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

Christopher N. Gross

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

Christopher N. Gross
University Center of Health Sciences at Klinikum Augsburg (UNIKA-T)
Department of Health Care Operations / Health Information Management
School of Business and Economics
University of Augsburg, Universitätsstraße 16, 86159 Augsburg, Germany

Tel.: +49-821-598 6452
Fax: +49-821-598 6442
E-mail: christopher.gross@unikat.uni-augsburg.de

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 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
$\bar{d}_{id}^{\text{duty}}$	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 (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
$x_{jw}^{\text{WE}} \in \{0,1\}$	1 if physician j is assigned to a duty on the weekend of week w, 0 otherwise
$\Delta_{iwd}^{\text{out-duty}} \in \mathbb{N}_0$	Missing physicians to cover demand of duty i on day d of week w
$\Delta_{jiwd}^{\text{req-on}} \in \{0,1\}$	1 if preference of physician j for duty i on day d of week w is not satisfied, 0 otherwise
$\Delta_{jwd}^{\text{req-off}} \in \{0,1\}$	1 if preference of physician j for being off duty on day d of week w is not satisfied, 0 otherwise

Minimize

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

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

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

subject to

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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 $\Delta^{\text{out-duty}}_{iwd}$ in case a duty is not covered. Constraints (3) and (4) set deviation variables

$\Delta^{\text{req-on}}_{jiwd}$ and $\Delta^{\text{req-off}}_{jwd}$ 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.

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

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^{work}_{ja} incurred by assigning a duties to physician j as follows.

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

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

$$ALV = \frac{1}{|\mathcal{J}|} \sum_{j \in \mathcal{J}} \underset{m \in \mathcal{M}}{\mathrm{Var}} (\lambda_{jm})$$

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[1]. Each of our instances can be solved to optimality within 4 seconds. We choose the weights for our computational 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[2] 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 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).

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

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

	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)

	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

indicators. We 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.

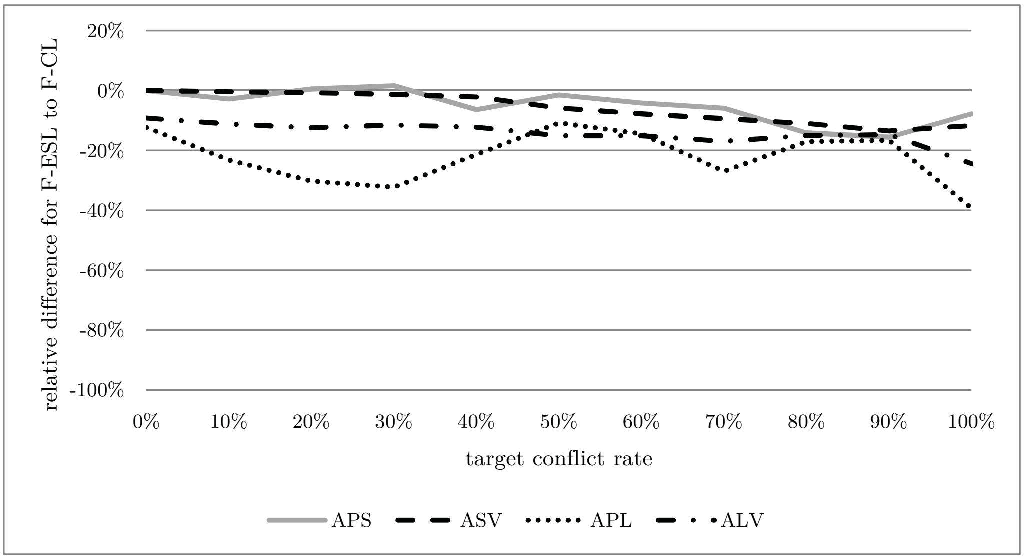

Figure 1 Relative difference in performance indicators between F-CL and F-ESL strategy for data with different target conflict rate

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 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

