Doctoral theses at NTNU, 2021:344

Kjartan Kastet Klyve

Personnel Planning in Health Care

Optimization of Nurse and Physician Rosters with real-life applications

Doctoral thesis

NTNU
Norwegian University of Science and Technology
Thesis for the Degree of
Philosophiae Doctor
Faculty of Economics and Management
Department of Industrial Economics and
Technology Management (IØT)

NTNU
Norwegian University of
Science and Technology

Kjartan Kastet Klyve

Personnel Planning in Health Care

Optimization of Nurse and Physician Rosters with real-life applications

Thesis for the Degree of Philosophiae Doctor

Trondheim, November 2021

Norwegian University of Science and Technology
Faculty of Economics and Management
Department of Industrial Economics and Technology Management (IØT)

NTNU
Norwegian University of Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the Degree of Philosophiae Doctor

Faculty of Economics and Management
Department of Industrial Economics and
Technology Management (IØT)

ISBN 978-82-326-6030-8 (printed ver.)
ISBN 978-82-326-5745-2 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

Doctoral theses at NTNU, 2021:344

Printed by NTNU Grafisk senter

Summary

This thesis considers personnel rostering in health care, a subfield of the more general research areas of Resource Management, Operations Research, and Optimization. Personnel rostering entails planning the utilization of employees; the most important resource in health care organizations.

Personnel rostering is challenging in multiple ways. Firstly, creating rosters is an inherently complex task from a combinatorial perspective. Secondly, as rosters reflect operations at a ward and affect the lives of staff, many different versions of real-life rostering problems exist. This means modelling, analyzing and implementation of decision-support are also very complicated tasks. This thesis attempts to answer relevant questions related to all of these challenges.

In the first article a comprehensive framework for robust personnel planning is developed at the Department of Neonatal Intensive Care at St. Olav's Hospital. We firstly perform nurse rostering in a detailed way. When a roster is established, uncertainty is realized through simulation using extensive historical data obtained at the department. We perform daily simulations of the supply of staff and the demand for health care, and perform rerostering to take into account any disruptions from the uncertainty realization. This enables analyses of robust rostering strategies such as strategic overstaffing, implementing shadow shifts to cover absent nurses and trading extra weekend work for time off.

The second article deals with physician rostering at the Clinic of Surgery at St. Olav's Hospital. Here, surgeons must work emergency shifts in a cyclic structure, while also ensuring an even and robust staffing level at the sections. The mathematical structure of this problem is novel and difficult to solve, and we develop a matheuristic to produce robust rosters of high quality.

The third article presents a generic ward with 24 hour staff demand, where we minimize nurse fatigue. We incorporate a model of human sleep in the Nurse Rostering Problem, and define biological profiles to analyze how rosters should be individualized to minimize fatigue. The approach is theoretical, but insights and a large potential for future research and possible implementation exists.

The fourth article presented is a formalization of the experiences from performing pilot projects of implementing an optimization-based rostering tool. Creating the tool entailed development of a detailed model customized to Maternity Ward West at St. Olav's Hospital in Trondheim. We discuss visions for how the implementation of decision support systems for rostering will affect future work life, and present the mathematical model at the core of our decision support tool.

Acknowledgements

During the last four years I have done research and duty work at the Department of Industrial Economics and Technology Management at the Norwegian University of Science and Technology. This thesis is the result of the research. In my work I have been privileged to have the support and guidance from several skilled persons in my research area and related fields.

Before beginning as a PhD student, I received my master's degree from the same department I have worked at these four years. One of the reasons I decided to do my PhD was the brilliant supervision I received during the work on my master's thesis. This continued in my PhD work, and for that I want to thank my supervisor, Professor Henrik Andersson. Your guidance and insight has been essential to my work. I am very grateful for how you have always been available for discussions and for the thorough feedback I have received on my work. I would also like to commend your support in pursuing my interests both in my academic work and duty work. I also want to give my co-supervisor Associate Professor Anders Gullhav praise and thanks, for contributing to the excellent supervision I have received.

I would also like to thank other colleagues at the department for support and cooperation, including Dr. Troels Range, Professor Marielle Christiansen, and Professor Kjetil Fagerholt. I thank Jone Hansen, Jørgen Skålnes, Professor Magnus Stålhane and my other colleagues for interesting discussions and making everyday life at our work place enjoyable.

Other key contributors to my work have been colleagues at St. Olav's Hospital. They have provided me with practical knowledge and insights to the complicated planning problems that exist in health care. It was a pleasure to get to know Vigdis Myhren at Maternity Ward West and Dr. Birger Endreseth at the Clinic of Surgery, as well as skilled managers and practitioners through the Center for Health Care Improvement.

I want to express my sincere gratitude to Professor Mark Wallace and Dr. Ilankaikone Senthooran at the University of Monash for welcoming me to a productive research stay. I will always remember my time with you in Melbourne fondly, and hope we can stay in touch.

I also want to thank my parents and my brother for always supporting me. Finally I want to thank my beloved wife Hanna-Emilie and our children for bringing joy to my life outside work. Hanna-Emilie, your support and encouragement made this thesis possible.

Contents

Introduction

Health research is one of four strategic research areas for the Norwegian University of Science and Technology. The aim is to create innovative solutions to complex health challenges, which includes research in IT-systems and organizing health care as well as the more traditional health research dealing with diagnostics and therapy etc. This is one of the reasons why the Department of Industrial Economy and Technology Management has increased its focus on applications of Operations Research and Optimization with applications within health care. This thesis is a part of this initiative.

Health expenditure has largely outpaced economic growth in the past, and despite a slowdown in recent years, is expected to do so in the future. New estimates point to health spending reaching 10.2% of GDP by 2030 across OECD countries, up from 8.8% in 2018 (OECD, 2019). Even disregarding costs, skilled personnel is a scarce resource. Projections developed by WHO and the World Bank point to the creation of approximately 40 million new health and social care jobs globally to 2030 and to the need for 18 million additional health workers. (WHO, 2016) Furthermore, skill-mix imbalances, maldistribution of personnel resources, barriers to inter-professional collaboration, inefficient use of resources, poor working conditions, a skewed gender distribution, limited availability of health workforce data, and an aging workforce further complicates the picture. (WHO, 2016).

This emphasizes the need for effective planning in health care. While planning of single hospital departments is at the core of this thesis, we provide a brief overview of how health care is organized in Norway to exemplify the link between the top and practitioner level, and to put into perspective the real-life problems dealt with in the thesis. We focus on the public health care sector, as the clear majority of Norwegian health expenditure occurs there (OECD, 2019).

Norwegian health care is organized through primary and specialty care. Each municipality is responsible for their inhabitants' access to primary health care services of good quality at appropriate times, and that each inhabitant is assigned one regular general practitioner (GP) (Lovdata.no, 2021). In practise, the GP is the first point of contact between patients with non-urgent sickness or medical questions and the health care system. Every municipality is also required by law to facilitate numerous other health care services and facilitate cooperation with other service providers (Lovdata.no, 2021).

The organization of specialty care is more complex, and we focus on how hospitals fit the bigger picture. At the state level, health care is managed by the Ministry of Health and Care Services(HOD). The HOD is responsible for providing good and equal health and care services for the population of Norway(regjeringen.no, 2021a).

The HOD is divided into 9 departments (regjeringen.no, 2021c). Each department's function is among other things to coordinate the HOD's work within their area of expertise. A significant department is the Department of Hospital Ownership. The principal task of the Department of Hospital Ownership is "governance of the four regional health authorities and the Norwegian Health Network" (regjeringen.no, 2021b). The regional health authorities are examples of subordinate enterprises owned by the HOD. These enterprises are run as if they were private companies by a board and a Chief Executive Officer. The subordinate enterprises are similar to joint-stock companies, although the state as the only shareholder is responsible for the subordinate enterprises' economy, and they cannot be bankrupted (Braut, 2019).

St. Olav's Hospital in Trondheim, our collaboratory hospital in Papers I, II, and IV, is owned by the Central Norway Regional Health Authority (stolav.no, 2021). Large health care organizations like hospitals are very complex when considering the scope of the services they provide. Hospitals often employ thousands of health care workers and service several wards around the clock. Logistically it is difficult to manage all these services and employees while ensuring cost efficiency and high quality health care.

Hans et al. (2011) point to several reasons for why health care organizations differ from organizations in other industries. Large health care providers like hospitals are typically made up of autonomously managed departments, and their managers tend not to properly consider their department as part of a greater interconnected planning environment. This makes planning and control more fragmented than what is typically the case in e.g. manufacturing, where the entire supply chain must be considered for maximizing profits. Even though health care managers are generally committed to provide the best possible services, they typically come from a background within health care, and lack sufficient training and knowledge to make optimal use of scarce resources in complex planning environments. An example of this is that ward managers are typically attained by promoting health care workers with long tenure, as opposed to hiring someone with a background in management and planning.

This thesis consists of four research papers. Paper I deals with development of a framework for robust rostering. The framework consists of a rostering model, a collection of simulation models, and a rerostering model. We use the framework to evaluate different rostering strategies in Chapter I. Paper II deals with rostering of physicians that must work emergency shifts in a weekly cyclic structure. This constrains rosters, and it is challenging to schedule physicians without large fluctuations in staffing levels at the sections. We present a two-step matheuristic that evens out staffing levels and increases robustness when absence occurs at emergency night shifts. Paper II is presented in Chapter II. Paper III deals with rostering nurses at a generic ward with round-the-clock activity. We incorporate a verified model of human sleep into our Nurse Rostering Problem to minimize the worst cases of fatigue. Introduction of biological profiles, sets of realistic parameter sets in the sleep model representing different humans, enables analyses and discussions of how rosters affect different nurses and raises questions as to what fairness in rostering means. This paper is presented in Chapter III. Finally we present Paper IV describing the implementation of a Nurse Rostering Model at a real-life maternity ward in two pilot projects. The paper discusses the potential and challenges from the perspectives of different stakeholders, and attempts to predict how rostering will be performed in thirty years time, assuming decision support tools are widely applied. This is presented in Chapter IV. Additionally, we present the Mixed Integer Program used in the pilot projects in Appendix A.

1.1 Background

In this section we present the research scope and background of the thesis. We review related literature to provide context to our work, focusing on aspects that are most relevant for the papers presented in this thesis. As a result, problem types and real-world issues are reviewed in more detail than e.g. solution methods.

1.1.1 Relevant terminology

The two terms *scheduling* and *rostering* are oftentimes used similarly in the research literature. However, while some authors use them interchangeably, others include slight distinctions. Ernst et al. (2004) define both personnel scheduling and rostering as "the process of constructing work timetables for its staff so that an organization can satisfy the demand for its goods or services." However, according to Burke et al. (2004), scientific literature tends to refer to short-term timetabling of staff when discussing rostering. Burke et al. (2004) use scheduling as a more general term than rostering, and while their definition of nurse rostering is also quite general ("the allocation of nurses to periods of work over several weeks"), they do not consider *staffing* as a part of nurse rostering. This means scaling of the workforce is not part of the rostering problem.

In Wren (1996), a very general definition of scheduling is given: "Scheduling may be seen as the arrangement of objects into a pattern in time or space in such a way that some goals are achieved, or nearly achieved, and that constraints on the way the objects may be arranged are satisfied, or nearly satisfied." However, a much more concrete definition of rostering is given; "the placing, subject to constraints, of resources into slots in a pattern. One may seek to minimize some objective, or simply to obtain a feasible allocation. Often the resources will rotate through a roster". Note that the Wren (1996) definition of rostering entails placing resources into slots, implying the length of shifts is predefined. This is not the case for the rostering definition in Erhard et al. (2018). While the Erhard et al. (2018) definition also uses a more general term for scheduling than rostering, they include shift flexibility in several papera that they consider work on physician rostering, see e.g. Kazemian et al. (2014).

In our work we typically use similar distinctions between scheduling and rostering as they do in Burke et al. (2004), Wren (1996), and Erhard et al. (2018). This means we use scheduling as a general term, while rostering excludes staffing. None of our papers deal with flexible shift definitions, and thus we normally assume shifts have fixed lengths in rostering problems if we do not explicitly state otherwise. However, when we cite related literature, definitions may very well diverge from ours.

As part of discussing the distinctions in the scheduling term, Erhard et al. (2018) classify different types of physician scheduling according to planning horizons. "Staffing problems focus on the strategic decision of determining the required size and composition of a workforce. These planning problems typically involve a long-term (e.g., annual) planning horizon. Rostering problems concentrate on the tactical or operational offline task of creating concrete or generic shift rosters. These problems may be classified as mid-term, as the planning horizon typically spans from weeks to a few months. Re-planning problems discuss short-term adjustments of the working schedule". The notion of placing rostering in the tactical or offline operational planning level is interesting. To define the hierarchical decision levels, including the distinction between offline and online operational planning, we use definitions provided in Hans et al. (2011).

- The *strategic* level has a long planning horizon and revolves around the structure of an organization. It involves defining the organization's missions, and making decisions to translate this mission into the design, dimensioning, and development of the health care delivery process. Examples of such decision areas are developing and implementing new medical protocols and mergers of nursing homes.

- The *tactical* level addresses the organization of the operations/execution of the health care delivery process. In this way, it is similar to operational planning. However, decisions are made with a longer planning horizon. Examples of decision areas are deciding staffing levels at wards.

- The *operational* level involves the short-term decision making related to the execution of the health care delivery process. The flexibility on this planning level is low as the higher levels has already set the scope for operational decision making. Furthermore, the operational level is divided into two categories:

  - *Offline operational* planning. All offline operational planning can be planned for ahead of incidents occurring. Examples are treatment selection and nurse rostering.

- *Online operational* planning. Online operational planning is done after sudden changes in circumstances and encompass e.g. finding substitute workers in case of sudden understaffing.

The difference between offline and online operational decision levels is interesting in our work, as we do work on robustness-enhancing rosters, especially in Paper I and II. The robustness term in our work is best understood as protection against unforeseen events (sometimes referred to as disruptions), and must not be confused with robust optimization, as known from e.g. Soyster (1973). This is discussed briefly in each of the relevant papers. We also differentiate between stability and flexibility when discussing robustness, where stability is the degree to which rosters can absorb disruptions, while a roster's flexibility is its capability to react efficiently to disruptions. The flexibility is thus dependent on the available options for reoptimizing rosters when disruptions have occurred.

Another relevant term in rostering is *cyclicity*. Rosters can be cyclic or acyclic in their structures. In Ernst et al. (2004), the following definition is provided: "In a cyclic roster all employees of the same class perform exactly the same line of work, but with different starting times for the first shift or duty. In acyclic rosters, the lines of work for individual employees are completely independent." Cyclic rosters can also be referred to as cyclical or fixed rosters in rostering literature. Cyclicity obviously implies restricting the feasible region of possible rosters, which can make it harder to comply with individuals' requests on specific days (Warner, 1976). On the other hand, Burke et al. (2004) point out that cyclic rostering tends to provide employees with predictable and even rosters without undesirable and unhealthy transitions between shifts, which can be a significant challenge in some cases. Burke et al. (2004) categorize nurse rostering problems as either cyclic, semi-cyclic, or non-cyclic. This distinction is relevant for Paper II.

Furthermore, applications of nurse rostering problems can be organized in three different administrative modes. The top down approach of centralized scheduling, the mixed approach of unit scheduling and the bottom up approach of self-scheduling. *Self-scheduling* is of particular interest, as it is the way nurse rostering is typically performed at St. Olav's Hospital, relevant for Papers I and IV. Self-scheduling entails employees themselves creating rosters manually through a process of cooperation and negotiation (Burke et al., 2004). At St. Olav's Hospital, this typically entails each nurse creating their preferred individual schedule given some guidelines. These preferred individual schedules are then aggregated to an initial roster, which is the starting point for a bartering process. Self-scheduling can be popular among nurses, as it gives them great influence on the planning process. However, self-scheduling is potentially very time-consuming (Burke et al., 2004), and can have other drawbacks including a tendency to cause over- or understaffing, that the schedule is made for the convenience of staff rather than patients, and that there are no formal procedures for conflict solving (Silvestro and Silvestro, 2000).

1.1.2 The Nurse Rostering Problem

The Nurse Rostering Problem (NRP) entails assigning nurses to shifts over a given planning period. Typically, the ward has around-the-clock activity, making several types of restrictions relevant to ensure that rosters have high quality. Burke et al. (2004) emphasize constraints related to coverage constraints, skill categories, time related constraints, and more. These constraints can be hard or soft depending on the problem at hand. We thus briefly review related literature modelling nurse rostering problems in light of these constraint types. We also note that several solution methods are used to solve different NRPs, and we mention some interesting techniques in relation to relevant works last in this section.

Some versions of coverage constraints exist in the majority of nurse rostering problems. They are normally demand coverage constraints, establishing the key balance between supply of personnel and the patients' demand, or an estimation thereof, for health care services. These constraints ensure some minimum staffing level in the NRP, and are most often modelled as hard constraints (Burke et al., 2004) (den Bergh et al., 2013). Examples are provided in Azaiez and Sharif (2005) and Liu et al. (2018). In reality, the demand for health care can fluctuate, and predefined hard minimum staffing levels thus reflect some established staffing norm in most nurse rostering problems. Not all authors consider the minimum staffing levels as absolute, and thus model this using soft constraints, see e.g. Bard and Purnomo (2005a) and Rahimian et al. (2017). Versions of coverage constraints are used in all papers in this thesis, and the most interesting ones are perhaps found in Paper I, where reestablishing the balance between supply and demand is key in the novel rostering framework.

Nurses can have different competencies, which can affect their ability to cover different types of demand. The competencies of the available staff is often referred to as the skill-mix. This is commonly modelled through hierarchical demand coverage where nurses can rank down to contribute. This is implemented in Lim and Mobasher (2011) and Aickelin and Dowsland (2004). In Gomes et al. (2017), the authors model ranking down as undesirable, penalizing this in the objective function. Others, e.g. Lim and Mobasher (2011), are indifferent to whether nurses rank down or not. Skill-mix is modelled in the problems in Papers I, II, and IV.

Time related constraints refer to all the restrictions on personal schedules. This includes constraints related to work regulations, which exist in most literature on NRPs. Such regulations typically ensure nurses sufficient rest times through avoiding double shifts and too many consecutive work days. These constraint types are included in e.g. Fügener et al. (2018), Knust and Xie (2019). Formulating work time constraints was very relevant for the problems in Papers I, II, and IV where problems are regulated by both governmental regulations, the

Norwegian Arbeidsmiljøloven §10 (Arbeids- og sosialdepartementet, 2017), by local agreements such as the rostering agreements, and by preferred practices and norms at the workplace.

In the categorization by Burke et al. (2004), nurse preferences and fairness were considered parts of the time related constraints. However, in work at St. Olav's Hospital, nurse preferences and different fairness measures were key characteristics of the nurse rostering problems, and so we provide additional insights on related literature. In real-life, the different needs and requests vary greatly from person to person, and this aspect of nurse rostering problems are likely formulated as a result of different work cultures in different wards, as well as the role decision support systems are meant to have in the planning process.

