

Bard, J. and Purnomo, H. (2005). Hospital-wide reactive scheduling of nurses with preference considerations. *IIE Transactions*, 37(7):589–608.

Baum, R., Bertsimas, D., and Kallus, N. (2014). Scheduling, revenue management, and fairness in an academic-hospital radiology division. *Academic radiology*, 21(10):1322–1330.

Bertsimas, D., Farias, V. F., and Trichakis, N. (2011). The price of fairness. *Operations research*, 59(1):17–31.

Bölt, U. (2015). Statistische Krankenhausdaten: Grund- und Kostendaten der Krankenhäuser. In Klauber, J., Geraedts, M., Friedrich, J., and Wasem, J., editors, *Krankenhaus-Report 2015*, chapter 20, pages 325–359. Schattauer GmbH, Hölderlinstraße 3, 70174 Stuttgart, Germany.

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

Burke, E. K., De Causmaecker, P., Berghe, G. V., and Van Landeghem, H. (2004). The state of the art of nurse rostering. *Journal of Scheduling*, 7(6):441–499.

Clark, A., Moule, P., Brodie, L., and Topping, A. (2014). The challenge of rescheduling nursing staff: Informing the development of a mathematical model decision tool. Project report, University of the West of England, Bristol.

Clark, A., Moule, P., Topping, A., and Serpell, M. (2015). Rescheduling nursing shifts: scoping the challenge and examining the potential of mathematical model based tools. *Journal of Nursing Management*, 23(4):411–420.

Clark, A. R. and Walker, H. (2011). Nurse rescheduling with shift preferences and minimal disruption. *Journal of Applied Operational Research*, 3(3):148–162.

Elomri, A., Elthlatiny, S., and Sidi Mohamed, Z. (2015). A goal programming model for fairly scheduling medicine residents. *International Journal of Supply Chain Management*, 4(2).

Erhard, M., Schoenfelder, J., Fügener, A., and Brunner, J. O. (2016). State of the art in physician scheduling in hospitals. Working Paper.

Fügener, A., Brunner, J. O., and Podtschaske, A. (2015). Duty and workstation rostering considering preferences and fairness: A case study at a department of anesthesiology. *International Journal of Production Research*, pages 1–23.

Gunawan, A. and Lau, C. H. (2013). Master physician scheduling problem. *Journal of the Operational Research Society*, 64(3):410–425.

Kitada, M., Morizawa, K., and Nagasawa, H. (2010). A heuristic method in nurse rerostering following a sudden absence of nurses. In *Proceedings of the 11th Asia Pacific Industrial Engineering and Management Systems Conference*.

Miller, R. B. (1968). Response time in man-computer conversational transactions. In *Proceedings of the December 9-11, 1968, Fall Joint Computer Conference, Part I*, AFIPS '68 (Fall, part I), pages 267–277, New York, NY, USA. ACM.

Moz, M. and Pato, M. V. (2003). An integer multicommodity flow model applied to the rerostering of nurse schedules. *Annals of Operations Research*, 119(1-4):285–301.

Moz, M. and Pato, M. V. (2004). Solving the problem of rerostering nurse schedules with hard constraints: New multicommodity flow models. *Annals of Operations Research*, 128(1-4):179–197.

Moz, M. and Pato, M. V. (2007). A genetic algorithm approach to a nurse rerostering problem. *Computers & Operations Research*, 34(3):667 – 691. Special Issue: Logistics of Health Care Management.

Petrovic, S. and Vanden Berghe, G. (2012). A comparison of two approaches to nurse rostering problems. *Annals of Operations Research*, 194(1):365–384.

Sherali, H. D., Ramahi, M. H., and Saifee, Q. J. (2002). Hospital resident scheduling problem. *Production Planning & Control*, 13(2):220–233.

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

Topaloglu, S. and Ozkarahan, I. (2011). A constraint programming-based solution approach for medical resident scheduling problems. *Computers & Operations Research*, 38(1):246 – 255.

van den Bergh, J., Beliën, J., Bruecker, P. D., Demeulemeester, E., and Boeck, L. D. (2013). Personnel scheduling: A literature review. *European Journal of Operational Research*, 226(3):367 – 385.

A Appendix

i	Mo	Tu	We	Th	Fr	Sa	Su
1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1
4	1	1	1	1	1	1	1
5	1	1	1	1	1	1	1
6	1	1	1	1	1	1	1
7	1	1	1	1	1	1	1

Table 3: Demand $\bar{d}_{it}^{\text{duty}}$ for overnight duties

i	Mo	Tu	We	Th	Fr	Sa	Su
8		1	1	1	1		
9		1	1	1	1		
10	1		1	1	1		
11	1	1		1	1		
12	1	1	1		1		
13	1	1	1	1			
14	1	1	1	1			
15		1	1	1	1		
16	1	1	1	1			
17		1	1	1	1		
18	1	1	1	1	1		
19	1		1	1	1		

Table 4: Demand $\bar{d}_{it}^{\text{duty}}$ for 4-day late duties

h	Mo	Tu	We	Th	Fr	h	Mo	Tu	We	Th	Fr
1	3	3	3	3	3	11	6	6	6	6	6
	2	2	2	2	2		5	5	5	5	5
2				3		12	5	5	5	5	5
				2			4	4	4	4	4
3	3	3	3	3	3	13	6	6	6	6	6
	2	2	2	2	2		5	5	5	5	5
4		3	3			14	7	7	7	7	7
		2	2				6	6	6	6	6
5	5	5	5	5	5	15	5	5	5	5	5
	4	4	4	4	4		4	4	4	4	4
6	5	5	5	5	5	16	6	6	6	6	6
	4	4	4	4	4		5	5	5	5	5
7	7	6	7	7	7	17	6	6	6	6	6
	6	5	6	6	6		5	5	5	5	5
8	5	6	5	5	5	18	3	3	3	3	3
	4	5	4	4	4		2	2	2	2	2
9	3	3	3	3	3	19	7	7	7	7	7
	2	2	2	2	2		6	6	6	6	6
10	5	5	5	5	5	20	5	5	5	5	5
	4	4	4	4	4		4	4	4	4	4

Table 5: Maximum ($\bar{d}^{\text{station}}_{ht}$, upper row) and minimum ($\underline{d}^{\text{station}}_{ht}$, lower row) demand of workstations h

Appendix B

Hospital physicians can't get no long-term satisfaction — an indicator for fairness in preference fulfillment on duty schedules

The following contribution (Gross et al., 2018a) has been published in "Health Care Management Science", which is ranked in category A in the VHB-JOURQUAL3 ranking (Verband der Hochschullehrer für Betriebswirtschaft e.V., 2015). The submitted version is reproduced below in its entirety.

Hospital physicians can't get no long-term satisfaction – An indicator for fairness in preference fulfillment on duty schedules

Christopher N. Gross[1] (corresponding author)

Jens O. Brunner[1]

Manfred Blobner[2]

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

Department of Health Care Operations / Health Information Management

Faculty of Business and Economics

University of Augsburg

Universitätsstraße 16, 86159 Augsburg, Germany

[2]Klinikum rechts der Isar der Technischen Universität München

Klinik für Anaesthesiologie

Ismaninger Straße 22, 81675 München, Germany

e-mail: christopher.gross@unikat.uni-augsburg.de

phone: +49 821 598 6452

Submission date: 28 February 2018

Revision date: 18 June 2018, 5 July 2018

This is a post-peer-review, pre-copyedit version of an article published in Health Care Management Science. The final authenticated version is available online at: https://doi.org/10.1007/s10729-018-9452-8

Hospital physicians can't get no long-term satisfaction – An indicator for fairness in preference fulfillment on duty schedules

Abstract

Physicians are a scarce resource in hospitals. In order to minimize physician attrition, schedulers incorporate individual physician preferences when creating the physicians' duty roster. The manual creation of a roster is very time-consuming and often produces suboptimal results. Many schedulers therefore use model-based software to assist in planning. The planning horizon for duty schedules is usually a single month. Many models optimize the plan for the current planning horizon, without taking into account data on preference fulfillment and work load distribution from previous months. It is therefore possible that, when looking at a longer time horizon, some physicians are disadvantaged in terms of preference fulfillment more often than their peers, simply because this generates better results for the individual months. This may be perceived as unfair by the disadvantaged physicians. In order to eliminate this imbalance, we introduce a satisfaction indicator for preference fulfillment in physician scheduling. This indicator is computed for each physician on each monthly plan and is then used to inform decisions regarding preference fulfillment on the current and future plans. As a result, a more equal distribution of preference fulfillment among physicians is achieved. We run a computational study with three different update strategies for our satisfaction indicator. Our study uses 24 months of data from a German university hospital and derives additional generated data from it. Results indicate that our satisfaction indicator, combined with the right update strategy, can achieve an equal distribution of satisfaction over all physicians within a peer group, as well as stable satisfaction levels for each individual physician over a longer time horizon. As our main contribution, we identify that our satisfaction indicator is more effective in creating equal distribution of long-term satisfaction the higher the rate of conflicting preferences is.

Keywords: OR in health services; mixed-integer program; long-term fairness; physician satisfaction

Submission: 28 February 2018

Revised: 18 June 2018, 5 July 2018

1 Introduction

Hospitals need to be staffed 24 hours a day, 7 days a week. During the day, physicians are present to perform procedures and other tasks. During the night, however, no elective procedures are scheduled. To ensure adequate medical assistance at any time, hospital schedulers create duty schedules which specify which physician is on duty during the night on which day. Being on duty mandates that the physician stays at the hospital throughout the night and takes care of medical emergencies as they occur. It follows that being on duty has a strong impact on the social life of physicians, as they have to stay at the hospital after their working day for the entire night. Many schedulers therefore let physicians specify preferences when they would like to be on or off duty and incorporate them in the schedule. This enables physicians to plan their duties and their social activities in advance and avoid collisions.

As physician scheduling is a complex task, there is a lot of research on automating this process. Duty schedules are usually created on a monthly basis. Many scientific methods for the creation of duty schedules therefore use a monthly planning horizon and generate an optimal plan for the given month. These models take physician preferences and availability as well as demand data into account and create a plan which conforms to physicians' preferences as much as possible. Additionally, many more constraints, such as working time regulations on maximum working hours or minimum rest times, are taken into account. A recent comprehensive overview on physician scheduling can be found in Erhard et al. [10]. Depending on the preferences and the constraints, however, it is often impossible to generate a plan which satisfies all preferences, meaning some physicians will not be granted their preferences to be on or off duty.

The importance of creating a schedule that physicians perceive as "fair" was already apparent to schedulers when designing the first computer-aided physician scheduling systems. Even though early systems could not yet create schedules on their own, they were already programmed to enable comparison of, e.g., weekend duties between colleagues

[11]. Later systems started incorporating rules to create fairness among physicians by equalizing work load and distribution of weekdays [15]. However, even today, many models for the creation of duty schedules do not keep track of fairness indicators, such as the amount of satisfied preferences per physician, over a time horizon longer than the planning horizon. Instead, they just create an optimal plan for each planning horizon, disregarding any information about how well individual physicians' preferences have been respected in the past. Depending on the preferences of the individual physicians, this might mean that it is often the same set of physicians whose preferences are violated while there is another set of physicians whose preferences are always fulfilled. As physicians have access to the finished plan and often also to the preferences of their colleagues, it is quite easy for them to deduce when colleagues have their preferences met while their own preferences are violated. This observation will lower the disadvantaged physicians' job satisfaction and may lead to other adverse effects such as decreased motivation, increased attrition, or higher turnover. Perceived unfairness has also been shown to contribute to absenteeism [4]. Unfairness makes affected physicians unhappy, which leads to a narrower attentional focus [21], meaning physicians will take more time in decision making and have more pessimistic assumptions about the decision outcomes. On the contrary, a plan respecting a physician's preferences might elicit positive emotions, which contribute to a wider attentional focus [12, 18], meaning a physician can rely more on their existing knowledge and arrive at a decision faster. Furthermore, positive emotions stack up and contribute to physicians' emotional well-being in the long term [13]. Being the highliest qualified personnel, physicians are very valuable to hospitals and high attrition rates due to job dissatisfaction and depression are very costly. Managers therefore want to prevent situations in which some physicians are treated unfairly. Instead, they want to ensure equal fulfillment of physicians' preferences.

In order to ensure fair consideration of physician preferences over several planning horizons, we propose an individual satisfaction indicator for each physician. This indicator models how well the physician's duty preferences have been incorporated in plans in the past. For the creation of new plans, this indicator is used to weight the preferences of each physician against those of the other physicians. Preferences of physicians with lower satisfaction levels (i.e., physicians who have not had many preferences fulfilled in the past) are given a higher weight than those of physicians with a higher satisfaction level.

The key contribution of this work is the aforementioned satisfaction indicator. To demonstrate the usefulness of this indicator, we develop a general scheduling model incorporating our satisfaction indicator. Our model is derived from a real-life scheduling model which is currently in use to create physician schedules at the department of anesthesiology at a German university hospital. We apply our model to data from this department and to generated data based on the real-world data in a study with three different strategies: no satisfaction tracking, satisfaction tracking after every planning horizon (past data), or satisfaction tracking during the planning horizon (past and current data). Depending on the strategy, we incorporate our indicator into a mixed-integer linear program (MILP) or a mixed-integer quadratic program (MIQP). We describe a linearization of the MIQP to solve it efficiently and prove that it is equivalent to the original program. Our experiments show that our satisfaction indicator, when used with the strategy which mandates updating during the planning horizon, achieves an equal distribution of fulfilled preferences among all physicians and also a comparatively stable level of satisfaction per physician over a time horizon of 24 months. The impact of our satisfaction indicator on equal distribution of preference fulfillment is more noticeable the higher the rate of conflicting preferences is.

The remainder of the paper is structured as follows. As we describe a general physician scheduling problem, we review related literature on fairness in personnel scheduling in the health care sector in section 2. This is followed by our model and a review of our different strategies for incorporation of the satisfaction indicators in section 3. Section 4 details the application of our model with satisfaction indicators to data from a German university hospital and to generated data based on the real-life data. We summarize our findings and give recommendations for application of long-term fairness in physician schedules in section 5.

2 Literature

There are many different interpretations of fairness when it comes to allocating resources to different actors. For a diverse overview of different measures of fairness and how they can be incorporated into Operations Research problems, we refer the reader to Bertsimas et al. [3] and Karsu and Morton [20]. In a recent review of scheduling literature for physicians in hospitals, Erhard et al. [10] conclude that fairness is often only implemented for shorter planning horizons. In the following, we briefly present some approaches to incorporate fairness into scheduling in the health care sector.

A very simple type of fairness is the ability of physicians to specify preferences for certain assignments. This approach is implemented by a multitude of papers. A recent example is presented by Niroumandrad and Lahrichi [24]. They propose a model for scheduling tasks and physicians at a radiotherapy center. Creating a schedule is complex because for each patient several tasks have to be performed in sequential order to prepare the patient for radiotherapy treatment. The optimal schedule assigns all tasks to physicians while minimizing patient stay at the center. Physicians can provide preferences for certain task on certain days. As the model is quite complex to solve, the authors propose a tabu search algorithm to efficiently find good solutions. An application of the algorithm to real life data from a radiotherapy center shows that the generated schedules provide shorter pretreatment phases for patients and therefore shorter patient stay than the manually created schedules in use at the center.

A more refined definition of fairness is its interpretation as equal distribution of preference fulfillment among scheduled personnel. Bard and Purnomo [2] schedule nurses using a column generation approach. They allow each nurse to specify individual preferences for when they would like to be scheduled to work. Instead of using a constant weight to penalize preference violations, they choose an exponential weight that increases with the amount of violations. This ensures that no single nurse has to endure too many violations and will instead make the model prefer solutions where several nurses have few violations over a solution where a few nurses have many violations. Instances with realistic data for a 4-week planning horizon could be solved in less than 4 minutes, obtaining schedules with on average less than 0.3 violations per nurse.

A different approach is taken by Bowers et al. [5]. They segment physicians into two groups: those who would like a schedule based on equal workload and those who would like a preference-based schedule. All physicians agree on the workload assigned to the duties which need to be scheduled. Afterwards, each physician can assign weights to the different duties based on personal preference. The proposed model then creates a schedule which distributes duties between the preference physician subset and the equality physician subset, according to the number of physicians in each subset. Distribution of the duties inside the respective subsets is then performed in such a way that duties are equally distributed based on the collaboratively agreed-upon workload for the equality subset of physicians and based on the individual preference values for the preference subset.

Ouelhadj et al. [26] explore three different objective formulations and their effect on fairness in nurse scheduling: minimizing all violations over all nurses, minimizing the violations on the individual roster with most violations, or adding individual objectives for each nurse to minimize violations on each individual roster. Additionally, they propose a method to incorporate nurses' individual weights for violations of different constraints and detail a cooperative heuristic to find solutions for their scheduling models. Martin et al. [23] improve upon their work and develop a new fairness objective measure which punishes the worst individual roster quadratically. In a computational study they also use a cooperative search heuristic, but combine all fairness formulations to get diverse solutions. The fairness of solutions is calculated using Jain's fairness index [19]. This index divides the squared sum of all allocations by the sum of the squared allocations. It follows that the value lies between 1 and 0, with 1 indicating a 100 % fair distribution of work. Results show that combining different fairness measures in the heuristic leads to fairer solutions.

Cohn et al. [7] create yearly on-call schedules for residents at Boston University School of Medicine. They focus on residents in their second and their post-graduation years. The on-call schedule therefore needs to respect educational goals of these residents by assigning them to different hospitals for a given amount of time. Note that this is different from our model, which schedules physicians of all seniority levels and does not assign physicians to a specific location to satisfy educational goals as the locations of our overnight duties are not relevant to physician education. The approach considers different measures of fairness, such as fixing the number of on-call assignments on holidays or number of denied vacation requests per physician. Due to the planning horizon of one year, the model manages to create a plan which the scheduled physicians perceive as fair. There is no mention in the paper of any data from past schedules being taken into account.

An approach for long-term fairness in physician scheduling is presented by Gierl et al. [16]. They measure the difference of duty assignment patterns between physicians. A schedule is considered more fair if the difference between the assignment patterns of different physicians is smaller. Or in other words: In an ideally fair schedule, each physician would have the same assignment pattern. This equality is impossible to achieve in a single planning horizon. Instead, information about the fairness of past rosters is stored in a computer system and incorporated into the creation of plans in the future. Their implementation is built on top of the existing "PEP" system [15] and enables a significant reduction in secretary and scheduler working hours spent on creation of fair plans.