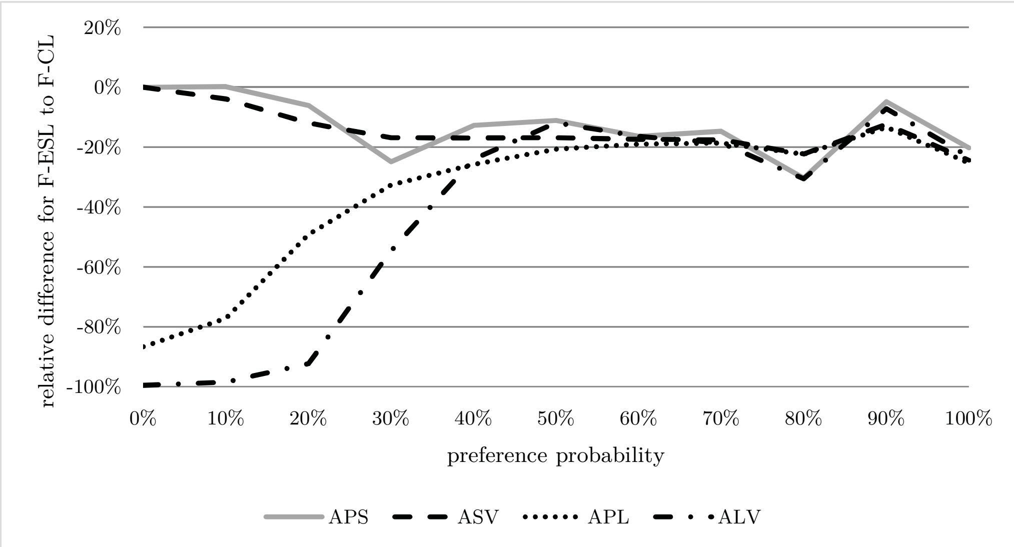

Figure 2 Relative difference in performance indicators between F-ESL and F-CL strategy for data with different preference probability

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.

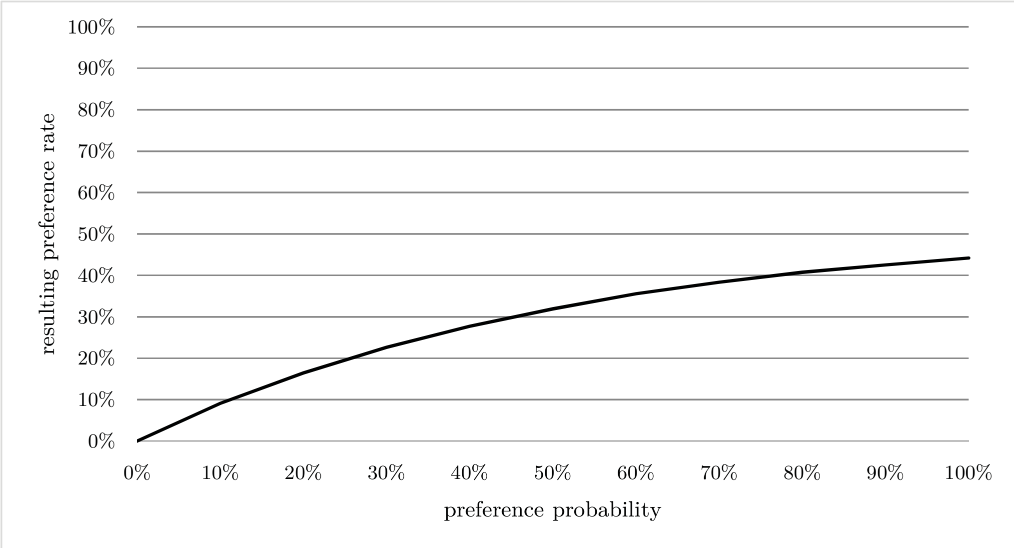

Figure 3 Resulting preference rate when running the preference generation algorithm with different preference probabilities

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 RC (2010) Long-term staff scheduling with regular temporal distribution. Computer Methods and Programs in Biomedicine 100(2):191–199, DOI 10.1016/j.cmpb.2010.03.015

Ceschia S, Dang NTT, Causmaecker PD, Haspeslagh S, Schaerf A (2015) Second international nurse rostering competition (INRC-II) – problem description and rules. URL http://mobiz.vives.be/inrc2/wp-content/uploads/2014/10/INRC2.pdf

Cheang B, Li H, Lim A, Rodrigues B (2003) Nurse rostering problems — a bibliographic survey. European Journal of Operational Research 151(3):447–460, DOI 10.1016/S0377-2217(03)00021-3

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

Ernst A, Jiang H, Krishnamoorthy M, Sier D (2004) Staff scheduling and rostering: A review of applications, methods and models. European Journal of Operational Research 153(1):3–27, DOI 10.1016/S0377-2217(03)00095-X

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

Gendreau M, Ferland J, Gendron B, Hail N, Jaumard B, Lapierre S, Pesant G, Soriano P (2006) Physician scheduling in emergency rooms. In: Practice and Theory of Automated Timetabling VI, Springer, pp 53–66, DOI 10.1007/978-3-540-77345-0_4

Gross CN, Brunner JO, 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 DOI 10.1007/s10729-018-9452-8

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

Haspeslagh S, De Causmaecker P, Schaerf A, Stølevik M (2014) The first international nurse rostering competition 2010. Annals of Operations Research 218(1):221–236, DOI 10.1007/s10479-012-1062-0

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

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