Numerous examples of preference enhancing constraints exist. Some authors model undesirable shift patterns, i.e. sequences of shifts, see e.g. Ruzzakiah et al. (2011) and Rönnberg and Larsson (2010). Multiple authors model weekends differently from weekdays due to nurse preferences. One example is ensuring that nurses either work a full weekend or have a full weekend off, see e.g. Clarissa and Suyanto (2019). Individual requests for working or having off-days on unique days is also modelled in some research, e.g. Smet et al. (2014), Ásgeirsson and Sigurðardóttir (2016), and Mischek and Musliu (2019).

While fairness is an ambiguous term, authors frequently propose constraints to enhance it in nurse rostering literature. Typically, it entails an even distribution of something considered desirable or undesirable. Ásgeirsson and Sigurðardóttir (2016) even out the number of granted requests by using a piecewise linear function of increasing penalties as the number of unfulfilled requests increases for a particular nurse. Furthermore, in Rönnberg and Larsson (2010) a preference score is calculated for each nurse, based on the number of respected preferences. The lowest amongst all the scores is maximized in the objective function, to enhance preferences and fairness in a balanced way. In Akbari et al. (2012) it is stated that "Our model objective tries to maximize preference of part time workers by a minimization objective while considering seniority, availability, and priority of employees." Akbari et al. (2012) thus considers fairness in a slightly different matter, where rank and seniority is included in the considerations of whose requests to prioritise.

In Paper IV, the ward manager stated that fairness implies equality in respect and influence, not necessarily equality in the individual schedules, as people have different health issues and needs. While this is not typical for constraints in NRPs, it is very much aligned with results presented in Paper III, where different individual nurses' fatigue is modelled as part of a NRP.

Multiple approaches are used to solve NRPs. Bergh et al. (2013) divide them coarsely into mathematical programming approaches and heuristic methods. Cross-checking the papers in Bergh et al. (2013) dealing with nurse scheduling and different solution methods, it is clear that mathematical programming approaches and heuristic methods are the most favored solution methods.

Popular mathematical approaches include Integer Programming (IP) and Mixed Integer Programming (MIP), see e.g. Valouxis et al. (2012), Mischek and Musliu (2019), and Ásgeirsson and Sigurðardóttir (2016). Some decomposition methods, most frequently branch-and-price, are also used by several authors (Bergh et al., 2013). Examples of papers using branch-and-price to solve the NRP are Purnomo and Bard (2007) and Beliën and Demeulemeester (2008). One possible reason for decomposition methods being used less frequently than the more straightforward IP and MIP approaches, is that variations in problem formulations make the time consuming effort of decomposition less appealing.

A typical distinction within heuristic methods are constructive heuristics, see e.g. Ásgeirsson (2014), and improvement heuristics. Popular types of improvement heuristics include the tabu search, e.g. Lü and Hao (2010), genetic algorithms, e.g. Moz and Pato (2007), simulated annealing, e.g. Liu et al. (2018) and Knust and Xie (2019), and several others. While the mentioned improvement heuristic types can in many cases be considered metaheuristics, metaheuristics are not explicitly listed in the categorization by Bergh et al. (2013). However, they are discussed by Burke et al. (2004), stating "We believe that metaheuristics are generally better suited than most other approaches for generating an acceptable solution in cases where the constraint load is extremely high and indeed in cases where even feasible solutions are very difficult (if not impossible) to find." Top results in terms of solution times in the two international nurse rostering competitions also tend to be well represented by metaheuristic methods (Haspeslagh et al., 2014, Ceschia et al., 2019).

1.1.3 The Physician Rostering Problem

The Physician Rostering Problem (PRP) has many characteristics differentiating it from general personnel rostering, as seen in Erhard et al. (2018). However, several of these characteristics are typical of a health care environment and are thus also present in nurse rostering, such as uncertain and fluctuating demand and a high degree of heterogeneity in patient needs. This can often lead to a mismatch in supply and demand, and is briefly discussed in in relation to rerostering in Section 1.1.4.

A characteristic that is more predominant in physician rostering literature than in nurse rostering is the planning of residency, i.e. a specialization period. In fact, Erhard et al. (2018) list 28 of a total of 68 reviewed papers on Physician Rostering as dealing with residents. Residency often entails working at several different departments to acquire relevant experience. Topaloglu (2006) presents a month-long rostering problem for emergency medicine residents. The problem includes residents of different seniorities and supervisory roles. As specialization periods are longer than typical rosters, rostering problems that deal with residency tend to have some link between rostering at the operational level and the longer term planning. Smalley and Keskinocak (2016) create two models, one for creating feasible assignment of residents to services over a one-year period and another that plans specific shifts to residents given the services they have been assigned to. Bard et al. (2016) present a problem of creating so-called rotations that constitute a month of a specializing physician's training in a particular department. Rotations are made from weekly scheduling templates stating each physician's duties. The goal of the problem is to distribute personnel evenly over the week and minimizing the number of changes to a scheduling template.

Another key characteristic of Physician Rostering is a high degree of power amongst individual physicians due to their expertise and them being a valuable asset to the health care organization. This typically leads to a high degree of autonomy. As a result, modeling of preferences and fairness issues is of great importance in physician scheduling (Erhard et al., 2018).

Preferences and fairness was discussed in light of projects with great nurse influence at St. Olav's Hospital previously, but we present two examples of authors modelling this in Physician Rostering as well. In Topaloglu (2006), rosters are planned to minimize the violation of eight different soft constraints, ensuring reasonable rest times and a different fairness measures. Fairness is considered by modelling fair distribution of supervising resident positions, Monday night shifts, and Tuesday night shifts. Stolletz and Brunner (2012) create fortnightly physician rosters taking preferences and fairness into account. They perform interviews with physicians in their collaborating hospital, and find that the most important fairness aspects that should be considered are a fair distribution of on-call services and fair assignment of working hours.

Solution methods in physician rostering resemble the overall picture of nurse rostering. The majority of problems studied in Erhard et al. (2018) are formulated using mathematical approaches, e.g. IPs and MIPs, but both exact methods and heuristic approaches are used.

Rostering problems are a subset of the papers reviewed in Erhard et al. (2018), so we briefly mention some physician rostering papers and their solution methods. Bruni and Detti (2014) present a MIP model of a realistic PRP that is tested in a case-study. They call it a flexible MIP, and present several extensions for modifications of the problem PRP. Puente et al. (2009) identify a PRP at a Spanish hospital emergency department with multiple soft constraints, including sequential shift constraints, and solve it using a genetic algorithm. The algorithm firstly creates an initial roster, focusing on key days such as holidays and weekends. The initial schedule is then improved iteratively using a genetic algorithm comprised of a crossover operator that incorporates the sequencial restrictions. A repair function is also introduced to avoid infeasibility in rosters. Fügener and Brunner (2019) use a MIP formulation and a heuristic solution approach based on a column generation decomposition to reduce unplanned overtime of physicians in a rostering problem with stochastic demand.

1.1.4 Rerostering

The planning of health care services is subject to uncertainty in various forms. Uncertainty incorporation in personnel scheduling problems is discussed in den Bergh et al. (2013), while Bai et al. (2018) review Operations Research in intensive care unit management, with a comprehensive discussion of stochastic methods. In Paper I, uncertainty in the supply of nurses and the demand for health care services was characteristic for the planning problem. When uncertainty is realized, disruptions can reduce the quality of rosters, making it necessary to make changes on the online operational decision level. This is the origin of the Nurse Rerostering Problem (NRRP). The NRRP is a relatively novel scheduling problem. In Burke et al. (2004), neither rerostering nor rescheduling is discussed specifically, although one of the first works clearly dealing with nurse rerostering, Moz and Pato (2004), is cited. In Clark et al. (2015), the authors review literature on NRRP and identify eight articles considered nurse rescheduling (nurse rerostering), and we briefly present some of them.

Moz and Pato (2003) present the first version of the NRRP identified in this thesis. The uncertainty realized prior to their rerostering problem is nurse absence, and they thus simply add an additional condition to the original rostering problem, that absent nurses cannot work. The NRRP is solved from the day of the first absence until the last day of the planning period (the rerostering period). This entails assuming all absence throughout the rerostering period is known on the first day. The objective is to minimize the difference between the original roster and the final roster after rerostering. If the problem proves infeasible, this is considered an administrative issue, but this does not happen in their test instances.

While the Moz and Pato (2003) rerostering problem may seem simplistic compared to practical cases, it is an interesting starting point for further research on the NRRP. New solution methods are presented in (Moz and Pato, 2004) and Moz and Pato (2007), and other authors present novel versions of the NRRP to increase realism. Bard and Purnomo (2006) deal with preference scheduling (self-scheduling), and minimize undercoverage and preference violations in rerostering by modelling the possibility of hiring temporary staff. Pato and Moz (2007) penalize allocation of assignments that individual nurses dislike. Kitada and Morizawa (2013) present a NRRP where nurse absence can last several consecutive days, further increasing the realism of the NRRP.

Recent and notable works include Maenhout and Vanhoucke (2011, 2013) increasing the focus on fairness of workloads in rerostering, and Schoenfelder et al. (2020) who include decisions of transferring and turning away patients with a rerostering problem to combine ideas of rerostering with research on patient flow. An interesting approach to rostering and rerostering is presented in Ingels and Maenhout (2015, 2017, 2018, 2019). They firstly construct a roster, then simulate the realization of uncertainty and perform rerostering every day throughout the roster's planning period. Such iterative rerostering is likely a more realistic representation of real-life practise, and Bard and Purnomo (2005b) present a model where rerostering occurs three times daily. For physicians, research on rerostering is even scarcer than for nurses. We have identified one such work, that of Gross et al. (2018), which states "we propose the first approach for rescheduling physicians in hospitals". Gross et al. (2018) consider nurse rerostering less complicated than physician rerostering, and mention modelling of preferences, fairness and training aspects as complicating factors in scheudling of physicians.

Due to the limited volume of rerostering literature, solution methods are not hugely varied. According to Clark et al. (2015), rerostering models are typically solved using mathematical optimization methods or some form of intelligent computerized search methods. The problem studied by Moz and Pato (2003) is formulated as an integer multicommodity flow model. Bard and Purnomo (2006) propose a column generation approach with a swapping heuristic that produces candidate rosters. Clark and Walker (2011) and Maenhout and Vanhoucke (2011) use mathematical programming and heuristic approaches respectively to solve new NRRPs. Kitada and Morizawa (2013) present a NRRP where nurse absence can last several consecutive days, further increasing the realism of the NRRP and proposes a heuristic method based on a recursive search technique.

1.1.5 Real-life implementation

Literature on personnel rostering, including work focused on nurses and physicians, has been focused on theoretical problems and the efficiency of solution methods. This is clear from e.g. den Bergh et al. (2013), where only a minority of works are considered applied in practice, which typically entails some kind of case study. This is notable, as many papers cite the potential for effective use of resources as a motivation for their work.

Burke et al. (2004) state the following: "There is a definite gap between much of the current state of the art in nurse scheduling research and the demanding and challenging requirements of today's hospital environments." This is further nuanced by Kellogg and Walczak (2007), who study the research-application gap specifically, and list ideas for how it can be bridged. This includes less focus on basic research for solving theoretical problems and more cooperative work with institutions practicing health care. They also suggest more focus on inclusion of self-choice and incorporating the advantages of self-scheduling. Furthermore, Kellogg and Walczak (2007) mention that scheduling systems should be aligned with third-party vendors, to make integration in larger systems possible. Lastly, they encourage researchers to share success stories on real-life implementation. However, in the years to follow, there is no sign of a revolution in practical use of computerized scheduling and rostering.

Drake (2014) forcefully states that "In practice, rostering nursing staff is often unrecognized, unrewarded and undervalued; yet, despite four decades of research, operations management has little to offer in terms of faster, safer, fairer or more effective rosters." Drake (2014) finds that many scheduling rules are created at each ward in a politicized environment with complex group dynamics. While this to some extent sounds similar to the situation observed at Maternity Ward West in Paper IV, wards studied in Drake (2014) seem to have a less structured approach to rostering. Drake (2014) found that few of the hospitals analyzed had formally documented policies for roster preparation, and experienced frequent violations of the informally created rules. This contrasts the structured process of creating rosters and the formally enforced scheduling rules in Paper IV, but is very similar to how we experienced the informally decided constraints in the rostering problem observed at Maternity Ward West.

In a recent study, Petrovic (2019) repeats the message of above mentioned authors, stating that "The use of computerised systems for personnel scheduling has been increasing, but they are still underutilised and often require considerable inputs from schedulers." Petrovic (2019) goes on to call for the research community to engage more with managers and schedulers in practice, in order to investigate complex real-world issues. However, while no evidence of a revolution in use of Operations Research to roster personnel in health care exist, there has been some progress.

There are multiple examples of inspiring work with successful case studies. Burke et al. (2006) present a tabu-search procedure for solving real-life problems at Belgian hospitals. Notably, the authors mention implementation of their system in more than 40 hospitals. However, Burke et al. (2006) do not elaborate on the implementation, but rather focus on the algorithm and its use. It can thus hardly be used for understanding implementation of real-life rostering tools. Bester et al. (2007) and Rönnberg and Larsson (2010) present inspiring case studies that are rich in detail on how the practical applications of their two different decision support systems for Nurse Rostering work. They both create versions of tabu-search algorithms, and the Bester et al. (2007) case study is from a psychiatric hospital in South Africa, while Rönnberg and Larsson (2010) deal with a "typical Swedish nursing ward" that practised self-scheduling at the time of their pilot study. Notably they take into account the existing planning process when developing their systems, which makes models more accessible to the personnel they plan for.

1.2 Purpose and outline

We discuss the purpose of this thesis in Section 1.2.1. In Sections 1.2.2 - 1.2.5, we outline each paper in the thesis and reflect briefly on each paper's purpose and contribution. Additionally, Appendix A is outlined in Section 1.2.5.

1.2.1 Purpose of thesis

The purpose of this thesis is to explore how Operations Research can be used to improve personnel planning in health care. In doing so, creating rosters for personnel is vital, and we ask the following key questions:

- What is a high-quality roster?
- How do we produce such a roster?

While our contributions to answering these questions are discussed in Section 1.3, the scope of the thesis is shortly discussed here. Most of the work presented in this thesis deals with variants of typical rostering problems. However, our work also implicitly questions the scope of some of the literature on personnel rostering in health care. Much of our work deals with observing real-life problems at St. Olav's Hospital and using mathematical modelling to formalize these problems. This entails grappling with implicit local rules and challenging how different considerations are included in the rostering process, such as preferences, fairness, and fatigue.

Modelling such rostering aspects using Operations Research techniques entails formalizing these tacit considerations to the best of our ability. In doing this, we implicitly assume that a collection of qualitative aspects related to the planning of people's lives can be quantified. While this may not be very controversial for researchers familiar with the techniques, practitioners may very well be skeptical.

After problems are formalized, different solution methods can be used to solve them. In some cases, MIP using commercial solvers proves sufficient to find rosters of useful quality. For other problems, development of novel algorithms are necessary to obtain high-quality solutions. While novel solution methods are interesting contributions to the research community in their own right, the main focus in this work is to find useful solutions on realistic problems rather than reducing run times on more general problems.

In Papers I and II we identify interesting real-life rostering problems observed at St. Olav's Hospital in Trondheim, Norway. We formalize the problems and develop solution methods that can solve them. Paper I deals with rostering, uncertainty modelling and rerostering of nurses at the Department of Neonatal Intensive Care. Paper II tackles rostering of surgeons at the Clinic of Surgery, where an interesting semi-cyclic structure greatly complicates planning. In Paper III we explore how fatigue modelling can be incorporated in Nurse Rostering to minimize the worst fatigue observed in the roster, aiming to enhance nurse health and reduce the risk of human errors. In Paper IV we describe a pilot project of using a nurse rostering model for planning at the Maternity Ward West at St. Olav's Hospital.

1.2.2 A modelling framework for evaluating proactive and reactive nurse rostering strategies - A case study from a Neonatal Intensive Care Unit

In this paper, we create a framework that combines rostering, uncertainty modelling and rerostering to enable evaluation of the robustness of rosters. Using the framework, we are able to evaluate the effects of multiple robustness-enhancing proactive strategies in the rostering problem. We further evaluate the effects of varying the rerostering period and analyze the effects of a strict policy against breaking any scheduling rules in the online operational planning level.

All models in the paper are developed to reflect the real-life problems observed at the Department of Neonatal Intensive Care (DNIC) at St. Olav's Hospital. We create a roster for 117 nurses of four different skill levels and solve it for a planning period of 105 days. The problem is formulated as a MIP model and solved using commercial software. The supply and demand is modelled irrespective of each other. Modeling supply entails modelling both the absence of nurses and the nurses' willingness to work extra shifts when disruptions occur. We develop a discrete time Markov chain for each nurse with three states that nurses transition between: non-absent, short-term absent or long-term absent. Nurses' willingness to work extra in rerostering is simulated using Monte Carlo-simulation. Demand is estimated by simulating the patient severity states of the different beds at the DNIC. The rostering model is run once, before daily simulations of uncertainty occurs as disruptions, and rerostering happens as a response.

The main contributions of this paper is the development of a framework that enables evaluation of the robustness of rosters. Our work stands out from related literature in its realism. The rostering model is developed in close cooperation with the scheduling manager at the DNIC, and all uncertainty modelling is based on real-life data. The rerostering model is based on a collection of rerostering actions, which reflect the daily online operations at the DNIC. The paper is co-authored with former master students Isabel Nordli Løyning and Line Maria Haugen Melby, Associate professor Anders Nordby Gullhav, and Professor Henrik Andersson.

1.2.3 Paper 2: Semi-cyclic rostering of ranked surgeons - a real-life case with stability and flexibility measures

In this paper, we study the problem of creating rosters for surgeons in specialization at the Clinic of Surgery at St. Olav's Hospital. The surgeons must be planned to work emergency shifts during day and night in a cyclic structure according to their rank, while also being assigned non-cyclic day shifts at sections essential for their training. The specializing surgeons are essential resources for covering both types of shifts, and in practice the rigid assignment of emergency shifts entail large fluctuations in available staff at sections from day to day. We formalize a rostering model that greatly improves rosters compared with those produced using manual methods, and also demonstrate that the use of shadow shifts reduce the probability that planners are forced to reroster because of absence during emergency night shifts.

The problem is solved by developing a two-step matheuristic designed specifically for the problem observed at the Clinic of Surgery. The matheuristic is based on Mixed Integer Programming, and while a simple implementation of the model in commercial software proves insufficient for this complex rostering problem, our matheuristic provides high-quality solutions for all test instances.

Our main contributions in this paper include describing and formalizing the novel Semi-Cyclic Ranked Physician Rostering Problem and developing a two-step matheuristic that solves the problem for real-life instances. We also demonstrate that we can produce high-quality rosters with superior quality to those created through manual planning and introduce shadow shifts that reduce the probability of rerostering. The paper is co-authored by Associate professor Anders Nordby Gullhav, Professor Henrik Andersson, and Head of the Surgery Clinic Birger Henning Endreseth.