Brunner et al. [6] introduce the flexible shift scheduling problem for physicians. While their model does not incorporate fairness, Stolletz and Brunner [27] build upon this work to enable the integration of fairness considerations. Their preprocessing algorithm creates shifts of varying length and inserts breaks into these shifts. A model then assigns physicians to shifts, taking into account labor restrictions and minimizing overtime. They show how the model can be extended to incorporate fairness. Fairness is considered in terms of stable shift starting times within each week, an even distribution of working hours, and an even assignment of on-call services. Computational results indicate that incorporating fairness is possible at small cost. However, it increases solution times considerably.

Fairness within each planning horizon is explored by Fügener et al. [14]. Their model assigns physicians to overnight duties, similar to our work. In terms of fairness, they set an upper limit of overnight duties per physician to keep the individual workload under a certain threshold. Applying their model to data from a German university hospital, they are able to generate schedules with less violations than those created by the experienced scheduling physician. Gross et al. [17] add rescheduling capability to this model. While both models respect physician preferences, they always just maximize preference fulfillment in the current planning horizon. None of them incorporate long-term fairness considerations for preference fulfillment.

Summarizing, there are many different types of fairness in scheduling in health care. Among them are the sole existence of preferences, equal fulfillment of preferences, stable shift starting times, and equal distribution of workload. However, existing literature mainly focuses on fairness in the current planning horizon. In the health care sector, work incorporating long-term fairness, i.e., fairness that takes into account data beyond the current planning horizon, does not exist.

3 Model

Our main contribution is a satisfaction indicator which calculates a physician's satisfaction with a roster in terms of how many of their preferences are fulfilled by the roster. Our satisfaction indicator is explained in detail in section 3.1. As the model is required for the understanding of the satisfaction indicator, we first present a model to create duty rosters (decision variables x_{jiwd}) for physicians $j \in \mathcal{J}$ who are assigned to duties $i \in \mathcal{I}$ during sequential weeks $w \in \mathcal{W}$, with each week having seven days $d \in \mathcal{D}$. Our model is similar in spirit to models proposed by Fügener et al. [14] and Gross et al. [17]. The model is always run for an entire month, i.e., for 4 or 5 weeks. Note that almost all data changes between the months. For reasons of comprehensibility, we do not add an index for the month, as we would have to add this index to each parameter. Each duty has a specific demand $\bar{d}_{id}^{\text{duty}}$ for each day of the week. Note that this is the target demand and is implemented as an upper bound on the number of assigned physicians, i.e., plans with undercoverage are possible.

We distinguish between different duties. This is necessary because we need to ensure an adequate experience mix during the night. To better understand what experiences are required, we provide a quick overview of selected stages of physician education in Germany. The 6 years of university education include a mandatory final unpaid clinical year, which is comparable to the internship of the US system. The first employment after university includes a training of 5 to 6 years in a clinical specialty. During this time, physicians are working comparable to US residents (first 2 years) and senior residents (following year). The training is completed with the respective board exam. Then, physicians work as specialists (e.g., anesthesiologists) and can practice independently. In hospitals, they can train in further subspecialties (e.g., intensive care, pain medicine), which is mostly a precondition to climb the greasy pole to a senior position with supervising duties, comparable to attending physicians in the US system. In our case, interns are not scheduled for overnight duties. We ensure that there is exactly one resident, one senior resident, another one who is in training for a subspecialty, one attending, and one senior attending physician present at the hospital. Additionally, there is an emergency medicine specialist on overnight duty who is on emergency call, i.e., responds to emergencies outside of the hospital. In case of overwhelming workload, there is an additional senior attending physician on standby (not present at the hospital) who can be called, whom we do not schedule using our model. Important to note, the German boards request that residents and senior residents are included in the overnight duties on a regular basis to be accepted for the board exams.

We model these requirements by creating a different type of duty for each of these types of physicians (except for the physician on standby). It follows that not every physician can be assigned to every duty. Instead, physicians must be qualified to be assigned to a duty. Additionally, schedulers do not always want to assign physicians by only their qualifications, but also by other criteria. For example, some physicians might be qualified to work as an emergency doctor, but are required at the hospital because of their qualifications. In this case, schedulers would like to not assign these physicians to the emergency doctor duty. We model this using binary parameter E^{pos}_{jiwd} to enable schedulers to specify whether it is possible to assign a certain physician to a certain duty on a given day.

We consider individual physicians' preferences for being assigned to a specific duty on a given day (parameters $g^{\text{req-on}}_{jiwd}$) and preferences for not being assigned to any duty on a given day (parameters $g^{\text{req-off}}_{jwd}$). In Germany, physicians working at hospitals are directly employed by the respective hospital. This is different from the system in the United States, where physicians can have admitting privileges at multiple hospitals [see e.g. 25]. Duties constitute overnight or weekend work, which is incentivized by additional payment for the time spent at the hospital. The payment scheme is very complex and defined by collective labor agreements with unions. It is possible for physicians to earn up to 20 or 30 % of their monthly salary by performing overnight or weekend work. Priorities vary between physicians. While some physicians are interested in the additional pay, others are more interested in having their evenings and weekends off. In our experience, most of the duties are requested by at least one physician. Consequently, only a few duties must be assigned to physicians who did not request them.

Physicians can also be absent (parameters D^{off}_{jwd}), e.g., on vacation, and thereby be unavailable for duties. Parameters $\hat{s}_j$ contain information based on our individual physician-specific satisfaction indicators from the previous planning horizon (see section 3.1). Our model is based on a real-life model we created for an anesthesiology department at a German university hospital. The real-life model is currently in use at the partner department to create monthly schedules. As the real-life model also assigns more than overnight duties, we have removed parts of it to arrive at the model presented below. Due to the generic nature of the resulting model, it is not restricted to anesthesiology departments, but can be used for any department.

Sets and indices	
$d \in \mathcal{D} = \{1, ..., 7\}$	Days of the week, starting with Monday = 1
$i \in \mathcal{I}$	Duties
$j \in \mathcal{J}$	Physicians
$w \in \mathcal{W}$	Weeks in the planning horizon
Parameters	
α_1	Weight for personnel demand coverage
α_2	Weight for preference fulfillment
γ	Smoothing constant for satisfaction indicator
$\hat{s}_j$	Pre-computed value as an input to calculate satisfaction-based weight for physician j based on the previous planning horizon
$\bar{d}_{id}^{\text{duty}}$	Demand of physicians for duty i on day d
$g_{jiwd}^{\text{req-on}}$	1 if physician j has a preference for duty i on day d of week w, 0 otherwise
$g_{jwd}^{\text{req-off}}$	1 if physician j has a preference for being off duty on day d of week w, 0 otherwise
E_{jiwd}^{pos}	1 if physician j can be assigned to duty i on day d of week w, 0 otherwise
D_{jwd}^{off}	1 if physician j is absent on day d of week w, 0 otherwise
Decision variables	
$x_{jiwd} \in \{0, 1\}$	1 if physician j is assigned to duty i on day d of week w, 0 otherwise

$s_j \in [0,1] \subset \mathbb{R}$	Satisfaction-based weight for preferences of physician j for the current planning horizon
$\sigma_j \in [0,1] \subset \mathbb{R}$	Satisfaction indicator for physician j. See section 3.1
$x^{\text{WE}}_{jw} \in \{0,1\}$	1 if physician j is assigned to a duty on the weekend of week w, 0 otherwise
$\Delta^{\text{out-duty}}_{iwd} \in \mathbb{N}^+_0$	Missing physicians to cover demand of duty i on day d of week w
$\Delta^{\text{req-on}}_{jiwd} \in \{0,1\}$	1 if preference of physician j for duty i on day d of week w is not satisfied, 0 otherwise
$\Delta^{\text{req-off}}_{jwd} \in \{0,1\}$	1 if preference of physician j for being off duty on day d of week w is not satisfied, 0 otherwise

Minimize

$$\alpha_1 \cdot \sum_{i\in\mathcal{I}} \sum_{w\in\mathcal{W}} \sum_{d\in\mathcal{D}} \Delta^{\text{out-duty}}_{iwd} + \tag{1a}$$

$$\alpha_2 \cdot \sum_{j\in\mathcal{J}} \left((2 - s_j) \cdot \left(\sum_{i\in\mathcal{I}} \sum_{w\in\mathcal{W}} \sum_{d\in\mathcal{D}} \Delta^{\text{req-on}}_{jiwd} + \sum_{w\in\mathcal{W}} \sum_{d\in\mathcal{D}} \Delta^{\text{req-off}}_{jwd} \right) \right) \tag{1b}$$

subject to

$$\sum_{j\in\mathcal{J}} x_{jiwd} + \Delta^{\text{out-duty}}_{iwd} = \bar{d}^{\text{duty}}_{id} \quad \forall i \in \mathcal{I}, w \in \mathcal{W}, d \in \mathcal{D} \tag{2}$$

$$\Delta^{\text{req-on}}_{jiwd} = g^{\text{req-on}}_{jiwd} \cdot (1 - x_{jiwd}) \quad \forall j \in \mathcal{J}, i \in \mathcal{I}, w \in \mathcal{W}, d \in \mathcal{D} \tag{3}$$

$$\Delta^{\text{req-off}}_{jwd} = g^{\text{req-off}}_{jwd} \cdot \sum_{i\in\mathcal{I}} x_{jiwd} \quad \forall j \in \mathcal{J}, w \in \mathcal{W}, d \in \mathcal{D} \tag{4}$$

$$\sum_{i\in\mathcal{I}} x_{jiwd} \leq 1 \quad \forall j \in \mathcal{J}, w \in \mathcal{W}, d \in \mathcal{D} \tag{5}$$

$$\sum_{j\in\mathcal{J}} \sum_{i\in\mathcal{I}} \sum_{w\in\mathcal{W}} \sum_{d\in\mathcal{D}: D^{\text{off}}_{jwd}=1} x_{jiwd} \leq 0 \tag{6}$$

$$\sum_{j\in\mathcal{J}} \sum_{i\in\mathcal{I}} \sum_{w\in\mathcal{W}} \sum_{d\in\mathcal{D}: E^{\text{pos}}_{jiwd}=0} x_{jiwd} \leq 0 \tag{7}$$

$$\sum_{i\in\mathcal{I}} x_{jiwd} + \sum_{i\in\mathcal{I}} x_{jiw(d-1)} \leq 1 \quad \forall j \in \mathcal{J}, w \in \mathcal{W}, d \in \mathcal{D}, d > 1 \tag{8a}$$

$$\sum_{i\in\mathcal{I}} x_{ji(w-1)7} + \sum_{i\in\mathcal{I}} x_{jiw,1} \leq 1 \quad \forall j \in \mathcal{J}, w \in \mathcal{W}, w > 1 \tag{8b}$$

$$x_{jw}^{\text{WE}} + x_{j(w-1)}^{\text{WE}} \leq 1 \quad \forall j \in \mathcal{J}, w \in \mathcal{W}, w > 1 \tag{9}$$

$$\sum_{i\in\mathcal{I}} \sum_{d\in\{6,7\}} x_{jiwd} \leq 2 \cdot x_{jw}^{\text{WE}} \quad \forall j \in \mathcal{J}, w \in \mathcal{W} \tag{10}$$

$$s_j = 1 \quad \forall j \in \mathcal{J} \tag{11a}$$

$$s_j = \hat{s}_j \quad \forall j \in \mathcal{J} \tag{11b}$$

$$s_j = \gamma \cdot \sigma_j + (1-\gamma) \cdot \hat{s}_j \quad \forall j \in \mathcal{J} \tag{11c}$$

$$\sigma_j = \frac{\sum\limits_{i\in\mathcal{I}} \sum\limits_{w\in\mathcal{W}} \sum\limits_{d\in\mathcal{D}} \left(g_{jiwd}^{\text{req-on}} - \Delta_{jiwd}^{\text{req-on}}\right) + \sum\limits_{w\in\mathcal{W}} \sum\limits_{d\in\mathcal{D}} \left(g_{jwd}^{\text{req-off}} - \Delta_{jwd}^{\text{req-off}}\right)}{|\mathcal{W}| \cdot |\mathcal{D}|} \quad \forall j \in \mathcal{J} \tag{12}$$

$$x_{jiwd}, x_{jw}^{\text{WE}}, \Delta_{jiwd}^{\text{req-on}}, \Delta_{jwd}^{\text{req-off}} \in \{0,1\} \quad \forall j \in \mathcal{J}, i \in \mathcal{I}, w \in \mathcal{W}, d \in \mathcal{D} \tag{13}$$

$$s_j, \sigma_j \in \mathbb{R}^+ \quad \forall j \in \mathcal{J} \tag{14}$$

$$\Delta_{iwd}^{\text{out-duty}} \in \mathbb{N}_0^+ \quad \forall i \in \mathcal{I}, w \in \mathcal{W}, d \in \mathcal{D} \tag{15}$$

Our objective function consists of two objectives: Coverage of physician demand (1a) and satisfaction of physician preferences (1b). We define a weight α_i for each of these objectives. This allows us to give different priorities to our objectives, which in our case have been selected to be lexicographic, i.e., $\alpha_1 \gg \alpha_2$. Constraints (2) ensure that we do not assign more physicians than required and also set our deviation variables $\Delta_{iwd}^{\text{out-duty}}$ in case of undercoverage. Similarly, constraints (3) ensure that $\Delta_{jiwd}^{\text{req-on}}$ is set in case of a violation of a duty preference while constraints (4) ensure that $\Delta_{jwd}^{\text{req-off}}$ is set in case of a violated preference to be off duty. As a duty runs through the entire night and might require the physician to work throughout, we cannot assign more than one duty per day to a physician. This is ensured by constraints (5). Constraints (6) prevent the model from assigning a duty to a physician who is not present at the day of the duty. Similarly, we never want to assign physicians to a duty for which they are not qualified or who the scheduler does not want to assign to this duty. We model this with constraints (7). Being an overnight assignment, physicians should be allowed to rest the day after each duty.

We therefore use constraints (8a) (for Tuesday through Sunday) and (8b) (for Monday, $d = 1$) to ensure that no physician is assigned to duties on consecutive days. Duties on the weekend (Saturday and Sunday) also put a lot of strain on physicians, so we do not want to assign duties on consecutive weekends to a physician. Constraints (9) ensure that we do not assign duties on consecutive weekends, while constraints (10) ensure that we set the variables x_{jw}^{WE} indicating whether a physician is assigned to a duty on a weekend in a given week. These constraints have been modeled based on real-life requirements at our partner hospital, but other implementations are possible. Constraints (11a), (11b), and (11c) are used for the different long-term fairness measures we propose. Note that these constraints are conflicting, so we always have to include only one of them in our model. We provide a more detailed explanation of these constraints and how they model our different fairness measures in sections 3.1 and 3.2. Constraints (12) calculate the satisfaction indicators for all physicians. These are only used by constraints (11c). For all other strategies, we only use the values of these decision variables as a model output for the evaluation of our results. Finally, constraints (13), (14), and (15) restrict the domains of our decision variables. Note that some of these domain constraints are redundant and could be dropped. However, we still include them in the paper to make the model more comprehensible.

The main purpose of our model is the evaluation of our satisfaction indicator. It is therefore kept intentionally simple. For example, we do not include constraints to ensure equal distribution of duties which are not requested. We limit our model to constraints which govern the distribution of requested duties, because we only evaluate the fulfillment of requests. Additional non-essential constraints would increase complexity of the model, hinder understanding by the reader and increase solution times. Nevertheless, our model is practical and includes most of the specifications of our partner hospital.

3.1 Physician-specific satisfaction indicator

In order to evaluate the plan in terms of granted requests for each physician, we propose a physician-specific satisfaction indicator, similar to Alsheddy and Tsang [1]. This indicator is calculated for a plan by dividing the number of granted requests by the number of days in the planning horizon, i.e., the number of possible requests for a duty or no duty. The satisfaction indicator for physician j is calculated as follows.

$$\sigma_j = \frac{\sum\limits_{i \in \mathcal{I}} \sum\limits_{w \in \mathcal{W}} \sum\limits_{d \in \mathcal{D}} \left(g_{jiwd}^{\text{req-on}} - \Delta_{jiwd}^{\text{req-on}} \right) + \sum\limits_{w \in \mathcal{W}} \sum\limits_{d \in \mathcal{D}} \left(g_{jwd}^{\text{req-off}} - \Delta_{jwd}^{\text{req-off}} \right)}{|\mathcal{W}| \cdot |\mathcal{D}|} \tag{16}$$

Theorem 1. *The satisfaction indicator is always a fractional number between 0 and 1.*

Proof. Physicians can only be assigned to at most one duty per day. Consequently, physicians can only submit at most either one duty preference or a preference to not be assigned to a duty on each day. The upper bound (UB) on requests of any kind is therefore one per day. This is accordingly also an UB on the number of satisfied requests. In the numerator of (16) we calculate the number of satisfied preferences over the entire planning horizon, i.e., on $|\mathcal{W}| \cdot |\mathcal{D}|$ days. We therefore know that the UB on the numerator is $|\mathcal{W}| \cdot |\mathcal{D}|$. The lower bound on the numerator is 0 as we can only satisfy ($g_{jiwd}^{\text{req-on}} - \Delta_{jiwd}^{\text{req-on}} = 1$) or not satisfy ($g_{jiwd}^{\text{req-on}} - \Delta_{jiwd}^{\text{req-on}} = 0$) each preference, or not have a preference at all ($g_{jiwd}^{\text{req-on}} = \Delta_{jiwd}^{\text{req-on}} = 0$), so there can never be a negative number of satisfied preferences (analogous for $g_{jwd}^{\text{req-off}} - \Delta_{jwd}^{\text{req-off}}$). It follows that $0 \leq \sigma_j \leq 1$. □

The higher a physician's satisfaction indicator, the higher his satisfaction with the respective plan. A satisfaction indicator of 1 indicates that the physician has entered a preference on each day and all of these preferences were satisfied. Note that this can only be the case when the preferences entered by the physician do not contain consecutive duty preferences. This means that a duty preference can be entered at most on alternating days and on all other days a preference to not be assigned to any duty must be entered. A satisfaction indicator of 0 indicates that not a single preference of the physician has been satisfied. This can be either because the physician entered preferences and they were ignored or because the physician did not enter any preferences. Note that our satisfaction indicator only counts satisfied preferences, so unsatisfied and missing preferences are treated the same way. By not adding a penalty to the satisfaction indicator for submitted but unsatisfied preferences, we prevent physicians from gaming the system. Otherwise, it would be possible for physicians to create requests which are hard or impossible to satisfy in order to decrease their satisfaction indicator and therefore get a higher weight for their preferences in the next planning horizon [1].