1.2.4 Paper 3: Nurse Rostering with Fatigue Modelling - Incorporating a Validated Sleep Model with Biological Variations in Nurse Rostering

In this paper we incorporate a validated model of human sleep in a Nurse Rostering Problem to formulate the Nurse Rostering Problem with Fatigue (NRPwF). We minimize the worst level of fatigue experienced by any nurse in the roster to produce rosters aiming to enhance nurse health and reduce human errors.

To solve the NRPwF we develop an algorithm combining Mixed integer Programming and Constraint Programming with a Large Neighbourhood Search. As the number of possible rosters is huge, all individual schedules of realistic lengths cannot be evaluated by the sleep model. We thus approximate the sleep model using a look-up table before performing a post-processing algorithm dealing with errors over a given threshold.

Our main contributions are approximating the sleep model and incorporating it in the NRP to create the NRPwF. The introduction of biological profiles in such a framework provides interesting new insights such as how nurses with different biological profiles must be treated differently to minimize the maximum observed fatigue. We also demonstrate how increasing staff levels enables a reduction in the worst cases of fatigue. Our algorithm for solving the NRPwF and performing post-processing represents another notable contribution. The paper is co-authored by PhD Ilankaikone Senthooran and Professor Mark Wallace.

1.2.5 Paper 4: Hospital rostering of the future - experiences with new technology

In this paper we describe a case study of implementing a decision support tool for nurse rostering at the Maternity Ward West at St. Olav's Hospital in two pilot projects. We survey employees at the ward to identify potentials and discuss whether they were fulfilled in our pilot projects. We go on to reflect on how future rostering systems will function in different health care organizations, provided systems such as ours are used to a large extent. We also draw on the experiences from similar initiatives elsewhere in Scandinavia to establish a broader foundation for these reflections.

The rostering model was developed in close co-operation with the Operations Coordinator at Maternity Ward West. The Operations Coordinator was responsible for personnel planning and rostering at the ward, and had worked there as a midwife prior to her current role. Multiple stakeholders were informed and gave their opinions on the project, including individual employees, the Norwegian Nurses' Association (NSF), and the hospital management. The model itself is not presented in the paper, but is formalized as a Mixed Integer Program in Appendix A.

The main contributions of this paper is demonstrating the realism of major potentials of implementing decision support software for Nurse Rostering at Maternity Ward West. Furthermore, we identify notable barriers for a large-scale implementation at MWW and likely similar wards. We discuss how rostering will change in the time up to the year 2050, and explain how the roles of different shareholders are likely to change. The paper is co-authored by Associate Professor Anders Nordby Gullhav.

Appendix A: Nurse Rostering Model developed for use at Maternity Ward West

The appendix presents a mathematical formulation of the Nurse Rostering Problem observed at Maternity Ward West. The mathematical formulation was not included in the original paper, as it was outside the scope of the anthology the paper was published in.

1.3 Contributions

This section presents the contributions made in this thesis to the research community and to the industry respectively. Lastly we briefly describe the author's contribution to each of the papers presented in the thesis.

1.3.1 Contributions to the research community

Each paper's individual contribution to the research community is already briefly discussed in Sections 1.2.2 - 1.2.5, and are further presented in the introductions of the papers. Here we discuss overarching contributions of the thesis to the research community, that help answer the two key research questions we asked in Section 1.2.1. Firstly, we discuss findings related to identifying what is a high-quality roster, structured using four important aspects of rosters. These are preferences, fairness, costs, and robustness.

Preferences

We contribute to broaden the perspectives on preference modelling in rostering in several ways. Papers I, II, and IV all deal with real-life problems, and thus entail modelling preferences that do not necessarily exist in related literature. In these projects, preference modelling required attempts to unveil and make explicit many different rules and norms of an implicit character that existed in the minds of planners and personnel. Furthermore, we found how the existing information system of registering requests and leaving comments was inadequate to fully capture staff preferences in Paper IV. We also believe that the practice of planning long rosters that was common at all wards we worked with at St. Olav's Hospital made it difficult for staff to know when they wanted different shifts, making the self-scheduling less effective.

Our findings demonstrate how it is not realistic to address the aspect of staff preferences in a sufficient manner simply by formulating the correct constraints in rostering models. For future automatic or semi-automatic rostering systems to be useful, the entire rostering process should ideally be evaluated, and outside stakeholders such as developers of information systems should be part of the process.

Fairness

Fairness is another aspect that is hard to assess directly in mathematical models. As mentioned in Section 1.1.2, most authors deal with this by modelling some form of equality, e.g. penalizing or disallowing large differences in unpopular shift types between staff. Versions of such constraints focusing on equality in objective roster characteristics exist in our papers as well, e.g. Paper II models equality in work hours among surgeons of similar ranks.

However, in Papers IV, a different perspective on fairness affects the rostering problem a great deal. While employees at Maternity Ward West should have a similar number of e.g. weekend shifts, rosters have a significant focus on the individual needs of employees. The underlying principle was stated clearly by the Operations Coordinator when working with Paper IV, saying that all employees' preferences are equally important to them, but that fairness implies equality in respect and influence, not necessarily equality in the individual rosters, as people have different health issues and needs. This focus on individual needs is endorsed through the inclusive working life agreement in Norway, a policy meant to increase individual customization and enhance employee health.

Notably, this perspective on fairness is aligned with the results in Paper III. When focusing on minimizing the maximum fatigue score for any nurse on any day to enhance nurse health and reduce the risk of human errors, results demonstrate that different individuals must be assigned very different rosters. As we write in Paper III, our results imply that managers must grapple with the idea of what fairness is in rostering. While it is easy and tempting to treat every nurse exactly the same irrespective of their reaction to working around-the-clock, this does not suffice if managers wish to create rosters that focus on nurse health and patient safety.

Costs

Costs are mostly considered sunk in our work. That is, while costs are an integral component of staffing decisions, our rostering problems tend to assume the available staff is fixed, making the majority of costs sunk. That does not mean our models cannot be used for analyses of costs, as well as other aspects, on a tactical level. By changing parameter values, effects of increased or reduced staff levels are available. Policy changes at wards are easily evaluated by introducing or removing constraints from the rostering models. At the end of the day, there are countless possible changes to staffing and policies at the tactical level that can be included in our rostering problems if budgets are increased, or that must be considered if budgets are reduced, but these decisions are generally not modelled in our work.

Furthermore, in the online operational decision level, costly actions such as asking staff to work overtime is performed. Monetary costs constitute part of the weights of rerostering actions in Paper I, but they are integrated in inconvenience costs to create weights useful in the planning problem.

The lack of focus on costs itself in this thesis may very well reflect the close co-operation with real-life practitioners and planners rather than top-level representatives at the hospital. Oftentimes, their focus is not on costs per se, but rather on performing the best possible planning and management within the restrictions provided by budgets. The lack of focus on costs in this thesis can also be a symptom of Norwegian work-life culture. Regardless, models tend to be formalized based on a given number of resources available, which reflects budgetary constraints in real-life. We believe this is a reasonable approach, but also note that it can be a weakness if it entails not identifying measures for saving costs. One example is that if preliminary testing demonstrated that feasible rosters holding high quality could be created in Paper II if hours were reduced for surgeons, but investigating the trade-off between roster quality, the number of section shifts given to specializing surgeons, and costs was not the focus of the study.

Robustness

Robustness is discussed in multiple papers in this thesis, most comprehensively in Papers I and II. While multiple different strategies are presented and evaluated in Paper I, only one is implemented in Paper II. It is interesting to note that the two papers both present different versions of strategies where extra personnel scheduled for an off-shift is called upon when needed. Furthermore, we conclude that the version implemented in Paper I was not able to improve robustness, but that the strategy implemented in Paper II gave promising results. This is interesting, as the two strategies seem very similar at first glance. However, there are notable differences between the two strategies and the two planning problems that affect the results.

In Paper I, ghost shifts are introduced. These entail assigning an off-shift such that a nurse can cover for absence during the night shift. The ghost shift strategy primarily fails to improve the robustness of the rosters for two reasons. Firstly, it turns out that covering night shifts is not very challenging without the ghost shifts at the DNIC. Secondly, when ghost shifts are introduced, we ensure that no time-related constraints are violated if a ghost shift is *realized*, i.e. the nurse has to step in and work the night shift. This makes rosters where ghost shifts are used frequently very rigid.

In Paper II, shadow shifts are introduced to cover for absence during emergency night shifts. Shadow shifts are similar to ghost shifts in Paper I, but we accept that surgeons finish their day shift prior to beginning a night shift (in practice, it would be up to managers to consider if the surgeon could be given time off or prolonged breaks etc. to ensure a safe schedule). As a result, shadow shifts are much less restricting to the rostering problem than ghost shifts, and results indicate that shadow shifts do not deteriorate other roster qualities notably. We are not able to evaluate the effects of the shadow shifts in the same way as we could with the framework provided in Paper I, but we know that covering emergency night shifts is a challenge in real-life at the ward and we calculate that the probability of being forced to reroster is smaller after introducing shadow shifts.

In Paper IV, one of many roster qualities is to limit overstaffing, thus enhancing robustness. In practice, there was little opportunity for the rostering model to prioritize robustness, because numerous other restrictions were prioritized. This indicates that in order to take advantage of robust measures in practice, advantages must be clearly communicated to different stakeholders. That is, it must be clear for employers how robust rosters can reduce overtime costs, while planners and staff should know how increased robustness can reduce the inconvenience related to last-minute changes in their schedules, such as in Papers I and II.

Producing high-quality rosters

Solution methods in this work all have in common that they are quite flexible, in the sense that models are easily adjusted to minor changes in problem descriptions. This has been important, as communication with practitioners often entail discussions and presentation of model output in an iterative process. This unveils new details in the problems and contributes to disclose misunderstandings. As a result, some of the methods presented in related literature for solving rostering problems would not be very practical for our use, e.g. some decomposition methods could entail a lot of extra work when changes are made iteratively to problem descriptions.

This is not to say structures of the problems observed are not relevant when developing solution methods for real-life problems. In the model used for Paper IV, frequent adjustments were proposed in meetings with the Operations Coordinator. In some cases, proposed changes led to infeasible problems, but the commercial software could not prove it within reasonable time limits. As a result, we introduced a feasibility check, which firstly checked the feasibility of the roster of each individual nurse when new constraints were introduced, before attempting to solve the problem. This was very helpful. Similarly, the structures of the problems were very relevant for the solution methods developed in Papers II and III. In our experience from this work, it is important to retain the flexibility in solution methods to easily make changes to problem descriptions until stakeholders agree on it. At that point, more efficient solution methods can be developed if it is desirable.

1.3.2 Contributions to the industry

This thesis demonstrates the potential for larger-scale implementation of a decision support system based on Operations Research for personnel planning at St. Olav's Hospital and similar health care organizations. Papers I and II show that real-life problems can realistically be solved, and that results produced outperform manual planning. Furthermore, the rostering models we develop all enable evaluation of policy changes on a tactical level, and give important managerial insights for planners. Paper III expands the scope of rostering in Operations Research, demonstrating how fatigue minimization can be incorporated into NRPs. Paper IV describes pilot projects of implementing such systems, identifies multiple potentials and challenges to realize them, and discusses how this will affect work life in the future. Combined, this thesis provides important insights on how Operations Research can be used for personnel planning in health care organizations.

An important motivation for this thesis is to create something useful for planners and practitioners. Thus we briefly discuss experiences with barriers and challenges in moving from observing a real-life problem to attempting to implement decision support systems for permanent use for practitioners. It is not clear to what extent these experiences can be generalized, but they are noteworthy for St. Olav's Hospital, and could motivate further research.

When models were developed and tests had been performed at the various departments we worked with at St. Olav's Hospital, we were faced with the task of documenting our work through scientific papers. At this point, it was unclear how projects should be continued. In the work at Maternity Ward West in Paper IV, results of surveys were presented for the Center for Health Care Improvement, an important stakeholder who showed a lot of interest in the project. However, we did not produce any clear plan for how to continue the incorporation of our model in planning, and in practice the project was shelved. In the project at the Department of Neonatal Intensive Care in Paper I, master's students had closest contact with planners, which may have reduced the chances of continuing co-operation after their graduation. In the project at the Clinic of Surgery in Paper II, adjustments were made to give output in a useful format for planners. However, due to possible upcoming restructuring, planners did not wish to commit to changes in the planning process.

While the above mentioned projects have a quite common lifespan for research projects, it would be remiss not to mention aspects that could help facilitate taking the projects from theory to practice. The overall impression is that while managers and planners are very interested in exploring possible improvements to personnel planning, there is no clear path from testing to permanent implementation of new decision support tools. To the author's knowledge, there were no specialists that could facilitate such a transition, and as researchers observing a planning problem from the outside, the mandate and funding for making changes to improve planning was lacking in our projects.

If future personnel planning projects are launched, they should ideally state a clear mandate, invite and demand cooperation from all relevant stakeholders and be prioritized in terms of personnel with specialties within both Operations Research and innovation. While most stakeholders were included in our projects, especially that at Maternity Ward West in Paper IV, it is clear that stakeholders' disagreements were not sufficiently addressed. That is, they were all given the chance to be heard in our process, but their opinions were never confronted with other stakeholders' views and priorities. In real-life, there is a thug-of-war between stakeholders in rostering, and a process without any conflict could be a sign that the core of trade-offs has not been properly discussed.

While the approach described above demands a significant and coordinated effort by a health care organization, we also believe there is more low-hanging fruit in our work. For details, see the individual papers. However, one useful example is trading extra weekend work for an additional off-day. For all the round-the-clock nurse rostering problems we encounter at St. Olav's Hospital, staffing levels at weekends are clear bottlenecks. For practitioners at the hospital, this will not come as a surprise. The weekend-bottlenecks are a result of nurses normally working every third weekend, as regulated for most nurses in roster agreements. This was well documented already in the Master's thesis by Beckmann and Klyve (2016), but Paper I also took into account uncertainty in evaluating this measure. The conclusion is clear. "There was a huge improvement in robustness when nurses were allowed to trade extra weekend work for extra off days in the initial rosters. The policy change led to a more stable roster during weekends, without any significant effects on the stability during weekdays although less work shifts where scheduled." and goes on to say "we believe that accepting this policy change would be very beneficial."

In practise, this policy change has not been attempted. Unless managers at St. Olav's Hospital believe that implementing the trade will make nurses' perception of covering for each other too transactional, this seems like a very promising and uncomplicated measure to implement.

1.3.3 The author's contribution to each paper in the thesis

This section describes the contribution of the author to each individual paper in the thesis. The contribution is divided in three categories; intellectual input, development and implementation of models and code, and writing of papers. To simplify, the author's contribution is rated on a scale of 1 to 3 in each category for each paper, as seen from Table 1.1. The rating 1 represents some contribution, 2 represents a significant contribution and 3 represents a majority of the contributions.

Paper	Intellectual input	Development and implementation	Writing
Paper I	3	1	3
Paper II	3	3	3
Paper III	3	3	3
Paper IV	3	3	3

Table 1.1: The author's contribution to each paper in the thesis.

As is clear from Table 1.1, the author contributed to the majority of all categories in each paper except Paper I. When the project was designed, master's students Isabel Nordli Løyning and Line Maria Haugen Melby were included, performing the majority of work on Development and Implementation of the code. It should also be noted that significant parts of development of the code in Paper III, related to coding the Nurse Rostering Problem with Fatigue in MiniZinc, was performed by Dr. Ilankaikone Senthooran.

Papers are written in American or British English depending on which authors that have co-operated in each project. Paper IV was originally published in Norwegian and has been translated by externals before proof-reading. In the translation, we have prioritized to avoid diverging communication in the two languages over the flow of the text in the English version.

1.4 Concluding remarks and future research

Personnel rostering is a central part of planning in health care organizations, and Operations Research offers promising potential for improved planning. This work verifies that Operations Research is a useful tool for rostering. The rosters we create contribute to answering what a high-quality roster is, but also highlights the complexities of modelling key aspects such as preferences, fairness, costs, robustness, and fatigue. We are able to solve the planning problems we are faced with throughout the thesis. While we develop novel solution methods when necessary, we do not focus on proving optimality, the way much rostering literature has done previously. Rather, our main concern is to gain insights on how to produce rosters of high quality.

The models we develop in this thesis reflect real-life problems, some with a very high level of detail. Furthermore, we develop solution methods that solve the problems for real-life instances. Our models aspire to fulfill two goals:

- Provide insights about real-life personnel planning
- Be the foundation for decision-support tools used in practice

Our models definitely meet the first goal. Insights acquired in each project are described in detail in the works themselves, but constitute an important part of our contributions both to the research community and to the industry. Evaluating whether we have reached the second goal is more unclear. Our projects face barriers for implementation, but identifying them should also be considered important findings. The future will unveil to what extent our work is a foundation to build on for practical implementation of decision-support tools, and this line of thought leads to our main suggestions for future research.

The overarching research idea that naturally stands out after our work is to make a more coordinated effort to implement a lasting system at a health care organization. This project should entail researchers across multiple research areas, and relevant stakeholders with different roles in the organization should be included in the project. The theoretical literature on personnel rostering, especially for nurses, has matured over the last decades, demonstrating that formalized problems can be solved. What is truly lacking at this point is the applied focus enabling it to culminate in a real breakthrough outside the literature on solution methods.

Bibliography

U. Aickelin and K. A. Dowsland. An indirect genetic algorithm for a nurse-scheduling problem. *Computers & Operations Research*, 31(5):761–778, 2004.

M. Akbari, M. Zandieh, and B. Dorri. Scheduling part-time and mixed-skilled workers to maximize employee satisfaction. *The International Journal of Advanced Manufacturing Technology*, 64(5):1017–1027, 2012.

Arbeids- og sosialdepartementet. Lov om arbeidsmiljø, arbeidstid og stillingsvern mv. (arbeidsmiljøloven). `https://lovdata.no/dokument/NL/lov/2005-06-17-62`, 2017.

E. I. Ásgeirsson. Bridging the gap between self schedules and feasible schedules in staff scheduling. *Annals of Operations Research*, 218(1):51–69, 2014.

E. I. Ásgeirsson and G. L. Sigurðardóttir. Near-optimal MIP solutions for preference based self-scheduling. *Annals of Operations Research*, 239(1):273–293, Apr 2016.

M. Azaiez and S. A. Sharif. A 0-1 goal programming model for nurse scheduling. *Computers & Operations Research*, 32(3):491–507, 2005.

J. Bai, A. Fügener, J. Schoenfelder, and J. O. Brunner. Operations research in intensive care unit management: a literature review. *Health Care Management Science*, 21(1):1–24, 2018.

J. Bard, Z. Shu, D. Morrice, L. Leykum, and R. Poursani. Annual block scheduling for family medicine residency programs with continuity clinic considerations. *IIE Transactions*, 48(9):797–811, 2016.

J. F. Bard and H. W. Purnomo. Preference scheduling for nurses using column generation. *European Journal of Operational Research*, 164(2):510–534, 2005a.