3.2 Updating strategies for satisfaction-based weights

The satisfaction indicator on its own is not helpful in generating fairer plans. We need to incorporate it into the planning process and then favor preferences of physicians with a lower satisfaction over preferences of physicians with a higher satisfaction. In order to do this, we punish each physician's preference violations with a cost based on the satisfaction indicator. In our model, we have chosen this cost as $\alpha_2 \cdot (2 - s_j)$, with s_j being a satisfaction-based weight for the preferences of physician j. Our different updating strategies describe different ways to calculate this weight. Note that this weight is based on the satisfaction indicator and all strategies ensure that $0 \leq s_j \leq 1$. We choose the constant 2 to evaluate different strategies. The only requirement for this constant is that it is greater than 1, because a constant smaller than 1 will result in zero or negative cost values for certain values of s_j. By adjusting this constant in combination with α_2, the scheduler can influence the impact of the satisfaction-based weight on the total weight for preference fulfillment. Higher values for the constant and lower values for α_2, for example, would keep the impact of preferences in comparison to other objectives about the same, but would decrease the impact of the satisfaction-based weight on the objective function.

When running the model, we supply a pre-calculated value for the historical satisfaction $\hat{s}_j$. Depending on the strategy, this value is calculated differently and usually depends on the historical satisfaction of the previous month's schedule. For all strategies, we initialize this value to 1 before the first planning horizon and test three separate updating strategies for the satisfaction-based weight:

I. Constant (C) Satisfaction-based weight is fixed at $s_j = 1$ for each physician, thereby giving all physicians' preferences the same weight in each planning horizon. Note that this means that the weight for each preference is $\alpha_2 \cdot (2 - s_j) = \alpha_2 = 1$ (see objective term (1b)). We use constraints (11a) for this experiment and do not supply parameters $\hat{s}_j$.

II. Exponential smoothing after the planning horizon (ESA) We calculate the satisfaction indicator σ_j for the generated plan after each planning horizon. Afterwards, we apply exponential smoothing to the satisfaction indicator and the historical satisfaction from the plan (i.e., the value we supplied as parameters $\hat{s}_j$ to the scheduling model to create the plan, now denoted as $\hat{s}_j^0$) using smoothing constant γ. The resulting value is used as the historical satisfaction $\hat{s}_j$ for the next planning horizon. For this experiment, we use constraints (11b) and calculate parameters $\hat{s}_j$ for the following planning horizon as follows.

$$\hat{s}_j = \gamma \cdot \sigma_j + (1 - \gamma) \cdot \hat{s}_j^0$$

III. Exponential smoothing during the planning horizon (ESD) We update our satisfaction-based weight directly in the model using exponential smoothing for the satisfaction indicator of the plan currently being generated and the satisfaction-based weight from the previous planning horizon. This will give us an updated satisfaction-based weight which we store in our decision variables s_j. This way, the satisfaction-based weight for preferences is already updated during assignment of the duties in the current planning horizon and the model will react to changes in the satisfaction indicator directly and not only after the planning horizon. For the next planning horizon, we look at the value of the satisfaction-based weight as output by the model in decision variables s_j. We now denote this value as s_j^0 and supply it as the historical satisfaction $\hat{s}_j$ to the model for the next planning horizon. This experiment is conducted using constraints (11c) and results in a quadratic decision model. We supply parameters $\hat{s}_j$ as follows.

$$\hat{s}_j = s_j^0$$

Note that the complexity of our model is different between our three updating strategies. For the C and ESA strategies, the value of our satisfaction-based weight does not change during the solving of the model. We can therefore model our satisfaction-based weights s_j as parameters, resulting in a linear decision model. For our ESD strategy, however, we change the value of the satisfaction-based weights based on the assigned duties during the solving of the model. Therefore, our satisfaction-based weights must also be modeled as decision variables. As we use our satisfaction indicator, which changes during model solving, to weight the fulfillment of preferences, we are left with a quadratic decision model.

3.3 Linearization of the ESD model

As solving a MIQP is much harder than solving a MILP, we provide a linearization of our quadratic model for the ESD strategy which yields equivalent results. For this linearization, we exploit properties of our problem. We know from (11c) that the value of the satisfaction-based weight s_j for the ESD strategy depends on some parameters and on the sum of decision variables $\sum_{i\in\mathcal{I}}\sum_{w\in\mathcal{W}}\sum_{d\in\mathcal{D}}\Delta_{jiwd}^{\text{req-on}} + \sum_{w\in\mathcal{W}}\sum_{d\in\mathcal{D}}\Delta_{jwd}^{\text{req-off}}$, i.e., the sum of preference violations in the planning horizon for physician j. As all other values are parameters and therefore fixed during model solving, we only need to think about what possible values the sum of the preference violations could assume. If we then pre-compute s_j for each of these values, we can supply these values of s_j as parameters and have the model select between one of these values.

We define that a physician can have at most one preference per day, either for a specific duty or for not being assigned to any duty. Allowing several preferences on a day does not make a lot of sense as most physicians are not qualified for more than one duty, meaning the only possible combination of preferences for them would be a preference for that duty and a preference for not being assigned to any duty. However, as those are also the only two possible outcomes, they would have then submitted preferences for any possible outcome, rendering the preference ineffective. The only case where multiple preferences could have a useful application would be a physician who is qualified for two duties A and B. They could then, for example, submit a preference for duty A and a preference for not being assigned to any duty, hoping to avoid being assigned to duty B. However, since we do not see the necessity for such a special way of using preferences, we decide to limit our preferences to one per day. Note that all further arguments made in this section are still valid for a higher limit of preferences per day. The only difference is that the upper bound on preference violations will need to be multiplied with the amount of preferences per day.

We first simplify the value for s_j from (11c). With some transformation, we can write (11c) as follows.

$$\begin{aligned} s_j = (-1) \cdot \frac{\gamma}{|\mathcal{W}| \cdot |\mathcal{D}|} \cdot \left(\sum_{i \in \mathcal{I}} \sum_{w \in \mathcal{W}} \sum_{d \in \mathcal{D}} \Delta^{\text{req-on}}_{jiwd} + \sum_{w \in \mathcal{W}} \sum_{d \in \mathcal{D}} \Delta^{\text{req-off}}_{jwd} \right) + \\ + \frac{\gamma}{|\mathcal{W}| \cdot |\mathcal{D}|} \cdot \left(\sum_{i \in \mathcal{I}} \sum_{w \in \mathcal{W}} \sum_{d \in \mathcal{D}} g^{\text{req-on}}_{jiwd} + \sum_{w \in \mathcal{W}} \sum_{d \in \mathcal{D}} g^{\text{req-off}}_{jwd} \right) + (1-\gamma) \cdot \hat{s}_j \end{aligned}$$

We now introduce a few auxiliary variables.

$$\tilde{h}_j = \frac{\gamma}{|\mathcal{W}| \cdot |\mathcal{D}|} \cdot \left(\sum_{i \in \mathcal{I}} \sum_{w \in \mathcal{W}} \sum_{d \in \mathcal{D}} g^{\text{req-on}}_{jiwd} + \sum_{w \in \mathcal{W}} \sum_{d \in \mathcal{D}} g^{\text{req-off}}_{jwd} \right) + (1-\gamma) \cdot \hat{s}_j$$

$$\tilde{k} = (-1) \cdot \frac{\gamma}{|\mathcal{W}| \cdot |\mathcal{D}|}$$

$$w_j = \sum_{i \in \mathcal{I}} \sum_{w \in \mathcal{W}} \sum_{d \in \mathcal{D}} \Delta^{\text{req-on}}_{jiwd} + \sum_{w \in \mathcal{W}} \sum_{d \in \mathcal{D}} \Delta^{\text{req-off}}_{jwd}$$

Substituting these, we arrive at the following equation.

$$s_j = \tilde{k} \cdot w_j + \tilde{h}_j \tag{17}$$

Note that $\tilde{k}$ and $\tilde{h}_j$ only consist of parameters and can be modeled as parameters, whereas w_j is an integer decision variable. Inserting this value for s_j into our cost term (1b) from the objective function, we get the following cost term.

$$\begin{aligned} \alpha_2 \cdot \sum_{j \in \mathcal{J}} ((2 - s_j) \cdot w_j) &= \\ \alpha_2 \cdot \sum_{j \in \mathcal{J}} \left(\left(2 - \tilde{k} \cdot w_j - \tilde{h}_j \right) \cdot w_j \right) &= \\ = \sum_{j \in \mathcal{J}} \left(\alpha_2 \cdot \left(\left(2 - \tilde{h}_j \right) \cdot w_j + \left(-\tilde{k} \right) \cdot w_j^2 \right) \right) & \end{aligned}$$

Recall that $\tilde{h}_j$ and $\tilde{k}$ are parameters, so they can be treated as constants as they will never change their values during model solving. As w_j is a sum of binary decision variables, it can only be a natural number, i.e., all possible values of w_j are countable. We can therefore introduce parameters c_{jw_j} for all physicians j and possible values of w_j, giving us the cost incurred by the preference violations of physician j. As w_j describes the number of violations of preferences for physician j, it is bounded by $|\mathcal{W}| \cdot |\mathcal{D}|$. This is easily explained: Assuming an upper limit of one preference per physician per day, our plan can violate at most this number of preferences for each physician. At the other extreme, the generated plan could violate no preferences at all. The possible values for the number of preference violations are therefore in $\mathcal{V} = \{0, \ldots, |\mathcal{W}| \cdot |\mathcal{D}|\}$. As we can always only satisfy or violate a preference, this set has to be a subset of the set of natural numbers. Using this set $\mathcal{V}$, we can now calculate $|\mathcal{V}|$ violation-based cost values c_{jv} for each physician j and denote the number of violations with $v \in \mathcal{V}$.

$$c_{jv} = \alpha_2 \cdot \left(\left(2 - \tilde{h}_j \right) \cdot v + \left(-\tilde{k} \right) \cdot v^2 \right) \tag{18}$$

These cost values can now be supplied as parameters to our model. In our objective function, we now only need to select the appropriate cost values based on the number of violations. To this end, we introduce a new binary decision variable y_{jv} which is 1 if physician j has exactly v violations of preferences in the current planning horizon and 0 otherwise. Using this new decision variable, we can substitute term (1b) in our objective function with the following.

$$\sum_{j\in\mathcal{J}}\sum_{v\in\mathcal{V}}(c_{jv}\cdot y_{jv}) \tag{19}$$

Additionally, we add the following constraints to the model.

$$\sum_{v\in\mathcal{V}} y_{jv} = 1 \quad \forall j\in\mathcal{J} \tag{20}$$

$$\sum_{i\in\mathcal{I}}\sum_{w\in\mathcal{W}}\sum_{d\in\mathcal{D}}\Delta^{\text{req-on}}_{jiwd} + \sum_{w\in\mathcal{D}}\sum_{d\in\mathcal{D}}\Delta^{\text{req-off}}_{jwd} = \sum_{v\in\mathcal{V}}(v\cdot y_{jv}) \quad \forall j\in\mathcal{J} \tag{21}$$

Term (19) adds the cost calculated for the satisfaction-based weight to the objective function and replaces the non-linear term (1b). Constraints (20) ensure that we set decision variables y_{jv} to 1 for exactly one number of violations v per physician. This ensures that we will always select exactly one amount of violations for each physician. Constraints (21) make sure we select y_{jv} in such a way that it describes the correct amount of violations.

Theorem 2. *For each physician, the selected pre-calculated cost values for preference violations c_{jv} are equal to the cost induced in the objective function of the quadratic model.*

$$\sum_{v\in\mathcal{V}}(c_{jv}\cdot y_{jv}) = \alpha_2\cdot(2-s_j)\cdot\left(\sum_{i\in\mathcal{I}}\sum_{w\in\mathcal{W}}\sum_{d\in\mathcal{D}}\Delta^{\text{req-on}}_{jiwd} + \sum_{w\in\mathcal{W}}\sum_{d\in\mathcal{D}}\Delta^{\text{req-off}}_{jwd}\right) \quad \forall j\in\mathcal{J}$$

Proof. We know from constraints (20) that for all values of v, there can only be one value of y_{jv} per physician j which is 1, while all others will be 0. We denote this value of v as w_j. With the same argument and with (21), we know that $w_j = \sum_{i \in \mathcal{I}} \sum_{w \in \mathcal{W}} \sum_{d \in \mathcal{D}} \Delta_{jiwd}^{\text{req-on}} + \sum_{w \in \mathcal{W}} \sum_{d \in \mathcal{D}} \Delta_{jwd}^{\text{req-off}}$ and can therefore write the equation as follows.

$$c_{jw_j} = \alpha_2 \cdot (2 - s_j) \cdot w_j \quad \forall j \in \mathcal{J}$$

Using equation 17 and equation 18, we can substitute to arrive at the following equation which is obviously true.

$$\alpha_2 \cdot \left(\left(2 - \tilde{h}_j\right) \cdot w_j + \left(-\tilde{k}\right) \cdot w_j^2 \right) = \alpha_2 \cdot \left(2 - \left(\tilde{k} \cdot w_j + \tilde{h}_j\right)\right) \cdot w_j \quad \forall j \in \mathcal{J}$$

□

Theorem 3. *The cost induced by preference violations in the linearized model is equal to that in the quadratic model.*

$$\sum_{j \in \mathcal{J}} \sum_{v \in \mathcal{V}} (c_{jv} \cdot y_{jv}) = \sum_{j \in \mathcal{J}} \alpha_2 \cdot \left((2 - s_j) \cdot \left(\sum_{i \in \mathcal{I}} \sum_{w \in \mathcal{W}} \sum_{d \in \mathcal{D}} \Delta_{jiwd}^{\text{req-on}} + \sum_{w \in \mathcal{W}} \sum_{d \in \mathcal{D}} \Delta_{jwd}^{\text{req-off}} \right) \right)$$

Proof. According to theorem 2, we know that the cost values for individual physicians in the linearized model are equal to those in the quadratic model. It is therefore trivial to see that the sum over all physicians j must also show this equality. □

With these modifications to our model, we now have an exact MILP for the ESD strategy. This model is provably equivalent in the objective function regarding physician preference violations. In summary, the models for all our strategies are now linear decision models and can be solved by any MILP solver.

4 Case Study

We use our satisfaction indicator σ_j as presented in (16) to quantify individual physicians' satisfaction with a schedule. As each schedule spans a month (4 or 5 weeks) and we need to calculate the satisfaction indicator for each schedule, we add an index $m \in \mathcal{M}$ to σ_j, with $\mathcal{M}$ being the set of all months in our study, and define σ_{jm} as the satisfaction of physician j with the schedule in month m. The focus of our evaluation is on two key performance indicators:

1. Variance of average satisfaction indicator per physician over all planning horizons, i.e., months m
$$APS = \operatorname*{Var}_{j \in \mathcal{J}} \left(\frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \sigma_{jm} \right)$$
2. Average of satisfaction indicator variance per physician between planning horizons, i.e., months m
$$ASV = \frac{1}{|\mathcal{J}|} \sum_{j \in \mathcal{J}} \operatorname*{Var}_{m \in \mathcal{M}} (\sigma_{jm})$$

For both key performance indicators, lower values indicate higher schedule quality. For the APS indicator, low values indicate that over all planning horizons all physicians are equally satisfied. Lower values for the ASV indicator signify that all physicians have a lower fluctuation of their personal satisfaction between planning horizons.

4.1 Real-life data and computational setup

To verify the impact of our model on long-term fairness, we apply it to real-world data from the anesthesiology department at our partner hospital. Note that our model is based on the scheduling model which is currently in use at this department and was also created by us. This department employs about 100 physicians at any time and is responsible for covering 6 overnight duties per day, requiring one physician per duty. The data spans a total of 24 months, starting in November 2015 and ending in October 2017. In total, the data contains 9 082 requests for a duty and for being off duty. During this time span, the data contains 158 distinct physicians employed for varying amounts of time at the department. However, only 47 of them are employed for the entire time horizon, accounting for 4 142 requests. As schedules are generated monthly, our data is also collected monthly, with each planning horizon spanning one month (4 or 5 weeks).

We run all three variants of our model implementing the different updating strategies for the satisfaction-based weights for each of the 24 months, resulting in 72 model iterations. Our model is implemented in CMPL 1.11.0, using the IBM ILOG CPLEX 12.7.1.0 solver on Xubuntu 16.04.3, kernel 4.10.0-40. The system is run as a VirtualBox 5.2.2 virtual machine with one virtual core of an Intel Core i5-4310M CPU and 4 GB of main memory. The software we use to conduct our experiments is written in Python 3 and can be found on GitHub[1], along with the data required to run our experiments. For privacy reasons, we do not provide the raw data from our partner hospital, but only generated data derived from it (see below).

	C	ESA	ESD
APS	0.006532	0.006493	0.006503
difference to C		−0.000039	−0.000029
difference in %		−0.60 %	−0.44 %
ASV	0.011292	0.011321	0.011191
difference to C		+0.000029	−0.000101
difference in %		+0.26 %	−0.89 %

Table 1: Values of the performance indicators for all three satisfaction indicator updating strategies using the real-life data set

For our smoothing constant γ, we need to choose a value between 0 and 1. Higher values let the satisfaction indicator depend more strongly on more recent values. Lower values for γ will lead to a stronger dependence on older data for the satisfaction indicator. However, since most physicians will likely remember their happiness with plans at most one or two months in the past, we recommend choosing values closer to 1 than to 0. For our case study, we select $\gamma = 0.8$. Preliminary testing with other values of γ indicated that the key results remain the same. We apply the raw data from our partner hospital to our model with all three different updating strategies and solve each run to optimality within 2 seconds. As only 47 physicians are employed continuously for the entire time horizon, our evaluations only take into account these physicians. For all other physicians

[1]https://github.com/chrisnig/long-term-fairness

newly joining our workforce in a month, we set their historic satisfaction to 1 but do not include them into the calculations for our performance indicators. We display the values of the *APS* and *ASV* indicators for the C, ESA, and ESD strategies for the continuously employed physicians in table 1. Only very small differences below ±1 % between our different satisfaction indicator updating strategies can be observed. This can be explained by the structure of preferences inherent in the data. Our cooperation hospital uses a web-based software system to track physician preferences. Each physician can enter their own preferences and simultaneously see all preferences of other physicians with the same qualification. The system counts the existing preferences and compares them with the demand for each duty, showing the physician which days of duty have not been requested yet. While physicians are not prevented by the system from requesting a duty that another physician has already requested, workplace culture at the department dictates that you do not submit requests competing with other physicians in general. In light of this preference elicitation method, it is not surprising that for all our updating strategies, 97 % of the preferences in the data can be fulfilled. Our model can therefore rarely make a trade-off between different physicians when fulfilling preferences because there are no competing preferences. In conclusion, respecting long-term satisfaction using our satisfaction indicator is not overly effective when used without competing preferences. Most real-world settings do not exhibit this feature. However, our model does not perform worse than strategies without long-term satisfaction. Our model respecting long-term fairness could therefore be used for scheduling problems with no conflicting preferences and will produce equally good or slightly better results than models without long-term fairness considerations. Nevertheless, in reality, competing preference requests are usual and so we show the superior behavior of our models for such situations next.

4.2 Generated data

In order to evaluate the differences between our updating strategies for satisfaction-based weights, we generate different instances of data for our model. We categorize our generated data in terms of preference rate and conflict rate. We define the preference rate as the number of preferences divided by the number of days of physician availability. As we only generate preferences for a duty and it is not possible for a physician to request two consecutive duties due to working time regulations, the preference rate for our generated data can be at most 50 %. A physician's preference is in conflict with another preference when another physician has a preference for the same duty on the same date. The conflict rate is then defined as the number of preferences which are in conflict with at least one other preference divided by the total number of preferences. Note that when generating data with a target conflict rate, this also bounds the preference rate. As the maximum number of total preferences is limited, it might be required to remove conflicting preferences to reach a target conflict rate.

Assumptions for our generated data are based on the first planning horizon of real-life data, which spans 5 weeks in the month of November 2015. Respecting absences, we derive a supply of 2 978 days of physician availability (including weekends, 7 days per week). This amount of days can be covered by $\frac{2\,978}{5\text{ weeks}\cdot 7\text{ days}} \approx 85$ physicians employed full-time without any absences. Furthermore, we find that 20 % of physicians have two qualifications, whereas the rest has only one qualification. Based on these observations, we generate data for 85 physicians who are each employed for the entire time horizon of 24 months so that each has at least one qualification for a duty, with a 20 % probability for each physician to hold an additional qualification for a different duty. This ensures that all physicians are employed continuously and that we can calculate our performance indicators over all physicians. Additionally, completely removing absences ensures that all physicians have the same chance of a high satisfaction in a given time span. The same holds true for generating qualifications which do not change during the planning horizon. Furthermore, we generate data for 6 duties with a demand of one physician per day (unchanged from the real-life data).

Our preference generation routine is described by algorithm 1 and operates as follows. On each day and for each physician, we assume a $p^{\text{on}} = 80$ % probability that the physician

Algorithm 1 Generating preferences for case study

```
procedure GENERATEPREFERENCES(𝒥, ℐ, 𝒲, 𝒟, D^off, E^pos, p^on)
    g^{req-on}_{jiwd} ← 0   ∀j ∈ 𝒥, i ∈ ℐ, w ∈ 𝒲, d ∈ 𝒟
    g^{req-off}_{jwd} ← 0   ∀j ∈ 𝒥, w ∈ 𝒲, d ∈ 𝒟
    for all j ∈ 𝒥, w ∈ 𝒲, d ∈ 𝒟 do                                  ▷ iterate randomly
        if D^off_{jwd} = 0 then
            r ← random rational number in [0, 1)
            if r < p^on and no preference on previous and next day exists then
                I ← {i ∈ ℐ | E^pos_{jiwd} = 1}
                if I ≠ ∅ then
                    i ← random value from I
                    g^{req-on}_{jiwd} ← 1
                end if
            end if
        end if
    end for
    return g^{req-on}, g^{req-off}
end procedure
```

Algorithm 2 Adjusting the conflict rate of preference data

```
procedure ADJUSTCONFLICTRATE(𝒥, ℐ, 𝒲, 𝒟, D^off, E^pos, g^{req-on}, g^{req-off}, g^{conflict})
    c ← 0                                       ▷ counter for number of existing conflicts
    C_{iwd} ← ∅   ∀i ∈ ℐ, w ∈ 𝒲, d ∈ 𝒟          ▷ set of conflicting preferences on each day
    t ← 0                                       ▷ counter for total number of requests
    for all i ∈ ℐ, w ∈ 𝒲, d ∈ 𝒟 do
        R̃ ← {(j, i', w', d') ∈ 𝒥 × ℐ × 𝒲 × 𝒟 | i = i' ∧ w = w' ∧ d = d' ∧ g^{req-on}_{jiwd} = 1}
        if |R̃| > 1 then
            c ← c + |R̃|
            C_{iwd} ← C_{iwd} ∪ R̃
        end if
        t ← t + |R̃|
    end for
    r^{conflict} ← c/t                          ▷ current conflict rate
    if r^{conflict} > g^{conflict} then
        g^{req-on} ← DECREASECONFLICTRATE       ▷ see algorithm 3
    else if r^{conflict} < g^{conflict} then
        g^{req-on} ← INCREASECONFLICTRATE       ▷ see algorithm 4
    end if
    return g^{req-on}
end procedure
```

Algorithm 3 Increasing the conflict rate of preferences

procedure IncreaseConflictRate
  $\bar{C} \leftarrow \{(j,i,w,d) \in \mathcal{J} \times \mathcal{I} \times \mathcal{W} \times \mathcal{D} \mid g^{\text{req-on}}_{jiwd} = 1 \wedge (j,i,w,d) \notin C_{iwd}\}$
  while $r^{\text{conflict}} < g^{\text{conflict}} \wedge |\bar{C}| > 0$ **do**
    randomly select $(j,i,w,d) \in \bar{C}$
    $\tilde{J} \leftarrow \{j' \in \mathcal{J} \mid E^{\text{pos}}_{j'iwd} = 1\}$ ▷ Candidate physicians for a request for this duty
    if $|\tilde{J}| = 1$ **then**
      $\bar{C} \leftarrow \bar{C} \setminus \{(j,i,w,d)\}$
    else
      for all $j' \in \tilde{J}$ **do**
        if $j \neq j' \wedge g^{\text{req-off}}_{j'wd} = 0 \wedge (\nexists i' \in \mathcal{I} : g^{\text{req-on}}_{j'i'wd} = 1) \wedge D^{\text{off}}_{j'wd} = 0$ **then**
          $g^{\text{req-on}}_{j'iwd} \leftarrow 1$
          $t \leftarrow t + 1$
          $\bar{C} \leftarrow \bar{C} \setminus \{(j,i,w,d)\}$
          $c \leftarrow c + 2$
          if $|\bar{C}| > 0$ **then**
            randomly select $(j'',i',w',d') \in \bar{C}$
            $\bar{C} \leftarrow \bar{C} \setminus \{(j'',i',w',d')\}$
            $g^{\text{req-on}}_{j''i'w'd'} \leftarrow 0$
            $t \leftarrow t - 1$
          end if
          break
        end if
      end for
      $r^{\text{conflict}} \leftarrow \frac{c}{t}$
    end if
  end while
  return $g^{\text{req-on}}$
end procedure

Algorithm 4 Decreasing the conflict rate of preferences

procedure DecreaseConflictRate
  while $r^{\text{conflict}} > g^{\text{conflict}} \wedge c > 0$ **do**
    randomly select $i \in \mathcal{I}$, $w \in \mathcal{W}$, $d \in \mathcal{D}$, such that $C_{iwd} \neq \emptyset$
    randomly select $j \in \mathcal{J}$, such that $(j, i, w, d) \in C_{iwd}$
    $g^{\text{req-on}}_{jiwd} \leftarrow 0$
    $c \leftarrow c - 1$
    $C_{iwd} \leftarrow C_{iwd} \setminus \{(j, i, w, d)\}$
    $t \leftarrow t - 1$
    for all $w' \in \mathcal{W}, d' \in \mathcal{D}$ **do** ▷ iterate randomly
      if $(w \neq w' \vee d \neq d') \wedge E^{\text{pos}}_{jiw'd'} = 1 \wedge (\nexists i' \in \mathcal{I} : g^{\text{req-on}}_{ji'w'd'} = 1) \wedge g^{\text{req-off}}_{jw'd'} = 0 \wedge (\nexists j' \in \mathcal{J} : g^{\text{req-on}}_{j'iw'd'} = 1) \wedge D^{\text{off}}_{jw'd'} = 0$ **then**
        $g^{\text{req-on}}_{jiw'd'} = 1$
        $t \leftarrow t + 1$
        break
      end if
    end for
    if $|C_{iwd}| = 1$ **then**
      $c \leftarrow c - 1$
      $C_{iwd} \leftarrow \emptyset$
    end if
    $r^{\text{conflict}} \leftarrow \frac{c}{t}$
  end while
  return $g^{\text{req-on}}$
end procedure

has a preference. If there is no preference on the previous day, we randomly select a preference based on the physician's qualifications and add it. After all the preferences have been added, we check whether the target conflict rate has been met and, if required, adjust it using algorithm 2. If the conflict rate is too high, we remove conflicting preferences using algorithm 4 and try to assign non-conflicting preferences to keep the total number of preferences equal, if possible. If the conflict rate is too low, we add conflicting preferences using algorithm 3 and remove the same amount of non-conflicting preferences in order to keep the total number of preferences equal. Note that increasing the conflict rate lets us preserve the total number of preferences, whereas decreasing the conflict rate might sometimes mandate that we remove previously added preferences from the data. Using this method, we generate in total 11 sets of data, with a target conflict rate between 0 % and 100 %, increasing in steps of 10 %. Note that we only generate duty preferences and do not generate preferences for no duty, because competition between duty preferences occurs much sooner than between preferences for no duty. This is easily illustrated: Consider the set of physicians holding a qualification for a certain duty. If there are several physicians with a preference for this duty on the same day, only one of them can be fulfilled. However, if all physicians in the set have a preference for no duty on the same day, then only one of these preferences must be violated as only one of these physicians must be on duty. Accordingly, as soon as two physicians submit a preference for the same duty, those preferences are in conflict. However, only if all physicians with the same qualification submit a preference to be off duty, they are in conflict.

After the data generation, we analyze our generated data for the actual preference and conflict rates. The results are given in table 2. The data generation routine as well as the generated data we use for our experiments is included in the published software code[2].

Using the generated data, we run the model for the 24 months of data for each satisfaction indicator updating strategy, resulting in an additional $24 \cdot 3 \cdot 11 = 792$ model

[2]see footnote on page 23

Target conflict rate	0 %	10 %	20 %	30 %	40 %	50 %
Preference rate	7.06 %	7.56 %	8.18 %	8.93 %	9.87 %	11.09 %
Conflict rate	0.00 %	9.68 %	19.73 %	29.77 %	39.78 %	49.91 %

Target conflict rate	60 %	70 %	80 %	90 %	100 %	(Real data)
Preference rate	12.57 %	14.54 %	17.86 %	23.43 %	44.99 %	6.35 %
Conflict rate	59.89 %	69.88 %	79.90 %	89.94 %	100 %	10.02 %

Table 2: Preference and conflict rate of data

runs. We solve the models for all strategies to optimality within 2 seconds of solver running time per model run. Because of the data generation process, all 85 physicians are employed continuously, so we can calculate our performance indicators on the entire workforce. Additionally, the qualifications of our generated physicians do not change during the 24 months and are equally distributed among the duties. Due to the duty preference generation strategy explained above, we have a varying number of competing requests for the solver to choose from (see table 2).

First, we take a look at the results for data with a 100 % conflict rate to find out whether the differences in the key performance indicators between the models are more pronounced than the ones we have seen for our real-life data in table 1. The values of the performance indicators for our data with a 100 % conflict rate are displayed in table 3. Results are interpreted and discussed in detail below.

	C	ESA	ESD
APS	0.004346	0.000433	0.000015
difference to C		-0.003913	-0.004331
difference in %		-90.04 %	-99.65 %
ASV	0.001876	0.012093	0.000423
difference to C		$+0.010217$	-0.001453
difference in %		$+544.62$ %	-77.45 %

Table 3: Values of the performance indicators for all three satisfaction indicator updating strategies using the generated data set with a 100 % target conflict rate

We display the average satisfaction over all 24 months by physician in figure 1. As this is the basis for the APS indicator, the chart already makes it obvious that the variance of the displayed data is much greater for the C strategy than for the other

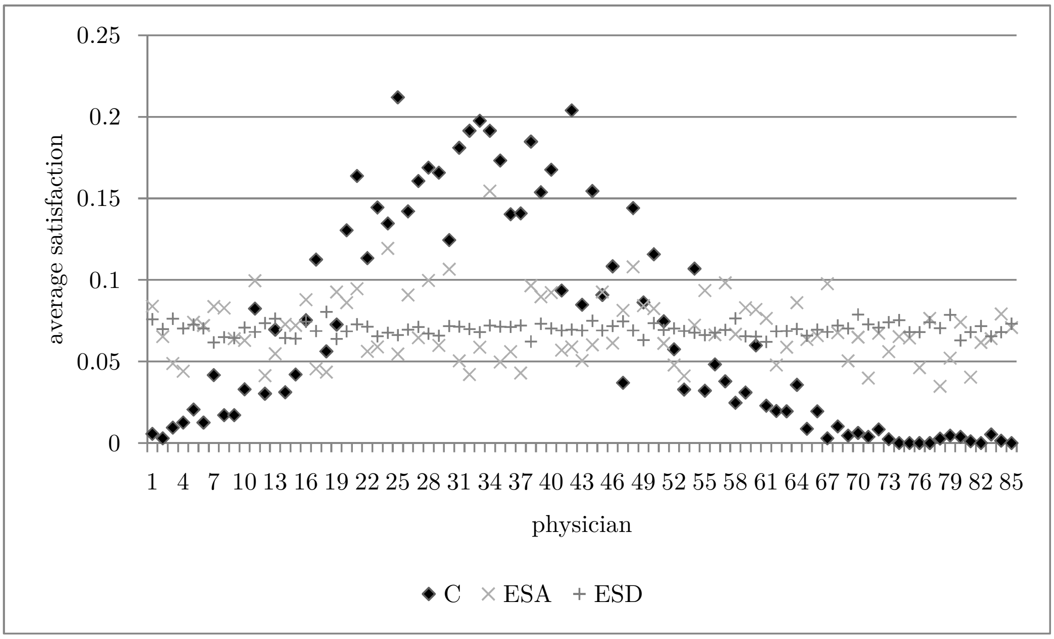


Figure 1: Average satisfaction $\frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}} s_{jm}$ over all 24 months per physician

strategies. In fact, the APS indicator for the ESA strategy is 90 % below the value for the C strategy, and for the ESD strategy 99 % below the C strategy (see table 3). This means that for the ESA and ESD strategies, the satisfaction is much more equally distributed among physicians than for the C strategy. The graph shows a noticeable peak in satisfaction around physician 34. Our model does not guide the solving process to solutions with this property, so we assume that underlying heuristics in the standard solver IBM ILOG CPLEX are responsible for this pattern. We verified this assumption about the underlying heuristics by running the model for the C strategy with the same data also with the CBC 2.9.8 solver and found a different pattern which showed a nearly linear increase in satisfaction with the physician number. As our model for the C strategy does not optimize for equal distribution of satisfaction over all physicians, these patterns can be seen in the results. Our ESA and ESD strategies, however, implicitly optimize for equal distribution of satisfaction among physicians and therefore score better APS values.

For the ASV indicator, we plot the variance per physician in figure 2. This measures

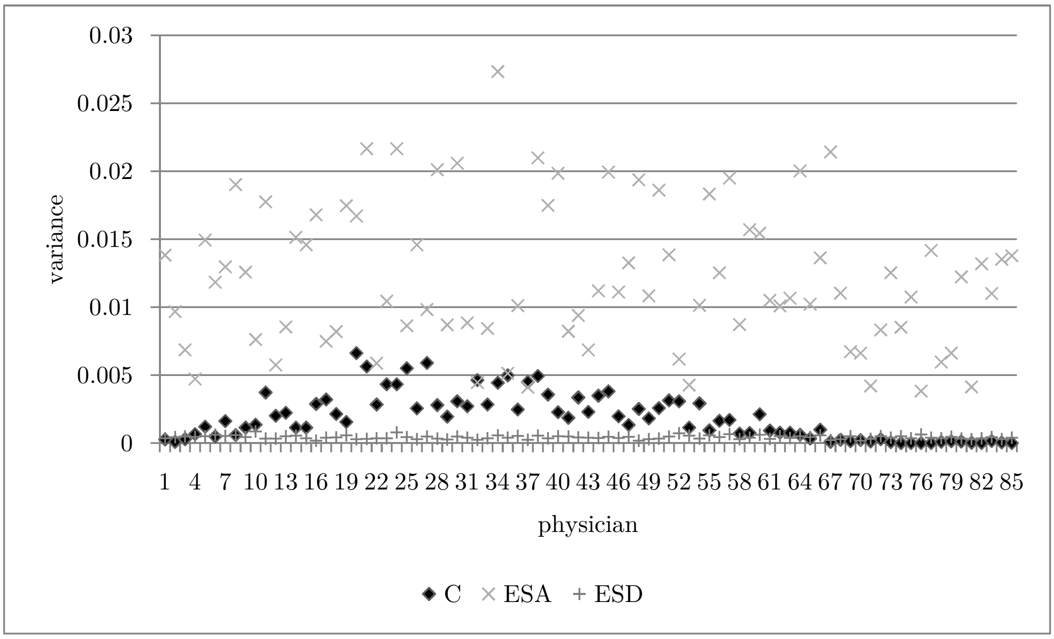


Figure 2: Variance between months $\mathrm{Var}_{m \in \mathcal{M}}\, s_{jm}$ per physician

how much the satisfaction for each physician varies between the months. The *ASV* indicator measures the average of the data displayed in the figure. Lower values indicate that the satisfaction of a physician remains closer to a physician-specific satisfaction average, whereas higher values indicate a high fluctuation of a physician's satisfaction between months. Looking at figure 2, we can see that the average variance of the ESA strategy is noticeably higher than that of the two other strategies. This can be explained by the properties of the strategy. The ESA strategy updates the weight given to each physician's preferences after each planning horizon. This leads to a pendulum-like fluctuation of the values. In one planning horizon, a physician does not get many preferences fulfilled, leading to a high weight for the next planning horizon. Consequently, due to the high weight, the physician is granted many preferences in the next planning horizon, so that the weight is greatly reduced after the planning horizon and the cycle repeats. To illustrate this, we include figure 3, which shows the satisfaction of physician 34 as an example. We select this physician because it is the physician with the highest variance of satisfaction for the ESA strategy between months (see figure 2). Here, the repeated increase and decrease