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

J. F. Bard and H. W. Purnomo. Incremental changes in the workforce to accommodate changes in demand. *Health Care Management Science*, 9(1):71–85, 2006.

F. R. Beckmann and K. K. Klyve. Optimisation-based nurse scheduling for real-life instances. Master's thesis, Norwegian University of Science and Technology, 2016.

J. Beliën and E. Demeulemeester. A branch-and-price approach for integrating nurse and surgery scheduling. *European journal of operational research*, 189(3): 652–668, 2008.

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

M. Bester, I. Nieuwoudt, and J. H. Van Vuuren. Finding good nurse duty schedules: a case study. *Journal of Scheduling*, 10(6):387–405, 2007.

G. S. Braut. Helseforetak. `https://sml.snl.no/helseforetak`, 2019.

R. Bruni and P. Detti. A flexible discrete optimization approach to the physician scheduling problem. *Operations Research for Health Care*, 3(4):191–199, 2014.

E. Burke, P. De Causmaecker, S. Petrovic, and G. Vanden Berghe. Metaheuristics for handling time interval coverage constraints in nurse scheduling. *Applied Artificial Intelligence*, 20:743–766, 12 2006.

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

S. Ceschia, N. Dang, P. De Causmaecker, S. Haspeslagh, and A. Schaerf. The second international nurse rostering competition. *Annals of Operations Research*, 274(1):171–186, 2019.

V. Clarissa and S. Suyanto. New reward-based movement to improve globally-evolved bco in nurse rostering problem. In *2019 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, pages 114–117. IEEE, 2019.

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

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

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

R. G. Drake. The nurse rostering problem: from operational research to organizational reality? *Journal of Advanced Nursing*, 70(4):800–810, 2014.

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

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

A. Fügener, A. Pahr, and J. O. Brunner. Mid-term nurse rostering considering cross-training effects. *International Journal of Production Economics*, 196:176–187, 2018.

A. Fügener and J. O. Brunner. Planning for overtime: The value of shift extensions in physician scheduling. *INFORMS Journal on Computing*, 31(4):732–744, 2019.

R. A. Gomes, T. A. Toffolo, and H. G. Santos. Variable neighborhood search accelerated column generation for the nurse rostering problem. *Electronic Notes in Discrete Mathematics*, 58:31–38, 2017. 4th International Conference on Variable Neighborhood Search.

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

E. W. Hans, M. Van Houdenhoven, and P. J. H. Hulshof. A framework for health care planning and control. Memorandum 1938, Department of Applied Mathematics, University of Twente, Enschede, February 2011.

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

J. Ingels and B. Maenhout. The impact of reserve duties on the robustness of a personnel shift roster: An empirical investigation. *Computers & Operations Research*, 61:153–169, 2015.

J. Ingels and B. Maenhout. Employee substitutability as a tool to improve the robustness in personnel scheduling. *OR Spectrum*, 39(3):623–658, Jul 2017.

J. Ingels and B. Maenhout. The impact of overtime as a time-based proactive scheduling and reactive allocation strategy on the robustness of a personnel shift roster. *Journal of Scheduling*, 21(2):143–165, 2018.

J. Ingels and B. Maenhout. Optimised buffer allocation to construct stable personnel shift rosters. *Omega*, 82:102–117, 2019.

P. Kazemian, Y. Dong, T. R. Rohleder, J. E. Helm, and M. P. Van Oyen. An ip-based healthcare provider shift design approach to minimize patient handoffs. *Health Care Management Science*, 17(1):1–14, 2014.

D. L. Kellogg and S. Walczak. Nurse scheduling: From academia to implementation or not? *Interfaces*, 37:355–369, 2007.

M. Kitada and K. Morizawa. A Heuristic Method for Nurse Rerostering Problem with a Sudden Absence for Several Consecutive Days. *International Journal of Emerging Technology and Advanced Engineering*, 3(11):353–361, Nov 2013.

F. Knust and L. Xie. Simulated annealing approach to nurse rostering benchmark and real-world instances. *Annals of Operations Research*, 272(1-2):187–216, 2019.

G. Lim and A. Mobasher. Robust Nurse Scheduling Problem. *IIE Annual Conference. Proceedings*, pages 1–8, 2011.

Z. Liu, Z. Liu, Z. Zhu, Y. Shen, and J. Dong. Simulated annealing for a multi-level nurse rostering problem in hemodialysis service. *Applied Soft Computing*, 64: 148–160, 2018.

Lovdata.no. Forskrift om fastlegeordning i kommunene, 2021. `https://lovdata.no/dokument/SF/forskrift/2012-08-29-842`.

Z. Lü and J.-K. Hao. Adaptive local search for the first international nurse rostering competition. *INRC2010 (http://www. kuleuven-kortrijk. be/nrpcompetition)*, 2010.

B. Maenhout and M. Vanhoucke. An evolutionary approach for the nurse rerostering problem. *Computers & Operations Research*, 38(10):1400–1411, 2011.

B. Maenhout and M. Vanhoucke. Reconstructing nurse schedules: Computational insights in the problem size parameters. *Omega*, 41(5):903–918, 2013.

F. Mischek and N. Musliu. Integer programming model extensions for a multi-stage nurse rostering problem. *Annals of Operations Research*, 275(1):123–143, Apr 2019.

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

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

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

OECD. *Health at a Glance 2019*. OECD, 2019. doi: https://doi.org/https://doi.org/10.1787/4dd50c09-en.

M. V. Pato and M. Moz. Solving a bi-objective nurse rerostering problem by using a utopic pareto genetic heuristic. *Journal of Heuristics*, 14(4):359–374, 2007.

S. Petrovic. "you have to get wet to learn how to swim" applied to bridging the gap between research into personnel scheduling and its implementation in practice. *Annals of Operations Research*, 275(1):161–179, 2019.

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

H. W. Purnomo and J. F. Bard. Cyclic preference scheduling for nurses using branch and price. *Naval Research Logistics (NRL)*, 54(2):200–220, 2007.

E. Rahimian, K. Akartunalı, and J. Levine. A hybrid integer programming and variable neighbourhood search algorithm to solve nurse rostering problems. *European Journal of Operational Research*, 258(2):411–423, 2017.

regjeringen.no. Ministry of health and care services. https://www.regjeringen.no/en/dep/hod/id421/, 2021a.

regjeringen.no. The department of hospital ownership. https://www.regjeringen.no/en/dep/hod/organisation-and-management-of-the-ministry-of-health-and-care-services/Departments/the-department-of-hospital-ownership/id1413/, 2021b.

regjeringen.no. Departments in the ministry of health and care services. https://www.regjeringen.no/en/dep/hod/organisation-and-management-of-the-ministry-of-health-and-care-services/Departments/id448/, 2021c.

J. Ruzzakiah, I. Wan Rosmanira, L. C. Yuen, and A. Oughalime. A cyclic nurse scheduling using goal programming. *A Cyclic Nurse Scheduling Using Goal Programming*, 43A:151–164, 2011.

E. Rönnberg and T. Larsson. Automating the self-scheduling process of nurses in swedish healthcare: a pilot study. *Health Care Management Science*, 13:35–53, 2010.

J. Schoenfelder, K. M. Bretthauer, P. D. Wright, and E. Coe. Nurse scheduling with quick-response methods: Improving hospital performance, nurse workload, and patient experience. *European Journal of Operational Research*, 283(1):390–403, 2020.

R. Silvestro and C. Silvestro. An evaluation of nurse rostering practices in the National Health Service. *Journal of Advanced Nursing*, 32(3):525–535, 2000.

H. K. Smalley and P. Keskinocak. Automated medical resident rotation and shift scheduling to ensure quality resident education and patient care. *Health Care Management Science*, 19(1):66–88, 2016.

P. Smet, B. Bilgin, P. De Causmaecker, and G. Vanden Berghe. Modelling and evaluation issues in nurse rostering. *Annals of Operations Research*, 218(1): 303–326, Jul 2014.

A. L. Soyster. Technical Note—Convex Programming with Set-Inclusive Constraints and Applications to Inexact Linear Programming. *Operations Research*, 21(5):1154–1157, 1973.

stolav.no. Om oss [about us]. `https://stolav.no/om-oss#om-helseforetaket`, 2021.

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

S. Topaloglu. A multi-objective programming model for scheduling emergency medicine residents. *Computers & Industrial Engineering*, 51(3):375–388, 2006. Special Issue on Selected Papers from the 34th. International Conference on Computers and Industrial Engineering (ICC&IE).

C. Valouxis, C. Gogos, G. Goulas, P. Alefragis, and E. Housos. A systematic two phase approach for the nurse rostering problem. *European Journal of Operational Research*, 219(2):425–433, 2012.

D. M. Warner. Scheduling nursing personnel according to nursing preference: A mathematical programming approach. *Operations Research*, 24(5):842–856, 1976.

WHO. Global strategy on human resources for health: workforce 2030. https://apps.who.int/iris/bitstream/handle/10665/250368/9789241511131-eng.pdf, 2016.

A. Wren. Scheduling, timetabling and rostering — a special relationship? In E. Burke and P. Ross, editors, *Practice and Theory of Automated Timetabling*, pages 46–75, Berlin, Heidelberg, 1996. Springer Berlin Heidelberg.

Paper I

K. K. Klyve, I. N. Løyning, L. M. H. Melby, H. Andersson, A. N. Gullhav:

A modelling framework for evaluating proactive and reactive nurse rostering strategies - A case study from a Neonatal Intensive Care Unit

Submitted to an international scientific journal

This paper is awaiting publication and in is not included in NTNU Open

Paper II

K. K. Klyve, H. Andersson, A. N. Gullhav, B. H. Endreseth:

Semi-cyclic rostering of ranked surgeons - a real-life case with stability and flexibility measures

Published in Operations Research for Health Care

Semi-cyclic rostering of ranked surgeons - a real-life case with stability and flexibility measures

Abstract

We consider the rostering problem for surgeons in residency at the Clinic of Surgery at St. Olav's Hospital, Trondheim University Hospital, in Trondheim, Norway. Each surgeon in residency has a rank depending on experience. An exact number of surgeons of each rank must work emergency shifts in a cyclic structure. Each surgeon is affiliated to a section, which has a minimum staffing level. Section shifts can be planned in an acyclic structure, thus establishing a semi-cyclic structure in the full roster. The addition of more typical rostering constraints establishes the novel Semi-Cyclic Ranked Physician Rostering Problem. In manually created rosters, the staffing at sections varies greatly, leading to frequent understaffing. With the addition of absence among staff when rosters are executed, this is problematic for the Clinic of Surgery. We present a two-step matheuristic based on mixed integer linear programming to solve the problem for five real-life instances. Comparing our results to a manually created roster demonstrates superior results in terms of staff availability at sections, greatly improving roster resilience to absence. We also introduce shadow shifts designed to increase the flexibility of rosters to cover for absence at emergency night shifts.

3.1 Introduction

Health care organizations have an integral role in modern society, accentuated by an average of 9% of gross domestic product in the Organization for Economic Cooperation and Development countries being spent on health, (OECD, 2017). Furthermore, hospitals account for nearly 40% of health spending, (OECD, 2017), including considerable staffing costs. The rostering of personnel is a recurring challenge at hospitals, having a significant effect on the personnel costs. This issue has been a topic of interest among researchers for many years (see e.g. Wolfe and Young (1965a), Wolfe and Young (1965b), Warner (1976)), but there still exists a large potential for improved utilization of personnel resources in hospitals worldwide. In this work, we focus on the scheduling of physicians, specifically a real-life problem of rostering surgeons at the Clinic of Surgery at St. Olav's Hospital, Trondheim University Hospital, Norway.

At the Clinic of Surgery, the rostering of different groups of employees is a highly complex task. One group of employees has proven particularly hard to schedule in a way that ensures reliable and stable staffing levels. This group is made up of specializing surgeons who are currently performing their medical *residency*. Residency refers to physicians training to become specialists in a medical field under the supervision of attending physicians within the same field. In this article, we simply refer to the specializing surgeons as surgeons.

As well as acquiring experience, the surgeons have two important functions at the Clinic of Surgery. The surgeons are an integral part of the staffing both at the emergency department and at a collection of sections at the Clinic of Surgery. Each surgeon has a *rank*, corresponding to his or her experience level, and a *section affiliation*. Different ranks need to meet the exact demand for emergency shifts at the emergency department at all times. Furthermore, sections have a defined minimum and preferred level of demand for staff every weekday, depending on the size of the section.

Another aspect to consider when creating the rosters, which greatly complicates the structure of this planning problem, is that the emergency shifts must be planned in a weekly cyclic structure. Other shifts are not constrained to a cyclic structure. Thus, as long as the emergency shifts are cyclic, the rest of the shifts can be assigned freely, assuming other requirements are satisfied.

We define a roster to be a plan that allocates employees to work shifts of predefined start and end times throughout a planning horizon. In the daily execution of the roster, surgeons may be absent from shifts they are planned to work. We define the robustness of a roster as a combination of the roster's ability to withstand disruption, i.e. *stability*, and the roster's potential to retain high quality after disruptions occur given a specific set of rerostering strategies, i.e. *flexibility*. These definitions are similar to those used in related work, see e.g. Ionescu and Kliewer (2011), Ingels and Maenhout (2017), and Ingels and Maenhout (2018).

In this paper we formalize the problem of finding a roster for surgeons at the Clinic of Surgery. The schedule for each surgeon must be legal and adhere to the work hour restrictions stated by the Norwegian Working Environment Act and the local collective agreement. In total, the roster needs to fulfill the demand at all shifts for the sections as well as the emergency department. The main objectives are to find the most robust roster and comply with fairness-related norms.

To the best of the authors' knowledge, no planning problem of a similar structure has been presented in literature on physician rostering before, and we thus propose calling this problem the Semi-Cyclic Ranked Physician Rostering Problem (SCRPRP). Our contributions in this paper are listed below.

- Describing and presenting the novel Semi-Cyclic Ranked Physician Rostering Problem.
- Presenting a matheuristic that solves the problem for real-life instances.
- Producing a high-quality roster for use at the Clinic of Surgery at St. Olav's Hospital, Trondheim University hospital.
- Analyzing the improvements of the roster compared with rosters from manual planning and the effect of shadow shifts on the robustness of the roster.

The outline of the paper is as follows. Section 3.3 provides background for, and briefly presents, the planning problem at the Clinic of Surgery. Relevant literature is presented in Section 3.2 while a concise problem description is given in Section 3.3.2. In Section 3.4, a two-step matheuristic to solve the problem is described. In the computational study, Section 3.5, the results for solving the instances are presented and evaluated and the effect of introducing shadow shifts is analyzed. Finally, Section 3.6 concludes the paper.

3.2 Related Literature

The problem presented here belongs to the class of personnel planning problems, see e.g. den Bergh et al. (2013) and Bruecker et al. (2015) for extensive reviews of personnel scheduling and workforce planning respectively. Within health care, much research has been focused on nurses, see e.g. Burke et al. (2004) for a review and Ceschia et al. (2019) for a recent report from the Second International Nurse Rostering Competition. As there is generally no difference in rostering a generic group of surgeons compared to rostering a generic group of physicians, we use the terms surgeon rostering/scheduling interchangeably with physician rostering/scheduling. Even though surgeon scheduling has a lot in common with nurse scheduling there are many aspects that differ, as pointed out by Erhard et al. (2018). The most prominent differences in our problem are that many surgeons are undergoing medical training, that they have different ranks based on experience, and that they are affiliated to different sections based on educational focus. In the excellent review by Erhard et al. (2018), physician scheduling is outlined and more than 60 papers are reviewed and classified. According to the classification in Erhard et al. (2018), the problem studied here is a physician rostering problem with residents and fairness aspects.

One dimension only briefly mentioned in Erhard et al. (2018) is cyclic plans, a key aspect in our problem. In the work of Ernst et al. (2004), the cyclic structure of many personnel scheduling problems are discussed extensively. They provide the following definition of cyclic rosters: "In a cyclic roster all employees of the same class perform exactly the same line of work, but with different starting times for the first shift or duty." In the review by Bergh et al. (2013), some articles dealing with cyclic scheduling problems are cited, but the cyclic structure of the problems is not discussed explicitly. Burke et al. (2004) discuss cyclic scheduling, and categorize nurse scheduling problems as either cyclic, semi-cyclic, or non-cyclic. This distinction is interesting when regarding the planning problem at the Clinic of Surgery as only some shifts must be scheduled as cyclic. Burke et al. (2004) do not define semi-cyclic scheduling explicitly, but mention examples of semi-cyclic scheduling; Burke et al. (2001), Warner (1976), Smith (1976), and Chan and Weil (2001). Smith (1976) shares the characteristics of having some, but not all, shifts scheduled in a cyclic structure, which means that the planning problem at the Clinic of Surgery would fit the semi-cyclic category as proposed by Burke et al. (2004). For newer literature dealing with cyclic scheduling, see for example the works of Becker et al. (2019) and Xie and Suhl (2015), where the applications are emergency medical services and public bus transit, respectively.

Operational variability can be handled by introducing either proactive or reactive mechanisms. Proactive mechanisms create rosters that can absorb the variability or that improve the possibility to handle the unexpected event by adjusting the roster, while reactive mechanisms focus on handling the variability once it has been realized. See Ingels and Maenhout (2017) for a longer discussion about different mechanisms. Fügener and Brunner (2019) study a physician scheduling problem with stochastic demand and allow variable shift extensions as a proactive measure, while EL-Rifai et al. (2015) focus on the staffing level. Dück et al. (2012) and Ionescu and Kliewer (2011) deal with stability and flexibility for airline crew schedules, while Ingels and Maenhout (2017) present employee substitutability as a mean to improve the robustness in personnel scheduling. Rescheduling, i.e. reactive mechanisms, within health care is found in the review by Clark et al. (2015) for nurses, and in Gross et al. (2018) for physicians.

This paper extends the existing literature by (i) combining section and emergency department scheduling where parts of the roster must be cyclic; (ii) including proactive measures to improve both the stability and flexibility of the roster; (iii) introducing shadow shifts as a strategy to increase the flexibility of the roster.

3.3 Surgeon rostering at the Clinic of Surgery at St. Olav's Hospital

This section describes the Clinic of Surgery at St. Olav's Hospital with a special emphasize on how the surgeon rostering is done and the considerations that must be taken into account. Section 3.3.1 presents the Clinic of Surgery and how rostering is done today. A formal description of the Semi-Cyclic Ranked Physician Rostering Problem is given in Section 3.3.2.

3.3.1 Background

The Clinic of Surgery at St. Olav's Hospital, Trondheim University Hospital, consists of several departments and sections. The sections are staffed at daytime during the weekdays, and surgeons consider section shifts their core activity, as it is here they get most of the training for their specialization. At the sections, several tasks are performed, including surgeries, patient consultations at the outpatient clinics, and visits to inpatient wards.

At the Clinic of Surgery, the planning horizon is 26 weeks. Planning much longer than this is unpractical due to potential changes in rank and section affiliation. The surgeons' ranks are based on how much experience they have acquired in their residency. Beginners are called *interns*, while *residents* are more experienced and *officers* are the most experienced. Every 6 months, surgeons may change rank and section affiliation as their training progress

The roster is an assignment of exactly one shift each day to each surgeon. The different shift types are shown in Table 3.1.

Table 3.1: The different shift types used as the Clinic of Surgery

Name	Description
emergency shifts	shifts at the emergency department, both day and night
section shifts	shifts at the section, only day
coursework shifts	shift for classes or coursework, only day
off-shifts	non-working shifts

There are two types of emergency shifts, the emergency day shift and the emergency night shift, and they must be planned in a weekly cyclic structure. This means that every surgeon of a given rank must work the exact same emergency shifts as another surgeon of the same rank did the previous week, creating a chain of cyclic emergency shifts throughout the planning horizon. If the minimum requirement of ranks allow for it to happen, surgeons can cycle in a chain of more than one rank, i.e. share the burden of a type of emergency shift between ranks. Other shifts are not constrained to a cyclic structure. Thus, as long as the emergency shifts are cyclic, the rest of the shifts can be assigned freely, assuming other constraints are satisfied. The cyclic structure is illustrated in Table 3.2, where D represents an emergency day shift, N represents an emergency night shift and an asterisk (*) symbolizes any other shift on or off duty.

Table 3.2: Illustration of the cyclic structure in the schedule. In this example, "Officer7" follows the pattern of emergency shifts worked the previous week by "Officer1" and "Officer4" follows "Officer7". An asterisk (*) symbolizes any shift that is not an emergency shift, including off-shifts.

Surgeon	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	...
Officer1	D	*	*	*	*	N	N	*	*	*	*	D	*	*	*	...
Officer7	*	*	D	D	*	*	*	D	*	*	*	*	N	N	*	...
Officer4	N	N	*	*	*	*	*	*	*	D	D	*	*	*	D	...

In addition to the cyclic structure of the emergency shifts, there are regulations stated by the Norwegian Working Environment Act and the local collective agreement regarding working hours that must be adhered to. The regulations state that all surgeons must have a *protected off-day* involving a prolonged rest period every week and minimum rest times between two shifts. More specifically, surgeons must not work two night shifts with only one day off in between. Working two consecutive night shifts are only accepted if the surgeon works short night shifts, if one of the night shifts is a shadow shift, or if the night shifts occur during the weekend. If a surgeon works emergency shifts during the weekend, the surgeon should work two similar emergency shifts.

Each surgeon must have a full week off every time its emergency shift cycle has completed twice. For example, if there are 8 surgeons of a given rank, they would have one week off every 16^{th} week. If it is advantageous, the week can be delayed by 1 week, as long as the delay does not prolong into the following cycles.

There are limits to the maximum number of hours a surgeon can work each week. There are also limits to the maximum and minimum number of hours each surgeon can work throughout the planning period. Among surgeons holding the same rank, the difference in number of emergency shifts worked throughout the planning period must be restricted.

The demand related to each type of working shift is handled differently. The demand for emergency shifts must be fulfilled exactly, and is defined for a single rank or groups of ranks, while the demand for section shifts is stated as preferred and absolute minimum levels. Each surgeon must also be assigned to coursework shifts and off-shifts.

The number of surgeons of different ranks affiliated to each section constitutes the foundation of the planning problem. This information is found in the Rank and Section

Affiliation-matrix (RSA-matrix), see Appendix A.5. Currently, manual methods are used to find useful rosters. Planners normally start by creating week-long patterns of shifts. As long as these are of high quality (i.e. not breaking any rules, regulations or norms), and the transition from one week-long pattern to the next is allowed, it is simple to produce a surgeon's schedule for the full 26 weeks. Furthermore, if another surgeon works exactly the same schedule, shifted by 7 days, the cyclicity of the roster is ensured. To ensure that the exact demand for emergency shifts is met, planners create week-long patterns that in total include emergency shifts on all weekdays. If the sum of emergency shifts in all the week-long patterns meets the demand for emergency shifts every day, the cyclic structure ensures this demand is met for all days in the initial roster. However, this planning approach neglects the section demand. Although the initial roster can be changed before it is considered final, the planning process entails that staffing levels at sections vary greatly from day to day.

During the execution of the roster, long-term absence can occur, e.g. in case of parental leave. This can hinder some surgeons in gaining work experience during a planning period. Because of this, the Clinic of Surgery cannot use a standardized track for which sections each surgeon should be affiliated to after a certain number of months at the clinic, and they must plan with a new RSA-matrix in every planning period.

There is a difference in quality between rosters that satisfy the restrictions mentioned. One key quality that rosters should have, is stability. A stable roster is able to absorb changes that occur as the roster is executed. At the Clinic of Surgery, absence among staff is the main challenge. To ensure high stability, it is advantageous that staffing levels at sections are high enough on all days such that absence does not entail understaffing. A section with five extra surgeons on a Wednesday can handle an absent surgeon on that Wednesday very well, but may be vulnerable to absence on the other weekdays. If measures are put in place so that the overstaffing is spread out on all five weekdays, one absent surgeon never causes understaffing and the roster is therefore more stable.

There is a similar challenge when surgeons scheduled to work emergency shifts are absent. However, demand must be met exactly for emergency shifts, making overstaffing impossible. Thus, we cannot improve the stability of the roster with regards to emergency shifts. We can, however, introduce measures to improve flexibility. Planners at the Clinic of Surgery experience that if a surgeon is absent from an emergency day shift, someone among the working staff is able to cover for him/her without the need for replanning. This is especially true if there is overstaffing at any of the sections. This means that reasonable overstaffing at sections is also a flexibility measure for emergency shifts during daytime. Finding someone to work the emergency night shift is harder. To create flexible rosters we introduce *shadow shifts*. Shadow shifts are shifts that reserve a surgeon for the following night in case absence occurs on the emergency night shift. Surgeons can work *shadow section shifts*, *shadow coursework shifts* or have a *shadow off-day*. However, if absence occurs at an emergency night shift, the surgeon assigned the shadow shift is designated to cover it. If so happens, we say that the shadow shift is realized. When a surgeon is assigned a shadow shift, we plan as if the surgeon is not available the day after. This means that we ensure the feasibility of the roster if the surgeon is absent during the daytime the day after, securing rest after a realized night shift. However, we accept breaking restrictions on weekly maximum work hours and more consecutive night shifts if it is due to a realized shadow shift.

Our measures to establish stability and flexibility in rosters are exemplified in Table 3.4, where a small week-long roster for the urology section is shown. Off-shifts are denoted

Off, section shifts are denoted Sec, and emergency day and night shifts are denoted D and N respectively. Shadow shifts are indicated by a following asterisk (e.g. Off* is a shadow off-shift). Shifts that are struck through represent shifts where surgeons are absent when the shift should be worked.

Table 3.4: Illustration of how overstaffing and shadow shifts are measures to improve stability and flexibility. Only surgeons affiliated to the small urology section during a single week are included.

Surgeon	M	T	W	T	F	S	S	...
Intern7	Off	~~N~~	Off	D	Sec	Off	Off	...
Resident6	Sec	Sec	Sec	~~Sec~~	Sec	Off	Off	...
Officer3	Off	Sec*	Sec	Sec	D	Off	Off	...
Officer6	Sec	Off	Sec	Off	N	N	Off	...
Sum Section	2	2	3	2	2	0	0	

As seen from Table 3.4, Resident6 is absent from his/her section shift on Thursday. The sufficient staffing levels at the section is one, and the overstaffing therefore gives a good quality roster despite absence from the section shift. This is an example of using tactical overstaffing as a measure to create stability. Furthermore, Intern7 is absent from an emergency night shift starting Tuesday evening and Officer3 is working a shadow section shift on Tuesday. Since Officer3 has a higher rank than Intern7, Officer3 may very well cover for Intern7. This illustrates how the inclusion of shadow shifts is one way of improving flexibility. It is up to planners to decide if the shadow shift should be realized when the absence is unveiled. We assume that Officer3 finishes the section shift on Tuesday if the shadow shift is realized, but in practise this is up to planners, and likely depends on the time they are notified of Intern7's absence. If they are notified some time in advance, they could choose to give Officer3 Tuesday off, but this is not explicitly implemented as part of the shadow shifts. If planners decide to realize Officer3's shadow shift on Tuesday, Officer3 will not work the section shift on Wednesday.

3.3.2 Problem Description

The Semi-Cyclic Ranked Physician Rostering Problem (SCRPRP) addresses the creation of rosters for surgeons at the Clinic of Surgery. The clinic is divided into sections, and surgeons have a responsibility to cover shifts at these sections and an emergency department in another clinic. Surgeons of different ranks and section affiliations are working at the clinic. There are different types of shifts: section shifts, emergency shifts (day and night), coursework shifts and off-shifts. Each surgeon must be assigned exactly one shift each day.

Surgeons need to meet the total demand for all types of emergency shifts exactly at all times, and also the minimum requirements of surgeons of certain ranks. All emergency shifts must be planned in a weekly cyclic structure. Every section has a defined minimum and preferred level of demand for staff every weekday, depending on its size.

There are regulations stated by the Norwegian Working Environment Act and the local collective agreement regarding working hours and off-periods that must be adhered to.

The quality of a roster is evaluated based on over- and understaffing at the sections, assigned shadow shifts, the allocation of emergency shifts and weekends, and the variations in working time between the surgeons. The goal of the SCRPRP is to create the roster with the highest quality, while satisfying all demand and adhering to all rules and regulations regarding working conditions.

3.4 Solution method

The full SCRPRP is far too complex to find high-quality integer solutions using standard commercial MIP solvers within reasonable time for our real-life instances, see Appendix D. Thus, we develop a two-step matheuristic, presented in Sections 3.4.1 and 3.4.2. Readers may find the symbol directory in Appendix A useful. No model of the full SCRPRP is given explicitly in this section, as it is practical to present the problem as part of our matheuristic. However, a formulation of the full SCRPRP is provided in Appendix B.

The general idea of the two-step matheuristic is to first formulate a simplified version of the SCRPRP, referred to as Step 1, where the main decisions are allocating the cyclic emergency shifts. This includes determining the order in which surgeons of a rank cycle. Furthermore, in Step 2, the remaining SCRPRP is solved, while the emergency shifts are fixed based on the solution found in Step 1.

The allocation of emergency shifts in Step 1 must be done carefully to be able to obtain good solutions in Step 2. The simplification in Step 1 entails only defining the emergency shifts and a generic shift representing all other shifts, referred to as an O-shift. As a result, fewer variables are defined and most restrictions that deal specifically with shift types that are comprised by the generic O-shift are disregarded in Step 1. However, not all restrictions in the full SCRPRP can simply be ignored in Step 1, and provide us with feasible allocation of emergency shifts. Therefore, some restrictions are introduced in Step 1, but then removed in Step 2.

In Step 2, the emergency shifts are fixed, and the remaining SCRPRP is solved. The rostering problem in Step 2 is thus more similar to a classic rostering problem, as the semi-cyclic structure of the roster is accounted for in Step 1. The set of shifts is considerably larger in Step 2, as O-shifts are substituted with section shifts, coursework shifts, and different kinds of off-shifts and shadow shifts. A collection of restrictions specific for these shift types are also introduced in Step 2.

The key ideas and decisions of the two-step matheuristic are illustrated in Figure 3.1, to help visualize the general idea of the matheuristic.

3.4.1 Step 1 - Solving the simplified SCRPRP

We define the set of emergency shifts $\mathcal{S}^E = \{s^{ED}, s^{EN}\}$ where s^{ED} is the emergency day shift and s^{EN} is the emergency night shift. Night shifts begin on the day they are assigned in the model, and end the day after. In Step 1, the set of all shifts $\mathcal{S} = \{s^{ED}, s^{EN}, s^O\}$ also includes the generic O-shift, s^O.

Also defined is the set of surgeons $\mathcal{E}$. Every surgeon has a rank r in the set of ranks $\mathcal{R} = \{r^I, r^R, r^O\}$, representing interns, residents and officers respectively. The set is ordered, such that $r^I < r^R < r^O$. We let $\mathcal{E}_r^R$ denote the set of surgeons with rank r. The set of days $\mathcal{T}$ and the set of weeks $\mathcal{W}$ both constitute the planning horizon, and subsets

Step 1 - Solving the simplified SCRPRP	Step 2 - Solving the remaining SCRPRP

Key ideas

• Create a simplified problem that retains the cyclic structure of the emergency shifts • Ensure sufficient space to facilitate feasibility, stability, and flexibility in Step 2 • Minimize undesirable assignment of emergency shifts	• Fix emergency shifts from Step 1, thus ensuring a semi-cyclic structure of the roster • Solve the remaining SCRPRP as a rostering problem • Maximize stability and flexibility, and minimize undesirable roster characteristics

Key decisions

• Decide the order surgeons cycle in • Assign emergency shifts • Assign space (O-shifts) for section shifts • Assign space (O-shifts) for off-weeks	• Assign section shifts, ensure stability • Assign shadow shifts, ensure flexibility • Assign off-shifts and remaining shift types, minimize undesirable roster characteristics

Figure 3.1: The key ideas and decisions in Step 1 and Step 2 of the matheuristic.

of these are introduced as they become relevant. We structure the presentation of the matheuristic similarly to the key ideas in Figure 3.1.

Cyclic emergency shifts

The required staffing levels D^{ET}_{s} for surgeons must be met exactly for all emergency shifts s on every day. However, there are also minimum requirements for the ranks of the surgeons working the emergency shifts at different times, given by D^{ER}_{rs}. The binary decision variable x_{est} is 1 if surgeon e is assigned shift s on day t, 0 otherwise.

$$\sum_{e \in \mathcal{E}} x_{est} = D^{ET}_{s}, \qquad s \in \mathcal{S}^E, t \in \mathcal{T} \tag{3.1}$$

$$\sum_{r \in \mathcal{R} | r \geq r_2} \sum_{e \in \mathcal{E}^R_r} x_{est} \geq D^{ER}_{r_2 s}, \qquad r_2 \in \mathcal{R}, s \in \mathcal{S}^E, t \in \mathcal{T} \tag{3.2}$$

$$x_{est} \in \{0, 1\}, \qquad e \in \mathcal{E}, s \in \mathcal{S}, t \in \mathcal{T} \tag{3.3}$$

Constraints (3.1) state that the total demand for emergency shifts is met every day. Constraints (3.2) ensure sufficient staffing of surgeons holding the required ranks and constraints (3.3) declare the binary variables.

The combination of cyclicity, constraints (3.1), and constraints (3.2) implies that surgeons of a given rank can cover emergency demand by sharing responsibility for demand within their rank or within a group of multiple ranks. E.g., assume the total demand for emergency nights shifts is two, $D^{ET}_{s^{EN}} = 2$, while the ranked demand for emergency night shifts is two for interns, but one for residents and officers ($D^{ER}_{r^I s^{EN}} = 2$, $D^{ER}_{r^R s^{EN}} = 1$, $D^{ER}_{r^O s^{EN}} = 1$). In this example, it is possible for the officers to share the responsibility of covering one emergency night shift each day, while interns and residents share the responsibility of the other emergency night shift each day.

Because of the surgeons' rank and section affiliation, the cycle order of the surgeons matters and is decided by the model. We call one surgeon of each rank r *first surgeon*, which we denote ϵ_r and then we define the lag of another surgeon to be the number of weeks after which this surgeon repeats the emergency shifts of the first surgeon. It should be noted that fixing a first surgeon rather than making this a decision in the model does not affect the optimal objective value. For any given roster not in conflict with the constraints that ensure the cyclic structure, anyone of the surgeons in a rank could be defined as the first surgeon. Following the same line of reasoning, for any given roster of a problem consisting of multiple ranks, there is a minimum of $\prod_{r \in \mathcal{R}} |\mathcal{E}_r^R|$ feasible combinations of first surgeons and lags. Defining first surgeons for each rank is thus a way to break symmetry in the problem.

We define the set of lags for rank r as $\mathcal{L}_r = \{0, \ldots, |\mathcal{E}_r^R| - 1\}$. The associated variable λ_{el} is 1 if surgeon e lags l weeks and $\lambda_{(\epsilon_r)0} = 1$ denote that the first surgeon lags 0 weeks.

$$\lambda_{el}(x_{(\epsilon_r)st} - x_{es(t+7l)}) = 0, \qquad r \in \mathcal{R}, e \in \mathcal{E}_r^R, s \in \mathcal{S}^E, t \in \mathcal{T}, l \in \mathcal{L}_r \quad |e \neq \epsilon_r, t \leq 7|\mathcal{E}_r^R| \tag{3.4}$$

$$x_{est} - x_{es(t+7|\mathcal{E}_r^R|)} = 0, \qquad r \in \mathcal{R}, e \in \mathcal{E}_r^R, s \in \mathcal{S}^E, t \in \mathcal{T} \quad |t \leq |\mathcal{T}| - 7|\mathcal{E}_r^R| \tag{3.5}$$

$$\sum_{e \in \mathcal{E}_r^R} \lambda_{el} = 1, \qquad r \in \mathcal{R}, l \in \mathcal{L}_r \tag{3.6}$$

$$\sum_{l \in \mathcal{L}_r} \lambda_{el} = 1, \qquad r \in \mathcal{R}, e \in \mathcal{E}_r^R \tag{3.7}$$

$$\lambda_{el} \in \{0, 1\}, \qquad r \in \mathcal{R}, e \in \mathcal{E}_r^R, l \in \mathcal{L}_r \tag{3.8}$$

Constraints (3.4) connect the emergency shifts worked by the first surgeon of any rank with its following surgeons. This implies connecting week 1 of the first surgeon with week 2 of the surgeon with lag 1, week 3 of the surgeon with lag 2, etc. However, for the cyclic structure to hold, we also need the emergency shifts to repeat themselves every time a cycle has finished. This is ensured by constraints (3.5). Unique combinations of surgeons and lags are established by constraints (3.6) - (3.7). As constraints (3.4) are non-linear, we provide a linearization, see constraints (124) and (125) in Appendix C.