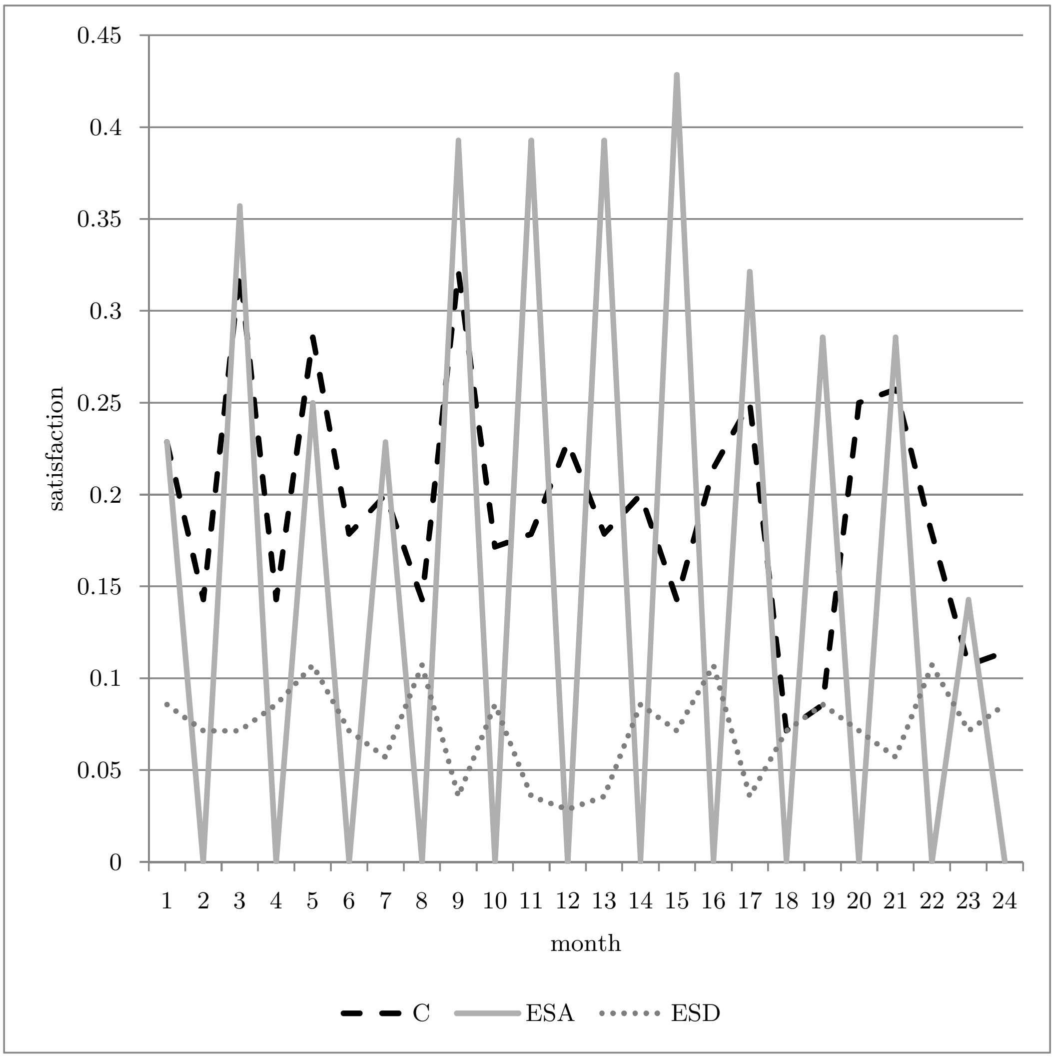


Figure 3: Satisfaction $s_{34,m}$ for physician 34 for the different updating strategies

in satisfaction for the ESA strategy is visibly more pronounced than fluctuation for both other strategies. This results in high values for the *ASV* indicator for strategy ESA (+545 % on the C strategy). For the C strategy, we again find the same peak around physician 34 which we already observed in figure 1. As we already know that, due to the underlying heuristics, physicians with higher numbers rarely get any requests fulfilled, it is not surprising that the variance of their satisfaction is close to zero when their average satisfaction is already close to zero. The ESD strategy performs best in terms of the *ASV* indicator with a 77 % decrease compared to the C strategy. This can be attributed to the fact that it already optimizes the satisfaction and the associated weights for preferences during the planning horizon, thereby smoothing out high variations in the satisfaction before they can occur.

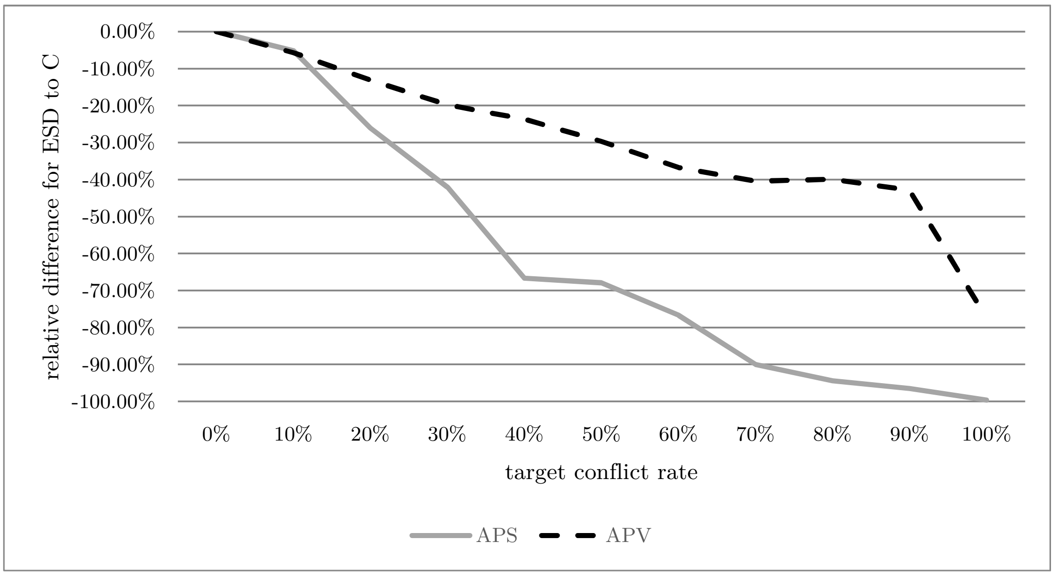


Figure 4: Relative difference in performance indicators between ESD and C strategy for data with different target conflict rate

As the possibility of high variance between the months makes the ESA strategy impractical to use, going forward we focus on comparing the ESD strategy with the C strategy. We have shown that a high conflict rate leads to a high difference in our performance indicators between those two strategies. In what follows, we discuss how the conflict and preference rates are related to the difference between the strategies. Looking at the relative change in our performance indicators displayed in figure 4, we find that with an increasing conflict rate, the decrease in our performance indicators for the ESD strategy in comparison to the C strategy rises. This means that our ESD strategy has more effect the higher the conflict rate inherent in the preferences is. The additional drop in the APV indicator between the target conflict rate of 90 % and 100 % can be explained by the significant difference in preference rate between those two data sets. Looking at table 2, we can see that the preference rate between those two data sets nearly doubles.

This is due to the preference rate being restricted by the conflict rate for the 90 % setting. For the 100 % setting, however, the preference rate is not limited by the conflict rate anymore, as we can theoretically add an unlimited amount of conflicting preferences and still keep the conflict rate the same. Therefore, the preference rate for the 100 % setting is only restricted by semantic constraints, such as not allowing consecutive preferences. This higher preference rate enables our algorithm to more evenly distribute preferences for a physician between the months, resulting in the sharp drop in the ASV indicator.

Summarizing, our results recommend using the ESD strategy in any case. Even for a small conflict rate, this strategy consistently scores better on both performance indicators. This means that physicians will be treated more fairly in terms of distribution of fulfilled requests: Over a longer time horizon with multiple planning horizons, physician satisfaction is more equally distributed than with the C strategy, which models what many existing physician scheduling models look like in terms of long-term fairness. Additionally, physician satisfaction fluctuates less between different planning horizons. Physicians can always expect a similar level of satisfaction for the entire time of their employment.

4.3 Managerial insights

We have shown that models without long-term fairness considerations can create plans that are disadvantageous to some physicians. When there are no long-term fairness constraints built into the model, fairness is not guaranteed and depends on heuristics built into the solver. This will agitate repeatedly disadvantaged physicians and create conflict within the work force, which may lead to increased turnover and decreased staff morale. In the absence of conflicting preferences in the data this is not a big issue. In this case, fairness needs to be ensured when entering preferences and not by the model. If, however, there are conflicting preferences in the data, the impact of long-term fairness rises with the rate of conflicting preferences. As the model performs better than conventional approaches without long-term fairness even for a small rate of conflicting preferences, we recommend that schedulers incorporate long-term fairness in their scheduling models, regardless of their current rate of conflicting preferences.

5 Conclusion

This paper proposes an indicator for measuring satisfaction of physicians with automatically generated duty rosters, focusing on the fulfillment of physicians' preferences for duties. We describe a physician scheduling model to assign physicians to overnight duties based on their preferences and incorporate fairness based on our satisfaction indicator. We discuss different strategies for updating this indicator over a time horizon spanning multiple planning horizons. One of our strategies requires a MIQP, which is hard to solve to optimality on consumer hardware. For this model, we provide an equivalent MILP which can be solved using standard software. We apply our models to real-life data from a German university hospital and to generated data based on the real-life data. Our results indicate that tracking fairness measures over several planning horizons is necessary, because otherwise unfair fulfillment of preferences can accumulate over time. This leads to unequal treatment of physicians and, if realized by the disadvantaged physicians, may lead to negative emotions directed towards colleagues or the employer. The proposed model using a satisfaction-based preference weight solves this problem by smoothing unequal treatment of different physicians over time.

Our experiments test three different strategies for satisfaction tracking and updating of the satisfaction-based weights for preference violations: Constant and equal satisfaction (C), satisfaction updates after the planning horizon (ESA), and updates during the planning horizon (ESD). Our results confirm that not tracking satisfaction, as implemented by the C strategy, can indeed lead to unfair distribution of preference fulfillment. With the C strategy, the distribution of preference fulfillment is dependent on the underlying solver heuristics. The ESA strategy mitigates this, but has a different drawback: Satisfaction for each physician oscillates between planning horizons with a relatively high amplitude, therefore physicians do not achieve a consistent satisfaction over several planning horizons. This problem is solved by the ESD strategy, which updates the satisfaction indicator during each planning horizon. It provides better distribution of preference fulfillment among all physicians over several planning horizons while also maintaining a steadier level of satisfaction for each individual physician. However, it comes at the cost of being implemented via a MIQP. We have shown how to transform this MIQP into a MILP to enable the application of linear programming solvers to the model.

We believe our insights are valuable to managers and schedulers creating duty rosters for peer groups of physicians who are interchangeable within the group (i.e., have the same qualifications). Taking into account long-term fairness is often neglected when creating automatic scheduling models. Instead, only data from the current planning horizon is taken into account to create fair rosters. As we have shown, this can lead to unfair distribution of preference fulfillment among physicians in a peer group. Our satisfaction-based weights remedy this problem and enable the creation of schedules which are fair over a longer time horizon. However, care needs to be taken when eliciting preferences. Our satisfaction-based weights are most effective when physicians are not restricted in their choice for preference submission by technical or social limitations. When there are only non-conflicting preferences in the data, it is impossible for the solver to create fairness by favoring disadvantaged physicians within the peer group and our satisfaction-based weights show only minor effects. However, our ESD strategy still performs slightly better than the C strategy, so it can also be used in this case.

Our work opens up possibilities for future research. We use the term "satisfaction" to denote individual preference fulfillment for a physician. Our assumption is that physicians are more satisfied with a plan when they have a higher rate of preference fulfillment. Future research might verify this relationship empirically. We believe our satisfaction indicator could be adapted to be used for fair distribution of workload, not unlike a working time account. Currently, we only distribute duties based on preferences. Future research could focus on an indicator combining fairness in preference fulfillment with workload distribution. For example, a physician who was unwillingly assigned to a lot of duties could be given a higher weight for off-duty requests in the following planning horizon.

The same approach could also be used for non-medical personnel in the health-care sector, such as logistics assistants. Current models for logistics assistant planning also lack fair workload distribution [28]. Additionally, we see a multitude of opportunities to which our satisfaction indicator could be applied. Previous approaches have, for example, tried to minimize the amount of unpopular shifts assigned to each physician. Instead of counting satisfied preferences, our satisfaction indicator could be adapted to count the instances in which unpopular shifts were not assigned. Alternatively, off-duty requests could be generated for each unpopular shift and the model could be used as presented.

While our model is based on experience levels for overnight/weekend staffing, it could also be adapted to handle situations which require physicians of different specialties to be present. This could be applied to hospitals which schedule physicians during nights and weekends according to their subspecialties, as proposed by Dexter et al. [8]. In this case, a separate duty should be created for each subspecialty (instead of for each experience level, as proposed by us) and physicians would be qualified for the respective subspecialty duty. With an appropriate demand setting for each duty, this would ensure that physicians of each specialty are available during the night or on weekends. Alternatively, the physicians could be partitioned into groups based on specialty and the model could then be applied to each group separately. However, the value added by subspecialties as perceived by physicians is mixed [9, 22].

References

[1] Alsheddy A, Tsang EPK (2011) Empowerment scheduling for a field workforce. Journal of Scheduling 14(6):639–654, DOI 10.1007/s10951-011-0232-2

[2] Bard JF, Purnomo HW (2005) Preference scheduling for nurses using column gen-

eration. European Journal of Operational Research 164(2):510–534, DOI 10.1016/j.ejor.2003.06.046

[3] Bertsimas D, Farias VF, Trichakis N (2011) The price of fairness. Operations research 59(1):17–31, DOI 10.1287/opre.1100.0865

[4] de Boer EM, Bakker AB, Syroit JE, Schaufeli WB (2002) Unfairness at work as a predictor of absenteeism. Journal of Organizational Behavior 23(2):181–197, DOI 10.1002/job.135

[5] Bowers MR, Noon CE, Wu W, Bass JK (2016) Neonatal physician scheduling at the University of Tennessee Medical Center. Interfaces 46(2):168–182, DOI 10.1287/inte.2015.0839

[6] Brunner JO, Bard JF, Kolisch R (2009) Flexible shift scheduling of physicians. Health Care Management Science 12(3):285–305, DOI 10.1007/s10729-008-9095-2

[7] 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, DOI 10.1287/inte.1080.0369

[8] Dexter F, Epstein RH, Dutton RP, Kordylewski H, Ledolter J, Rosenberg H, Hindman BJ (2016) Diversity and similarity of anesthesia procedures in the United States during and among regular work hours, evenings, and weekends. Anesthesia & Analgesia 123(6):1567–1573, DOI 10.1213/ane.0000000000001558

[9] Dexter F, Ledolter J, Epstein RH, Hindman BJ (2017) Operating room anesthesia subspecialization is not associated with significantly greater quality of supervision of anesthesia residents and nurse anesthetists. Anesthesia & Analgesia 124(4):1253–1260, DOI 10.1213/ane.0000000000001671

[10] Erhard M, Schoenfelder J, Fügener A, Brunner JO (2018) State of the art in

physician scheduling. European Journal of Operational Research 265(1):1–18, DOI 10.1016/j.ejor.2017.06.037

[11] Ernst EA, Matlak EW (1974) On-line computer scheduling of anesthesiologists. Anesthesia & Analgesia 53(6):854–858, URL http://journals.lww.com/anesthesia-analgesia/Fulltext/1974/53060/On_Line_Computer_Scheduling_of_Anesthesiologists.9.aspx

[12] Fredrickson BL (2001) The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions. American Psychologist 56(3):218–226, DOI 10.1037/0003-066x.56.3.218

[13] Fredrickson BL, Joiner T (2002) Positive emotions trigger upward spirals toward emotional well-being. Psychological Science 13(2):172–175, DOI 10.1111/1467-9280.00431

[14] Fügener A, Brunner JO, Podtschaske A (2015) Duty and workstation rostering considering preferences and fairness: A case study at a department of anesthesiology. International Journal of Production Research 53(24):7465–7487, DOI 10.1080/00207543.2015.1082667

[15] Gierl L, Pollwein B, Heidenreich D, Heyde G (1986) Personaleinsatzplanung für die Anästhesiologie mit Hilfe eines Expertensystems (PEP). In: Horbach L (ed) Medizinische Informatik in der Schweiz, Schwabe, Basel u.a., pp 211–217

[16] Gierl L, Pollwein B, Heyde G, Kurt H (1993) Knowledge-based scheduling of duty rosters for physicians. Medical Informatics 18(4):355–366, DOI 10.3109/14639239309025324

[17] Gross CN, Fügener A, Brunner JO (2018) Online rescheduling of physicians in hospitals. Flexible Services and Manufacturing Journal 30(1):296–328, DOI 10.1007/s10696-016-9274-2

[18] Jäger DT, Rüsseler J (2016) Low arousing positive affect broadens visual attention and alters the thought-action repertoire while broadened visual attention does not. Frontiers in Psychology 7, DOI 10.3389/fpsyg.2016.01652

[19] Jain RK, Chiu DMW, Hawe WR (1984) A quantitative measure of fairness and discrimination for resource allocation in shared computer system. Tech. rep., Eastern Research Lab

[20] Karsu Ö, Morton A (2015) Inequity averse optimization in operational research. European Journal of Operational Research 245(2):343–359, DOI 10.1016/j.ejor.2015.02.035

[21] Loewenstein G, Lerner JS (2003) The role of affect in decision making. In: Davidson RJ, Scherer KR, Goldsmith HH (eds) Handbook of Affective Sciences, Oxford University Press, chap 31, pp 619–642

[22] Lubarsky DA, Reves JG (2005) Effect of subspecialty organization of an academic department of anesthesiology on faculty perceptions of the workplace. Journal of the American College of Surgeons 201(3):434–437, DOI 10.1016/j.jamcollsurg.2005.04.014

[23] Martin S, Ouelhadj D, Smet P, Berghe GV, zcan E (2013) Cooperative search for fair nurse rosters. Expert Systems with Applications 40(16):6674–6683, DOI 10.1016/j.eswa.2013.06.019

[24] Niroumandrad N, Lahrichi N (2017) A stochastic tabu search algorithm to align physician schedule with patient flow. Health Care Management Science DOI 10.1007/s10729-017-9427-1

[25] O'Neill L, Hartz AJ (2012) Lower mortality rates at cardiac specialty hospitals traceable to healthier patients and to doctors' performing more procedures. Health Affairs 31(4):806–815, DOI 10.1377/hlthaff.2011.0624

[26] Ouelhadj D, Martin S, Smet P, Ozcan E, Vanden Berghe G (2012) Fairness in nurse rostering. Working Paper, URL `https://researchportal.port.ac.uk/portal/en/publications/fairness-in-nurse-rostering(a87fbb68-09ef-446f-9084-1266e82ccc53)/export.html`

[27] Stolletz R, Brunner JO (2012) Fair optimization of fortnightly physician schedules with flexible shifts. European Journal of Operational Research 219(3):622–629, DOI 10.1016/j.ejor.2011.10.038

[28] Volland J, Fügener A, Brunner JO (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, DOI 10.1016/j.ejor.2016.12.026

Appendix C

Long-term workload equality on duty schedules for physicians in hospitals

The following contribution (Gross, 2018) has been published in "PATAT2018: Proceedings of the 12th International Conference on the Practice and Theory of Automated Timetabling", which is not ranked in the VHB-JOURQUAL3 ranking (Verband der Hochschullehrer für Betriebswirtschaft e.V., 2015). Additionally, it has been submitted to "Annals of Operations Research" (category B) for inclusion in a PATAT special issue. The submitted version is reproduced below in its entirety.

Long-term workload equality on duty schedules for physicians in hospitals

Christopher N. Gross

ORCID 0000-0002-7306-5767

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

Department of Health Care Operations / Health Information Management

Faculty of Business and Economics

University of Augsburg

Universitätsstraße 16, 86159 Augsburg, Germany

e-mail: christopher.gross@unikat.uni-augsburg.de

Long-term workload equality on duty schedules for physicians in hospitals

Abstract

Staffing hospitals 24 hours a day requires some physicians to be assigned to overnight duties via duty schedules. As overnight duties have an impact on physicians' personal life, physicians can submit preferences indicating when they would prefer to perform duties and when they would prefer not to be assigned to a duty. The created schedule then tries to respect those preferences. However, some duties are assigned to physicians who have not requested them, simply because nobody requested these duties and they have to be covered. This workload should be evenly distributed. We propose a workload indicator that tracks how much duties physicians perform over several planning horizons. The workload from the past is then used to inform decisions on the current plan. Our workload indicator is integrated into a scheduling model for physicians in hospitals. The application of our model to test data shows that our workload indicator helps to spread workload over all physicians more evenly.

Keywords: OR in health services; mixed-integer program; physician scheduling; long-term fairness; workload distribution

1 Introduction

Patients in hospitals require around-the-clock care. To provide this, some physicians have to stay overnight at the hospital and perform so-called overnight duties. Identifying which physician should stay over night is a complex task. Physicians are assigned to overnight duties by duty rosters, which need to satisfy a multitude of constraints, such as working time regulations, minimum staffing levels, and required experience levels. As duties span the entire night, physicians on duty need to be present at the hospital throughout the night and need to plan their private lives accordingly. This makes performing duties quite demanding on physicians. The scheduling process should therefore incorporate physicians' preferences for which duties they want to perform. Physicians can request to be assigned to certain duties and can also request to not be assigned to any duties on a specific date. Additionally, the workload of duties should be evenly distributed among all physicians.

Duty rosters for physicians are usually created monthly. Models creating duty rosters therefore consider a planning horizon of 4 to 5 weeks. Many models in the existing physician scheduling literature only optimize for the current planning horizon and do not take into account data from previous planning horizons. When thinking about physician satisfaction, this might mean that some physicians are repeatedly disadvantaged in sequential months. In terms of preference fulfillment, it is possible that some physicians are repeatedly denied their duty requests whereas other physicians are repeatedly granted their requests. Gross et al. (2018a) show that using a model that does not equalize preference fulfillment over several planning horizons creates unequal preference fulfillment between physicians. With these models, the fulfillment of preferences for an individual physician – and therefore this physician's satisfaction – is based on solver implementation details. Gross et al. (2018a) propose a satisfaction indicator for preference fulfillment to mitigate this problem and equalize preference fulfillment over all physicians over several planning horizons. However, they do not take into account duties which have not been requested via a preference but must still be assigned to ensure adequate staffing of the hospital. For these duties, the problem is similar to the one with preference satisfaction: It is possible that some physicians are repeatedly assigned to many duties they did not request while others are not.

This work has two main contributions. First, we propose a workload indicator, modeled similarly to the satisfaction indicator by Gross et al. (2018a). Second, we evaluate the impact of our workload indicator. We incorporate both the existing satisfaction indicator and our proposed workload indicator into a physician scheduling model. The model is then compared to the model with only the satisfaction indicator and to a model with only the workload indicator by applying it to the data used by Gross et al. (2018a). We generate additional data with a varying number of preferences and find that the effectiveness of the workload indicator is tied to the number of preferences when used together with the satisfaction indicator. Results indicate that our workload indicator succeeds in equalizing workload among physicians over several planning horizons.

The remainder of this paper is structured as follows. We review related literature on equal workload distribution in personnel scheduling in section 2. Afterwards, we provide a description of long-term equality considerations in physician scheduling and our workload indicator in section 3. A physician scheduling model for equal workload distribution is presented in section 4. Section 5 describes the application of our model to data and its results. Our work is concluded by section 6, which summarizes our findings.

2 Literature

We review literature on scheduling which considers equal workload distribution, with a focus on scheduling in the health care sector. Most of the reviewed literature does not incorporate long-term equality of workload over several planning horizons. For more literature on the topics of this work, we refer to respective literature reviews on inequity averse optimization (Karsu and Morton, 2015), staff scheduling (Ernst et al., 2004), nurse scheduling (Cheang et al., 2003), and physician scheduling (Erhard et al., 2018). Further literature on fairness in preference fulfillment can also be found in Gross et al. (2018a).

A simple approach to ensure equal workload distribution is setting an upper limit on the individual workload. The international nurse rostering competition (Haspeslagh et al., 2014) defines constraints for fairness, such as maximum and minimum number of shifts assigned to a nurse or the maximum and minimum number of consecutive days on which a nurse does not have a shift assigned. Additionally, nurses can request to be assigned to a specific shift or to not be assigned to shifts on a given day. A solution is always created for the current planning horizon without taking into account fairness data from the previous planning horizon. The second iteration of the international nurse rostering competition (Ceschia et al., 2015) introduces a multi-stage scheduling problem. Limits, such as on the amount of shifts per nurse, were defined as a sum over several planning horizons. Scheduling therefore requires data from previous plans to make decisions in the current planning horizon, Furthermore, forecasts for plans in the future are required to assign shifts in the current planning horizon in such a way that plans in the future are feasible.

A similar method of restricting individual workload is employed by Fügener et al. (2015) and Gross et al. (2018b). They assign overnight duties to physicians in a one-month planning horizon and set an upper limit on the duties which should be assigned to any individual physician. Any duty above this limit is then penalized with a constant weight, encouraging the model to assign this duty to another physician whose assignments are still below the limit. This approach, however, cannot ensure equal workload distribution above or under the limit. Let (a, b) be an assignment of duties to two physicians where the first physician is assigned to a duties and the second physician is assigned to b duties. If the limit is, e.g., 2 duties per planning horizon, then the assignments (1, 1) and (2, 0) are considered equal. Similarly, the assignments (2, 4) and (3, 3) are also considered equal. Both properties are obviously undesirable. Additionally, the number of assigned duties above or below the limit is not carried over into the next planning horizon.

Stolletz and Brunner (2012) create fortnightly physician schedules by using flexible shift start and end times. Workload is measured in terms of over- and undertime as well as number of assigned overnight duties. Their proposed model succeeds in creating schedules which assign exactly the same amount of over- and undertime to each physician. Due to the number of required overnight duties not necessarily being divisible by the number of available physicians, the model cannot assign the same number of overnight duties to each physician. However, it creates an assignment where differences between physicians are minimized. The amount of assigned over-/undertime and the amount of assigned overnight duties are not carried over into the next planning horizon.

A survey of constraints on physician scheduling in practice is conducted by Gendreau et al. (2006). They study the scheduling process in five hospitals in the area of Montreal, Canada. The study categorizes scheduling constraints into, among others, “workload constraints” and “fairness constraints”, both with the effect of limiting the amount of work that is assigned to a single physician and distributing the work equally among all physicians. In the first category, the authors find limits on the amount of working hours or number of assigned shifts per physician. This is the same approach as used by Gross et al. (2018b) and Fügener et al. (2015). The second category describes constraints such as a fixed number of shifts that needs to be assigned to all physicians with the same experience level, or a maximum number of weekends shifts in a certain period. There is no mention that deviations from these constraints are carried over into the next planning horizon.

A different approach to long-term fairness is taken by Carrasco (2010). Instead of carrying inequalities into the next planning horizon, he chooses a sufficiently large planning horizon to be able to create equal assignments during the planning horizon. While most scheduling models use a one-month planning horizon, this approach uses 12 months and creates overnight duty rosters for a Spanish hospital. As such large instances cannot be solved using linear optimization in acceptable time and memory, an algorithm is proposed that creates schedules while selecting employees for shifts in such a way that the workload is equalized throughout the planning horizon. However, this schedule cannot incorporate midterm changes which are not known at the time of schedule creation, such as personnel turnover. Vacation also has to be planned either ahead of schedule creation or by exchanging shifts between physicians to keep the global workload balance intact.

Summarizing, current work on fairness in physician scheduling largely does not consider long-term fairness over several months. Most approaches just provide fairness during the planning horizon and do not take into account data from the past. As planning horizons in physician scheduling are usually one to two months, this is not a sufficient time span to create total fairness among physicians. Some existing approaches solve this problem by extending the planning horizon to a year but this creates new problems as required changes to such a plan will be abundant. Research on how to provide long-term fairness on physician duty schedules between several planning horizons of one month is currently lacking.

3 Physician scheduling and long-term equality considerations

We now provide a more detailed description of the physician scheduling problem with a focus on considerations of long-term equality among physicians over several planning horizons. In hospitals, patients need to be cared for around the clock. During the day, a sufficient number of physicians is always present to perform scheduled procedures. During the night, no procedures are scheduled and physicians are only present to ensure adequate care in case of emergencies. Therefore, there are only a few physicians present during night hours. To ensure that a sufficient number of physicians is always present, physicians are assigned to overnight duties via a roster. These rosters have to fulfill many requirements, which makes their creation quite complex. For a sufficiently large number of physicians the number of possible schedules is virtually endless. This makes it hard for human schedulers to create a schedule respecting all the constraints. Often, software-assisted scheduling is employed to create optimal schedules. Research on physician scheduling is abundant. A recent review of this area of research was published by Erhard et al. (2018). They find that many works on rostering problems, such as the one we describe, create rosters for planning horizons of up to six weeks. This is a comparatively short time span when the goal is to equalize workload or fulfillment of physician preferences. However, not many works take into account data from the previous planning horizon to create the next roster. This opens up the possibility that some physicians are disadvantaged by the plan repeatedly.

Gross et al. (2018a) show that this is not only a theoretical possibility but rather a real problem which can occur in practice. They calculate the satisfaction of physicians in terms of preference fulfillment and propose several strategies for equalizing physician satisfaction over all physicians. The constant strategy, denoted by C, only maximizes satisfaction over all physicians in the current planning horizon. This is what many similar physician rostering models implement. The second strategy, denoted by ESA, calculates a satisfaction indicator for each physician for each planning horizon and then uses that satisfaction indicator to calculate the individual physician's preference weight for the next planning horizon. Their third strategy, denoted by ESD, calculates the satisfaction not only after planning horizons but continuously updates the satisfaction and physicians' individual weights online during the rostering process. In this case, the preference weight is a decision variable based on the amount of satisfied preferences, i.e., when the model satisfies a preference it simultaneously changes the physician's preference weight. An application of their models to 24 months of data based on a real-life problem shows that the C strategy indeed disadvantages certain physicians in the long run. In contrast, the ESD strategy performs best when considering the equal distribution of satisfaction over all physicians after all 24 planning horizons as well as the variance of satisfaction between planning horizons for each physician. Or in other words: After 24 planning horizons, schedules created with the ESD strategy distributed satisfaction among physician more equally than the C strategy. Additionally, the ESD strategy achieved a more stable level of satisfaction between months for each physician than the ESA strategy.

An additional factor in physician happiness is the distribution of the workload. The problem here is similar: slightly unequal distribution of workload during one planning horizon does not have a huge impact, but if the same set of physicians is repeatedly assigned a higher workload over several planning horizons, this unequal treatment adds up and can lead to physician attrition. The focus of this work is therefore the equal distribution of workload over many planning horizons. Our main contribution is the introduction of a workload indicator for physician scheduling. We integrate this workload indicator into a scheduling model and combine it with the satisfaction indicator introduced by Gross et al. (2018a).

3.1 Physician-specific workload indicator

To measure the individual workload of each physician, we define a workload indicator. This workload indicator can be calculated for each duty roster and for each physician. Intuitively, we define the workload indicator as the amount of overnight duties a physician is assigned on a roster divided by the number of days in the roster. This gives us an approximation of the number of overnight duties this physician performs per day. Note that the division by the number of days is unavoidable as duty rosters have different lengths (4 or 5 weeks) in different planning horizons. We define the workload indicator on a roster which assigns physicians $\mathcal{J}$ to duties $\mathcal{I}$ on days $\mathcal{D}$ in weeks $\mathcal{W}$. x_{jiwd} is a binary variable which is 1 if physician j is assigned to duty i on day d of week w and 0 otherwise. The workload indicator λ_j of physician j is then defined as follows.

$$\lambda_j = \frac{\sum_{i\in\mathcal{I}} \sum_{w\in\mathcal{W}} \sum_{d\in\mathcal{D}} x_{jiwd}}{|\mathcal{W}| \cdot |\mathcal{D}|}$$

Our approach differs from existing work by its focus on the individual physician. Existing approaches discuss indicators for groups, such as minimizing the maximum deviation from the average of a group's working hours. Our workload indicator, however, tracks the workload for each physician individually, regardless of other physicians' workload. This individual indicator can then be used to calculate weights which can compare physicians in a group with each other without requiring some sort of group baseline for working hours or number of assigned duties.

4 Model

We propose a model to assign overnight duties to physicians. Our model is derived from the model proposed by Gross et al. (2018a). It assigns physicians $j \in \mathcal{J}$ to duties $i \in \mathcal{I}$ on days $d \in \mathcal{D}$ of weeks $w \in \mathcal{W}$. Each physician can be assigned to at most one duty per day and needs to be given the next day off, i.e., physicians cannot be assigned to duties on consecutive days. Similarly, physicians cannot be assigned to duties on consecutive weekends. The number of required physicians for a duty i on the day of the week d is given by $\bar{d}_{id}^{duty}$. We implement this demand as an upper bound, so we can always find feasible solutions even if there is an insufficient number of physicians available to perform a duty. This undercoverage can then be seen in decision variables $\Delta_{iwd}^{\text{out-duty}}$ and will need to be handled by human schedulers. Not every physician should be assigned to every duty. Our model requires parameters E_{jiwd}^{pos} to specify which physician can be assigned to which duty on which day. As we also do not want to assign physicians to duties when they are absent, our model respects absences supplied in parameters D_{jwd}^{off}. Physicians want to have a say in which duties they are assigned to. We provide physicians' preferences to be assigned to a duty in parameters $g_{jiwd}^{\text{req-on}}$ and the preferences to not be assigned to a duty in parameters $g_{jwd}^{\text{req-off}}$. Our model tracks the violations of these preferences with variables $\Delta_{jiwd}^{\text{req-on}}$ and $\Delta_{jwd}^{\text{req-off}}$. As we consider preference fulfillment and workload distribution from past planning horizons, we provide the historic satisfaction with preference fulfillment in parameters $\hat{s}_j$ and the historic workload in parameters $\hat{l}_j$.

The model we introduce has two sets of conflicting constraints. The first set consists of $\{(11a), (11b)\}$, and the second set consists of $\{(13a), (13b)\}$. Note that these numbers refer to constraints introduced below. We choose not to repeat them here to avoid duplication. When implementing the model, only one constraint out of each set can be used, otherwise the model becomes infeasible. As our model is geared towards evaluating the impact of satisfaction and workload indicators, these constraints each implement a different strategy for updating these indicators. Regarding equal distribution of preference fulfillment, we define two different strategies:

I. No preference fulfillment (unfair, U) For this strategy, we use constraints (11a). This effectively disables preferences in the model, meaning the solver will not optimize for the fulfillment of preferences. With this strategy, physicians' preferences are completely ignored.

II. Long-term fair preference fulfillment (fair, F) This strategy is implemented by constraints (11b). These constraints implement the ESD strategy as proposed by Gross et al. (2018a). They use a physician-specific satisfaction indicator to ensure long-term equal preference fulfillment among all physicians while updating the satisfaction continuously during the solving of the model. This is achieved by basing the satisfaction-based weight directly on the satisfaction indicator of the current planning horizon, thereby requiring a quadratic decision model. They describe how this can be transformed into a linear decision model. Note that we do not implement the other strategies for equal preference fulfillment proposed by Gross et al. (2018a) because they find that the ESD strategy creates superior results for the APS and ASV performance indicators in comparison to all other tested strategies.

To achieve equal distribution of workload, we define two different strategies:

I. Constant workload-based weight (CL) We set the workload-based weight to 0 for all physicians. This effectively disables tracking the workload among physicians. For this strategy, we use constraints (13a).

II. Exponential smoothing for workload-based weight during the planning horizon (ESL) We update the workload-based weight in the model, i.e., during the planning horizon, depending on the amount of duties assigned to the respective physician. This strategy requires the use of constraints (13b). Note that this results in a quadratic decision model as well, for which we describe an equivalent linear formulation in section 4.1.

In their study of the satisfaction indicator, Gross et al. (2018a) also evaluate a strategy with exponential smoothing of the satisfaction indicator after each planning horizon and no updating during the planning horizon. Based on the results found in their evaluation, we refrain from applying this strategy to the workload indicator and defining a model with exponential smoothing of the workload indicator after the planning horizon. This strategy has the downside of alternating between high and low values for the smoothed values, leading to a high fluctuation of the weights derived from them between planning horizons. As these results discovered for the satisfaction indicator can be generalized, we expect the same results for a similar strategy for the workload-based weights. We therefore only incorporate the ESL strategy with exponential smoothing in the model itself.