Create space for section shifts

In Step 1, the model does not explicitly assign section shifts. Rather, we allocate O-shifts that can potentially become section shifts in Step 2. We introduce the set $\mathcal{T}'$ as all weekdays. Furthermore, we define the sets $\mathcal{A}$ and $\mathcal{A}^L$ representing all sections and all large sections respectively, and set $\mathcal{E}_a^A$ as the set of surgeons affiliated with section a. To track the staffing relative to the section demand, a binary variable δ_{ew} that is 1 if no emergency shift is planned during week w is introduced together with y_{at}^{O1}, y_{at}^{O2} and y_{at}^{U}, which are 1 if section a is overstaffed by one or two surgeons or understaffed by one day t. Finally, the variable u_{et}^{NC} is 1 if surgeon e works a night shift day $t-1$.

$$\sum_{e \in \mathcal{E}_a^A} x_{es^O t} - \sum_{e \in \mathcal{E}_a^A} u_{et}^{NC} - \sum_{e \in \mathcal{E}_a^A} \delta_{ew} - y_{at}^{O1} - y_{at}^{O2} + y_{at}^{U} \geq D_a^A, \qquad a \in \mathcal{A}, t \in \mathcal{T}' \tag{3.9}$$

$$x_{es^O t} + x_{es^{EN}(t-1)} - u_{et}^{NC} \leq 1, \qquad e \in \mathcal{E}, t \in \mathcal{T}' \tag{3.10}$$

$$u_{et}^{NC} \in \{0, 1\}, \qquad e \in \mathcal{E}, t \in \mathcal{T}' \tag{3.11}$$

$$y_{at}^{O1} \in \{0,1\} \qquad a \in \mathcal{A}, t \in \mathcal{T}' \tag{3.12}$$

$$y_{at}^{O2} \in \{0,1\}, \qquad a \in \mathcal{A}^L, t \in \mathcal{T}' \tag{3.13}$$

$$y_{at}^{U} \in \{0,1\}, \qquad a \in \mathcal{A}, t \in \mathcal{T}' \tag{3.14}$$

Constraints (3.9) ensure sufficient surgeons affiliated to each section working O-shifts, D_a^A being the required staffing level at section a. Surgeons working night shifts the day before or having off-weeks are subtracted since these O-shifts most likely cannot be turned into section shifts. Over- and understaffing is handled, a high penalty for understaffing and decreasing marginal utility for overstaffing guaranty correct values. In Step 2, most of the O-shifts will be specified to section shifts, thus ensuring sufficient staffing levels in the final roster. The variable u_{et}^{NC} is set to 1 when surgeon e works a night shift the day before an O-shift in constraints (3.10). As Constraints (3.11) - (3.14) declare the variables as binary, we have laid the groundwork for good staffing levels at the sections in Step 2.

Create space for off-weeks

Every surgeon should have an have a full week off every time the surgeon has worked for two full emergency shift cycles (e.g. 16 weeks for a person in a rank of 8 surgeons). We define a binary variable δ_{ew} that is 1 if no emergency shifts are allocated to surgeon e week w.

$$\sum_{t \in \mathcal{T}_w^W} x_{es^O t} \geq 7\delta_{ew}, \qquad r \in \mathcal{R}, e \in \mathcal{E}_r^R, w \in \mathcal{W} \tag{3.15}$$

$$\sum_{e \in \mathcal{E}_r^R} (\delta_{e(w-1)} + \delta_{ew}) = 1, \qquad r \in \mathcal{R}, w \in \mathcal{W} \quad |mod(w,2) = 0 \tag{3.16}$$

$$\delta_{e(w-1)} + \delta_{ew} = \delta_{e(w-1+2|\mathcal{E}_r^R|)} + \delta_{e(w+2|\mathcal{E}_r^R|)}, \quad r \in \mathcal{R}, e \in \mathcal{E}_r^R, w \in \mathcal{W} \quad |mod(w,2) = 0 \tag{3.17}$$

$$\delta_{ew} \in \{0,1\}, \qquad e \in \mathcal{E}, w \in \mathcal{W} \tag{3.18}$$

Constraints (3.15) force surgeon e to only have O-shifts when $\delta_{ew} = 1$ for week w. Constraints (3.16) make sure that during any pair of weeks (1 and 2, 3 and 4, etc.) there is exactly one surgeon of each rank that has a week without any emergency shifts. By letting the model choose one week without emergency shifts from a pair of weeks, we allow some flexibility in deciding when the surgeons will be unavailable for section work in Step 2. Constraints (3.17) ensure the pair of potential off-weeks is repeated every time the surgeons of rank r have rotated fully twice, i.e. every $2|\mathcal{E}_r^R|$ weeks. This means off-weeks are awarded regularly to all surgeons throughout the planning period.

Regulations and norms

Many norms depend on the different days of the week. We therefore note that any day $t \in \mathcal{T}$ where $mod(t,7) = 1$ is a Monday, $t \in \mathcal{T}$ where $mod(t,7) = 2$ is a Tuesday, etc. We also define the sets $\mathcal{T}^{Sat}$ and $\mathcal{T}^{Sun}$ to include all Saturdays and Sundays in the planning period, respectively.

$$\sum_{s\in\mathcal{S}} x_{est} = 1, \qquad e\in\mathcal{E}, t\in\mathcal{T} \quad (3.19)$$

$$x_{es^{EN}(t-1)} + x_{es^{ED}t} \leq 1, \qquad e\in\mathcal{E}, t\in\mathcal{T} \quad |t>1 \quad (3.20)$$

$$x_{es^{EN}(t-2)} + x_{es^{O}(t-1)} + x_{es^{EN}t} \leq 2, \qquad e\in\mathcal{E}, t\in\mathcal{T} \quad |t>2 \quad (3.21)$$

$$\sum_{\tau=t-\overline{P}^N}^{t} x_{es^{EN}\tau} - z^{CN}_{et} \leq \overline{P}^N, \qquad r\in\mathcal{R}, e\in\mathcal{E}^R_r, t\in\mathcal{T} \quad |t\notin\mathcal{T}^{Sat}, t>\overline{P}^N \quad (3.22)$$

$$\sum_{\tau=t-\overline{P}^N-1}^{t} x_{es^{EN}\tau} \leq \overline{P}^N + 1, \qquad e\in\mathcal{E}, t\in\mathcal{T} \quad |t\notin\mathcal{T}^{Sat}, t>\overline{P}^N+1 \quad (3.23)$$

$$x_{es^{ED}t} - x_{es^{ED}(t-1)} = 0, \qquad e\in\mathcal{E}, t\in\mathcal{T}^{Sun} \quad (3.24)$$

$$x_{es^{EN}t} - x_{es^{EN}(t-1)} = 0, \qquad e\in\mathcal{E}, t\in\mathcal{T}^{Sat} \quad (3.25)$$

$$\sum_{s\in\mathcal{S}^E}\sum_{\tau=1}^{\overline{P}^{CW}+1} x_{es(t-7(\tau-1))} - z^{CW}_{et} \leq \overline{P}^{CW}, \qquad e\in\mathcal{E}, t\in\mathcal{T}^{Sat} \quad |t>7\overline{P}^{CW} \quad (3.26)$$

$$\sum_{t\in\mathcal{T}^W_w} x_{es^{ED}t} - z^{CE}_{rt} \leq P^{EDS}, \qquad r\in\mathcal{R}, e\in\mathcal{E}^R_r, w\in\mathcal{W} \quad |r\neq r^O \quad (3.27)$$

$$z^{CN}_{et} \in \{0,1\}, \qquad r\in\mathcal{R}, e\in\mathcal{E}^R_r, t\in\mathcal{T} \quad |r\neq r^O \quad (3.28)$$

$$z^{CW}_{et} \in \{0,1\}, \qquad e\in\mathcal{E}, t\in\mathcal{T}^{Sat} \quad (3.29)$$

$$z^{CE}_{rt} \in \{0,1\}, \qquad r\in\mathcal{R}, t\in\mathcal{T} \quad |r\neq r^O \quad (3.30)$$

Constraints (3.19) ensure that each surgeon e should be allocated exactly one shift s per day t. Constraints (3.20) establish sufficient rest between two shifts, by stating that a surgeon e cannot work a night shift followed by a day shift, and constraints (3.21) ensure that no surgeon works the very disfavoured work pattern of a night shift on day $t-2$, an O-shift on day $t-1$ and a night shift again on day t. Constraints (3.22) penalize employees e of rank r working more than $\overline{P}^N$ consecutive night shifts through assigning binary variable $z^{CN}_{et} = 1$, while constraints (3.23) prevent the pattern associated with such a relaxation on two consecutive days for the same surgeon. The constraints are not active when ending on Saturdays, as surgeons prefer combining working the two nights beginning Friday and Saturday. As seen from Constraints (3.28), z^{CN}_{et} is not defined for surgeons of rank r^O. The relaxation in constraints (3.22) is thus impossible in cases involving surgeons holding the rank officer. Constraints (3.24) and (3.25) ensure that every surgeon works the same shifts during the weekend (night shifts begin on the day they are assigned). Emergency shifts are the only work shifts that can occur during weekends, which implies that an O-shift during the weekend will be an off-day or a shadow shift in Step 2. Constraints (3.26) establish that no surgeon works $\overline{P}^{CW}$ consecutive weekends without penalizing the objective function through binary variable $z^{CW}_{et} = 1$ for surgeon e on Saturday t. The set

$\mathcal{T}_w^W$ includes all days during week w. To avoid congestion of long emergency day shifts for a surgeon during a single week, constraints (3.27) impose soft constraints for interns and residents working more than P^{EDS} emergency day shifts per week. This is penalized through variables z_{rt}^{CE}, which is only defined for interns and residents, as they have long emergency day shifts. Constraints (3.28) - (3.30) declare several key variables.

$$\sum_{s \in \mathcal{S}^E} \sum_{t \in \mathcal{T}_w^W | t \notin \mathcal{T}^{Sat}} x_{est} \leq \overline{P}^{WE}, \qquad e \in \mathcal{E}, w \in \mathcal{W} \tag{3.31}$$

$$\sum_{s \in \mathcal{S}^E} \sum_{t \in \mathcal{T}} x_{est} - \overline{v}_r^E \leq 0, \qquad r \in \mathcal{R}, e \in \mathcal{E}_r^R \tag{3.32}$$

$$\sum_{s \in \mathcal{S}^E} \sum_{t \in \mathcal{T}} x_{est} - \underline{v}_r^E \geq 0, \qquad r \in \mathcal{R}, e \in \mathcal{E}_r^R \tag{3.33}$$

$$\overline{v}_r^E - \underline{v}_r^E - v_r^E \leq \overline{P}^{NES}, \qquad r \in \mathcal{R} \tag{3.34}$$

$$\overline{v}_r^E, \underline{v}_r^E, v_r^E \geq 0, \qquad r \in \mathcal{R} \tag{3.35}$$

Constraints (3.31) limit the number of emergency shifts each week (excluding Saturdays) to a maximum of $\overline{P}^{WE}$, thus spreading out unpopular emergency shifts. Constraints (3.32), (3.33), and (3.34) ensure that if the difference in number of emergency shifts worked by any two employees of rank r is larger than $\overline{P}^{NES}$, it is penalized through variables v_r^E. The introduction of v_r^E in constraints (3.34) and having a positive right hand side 1 is relevant if the planning period ends without finalizing all emergency shift cycles. E.g. if we plan for 26 weeks with one group of surgeons of 10 employees of the same rank, they have only finished 6 of the last 10 weeks in their cycle, meaning they may not have worked an identical number of emergency shifts.

Symmetry breaking constraints

We introduce constraints to reduce symmetry in the problem.

$$\sum_{l_2=1}^{l_1} \lambda_{e_1 l_2} - \sum_{l_2=1}^{l_1} \lambda_{e_2 l_2} \leq 0,$$
$$a \in \mathcal{A}, r \in \mathcal{R}, e_1, e_2 \in (\mathcal{E}_a^A \cap \mathcal{E}_r^R), t \in \mathcal{T}, l_1 \in \mathcal{L}_r \quad | e_1 \neq e_2, e_1 \neq \epsilon_r, e_2 \neq \epsilon_r \tag{3.36}$$

For homogeneous surgeons, the schedules could be swapped between them for the entire planning period without affecting the quality of the roster. To avoid this symmetry, constraints (3.36) reduce the solution space by sorting homogeneous surgeons by lag indices. Preliminary testing demonstrates that this reduces the computation time for Step 1.

Objective function

The objective function is extensive due to the many different considerations to handle when creating high-quality rosters at the Clinic of Surgery. All variables in the objective function are given different base letters dependent on their role in the formulation. Variables related to the quality of staffing levels at the sections are given base letter y, while those related to consecutive shifts or congestion of long shifts have base letter z. The base letter v implies that the variable affects how shifts are divided between surgeons. All parameters in the objective function are given the base letter W, representing the weight given to each decision variable.

$$Max\ Z = \sum_{a \in \mathcal{A}} \sum_{t \in \mathcal{T}'} W^{O1} y^{O1}_{at} + \sum_{a \in \mathcal{A}^L} \sum_{t \in \mathcal{T}'} W^{O2} y^{O2}_{at} - \sum_{a \in \mathcal{A}} \sum_{t \in \mathcal{T}'} W^{U} y^{U}_{at}$$

$$- \sum_{e \in \mathcal{E}} \sum_{t \in \mathcal{T}} W^{CN} z^{CN}_{et} - \sum_{r \in \mathcal{R}} \sum_{t \in \mathcal{T}} W^{CE} z^{CE}_{rt} - \sum_{e \in \mathcal{E}} \sum_{t \in \mathcal{T}^{Sat}} W^{CW} z^{CW}_{et} - \sum_{r \in \mathcal{R}} W^{E} v^{E}_{r} \quad (3.37)$$

The objective function terms including variables with base letter y reward overstaffing and penalize understaffing. Objective function terms with base letter z penalize working too many consecutive nights, working more than a given number of emergency day shifts in a single week and working too many consecutive weekends. The objective function term with base letter v penalizes variations in the number of emergency shifts between surgeons of the same rank. To provide insights to the relative weights of the objective function terms, see Table 13 in Appendix A.4.

3.4.2 Step 2 - Solving the full SCRPRP

In Step 1 we assigned the cyclic emergency shifts. In Step 2, the matheuristic must assign specific shifts where O-shifts were assigned in Step 1, to create full rosters.

Cyclic emergency shifts

Assume the parameter X^{S1}_{est} is the value of the x-variables of the emergency shifts for the best solution found in Step 1, then we can add the following constraints.

$$x_{est} = X^{S1}_{est}, \qquad e \in \mathcal{E}, s \in \mathcal{S}^E, t \in \mathcal{T} \quad (3.38)$$

Constraints (3.38) fix all emergency shifts from the solution obtained in Step 1, thus fixing significant parts of the roster. This means that solving Step 2 is a more classic nurse rostering problem.

The O-shift is now replaced by off-shifts, section shifts and coursework shifts, as well as shadow shifts. This means that the set of all shifts, $\mathcal{S}$, now includes all shifts defined in Step 2, and several Constraints in Step 1 remain unchanged in Step 2 if we disregard this change. There are two types of off-shifts, s^{NOff}, is the normal off-shift, while the protected off-shift, s^{POff} implies a minimum number of rest hours. We define $\mathcal{S}^{Off} = \{s^{NOff}, s^{POff}\}$. Finally, we have the section shift, s^{Sec}, and the coursework shift s^{CW}.

Shadow shifts are defined to have one realized and one unrealized mode. The set of shadow shifts $\mathcal{S}^{SS}$ consists of s^{SO}, s^{SCW} and s^{SSec} representing shadow shifts that are off-shifts if unrealized, coursework shifts if unrealized, and section shifts if unrealized respectively, i.e. $\mathcal{S}^{SS} = \{s^{SO}, s^{SCW}, s^{SSec}\}$.

If realized, a shadow shift becomes an emergency night shift. In practise, this could happen on short notice. It is therefore important that rosters are created to maintain high quality if the surgeon working the shadow shift becomes unavailable for his or her section shift the following day. Because of the potentially short notice, we assume that surgeons working shadow section shifts on a given day will work the entire section shift and then continue to work emergency night shifts starting later that same day. Therefore, we do not accept assigning any kind of day shift the day after working a shadow shift, thus ensuring surgeons get rest after their shadow shift is potentially realized. We define the set $\mathcal{S}^N$ as all shifts that can entail night work, which includes emergency night shifts and shadow shifts, i.e. $\mathcal{S}^N = \{s^{EN}\} \bigcup \mathcal{S}^{SS}$. It should be noted that the inclusion of shadow shifts does not affect the feasibility of the full SCRPRP, as it is possible not to realize any of the shadow shifts.

Constraints (3.1), (3.2), (3.4) - (3.8), and (3.36) are redundant in Step 2, as they must hold due to the fixation in Constraints (3.38). Constraints (3.3) are still active, and set definitions are updated as explained in the previous paragraphs.

Assign section shifts

When assigning section shifts we introduce the binary variables u_{et}^{SecA}, which is 1 if surgeon e works a section shift on day t, the day after a shadow shift. This implies that although the surgeon may work at the section that day, we have no guarantee that the surgeon is available for work at the section when the day arrives. To keep track of the shadow shifts, we also introduce y_{rt}^{SS} being 1 if a shadow shift is assigned to surgeons of rank r on day t.

$$\sum_{e \in \mathcal{E}_a^A} x_{es^{Sec}t} - y_{at}^{O1} - y_{at}^{O2} + y_{at}^{U} - u_{et}^{SecA} \geq D_a^A, \qquad a \in \mathcal{A}, t \in \mathcal{T}' \tag{3.39}$$

$$\sum_{s \in \mathcal{S}^{SS}} x_{es(t-1)} + x_{es^{Sec}t} - u_{et}^{SecA} \leq 1, \qquad e \in \mathcal{E}, t \in \mathcal{T}' \tag{3.40}$$

$$\sum_{e \in \mathcal{E}_r^R} \sum_{s \in \mathcal{S}^{SS}} x_{est} - y_{rt}^{SS} \geq 0, \qquad r \in \mathcal{R}, t \in \mathcal{T} \tag{3.41}$$

$$\sum_{r \in \mathcal{R}} y_{rt}^{SS} \leq \overline{P}^{SR}, \qquad t \in \mathcal{T} \tag{3.42}$$

$$\sum_{t \in \mathcal{T}} (x_{es^{CW}t} + x_{es^{SCW}t}) = |\mathcal{W}|/2, \qquad e \in \mathcal{E} \tag{3.43}$$

$$u_{et}^{SecA} \in \{0,1\}, \qquad e \in \mathcal{E}, t \in \mathcal{T}' \tag{3.44}$$

$$y_{rt}^{SS} \in \{0,1\}, \qquad r \in \mathcal{R}, t \in \mathcal{T}' \tag{3.45}$$

Constraints (3.9) are substituted for Constraints (3.39). Constraints (3.40) connect u_{et}^{SecA} with the shift variables and constraints (3.41) handle the shadow shifts. Constraints (3.42) limit the number of shadow shifts each day to $\overline{P}^{SR}$. Every surgeon should do coursework on average every second week, as enforced by constraints (3.43). It should be noted, that if a surgeon is assigned to a shadow coursework shift and this is realized, constraints (3.43) are not complied with when the roster is executed. This is considered acceptable, however, and is a practical issue that is typically resolved easily at the Clinic of Surgery.

Constraints (3.10) and (3.11), as well as variables u_{et}^{NC} are not part of Step 2. The declaration of y-variables in Constraints (3.12) - (3.14) remain active in Step 2.

Assign off-weeks

Ensuring a full week off for all employees, as stated in constraints (3.15) - (3.18) does not need much adjustment.

$$\sum_{s \in \mathcal{S}^{Off}} \sum_{t \in \mathcal{T}_w^W} x_{est} \geq 7\delta_{ew}, \qquad r \in \mathcal{R}, e \in \mathcal{E}_r^R, w \in \mathcal{W} \tag{3.46}$$

Constraints (3.46) allow for δ_{ew} to take value 1 if the shifts assigned are off-shifts, rather than any choice of O-shift, as presented in Constraints (3.15). Constraints (3.16) - (3.18) remain active in Step 2.