Sets and indices	
$d \in \mathcal{D} = \{1, ..., 7\}$	Days of the week, starting with Monday = 1
$i \in \mathcal{I}$	Duties
$j \in \mathcal{J}$	Physicians
$w \in \mathcal{W}$	Weeks in the planning horizon
Parameters	
α_1	Weight for personnel demand coverage
α_2	Weight for preference fulfillment
α_3	Weight for workload distribution
γ_1	Smoothing constant for satisfaction indicator
γ_2	Smoothing constant for workload indicator
$\hat{s}_j$	Pre-computed value as an input to calculate satisfaction-based weight for physician j based on the previous planning horizon

$\hat{l}_j$	Pre-computed value as input to calculate workload-based weight for physician j based on the previous planning horizon
$\bar{d}^{\text{duty}}_{id}$	Demand of physicians for duty i on day d
$g^{\text{req-on}}_{jiwd}$	1 if physician j has a preference for duty i on day d of week w, 0 otherwise
$g^{\text{req-off}}_{jwd}$	1 if physician j has a preference for being off duty on day d of week w, 0 otherwise
E^{pos}_{jiwd}	1 if physician j can be assigned to duty i on day d of week w, 0 otherwise
D^{off}_{jwd}	1 if physician j is absent on day d of week w, 0 otherwise
Decision variables	
$x_{jiwd} \in \{0,1\}$	1 if physician j is assigned to duty i on day d of week w, 0 otherwise
$s_j \in [0,1] \subset \mathbb{R}$	Satisfaction-based weight for preferences of physician j for the current planning horizon
$\sigma_j \in [0,1] \subset \mathbb{R}$	Satisfaction indicator for physician j (Gross et al., 2018a)
$l_j \in [0,1] \subset \mathbb{R}$	Workload-based weight for assignment of duties to physician j for the current planning horizon
$\lambda_j \in [0,1] \subset \mathbb{N}_0$	Workload indicator for physician j
$x^{\text{WE}}_{jw} \in \{0,1\}$	1 if physician j is assigned to a duty on the weekend of week w, 0 otherwise
$\Delta^{\text{out-duty}}_{iwd} \in \mathbb{N}_0$	Missing physicians to cover demand of duty i on day d of week w
$\Delta^{\text{req-on}}_{jiwd} \in \{0,1\}$	1 if preference of physician j for duty i on day d of week w is not satisfied, 0 otherwise
$\Delta^{\text{req-off}}_{jwd} \in \{0,1\}$	1 if preference of physician j for being off duty on day d of week w is not satisfied, 0 otherwise

Minimize

$$\alpha_1 \cdot \sum_{i\in\mathcal{I}} \sum_{w\in\mathcal{W}} \sum_{d\in\mathcal{D}} \Delta^{\text{out-duty}}_{iwd} + \tag{1a}$$

$$\alpha_2 \cdot \sum_{j \in \mathcal{J}} \left((2 - s_j) \cdot \left(\sum_{i \in \mathcal{I}} \sum_{w \in \mathcal{W}} \sum_{d \in \mathcal{D}} \Delta^{\text{req-on}}_{jiwd} + \sum_{w \in \mathcal{W}} \sum_{d \in \mathcal{D}} \Delta^{\text{req-off}}_{jwd} \right) \right) + \tag{1b}$$

$$\alpha_3 \cdot \sum_{j \in \mathcal{J}} \left(l_j \cdot \sum_{i \in \mathcal{I}} \sum_{w \in \mathcal{W}} \sum_{d \in \mathcal{D}} x_{jiwd} \right) \tag{1c}$$

subject to

$$\sum_{j \in \mathcal{J}} x_{jiwd} + \Delta^{\text{out-duty}}_{iwd} = \bar{d}^{\text{duty}}_{id} \quad \forall i \in \mathcal{I}, w \in \mathcal{W}, d \in \mathcal{D} \tag{2}$$

$$\Delta^{\text{req-on}}_{jiwd} = g^{\text{req-on}}_{jiwd} \cdot (1 - x_{jiwd}) \quad \forall j \in \mathcal{J}, i \in \mathcal{I}, w \in \mathcal{W}, d \in \mathcal{D} \tag{3}$$

$$\Delta^{\text{req-off}}_{jwd} = g^{\text{req-off}}_{jwd} \cdot \sum_{i \in \mathcal{I}} x_{jiwd} \quad \forall j \in \mathcal{J}, w \in \mathcal{W}, d \in \mathcal{D} \tag{4}$$

$$\sum_{i \in \mathcal{I}} x_{jiwd} \leq 1 \quad \forall j \in \mathcal{J}, w \in \mathcal{W}, d \in \mathcal{D} \tag{5}$$

$$\sum_{j \in \mathcal{J}} \sum_{i \in \mathcal{I}} \sum_{w \in \mathcal{W}} \sum_{d \in \mathcal{D}: D^{\text{off}}_{jwd} = 1} x_{jiwd} \leq 0 \tag{6}$$

$$\sum_{j \in \mathcal{J}} \sum_{i \in \mathcal{I}} \sum_{w \in \mathcal{W}} \sum_{d \in \mathcal{D}: E^{\text{pos}}_{jiwd} = 0} x_{jiwd} \leq 0 \tag{7}$$

$$\sum_{i \in \mathcal{I}} x_{jiwd} + \sum_{i \in \mathcal{I}} x_{jiw(d-1)} \leq 1 \quad \forall j \in \mathcal{J}, w \in \mathcal{W}, d \in \mathcal{D}, d > 1 \tag{8a}$$

$$\sum_{i \in \mathcal{I}} x_{ji(w-1)7} + \sum_{i \in \mathcal{I}} x_{jiw,1} \leq 1 \quad \forall j \in \mathcal{J}, w \in \mathcal{W}, w > 1 \tag{8b}$$

$$x^{\text{WE}}_{jw} + x^{\text{WE}}_{j(w-1)} \leq 1 \quad \forall j \in \mathcal{J}, w \in \mathcal{W}, w > 1 \tag{9}$$

$$\sum_{i \in \mathcal{I}} \sum_{d \in \{6,7\}} x_{jiwd} \leq 2 \cdot x^{\text{WE}}_{jw} \quad \forall j \in \mathcal{J}, w \in \mathcal{W} \tag{10}$$

$$s_j = 2 \quad \forall j \in \mathcal{J} \tag{11a}$$

$$s_j = \gamma_1 \cdot \sigma_j + (1 - \gamma_1) \cdot \hat{s}_j \quad \forall j \in \mathcal{J} \tag{11b}$$

$$\sigma_j = \frac{\sum_{i \in \mathcal{I}} \sum_{w \in \mathcal{W}} \sum_{d \in \mathcal{D}} \left(g^{\text{req-on}}_{jiwd} - \Delta^{\text{req-on}}_{jiwd} \right) + \sum_{w \in \mathcal{W}} \sum_{d \in \mathcal{D}} \left(g^{\text{req-off}}_{jwd} - \Delta^{\text{req-off}}_{jwd} \right)}{|\mathcal{W}| \cdot |\mathcal{D}|} \quad \forall j \in \mathcal{J} \tag{12}$$

$$l_j = 0 \quad \forall j \in \mathcal{J} \tag{13a}$$

$$l_j = \gamma_2 \cdot \lambda_j + (1 - \gamma_2) \cdot \hat{l}_j \quad \forall j \in \mathcal{J} \tag{13b}$$

$$\lambda_j = \frac{\sum\limits_{i \in \mathcal{I}} \sum\limits_{w \in \mathcal{W}} \sum\limits_{d \in \mathcal{D}} x_{jiwd}}{|\mathcal{W}| \cdot |\mathcal{D}|} \quad \forall j \in \mathcal{J} \tag{14}$$

$$x_{jiwd}, x_{jw}^{\text{WE}}, \Delta_{jiwd}^{\text{req-on}}, \Delta_{jwd}^{\text{req-off}} \in \{0, 1\} \quad \forall j \in \mathcal{J}, i \in \mathcal{I}, w \in \mathcal{W}, d \in \mathcal{D} \tag{15}$$

$$s_j, \sigma_j, l_j, \lambda_j \in \mathbb{R}^+ \quad \forall j \in \mathcal{J} \tag{16}$$

$$\Delta_{iwd}^{\text{out-duty}} \in \mathbb{N}_0 \quad \forall i \in \mathcal{I}, w \in \mathcal{W}, d \in \mathcal{D} \tag{17}$$

Our objective function describes our three main objectives. Term (1a) penalizes undercoverage, i.e., assigning an insufficient number of physicians to duties. Term (1b) penalizes preference violations based on physician-specific satisfaction-based weights. We weight the preference violations with $\alpha_2 \cdot (2 - s_j)$, as recommended by Gross et al. (2018a). This is required because $0 \leq s_j \leq 1$ and smaller values for s_j should result in a higher weight. The strategy to calculate the satisfaction-based weight s_j based on the satisfaction indicator is implemented by constraints (11a) or (11b). Finally, term (1c) punishes assigning duties to physicians. This term contains the individual physician's workload-based weight. Our model will therefore incur different penalties for assigning a duty, based on the workload of the physician to which we are assigning the duty. This guides the model towards assigning duties to physicians with a lower workload-based weight and therefore results in a more equal distribution of assigned duties among physicians. The strategy to calculate the workload-based weight l_j based on the workload indicator is implemented by constraints (13a) or (13b).

To ensure that an adequate number of physicians is assigned to each duty, constraints (2) set deviation variables $\Delta_{iwd}^{\text{out-duty}}$ in case a duty is not covered. Constraints (3) and (4) set deviation variables $\Delta_{jiwd}^{\text{req-on}}$ and $\Delta_{jwd}^{\text{req-off}}$ when physician preferences for a certain duty or for not being assigned to any duty are violated. To ensure that a physician is not assigned to more than one duty per day, we use constraints (5). A valid duty roster cannot assign duties to physicians who are either not present or not qualified for the assigned duty. This is ensured by constraints (6) and (7). As duties span the entire night, physicians need to be given a day off after a duty and cannot be assigned to a duty on the following day. Constraints (8a) ensure this for Tuesday through Sunday and constraints (8b) for Monday. As physicians are unwilling to work duties on consecutive weekends, we track whether a physician is working on a weekend with constraints (10) and prevent assigning duties on consecutive weekends with constraints (9). Constraints (11a) and (11b) specify how our satisfaction-based weights for preference fulfillment are calculated. These constraints are conflicting, so only one of them can be included in the model at any time. Constraints (11a) effectively disable objective (1b), whereas constraints (11b) calculate the satisfaction-based weight using exponential smoothing. The satisfaction indicator for each physician (see Gross et al., 2018a) is calculated for the current planning horizon using constraints (12). Constraints (13a) and (13b) set our workload-based weights. These constraints are conflicting and we can include only one at a time in the model. Constraints (13a) disable physician-specific workload-based distribution of duties (1c), and constraints (13b) calculate the workload-based weight using exponential smoothing on the current workload and the historic workload. See section 5 for how we use constraints (11a), (11b), (13a), and (13b) to evaluate the effects of satisfaction- and workload-based weights. Constraints (14) calculate the workload indicator (see section 3.1) for each physician for the current planning horizon. Finally, constraints (15), (16), and (17) restrict the domains of our decision variables.

4.1 Linearization of the model

In the form above, our decision model is a quadratic decision model. This can be seen in objective terms (1b) and (1c). For term (1b), we apply the linearization described by Gross et al. (2018a). For term (1c), we can identify two cases: In the first case, we use constraints (13a). This means that l_j is set to a fixed value of 0 and can be modeled as a parameter, making the model linear. For the second case, we use constraints (13b). In this case, l_j must be modeled as a decision variable and our decision model is quadratic. We now describe a linearization for the second case to transform the model with constraints (13b) into a linear decision model.

Looking at the workload indicator λ_j for physician j, we find that it depends on the sum of assignments to this physician ($\sum_{i\in\mathcal{I}}\sum_{w\in\mathcal{W}}\sum_{d\in\mathcal{D}} x_{jiwd}$), the amount of days per week $|\mathcal{D}|$, and the amount of weeks in the planning horizon $|\mathcal{W}|$. As the amount of days and weeks are parameters, these values can never change and are always constant for all physicians. The sum of assignments, however, is described by decision variables and is different for each physician. When thinking about what possible values our workload indicator can assume, we therefore need to identify all possible values the sum of assignments can assume. We know that physicians cannot be assigned to duties on consecutive days as they need to rest on a day after a duty. It follows that physicians can at most be assigned to a duty on half of the days in the planning horizon. Additionally, we know that physicians cannot be assigned to more than one duty per day. The upper bound (UB) on the duty assignments for each physician is therefore $\left\lceil \frac{|\mathcal{W}|\cdot|\mathcal{D}|}{2} \right\rceil$.

As x is a binary variable and therefore integer, the sum over x must always be integer. We can therefore enumerate all integers between 0 and the UB to find all possible values for the sum of duty assignments for physician j. We define the set $\mathcal{A}$ as all possible amounts of duty assignments for each physician.

$$\mathcal{A} = \left\{ n \,\middle|\, n \in \mathbb{N}_0 \land n \leq \left\lceil \frac{|\mathcal{W}|\cdot|\mathcal{D}|}{2} \right\rceil \right\}$$

Using these values, we can then pre-calculate the workload indicator λ_j for physician j for all possible values $a \in \mathcal{A}$ and in consequence the workload-based weight l_j for physician j. Using these, we can calculate the workload-based cost c_{ja}^{work} incurred by assigning a duties to physician j as follows.

$$\begin{aligned} c_{ja}^{\text{work}} &= \alpha_3 \cdot l_j \cdot a \\ &= \alpha_3 \cdot \left(\gamma_2 \cdot \lambda_j + (1-\gamma_2) \cdot \hat{l}_j \right) \cdot a \end{aligned}$$

$$= \alpha_3 \cdot \left(\gamma_2 \cdot \frac{a}{|\mathcal{W}| \cdot |\mathcal{D}|} + (1 - \gamma_2) \cdot \hat{l}_j \right) \cdot a$$

As can be seen, c_{ja}^{work} does not depend on any decision variables and can therefore be supplied as parameters. For our objective function, we want to replace the quadratic formulation. In its stead, we want to select the appropriate cost values c_{ja}^{work}. To achieve this, we introduce an additional binary decision variable z_{ja} which is 1 if physician j is assigned to exactly a duties and 0 otherwise. We can then replace term (1c) with the following.

$$\sum_{j \in \mathcal{J}} \sum_{a \in \mathcal{A}} \left(c_{ja}^{\text{work}} \cdot z_{ja} \right) \tag{18}$$

Additionally, we add the following constraints to the model.

$$\sum_{a \in \mathcal{A}} z_{ja} = 1 \quad \forall j \in \mathcal{J} \tag{19}$$

$$\sum_{i \in \mathcal{I}} \sum_{w \in \mathcal{W}} \sum_{d \in \mathcal{D}} x_{jiwd} = \sum_{a \in \mathcal{A}} (a \cdot z_{ja}) \quad \forall j \in \mathcal{J} \tag{20}$$

The quadratic term (1c) in the objective function has now been replaced by the linear term (18). Constraints (19) are linear and ensure that exactly one number of assignments a is selected via z for each physician. Constraints (20) are also linear and ensure that the selected number of assignments a is equal to the actual number of assignments. Therefore, we now have a linear model that we can solve using any MILP solver.

5 Computational Study

To enable a comparison of physician satisfaction and workload between planning horizons, we add an index $m \in \mathcal{M}$ to the satisfaction and workload indicators, with $\mathcal{M}$ being the set of all the months for which we create duty rosters. σ_{jm} and λ_{jm} then describe the satisfaction and workload indicators for physician j on the duty roster for month m, respectively.

We evaluate our results using the APS and ASV performance indicators introduced by Gross et al. (2018a):

1. Variance of average satisfaction indicator per physician over all planning horizons, i.e., months m
$$APS = \operatorname*{Var}_{j \in \mathcal{J}} \left(\frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \sigma_{jm} \right)$$
2. Average of satisfaction indicator variance per physician between planning horizons, i.e., months m
$$ASV = \frac{1}{|\mathcal{J}|} \sum_{j \in \mathcal{J}} \operatorname*{Var}_{m \in \mathcal{M}} (\sigma_{jm})$$

Additionally, we define the following two performance indicators for plan quality in terms of equal distribution of workload.

1. Variance of average workload indicator per physician over all planning horizons, i.e., months m
$$APL = \operatorname*{Var}_{j \in \mathcal{J}} \left(\frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \lambda_{jm} \right)$$
2. Average of workload indicator variance per physician between planning horizons, i.e., months m
$$ALV = \frac{1}{|\mathcal{J}|} \sum_{j \in \mathcal{J}} \operatorname*{Var}_{m \in \mathcal{M}} (\lambda_{jm})$$

All our experiments are run on a VirtualBox 5.2.6 virtual machine with 4 GB of RAM and one virtual core of an Intel i5-4310M CPU, running Xubuntu 16.04.3 on kernel 4.13.0-32. Our model is implemented in CMPL 1.11.0 and is solved with IBM ILOG CPLEX 12.7.1.0. The software used for our experiments is derived from the published code by Gross et al. (2018a) and can be found on GitHub*. Each of our instances can be solved to optimality within 4 seconds. We choose the weights for our computational

*`https://github.com/chrisnig/long-term-workload`

study lexicographically, i.e., $\alpha_1 \gg \alpha_2 \gg \alpha_3$ ($\alpha_1 = 100$, $\alpha_2 = 10$, $\alpha_3 = 1$). This is based on our assumption that the coverage of all duties is much more important than the equal fulfillment of preferences, which in turn is much more important than the equal distribution of workload. Our smoothing constants γ_1 and γ_2 are both set to 0.8. Setting these smoothing constants to a value close to 1 puts more weight on data from the more recent planning horizons in comparison to data from the past. As physicians' happiness depends more on their treatment in the more recent past, we put a high emphasis on the more recent satisfaction and workload data.

5.1 Data with varying conflict rate

First, we apply our model to the data presented by Gross et al. (2018a) on GitHub† which is based on real world data from a German university hospital. This data contains 85 physicians who are employed throughout the time horizon, which consists of 24 months with 4 to 5 weeks each. There are 6 duties to be covered with a demand of 1 physician per day. In the data, no physicians are absent. A preference is defined as being in conflict when there is at least one other preference by a different physician for the same duty on the same date. The *conflict rate* is then defined as the number of conflicting preferences divided by the total number of preferences. The *preference probability* is defined as the probability that a physician is assigned a preference during data generation on any given day. A preference probability of 80 % indicates that for any given physician on any given day the preference generation algorithm assigns a preference with a probability of 80 %. The *preference rate* is then defined as the number of actual preferences divided by the number of days. It follows that the preference rate is always bounded by the preference probability. The data exhibits duty preferences with a preference probability of 80 % and different target conflict rates between 0 and 100 %.