Regulations and norms

The constraints handling regulations and norms are updated to reflect the assignment of specific shifts and not the generic O-shifts as in Step 1.

$$x_{es^{EN}(t-1)} - \sum_{s_2 \in \{s^{EN}, s^{SO}\} \bigcup \mathcal{S}^{Off}} x_{es_2t} \leq 0, \qquad e \in \mathcal{E}, t \in \mathcal{T} \tag{3.47}$$

$$x_{es_1(t-2)} + \sum_{s_2 \in \mathcal{S}^{Off}} x_{es_2(t-1)} + x_{es_1t} \leq 2, \qquad e \in \mathcal{E}, s_1 \in \mathcal{S}^N, t \in \mathcal{T} \tag{3.48}$$

$$x_{es_1(t-1)} + x_{es^{POff}t} \leq 1, \qquad e \in \mathcal{E}, s_1 \in \mathcal{S}^N, t \in \mathcal{T} \tag{3.49}$$

$$\sum_{t \in \mathcal{T}_w^W} x_{es^{POff}t} = 1, \qquad e \in \mathcal{E}, w \in \mathcal{W} \tag{3.50}$$

$$\sum_{s \in \mathcal{S}^N} \sum_{\tau = t - \overline{P}^{NS}}^{t} x_{es\tau} \leq \overline{P}^{NS}, \qquad r \in \mathcal{R}, e \in \mathcal{E}_r^R, t \in \mathcal{T} | t \notin \mathcal{T}^{Sat}, t > \overline{P}^{NS} \tag{3.51}$$

$$x_{est} - x_{es(t-1)} = 0, \qquad e \in \mathcal{E}, s \in \mathcal{S}^{SS}, t \in \mathcal{T}^{Sat} \tag{3.52}$$

$$\sum_{s \in \mathcal{S}^E \bigcup \{s^{SO}\}} \sum_{\tau=1}^{\overline{P}^{CW}+1} x_{es(t-7(\tau-1))} - z_{et}^{CW} \leq \overline{P}^{CW}, \qquad e \in \mathcal{E}, t \in \mathcal{T}^{Sat} \tag{3.53}$$

$$\sum_{s \in \mathcal{S}^{Off}} \sum_{\tau = t - P^{CD}}^{t} x_{est} \geq 1 - z_{et}^{CD}, \qquad e \in \mathcal{E}, t \in \mathcal{T} \quad | t > P^{CD} \tag{3.54}$$

$$\sum_{s_1 \in \mathcal{S}^{SS}} x_{es_1(t-1)} + x_{es^{ED}t} \leq 1, \qquad e \in \mathcal{E}, t \in \mathcal{T} \tag{3.55}$$

$$z_{et}^{CD} \in \{0, 1\}, \qquad e \in \mathcal{E}, t \in \mathcal{T} \tag{3.56}$$

Constraints (3.47) replace constraints (3.20) to ensure an emergency night shift is only succeeded by another emergency night shift or an off-shift. This includes shadow shifts that are off-days if unrealized. Constraints (3.48) replace constraints (3.21) to prevent off-shifts from being placed between night shifts or shadow shifts, as this is deemed an unreasonable work pattern by surgeons. Constraints (3.49) establish that protected off-days are not allocated the day after a night shift, as the minimum rest time rule would be violated. Furthermore, constraints (3.50) enforce weekly protected off-days. A maximum number of night shifts $\overline{P}^{NS}$ including shadow shifts is given in constraints (3.51), while constraints (3.52) ensure that surgeons working shadow shifts during the weekend do so both days. Constraints (3.51) and (3.52) are not conflicting as $\overline{P}^{NS} > 1$. Constraints (3.53) is updated from constraints (3.26), and now include shadow shifts when penalizing too many consecutive work weekends. Constraints (3.54) allocate a penalty for working P^{CD} consecutive days.

There is an important difference in the flexibility of shift patterns including emergency night shifts and shadow shifts. Shadow shifts can very well be succeeded by section shifts or coursework shifts. This presupposes that the surgeon working the shadow shift is not integral to the staffing at sections the day after in case the shadow shift is realized. However, shadow shifts cannot be followed by an emergency day shift. This is ensured by constraints (3.55).

Further, constraints dealing with regulations and norms in Step 1 are handled as follows. Constraints (3.19) remain active with new set definitions in Step 2. Constraints (3.22), (3.23), (3.24), (3.25), (3.27), (3.28), (3.29), and (3.30) remain active in Step 2. The same is true for constraints (3.31) - (3.35).

Work time constraints

When specifying the O-shift, weekly and total work time is relevant.

$$\sum_{s\in\mathcal{S}}\sum_{t\in\mathcal{T}^W_w} P^H_{rst}x_{est} \le \overline{H}^W, \qquad r\in\mathcal{R}, e\in\mathcal{E}^R_r, w\in\mathcal{W} \tag{3.57}$$

$$\sum_{s\in\mathcal{S}}\sum_{t\in\mathcal{T}} P^H_{rst}x_{est} - |\mathcal{W}|\overline{v} \le |\mathcal{W}|P^{TH}, \qquad r\in\mathcal{R}, e\in\mathcal{E}^R_r \tag{3.58}$$

$$\sum_{s\in\mathcal{S}}\sum_{t\in\mathcal{T}} P^H_{rst}x_{est} + |\mathcal{W}|\underline{v} \ge |\mathcal{W}|P^{TH}, \qquad r\in\mathcal{R}, e\in\mathcal{E}^R_r \tag{3.59}$$

$$\sum_{s\in\mathcal{S}}\sum_{t\in\mathcal{T}} P^H_{rst}x_{est} - \overline{v}^{RH}_r \le 0, \qquad r\in\mathcal{R}, e\in\mathcal{E}^R_r \tag{3.60}$$

$$\sum_{s\in\mathcal{S}}\sum_{t\in\mathcal{T}} P^H_{rst}x_{est} - \underline{v}^{RH}_r \ge 0, \qquad r\in\mathcal{R}, e\in\mathcal{E}^R_r \tag{3.61}$$

$$\overline{v}^{RH}_r - \underline{v}^{RH}_r - v^{RH}_r \le 0, \qquad r\in\mathcal{R} \tag{3.62}$$

$$\overline{v}^{RH}_{r_1} - \overline{v}^{RH}_{r_2} - v^{R_2H}_{r_1} \le 0, \qquad r_1, r_2\in\mathcal{R}|r_1 < r_2 \tag{3.63}$$

$$\overline{v}, \underline{v} \ge 0 \tag{3.64}$$

$$\overline{v}^{RH}_r, \underline{v}^{RH}_r, v^{RH}_r, v^{R_2H}_r \ge 0, \qquad r\in\mathcal{R} \tag{3.65}$$

Constraints (3.57) limit the weekly maximum work hours to $\overline{H}^W$ when disregarding any realized shadow shifts in compliance with regulations. Variables $\overline{v}$ and $\underline{v}$ represent the maximum positive and negative deviations from the ideal number of weekly hours P^{TH} worked on average throughout the planning horizon by any surgeon. This is ensured by constraints (3.58), that represent budgetary restrictions, and constraints (3.59), that establish that surgeons get necessary experience and match normative salary levels. The deviation is later penalized in the objective function. As a consequence, work hours should be somewhat evenly divided between surgeons of the same rank. Constraints (3.60), (3.61), and (3.62) ensure that the difference in total work hours for any two employees of rank r is penalized through variable v_r^{RH}. Furthermore, it is deemed unfair for surgeons of a lower rank to be given significantly more work than surgeons of higher ranks. As constraints (3.62) already establish a quite even distribution of hours between surgeons of the same rank, it is sufficient to ensure that the most working surgeon of higher ranks are working at least as much as the most working surgeons of lower ranks. This is enforced in constraints (3.63), and penalized by variables $v_r^{R_2H}$.

Objective function

The objective function in Step 2 is the total objective function with all aspects included.

$$\begin{aligned}
Max\ Z = & \sum_{a\in\mathcal{A}}\sum_{t\in\mathcal{T}'} W^{O1} y_{at}^{O1} + \sum_{a\in\mathcal{A}}\sum_{t\in\mathcal{T}'} W^{O2} y_{at}^{O2} - \sum_{a\in\mathcal{A}}\sum_{t\in\mathcal{T}'} W^{U} y_{at}^{U} + \sum_{r\in\mathcal{R}}\sum_{t\in\mathcal{T}} W_r^{SS} y_{rt}^{SS} \\
& - \sum_{e\in\mathcal{E}}\sum_{t\in\mathcal{T}} W^{CN} z_{et}^{CN} - \sum_{r\in\mathcal{R}}\sum_{t\in\mathcal{T}} W^{CE} z_{rt}^{CE} - \sum_{e\in\mathcal{E}}\sum_{t\in\mathcal{T}^{Sat}} W^{CW} z_{ew}^{CW} - \sum_{e\in\mathcal{E}}\sum_{t\in\mathcal{T}} W^{CD} z_{et}^{CD} \\
& - \sum_{r\in\mathcal{R}} W^{E} v_r^{E} - W(\overline{v} + \underline{v}) - \sum_{r\in\mathcal{R}} W^{RH} v_r^{RH} - \sum_{r\in\mathcal{R}|r\neq r^O} W^{R_2H} v_r^{R_2H}
\end{aligned} \tag{3.66}$$

The objective function terms with base letter y are unchanged, with the addition of a term rewarding the use of shadow shifts. Objective function terms with base letter z are identical to those in Step 1, with the addition of penalizing too many consecutive work shifts using variables z_{et}^{CD}. There are multiple additions to the objective function in Step 1, in terms of variables with base letter v. The objective function in Step 2 also penalizes deviation from the total number of appropriate work hours, differences in work hours between surgeons of the same rank and surgeons of higher ranks working less than those of lower ranks. To provide insights to the relative weights of the objective function terms, see Table 13 in Appendix A.4.

3.5 Computational study

The matheuristic is implemented in Mosel and the models are solved using Xpress-Optimizer Version 8.8.1. The software is run on Lenovo NextScale nx360 M5 computers with the specifications below.

CPU:	2x 3.4GHz Intel E5-2643v3 – 6 core
RAM:	512Gb
Disk:	120Gb SATA SSD

During preliminary testing the full SCRPRP was run for up to 10 hours using Xpress-Optimizer Version 8.8.1 on standard settings. These results can be found in Appendix D.

When running our two-step matheuristic, each step was run for one hour. The automatic cut generation of the Xpress-Optimizer was turned off in Step 1, as preliminary testing showed that the cuts generated did not improve the dual bound significantly.

The computational study is divided into two parts. In the first part, shadow shifts are not included. Since the current manual planning at the Clinic of Surgery does not support the introduction of shadow shifts, this gives a fair comparison between the manual planning and the matheuristic under the same assumptions. The second part compares solutions from the matheuristic with and without shadow shifts to evaluate the potential of including them in the planning. We also present an excerpt of a roster produced by the matheuristic in Appendix F, which can provide some intuition to what a solution to the SCRPRP looks like.

3.5.1 Instances and parameters

Some parameter values are provided by rules and regulations, while others are decided in cooperation with a group of planners, the Head of Department of Physicians and the Head of the Clinic of Surgery. Preliminary test results have been presented to them to facilitate discussion, leading to an iterative process of testing different parameter values.

The objective function weights are examples of values that have been discussed extensively. When discussing over- and understaffing, for example, the relative magnitude of weights provide insights to their relative priority. However, when discussing the weight of objective function terms that penalize the relaxation of constraints, we have had better experiences comparing excerpts of rosters produced by the model with suggested parameter values. All parameters values used in the computational study are presented in Appendix A.4.

Real-life data in the form of five RSA-matrices from Spring 2017 to Spring 2019 constitute our five instances, see Appendix A.5. The instances are named 'S' or 'A' for spring and autumn followed by the year, e.g. S18 means Spring 2018. When discussing these instances, they are referred to as e.g. the S18 instance.

As discussed in Section 3.3, rosters are subject to disruptions during the online operational planning. Thus, the real-life rosters best suited for comparison are the original ones, produced before the period it is planned for. In practise, planners edit the rosters as disruptions occur without saving the original rosters, and these were impossible to retrieve. The last 16 weeks of the real-life roster for Spring 2019 was still unedited, and is used as basis of comparison.

In the case study, the total demand for emergency night shifts is two. While one surgeon must hold the rank officer, the other can hold any rank. This means that interns and residents can share the responsibility for covering one emergency night shift, as exemplified in Section 3.4. When running the matheuristic, this is simplified by defining lag numbers for the combined group of interns and residents sharing the responsibility for one nightly emergency shift.

3.5.2 Creating rosters using the matheuristic

In this section, we run all five instances without shadow shifts. This is interesting as no shadow shifts are currently used in real-life planning, meaning we get a chance to see how our matheuristic performs under the same assumptions.

In Table 3.6, we present results for each instance. For both steps, we present the objective value of the best feasible solution found, Z, the best dual bound, BB, and the objective value of the linear relaxation, LP-bound. We also present the value of each part of the objective function, here denoted by the respective variables. For all variables in the objective function, we present values of both their weighted (W) and non-weighted (NW) sums (e.g. both $\sum_{a\in\mathcal{A}}\sum_{t\in\mathcal{T}'} W^{O1}y_{at}^{O1}$ and $\sum_{a\in\mathcal{A}}\sum_{t\in\mathcal{T}'} y_{at}^{O1}$).

Table 3.6: Results after running Step 1 and Step 2 of the five instances without defining shadow shifts. For each instance we present weighted (W) and non-weighted (NW) sums of variables in the objective function. The second column indicates if a positive value of a variable has a positive (+) or negative (-) contribution to the objective function value.

Instance	+/−	S17		A17		S18		A18		S19	
		W	NW	W	NW	W	NW	W	NW	W	NW
Step 1											
Z		1390.0		1388.0		1328.0		1382.0		1321.0	
BB		1430.0		1430.0		1430.0		1430.0		1430.0	
LP-bound		1430.0		1430.0		1430.0		1430.0		1430.0	
Step 2											
Z		1208.3		1216.4		1109.7		1138.7		1055.8	
BB		1247.7		1243.3		1137.0		1172.4		1084.1	
LP-bound		1251.6		1250.0		1139.1		1173.4		1091.0	
y_{at}^{O1}	+	1016.0	508.000	1014.0	507.000	1024.0	512.000	1018.0	509.000	996.0	498.000
y_{at}^{O2}	+	226.0	226.000	232.0	232.000	172.0	172.000	182.0	182.000	131.0	131.000
y_{at}^{U}	−	0.0	0.000	0.0	0.000	0.0	0.000	0.0	0.000	0.0	0.000
z_{et}^{CN}	−	0.0	0.000	0.0	0.000	0.0	0.000	0.0	0.000	0.0	0.000
z_{rt}^{CE}	−	0.0	0.000	0.0	0.000	0.0	0.000	0.0	0.000	0.0	0.000
z_{ew}^{CW}	−	0.0	0.000	0.0	0.000	0.0	0.000	0.0	0.000	0.0	0.000
z_{et}^{CD}	−	−5.0	1.000	−5.0	1.000	0.0	0.000	0.0	0.000	−5.0	1.000
v_r^E	−	0.0	0.000	0.0	0.000	−60.0	3.000	−20.0	1.000	−40.0	2.000
$\overline{v}$	−	−11.5	0.115	−12.5	0.125	−14.4	0.144	−28.8	0.288	−10.6	0.106
$\underline{v}$	−	−8.7	0.087	−6.7	0.067	−2.9	0.029	0.0	0.000	−6.7	0.067
v_r^{RH}	−	−8.5	0.423	−5.4	0.269	−9.0	0.452	−12.5	0.625	−8.8	0.442
$v_r^{R_2H}$	−	0.0	0.000	0.0	0.000	0.0	0.000	0.0	0.000	0.0	0.000

The LP-bound in Step 1 is identical for all five instances and the BB is equal to them in all cases. This implies that the problem is highly symmetric. Despite identical LP-bounds and BBs, there is considerable difference in the Z values for the five instances in Step 1, and this is transmitted to Step 2. As can be expected, Z in the first step solutions in Table 3.6 seem to affect the LP-bound values in Step 2, which again relate to the final objective function values found in Step 2. The instances are not solved to optimality, as seen from the differences between the Z and BB-values for Step 2 in Table 3.6.

The quality of staffing levels seems to be very good for all instances. Firstly, there are no cases of understaffing (y_{at}^{U}). Furthermore, the planning period of 26 weeks with four sections to cover five times per week entails a maximum of 520 cases of single-surgeon overcoverage (y_{at}^{O1}). We are able to reach between 498 and 512 in all instances, thus ensuring overcoverage on almost all sections on all days. Similarly, there are three large sections that can be overstaffed with two surgeons (y_{at}^{O1}), entailing a maximum of 390 double overcoverages (note that overcoverage by two surgeons implies overcoverage by one surgeon as well). We are able to reach between 131 and 232 in all instances, as presented in Table 3.6.

There are no unwanted consecutive shifts (z_{et}^{CN}, z_{rt}^{CE}) nor congestion of emergency day shifts (z_{ew}^{CW}). However, in three instances we get an occurrence of seven consecutive workdays (z_{et}^{CD}).

The fairness regarding emergency shifts (v_r^E) and deviations from the ideal number of weekly hours ($\overline{v}$, $\underline{v}$) is high. The largest difference in number of emergency shifts over the planning period is three and the largest deviations from the ideal number of weekly hours are around 17 minutes too much and around 5 minutes too little. This amounts to 7.75 and 2.3 hours for the full planning horizon, respectively. As the shortest shifts are 7.75 hours, this is considered acceptable. There are differences in working hours among surgeons with the same rank (v_r^{RH}), around 37 minutes per week at most, but higher ranked surgeons are working more than lower ranked ($v_r^{R_2H}$).

When comparing the results from the matheuristic with the real-life case, a precondition is that the last 16 weeks of the real-life roster for Spring 2019 is representative of a typical roster produced at the Clinic of Surgery, see Table 3.7. We believe this precondition is supported since all emergency shift cycles have finished at least once, including the combined cycle of shared emergency night shifts for interns and residents. We present the values from the real-life roster from Spring 2019, Real-life, and the corresponding values from the matheuristic run on the same RSA-matrix, S19. We also include the average values from Spring 2017 to Autumn 2018, Avg. S17-A18. To compare the quality of our produced rosters with that of the real-life roster, we divide all objective function values with the number of weeks it has been planned for to get comparable results in Table 3.7.

In the real-life roster there are 3 cases of section understaffing so severe that they could not be compensated for by the y_{at}^U variable (0 surgeons at a large section), implying that the real-life roster is infeasible in our formulation of the SCRPRP. Results in Table 3.7 are calculated as if there was one more surgeon working at the relevant section on the days of infeasibility. Results of the comparison are presented in Table 3.7.