The different strategies for equal preference fulfillment and equal distribution of workload are introduced in section 4. As we need to choose a preference fulfillment strategy

† `https://github.com/chrisnig/long-term-fairness`

and a workload distribution strategy for each experiment, we will denote the combination of the two strategies as A-B, where A is the strategy for preference fulfillment (U or F) and B is the strategy for workload distribution (CL or ESL). See table 2 for an overview of the combinations of strategies used in our study. Initially, we compare the U-CL strategy and the U-ESL strategy and apply both strategies to the data set with a 0 % conflict rate. This means we ignore physician preferences completely and just optimize for coverage (U-CL) or for coverage and equal workload distribution (U-ESL). As we do not consider preferences in either strategy, evaluating the results by the APS and ASV performance indicators is not meaningful, because those only evaluate preference fulfillment. Instead, we only evaluate the results based on the APL and ALV indicators. The results can be found in table 3. For both indicators, we find a dramatic decrease of about 99 % for the U-ESL strategy in comparison with the U-CL strategy. This proves that our workload indicator is effective in influencing our key performance indicators for workload distribution and shows that it can achieve a more equal distribution of workload among physicians (APL) and a more equal distribution of workload between months for each individual physician (ALV).

	preferences		workload	
label	U	F	CL	ESL
U-CL	x		x	
U-ESL	x			x
F-CL		x	x	
F-ESL		x		x

Table 2: Overview of used combinations of preference fulfillment strategies (U/F) and workload distribution strategies (CL/ESL)

The incorporation and fulfillment of physician preferences in the scheduling process is very important, usually even more important than the equal distribution of workload. In order to demonstrate that our workload indicator is also effective when used in conjunction with physician preferences, we now use the F strategy for equal preference fulfillment for all further experiments. We now apply the F-CL and F-ESL strategies for our model to the data and calculate the APS, ASV, APL, and ALV performance indicators. We

	U-CL	U-ESL
APL	0.006075	0.000012
difference to U-CL		-0.006063
difference in %		-99.80 %
ALV	0.006694	0.000082
difference to U-CL		-0.006612
difference in %		-98.78 %

Table 3: Values of the performance indicators for the U-CL and U-ESL strategies using the data set with a 0 % conflict rate

compare each indicator for the F-CL model with the indicator for the F-ESL model and report the change as a percentage in figure 1. Note that the model for the F-CL strategy is identical to the model for the ESD strategy proposed by Gross et al. (2018a) as it only considers equal preference fulfillment and no equal workload distribution.

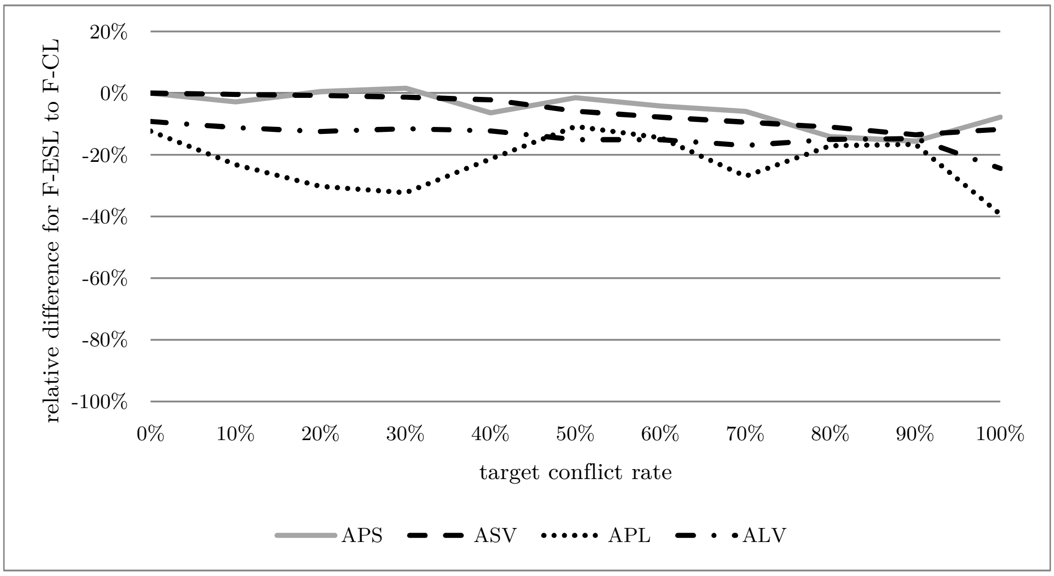


Figure 1: Relative difference in performance indicators between F-CL and F-ESL strategy for data with different target conflict rate

As can be seen from the graph, none of the key performance indicators change a lot between the data sets for different target conflict rates. The improvement in the workload performance indicators APL and ALV fluctuates at -20 % with a slight downward trend with increasing target conflict rate. This can be attributed to the higher preference rate which comes with the higher conflict rate. Data with higher conflict rates have higher preference rates. This can be attributed to the preference/conflict generation algorithm used by Gross et al. (2018a): First, preferences are generated at random with a given preference probability. Second, in order to meet the target conflict rate, if the conflict rate inherent in the generated preferences is too high, conflicting preferences are removed until the target conflict rate is reached. This reduces the preference rate for data with a low target conflict rate. When there are more preferences, the ideal plan is not as highly constrained as it is for lower preference rates: With lower preference rates, there are not that many physicians who have a preference for the same duty on a day. As we penalize preference violations more harshly than unequal distribution of duties, the solver will always try to fulfill these preferences, even if it results in a more unequal distribution of duties. The higher the preference rate, the more likely it is that several physicians enter a preference for the same duty on the same day. The solver can then choose the physician to assign to the duty in such a way that duties are more equally distributed, without incurring a penalty for violating a duty preference.

For the fairness indicators APS and ASV, we can identify a similarly small downward trend with higher target conflict rate. It may seem surprising that improvements, i.e., a negative relative change, is even possible for these performance indicators as the F-CL and F-ESL strategy both contain the same constraints for equal distribution of preference fulfillment. One might assume that optimizing for equal preference fulfillment in the F-CL model would already yield the optimal result for the APS and ASV performance indicators, so that the F-ESL model could not possibly improve on this. However, there are some situations where several optimal solutions for the F-CL model exist, which are different in terms of the APS and ASV performance indicators. The F-ESL strategy will therefore be able to distribute some duties in a different way without achieving a worse score for our fairness objective (1b), but at the same time improving the objective of equal distribution of duties (1c). As the preferences in our test data are also equally distributed, this implicitly improves the fairness performance indicators APS and ASV.

Because the differences between all optimal solutions for the F-CL strategy are small, reductions in the fairness performance indicators are always below 20 %. In some cases, improvements are even impossible and the key performance indicators for fairness are worse when taking into account equal workload distribution in the model, e.g., for a target conflict rate of 30 %.

5.2 Data with varying preference probability

Next, we take a look at how our performance indicators behave for different preference rates. We use the same 24 months of base data and generate preference data with the preference generation algorithm proposed by Gross et al. (2018a). The only difference in our generation strategy can be found in the preference probability, i.e., the probability that a physician submits a preference on any given day in the planning horizon. While Gross et al. (2018a) always assume a preference probability $p^{\text{on}} = 80$ % and adjust the generated data to match a target conflict probability, we vary this probability between 0 % and 100 % and skip the adjustment for the conflict probability. We generate data for each preference probability in increments of 10 percentage points. For each generated data set with a different preference probability, we run the models for both our strategies and calculate the key performance indicators for fairness and equal workload distribution of duties. The graph in figure 2 shows the relative change in key performance indicators between the F-ESL and the F-CL strategy. This table shows the value of the respective key performance indicator for the F-CL strategy subtracted from the value for the F-ESL strategy and the result then divided by the value for the F-CL strategy.

For the *APL* and *ALV* indicators, we see that the biggest improvement can be achieved with a preference probability of 0 %. This is not surprising, as having no preferences at all essentially prevents the model from optimizing the roster according to submitted preferences. Note that this means that the model for the F-CL strategy with a 0 % preference rate is identical to the U-CL model. It is therefore not surprising that we exhibit similar differences in performance indicators as in table 3. The only

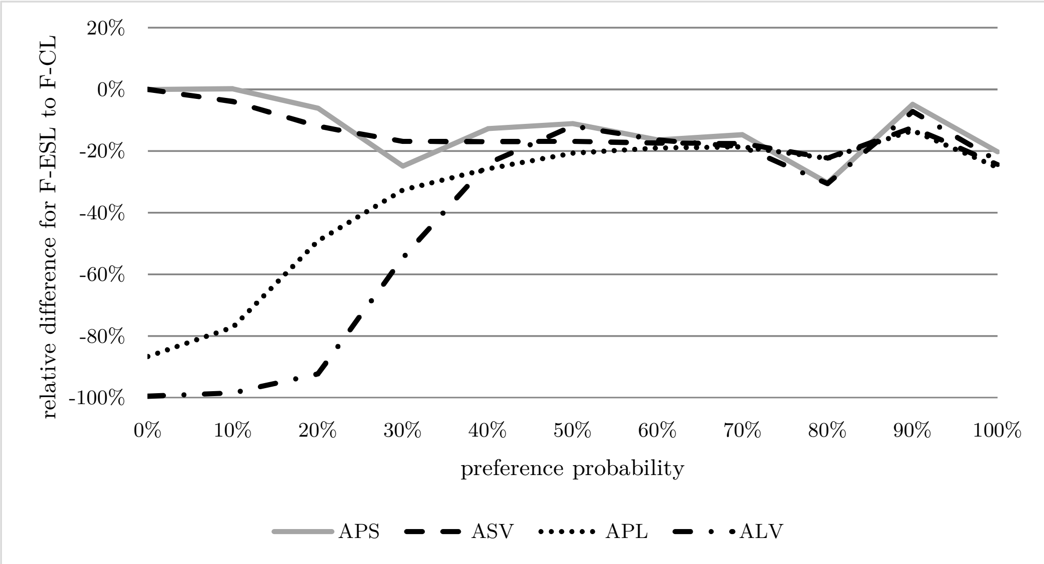


Figure 2: Relative difference in performance indicators between F-ESL and F-CL strategy for data with different preference probability

objective that is weighted higher than the equal distribution of duties is the coverage of all duties. As the coverage is easily fulfilled and does not interfere with how the duties are distributed, the model will distribute the duties only according to our constraints for equal workload distribution. As soon as we start introducing preferences, we can see that the improvement in the APL and APV performance indicators decreases. This is because our model will distribute duties in such a way that preferences are fulfilled first, and only then optimize for the equal distribution of duties among physicians. This means our model will only move duties between physicians to create a more equal distribution if this is not in conflict with the objective of satisfying preferences equally. Essentially, the model can only move duties between physicians who have both submitted a preference for the same duty on the same day or duties which have not been requested by any physician, severely limiting the possibilities for achieving an equal distribution. In consequence, the improvement in the performance indicators for equal distribution (APL, ALV) declines with a rising preference probability.

When looking at the APS and ASV indicators for physician satisfaction, we can see that these start at a 0 % improvement and the improvement then gradually increases until it starts fluctuating around 20 % for a preference probability of 50 % and above. The 0 % improvement for a preference probability of 0 % can be easily explained: This data set does not contain any preferences at all, so no preferences can be fulfilled. In consequence, regardless of what strategy or model we use, the preference fulfillment will always be at 0 and the difference between the performance indicators will also always be 0. As the preference rate increases, the model can start fulfilling preferences. As already explained for the data set with varying conflict rates, our F-ESL strategy will now resolve stalemates in the preference fulfillment objective with a bias to distribute duties more evenly. This, in turn, will also lead to an improvement in the satisfaction performance indicators because our duty preferences are also distributed evenly.

All our performance indicators converge to a constant improvement value around which they fluctuate for preference probabilities over 50 %. For preference rates higher than 50 %, no noticeable change is recognizable in the improvement in the performance indicators. This phenomenon results from the upper limit on preferences we implicitly define. As a physician cannot work two consecutive duties, we also do not allow specifying consecutive preferences. Therefore, the maximum number of preferences per physician is $\left\lceil \frac{|\mathcal{W}| \cdot |\mathcal{D}|}{2} \right\rceil$. The maximum preference rate, i.e., the number of days with a preference divided by the total number of days in the planning horizon, is therefore close to 50 %. The actual preference rate, however, is somewhat lower as the random iteration of the preference generation algorithm makes it unlikely to distribute the preferences in such a way that there is a preference on every other day. It follows that preference probabilities above 50 % will start converging towards a preference rate of less than 50 %, meaning the change in the preference rate between a preference probability of 50 % and 100 % will be smaller than the change in preference rate for a preference probability between 10 % and 50 %. To illustrate this, we show the actual preference rate compared to the preference probability in figure 3, which shows the flattening of the preference rate with increasing preference probability.

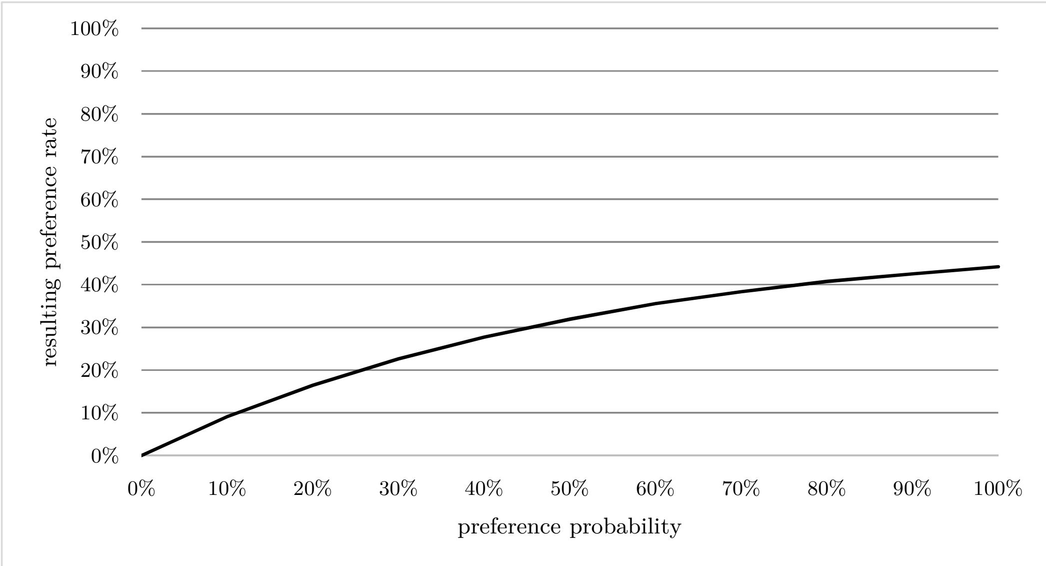


Figure 3: Resulting preference rate when running the preference generation algorithm with different preference probabilities

5.3 Managerial results

Our results indicate that our workload indicator can improve the equal distribution of workload among physicians in all settings. As we reward the equal distribution of preference fulfillment more than the equal distribution of workload, our workload indicator is especially effective when there are no preferences or only a small number of preferences. For a higher number of preferences, our workload indicator can still achieve a more equal distribution of workload, but the improvements are smaller. We therefore recommend implementing our workload indicator to schedulers, regardless of the preference rate in their data. Improvements in the equal distribution of workload will be more noticeable on rosters with less preferences. Additionally, schedulers should identify in which order the objectives of equal preference fulfillment and equal workload distribution are important for satisfaction of their workforce. In our experiments, we assume that equal preference fulfillment is more important than equal workload distribution, which is reflected in our α-weights. These weights (α_2 for equal preference fulfillment and α_3 for equal workload distribution) should be adjusted in accordance with the priorities of the physicians to be scheduled.

6 Conclusion

Our work describes a workload indicator to measure an individual physician's workload based on a given duty roster. We incorporate the workload indicator into a scheduling model which creates duty rosters for physicians. Our model takes into account the individual workload of each physician based on previous rosters and distributes the workload in the new roster in such a way that the workload is equalized over all physicians in the long term. To track this workload over several planning horizons, we require a quadratic decision model. We describe a linearization of this model so it can be transformed into an equivalent linear decision model. Our model with workload tracking is evaluated in comparison to a model without workload tracking. For our study, we apply generated data for 24 months of physician schedules to both models and compare the results using performance indicators. Our results indicate that the workload indicator is effective in ensuring a more equal distribution of workload (i.e., overnight duties) among physicians. The effectiveness of our workload indicator is higher when it does not have to compete with higher ranked objectives, such as, e.g., equal fulfillment of preferences.

Our results provide valuable insights for managers and schedulers. We show that without our workload indicator the distribution of workload is not necessarily equal among physicians. This inequality can accumulate over several planning horizons and lead to dissatisfaction for physicians who are repeatedly burdened with a high workload. For all test instances, using our workload indicator leads to a more equal distribution of workload among physicians over several planning horizons. We therefore recommend managers incorporate our workload indicator into their scheduling process.

References

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

Ceschia, S., Dang, N. T. T., Causmaecker, P. D., Haspeslagh, S., and Schaerf, A. (2015). Second international nurse rostering competition (INRC-II) – problem description and rules.

Cheang, B., Li, H., Lim, A., and Rodrigues, B. (2003). Nurse rostering problems — a bibliographic survey. *European Journal of Operational Research*, 151(3):447–460.

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

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

Fügener, A., Brunner, J. O., and Podtschaske, A. (2015). Duty and workstation rostering considering preferences and fairness: A case study at a department of anesthesiology. *International Journal of Production Research*, 53(24):7465–7487.

Gendreau, M., Ferland, J., Gendron, B., Hail, N., Jaumard, B., Lapierre, S., Pesant, G., and Soriano, P. (2006). Physician scheduling in emergency rooms. In *Practice and Theory of Automated Timetabling VI*, pages 53–66. Springer.

Gross, C. N., Brunner, J. O., and Blobner, M. (2018a). Hospital physicians can't get no long-term satisfaction — an indicator for fairness in preference fulfillment on duty schedules. *Health Care Management Science*.

Gross, C. N., Fügener, A., and Brunner, J. O. (2018b). Online rescheduling of physicians in hospitals. *Flexible Services and Manufacturing Journal*, 30(1):296–328.

Haspeslagh, S., De Causmaecker, P., Schaerf, A., and Stølevik, M. (2014). The first international nurse rostering competition 2010. *Annals of Operations Research*, 218(1):221–236.

Karsu, Ö. and Morton, A. (2015). Inequity averse optimization in operational research. *European Journal of Operational Research*, 245(2):343–359.

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