The S19 instance has a somewhat lower objective function value than the Avg. S17-A18 instance, implying the RSA-matrix entails a somewhat more challenging problem than in the other instances. The roster produced in real-life is best compared with the instance S19, that has the same RSA-matrix.

It is clear that the matheuristic outperforms the roster created in real-life for S19. As we have no real-life roster available for earlier planning periods, we cannot directly compare the performance of our matheuristic for the Avg. S17-A18 instance. We do, however, note that staffing levels are evenly high for these instances and that understaffing does not occur. As understaffing and fluctuations in staffing levels have been major issues at the Clinic of Surgery previously, this indicates that our matheuristic performs well for the Avg. S17-A18 instance.

The largest contribution to the difference between the results of the matheuristic and the real-life results in Table 3.7 is from understaffing. Even when disregarding the infeasibility in the real-life case, $y_{at}^U = 1.25$ indicates that there is understaffing more than once per week in one of the sections. In sharp contrast, there is no understaffing for any roster produced by the matheuristic. Another key difference is the relatively unfair schedules found in the real-life roster.

To provide a clearer overview of how staff is divided in the roster produced by our

Table 3.7: Objective function scores divided by the number of weeks in the roster. For each instance we present weighted ($\overline{\text{W}}$) and non-weighted ($\overline{\text{NW}}$) sums of variables in the objective function. The second column indicates if a positive value of a variable has a positive (+) or negative (-) contribution to the objective function value.

Instance	+/−	Avg. S17-A18		S19		Real-life	
		$\overline{\text{W}}$	$\overline{\text{NW}}$	$\overline{\text{W}}$	$\overline{\text{NW}}$	$\overline{\text{W}}$	$\overline{\text{NW}}$
Z		44.9		40.6		−132.8	
y_{at}^{O1}	+	39.2	19.577	38.3	19.154	30.6	15.313
y_{at}^{O2}	+	7.8	7.808	5.0	5.038	7.9	7.938
y_{at}^{U}	−	0.0	0.000	0.0	0.000	−125.0	−1.250
z_{et}^{CN}	−	0.0	0.000	0.0	0.000	0.0	0.000
z_{rt}^{CE}	−	0.0	0.000	0.0	0.000	0.0	0.000
z_{ew}^{CW}	−	0.0	0.000	0.0	0.000	0.0	0.000
z_{et}^{CD}	−	−0.1	0.019	−0.2	0.038	−4.7	0.938
v_{r}^{E}	−	−0.8	0.038	−1.5	0.077	−5.0	0.250
$\overline{v}$	−	−0.6	0.006	−0.4	0.004	−27.7	0.277
$\underline{v}$	−	−0.2	0.002	−0.3	0.003	−2.8	0.028
v_{r}^{RH}	−	−0.3	0.017	−0.3	0.017	−6.1	0.306
$v_{r}^{R_2H}$	−	0.0	0.000	0.0	0.000	0.0	0.000

matheuristic and the real-life roster, we present frequencies of overstaffing during weekdays at the large section upper gastric, see Figure 3.2. The 16 weeks we have real-life data for (blue hashed bars) are compared to the roster produced by the matheuristic for the same 16 weeks (red bars). Note that there are only six surgeons affiliated to the upper gastric section in the real-life instance and the S19 instance, while the required staffing level (overstaffing of 0) is 2.

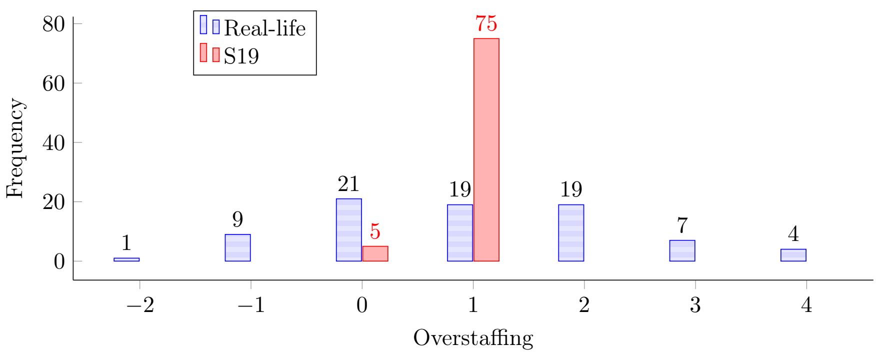

Figure 3.2: Frequency of overstaffing for the first 16 weeks of the real-life roster and the S19 roster.

From Figure 3.2 it is clear that staff is far more evenly divided between weekdays in the roster produced by the matheuristic. Despite only having 6 surgeons to choose from

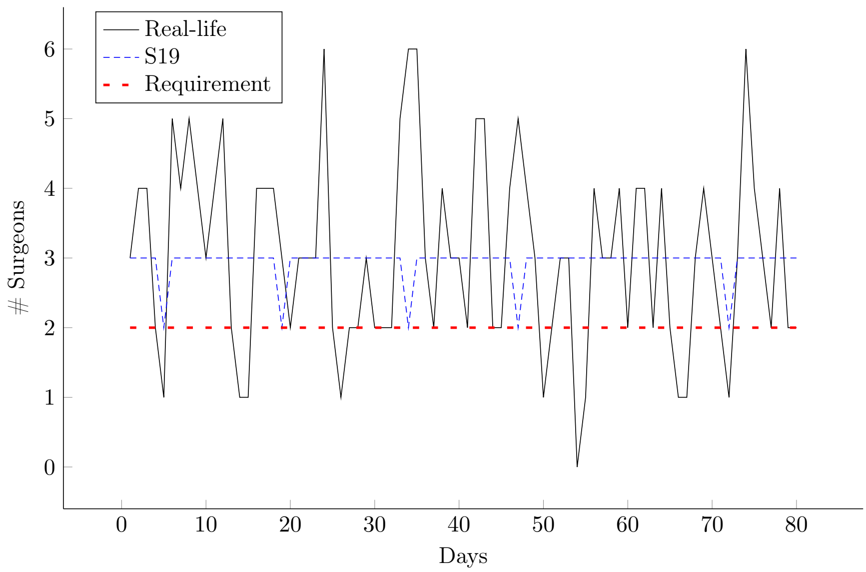

Figure 3.3: The number of surgeons working section shifts at the upper gastric section from weekdays 1 to 80. The black solid line represents the real-life roster, while the blue dashed line represent the roster created by the matheuristic, both for the RSA-matrix in instance S19. The thick red dashed line marks the minimum required number of surgeons.

every day, the matheuristic is able to overstaff the upper gastric section 75 times out of 80. However, the matheuristic does not provide double overstaffing once during the 16 weeks.

There is a considerable variation in staffing levels at the upper gastric section in the real-life roster. In fact, there is one section shift without surgeon, which is infeasible in our matheuristic. At the other end of the spectrum, the real-life instance assigns all six surgeons, i.e. quadruple overstaffing, to work section shift on four occasions. This demonstrates one of the key differences in solutions produced by the matheuristic and the real-life solution. The matheuristic is able to divide staff resources evenly for all sections over time, ensuring many occurrences of single overstaffing and few of understaffing.

While Figure 3.2 demonstrates the accumulated cases of different levels of overstaffing, it does not provide insights into how the staffing at a section changes from day to day.

In Figure 3.3, we plot the 16 weeks (weekends are still excluded). The black solid line represents the number of surgeons working a section shift each day at the upper gastric section in the roster created in real-life. The blue dashed line represents the same data for the S19-case. The thick red dashed line represents the staffing requirement.

Notice how the staffing levels in the real-life roster fluctuates. For staff at sections, the over- and understaffing likely seems random and sporadic. This makes it challenging to plan patient flow and the use of complementary resources. On the other hand, the staffing level in the S19-case is stable, with exceptions in the form of occasional dips in the otherwise constant single overstaffing. These results indicate that the matheuristic is able to produce rosters that have far better stability than they are currently able to do through manual planning at the Clinic of Surgery. Assuming the stability does not counteract the flexibility of rosters, the rosters produced by our matheuristic are likely considerably more robust than manually produced ones. The flexibility of rosters is discussed in relation to shadow shifts in Section 3.5.3.

From the results in Table 3.7 it is apparent that the rosters created by the matheuristic and the real-life rosters are able to reduce unwanted consecutive shifts, congestion of emergency day shifts for interns and residents, and consecutive weekend work. This is likely due to the way planning of emergency shifts is performed in practise, as elaborated upon in Section 3.3. By creating week-long shift patterns of high-quality, consecutive shifts and congestion within a week is avoided. Lastly, the matheuristic plans very few occurrences of seven consecutive work days for instances S17-A18 (0.019 per week) and S19 (0.038 per week). The real-life case includes such a pattern almost once per week on average (0.938).

3.5.3 Including shadow shifts

We include the possibility of adding shadow shifts and run the two-step matheuristic on the five instances. Key results are presented in Table 3.8, for comparison with the rosters that do not include shadow shifts. All information in Table 3.8 is given as non-weighted (NW), except for the top three rows presenting Z, Gap, and $\hat{Z}$. Z is the objective function value of the best solution found after the standard run of 2 hours using the matheuristic. Gap is defined as $BB_{e_{10}} - Z$ for each instance with and without shadow shifts, where $BB_{e_{10}}$ is the best bound found after running the full SCRPRP for 10 hours using a commercial solver at standard settings, as presented in Appendix D. The shadow-reduced objective function value ($\hat{Z}$) is defined by disregarding the effect of shadow shifts on the objective function, so that $\hat{Z} = Z - \sum_{r \in \mathcal{R}} \sum_{t \in \mathcal{T}} W_r^{SS} y_{rt}^{SS}$. The shadow-reduced objective function value $\hat{Z}$ shows the objective function value without the contribution from the shadow shifts and is therefore comparable with the objective function value Z in the case without shadow shifts. A more comprehensive table of information regarding the rosters produced with shadow shifts can be found in Appendix E.

Table 3.8: Results after running Step 1 and Step 2 of the five real instances with and without shadow shifts. All information is non-weighted, except for the top three rows presenting Z, Gap, and $\hat{Z}$. The second column indicates if a positive value of a variable has a positive (+) or negative (-) contribution to the objective function value.

Instance	+/-	S17		A17		S18		A18		S19	
		No Shadow	Shadow	No Shadow	Shadow	No Shadow	Shadow	No Shadow	Shadow	No Shadow	Shadow
Z		1208.300	2111.000	1216.400	2103.000	1109.700	2013.900	1138.700	2048.800	1055.800	1949.600
Gap		52.300	59.600	44.200	67.600	150.900	156.700	121.900	121.800	192.100	208.000
$\hat{Z}$		1208.300	1202.700	1216.400	1194.700	1109.700	1103.900	1138.700	1138.800	1055.800	1039.600
y_{at}^{O1}	+	508.000	509.000	507.000	508.000	512.000	516.000	509.000	511.000	498.000	499.000
y_{at}^{O2}	+	226.000	216.000	232.000	220.000	172.000	166.000	182.000	178.000	131.000	133.000
y_{at}^{U}	−	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
y_{rt}^{SS}	+		182.000		182.000		182.000		182.000		182.000
z_{et}^{CN}	−	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
z_{rt}^{CE}	−	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
z_{ew}^{CW}	−	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	3.000
z_{et}^{CD}	−	1.000	0.000	1.000	1.000	0.000	0.000	0.000	0.000	1.000	1.000
v_r^{E}	−	0.000	0.000	0.000	0.000	3.000	3.000	1.000	1.000	2.000	2.000
$\overline{v}$	−	0.115	0.096	0.125	0.125	0.144	0.144	0.288	0.221	0.106	0.231
$\underline{v}$	−	0.087	0.115	0.067	0.115	0.029	0.067	0.000	0.048	0.067	0.038
v_r^{RH}	−	0.423	0.510	0.269	0.615	0.452	0.548	0.625	0.712	0.442	0.673
$v_r^{R_2H}$	−	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ShadowSec			0.000		0.000		22.000		30.000		42.000
ShadowCourse			109.000		98.000		88.000		89.000		72.000
ShadowOff			145.000		162.000		155.000		154.000		124.000

Recall that Step 1 of the matheuristic is identical for the case with and without shadow shifts. As a result, all technical values for Step 1 are identical to the values presented in Table 3.6. Furthermore, values of variables z_{et}^{CN}, z_{rt}^{CE} and v_r^E are decided in Step 1, and are thus equal for the instances with and without shadow shifts.

In Table 3.8, it is clear that the cases without shadow shifts perform better than the cases with them in all instances except for A18, where the shadow-reduced objective function value is 0.1 larger than the objective function for the case without shadow shifts. However, for the other instances there is a deterioration in the quality of between 5.6 (S17) and 21.7 (A17), which is still significantly better than the values for the real-life instance in Table 3.7. If the model in Step 2 was solved to optimality, the objective function value from the cases without shadow shifts would always be greater than or equal to the shadow-reduced objective function values for the comparable cases with shadow shifts, but due to suboptimal solutions this relation does not necessarily hold.

The value of Gap is very much dependent on the quality of the objective function value Z. This is because the best bounds do not improve from the LP-solution for most instances, and very little when it does (largest improvement is 1.1). We argued that the limited improvement of the best bounds observed in Table 3.6 was likely due to symmetry, and we believe symmetry plays a similar role in suppressing improvements of the best bounds when using commercial solvers. Thus, a low Gap in Table 3.8 is primarily a testament to the model's ability to find good solutions, not the ability to improve the optimistic bounds of the problem.

Introducing shadow shifts has a very small negative effect on the staffing quality in total. With shadow shifts, we still get very high-quality values for both single and double overstaffing (y_{at}^{O1}, y_{at}^{O2}), and the understaffing (y_{at}^{U}) remains at the ideal value 0. Looking at the total number of days where one or more surgeons are assigned to work shadow shifts (y_{rt}^{SS}), we see that the matheuristic is able to find one or more surgeons working a shadow shift every day. It simply seems that the matheuristic is able to increase the flexibility of the roster without reducing the stability very much.

While the introduction of shadow shifts ensure a viable strategy for covering for a single surgeon being absent from emergency night shifts, it is hard to quantify the improvement in flexibility compared to real-life rosters. Our flexibility measure assumes a given set of strategies for rerostering, while in real-life this can be done in many ways. In practise it is likely that managers hope to find an available surgeon to cover for the absentee or alternatively to find a surgeon who is planned to work an overstaffed shift the day after, so that the surgeon can cover the night shift and be absent the day after. When no such surgeon exists it could also be possible to use more experienced surgeons considered outside the scope of ouing problem, bringing in temps (if available) or breaking shift work rules that are normally considered hard restrictions.

With this in mind, the exact value of introducing shadow shifts is hard to estimate, and we cannot compare the flexibility of real-life rosters with the flexibility in rosters produced by the matheuristic directly. However, when a surgeon works a shadow shift, he/she represents a safety measure against rerostering that causes challenging administrative work and often poor schedules for the surgeon who fills in. A simple estimate of the effect of the shadow shifts is to assume a probability α of any form of absence for each surgeon every day. We can then derive the probabilities of rerostering depending on the availability of a surgeon working a shadow shift and his/her rank.

In our runs, $\overline{P}^{SR} = 1$, meaning that only one shadow shift is rewarded by the objective function. This means that there is only one surgeon readily available as a backup (if the algorithm has assigned one). Furthermore, the demand for emergency night shifts is 1 for officers and 1 for interns and residents combined. If there is no shadow shift, the probability of rerostering, P^N, is 1 minus the probability that both surgeons are *not* present ($(1-\alpha)^2$), resulting in $P^N = 1-(1-\alpha)^2$.

This changes when a backup is readily available to fill in. Depending on the rank of the surgeon working the shadow shift (officers can rank down), we get different probabilities of rerostering. Note that the backup can also be absent with a probability of α. If an intern or resident is backup, the probability of rerostering, P^{IR} is 1 minus the probability that at least one of the intern/residents are present times the probability that the offiecer is present, resulting in $P^{IR} = 1-(1-\alpha^2)(1-\alpha)$.

Finally, the probability of rerostering when an officer is backup, P^O, is 1 minus the probability that no more than one surgeon is absent, resulting in $P^O = 1-((1-\alpha)^3 + 3\alpha(1-\alpha)^2)$.

Table 3.9: Probabilities of rerostering any given day depending on the rank of the surgeon working the shadow shift and the probability of absence.

Rank of surgeon	Probability of absence α														
working shadow shift	1.00%	1.25%	1.50%	1.75%	2.00%	2.25%	2.50%	2.75%	3.00%	3.25%	3.50%	3.75%	4.00%	4.25%	4.50%
None	1.99%	2.48%	2.98%	3.47%	3.96%	4.45%	4.94%	5.42%	5.91%	6.39%	6.88%	7.36%	7.84%	8.32%	8.80%
Intern/Resident	1.01%	1.27%	1.52%	1.78%	2.04%	2.30%	2.56%	2.82%	3.09%	3.35%	3.62%	3.89%	4.15%	4.42%	4.69%
Officer	0.03%	0.05%	0.07%	0.09%	0.12%	0.15%	0.18%	0.22%	0.26%	0.31%	0.36%	0.41%	0.47%	0.53%	0.59%

Table 3.9 shows the probability for rerostering given no shadow shift, an intern/resident backup and an officer backup for different probabilities of absence. The probabilities of rerostering on any given day are greatly reduced when a surgeon works shadow shifts, especially when the backup surgeon holds the rank officer. For the lowest absence rate in Table 3.9, the probability of rerostering is 1.99% without the introduction of shadow shifts, but is reduced to approximately half (1.01%) when an intern/resident works a shadow shift and 0.03% when the backup surgeon is an officer. As can be expected, the probability of rerostering increases with absence rates. The differences between assigning the shadow shift to different ranks also vary with absence rates, but are consistently significant.

In our five instances, the matheuristic assigns a shadow shift every day. In instances S17 and A17, all shadow shifts except one is assigned to an officer, and the last shift is assigned to a resident. In instances S18, A18 and S19, an officer is assigned a shadow shift on all days. This can be verified by comparing weighted and non-weighted values of the row y^{SS}_{rt} in Table 21 in Appendix E, given the parameter values in Table 13 in Appendix A.4. Results in Table 3.9 imply that the introduction of shadow shifts in all five instances significantly reduces the probability that rerostering must occur due to surgeon absence. As there are officers assigned to shadow shifts on close to all days in the five instances, the probabilities in the officer-row in Table 3.9 give a good indication of the frequency such rerostering would occur with, depending on the probability of absence. Furthermore, the None-row gives a good indication of the frequency of rerostering if no shadow shifts are included in the matheuristic.

In real-life, managers at the Clinic of Surgery estimate that a surgeon is absent from the emergency night shift once or twice every month. One or two cases of absence among two surgeons in approximately 30 days gives an estimated range for the probability of absence of 1.67% to 3.33%. In the relevant range of probabilities, $1.50 \leq \alpha \leq 3.50$, in Table 3.9, the probability of rerostering ranges from 2.98% to 6.88% when no shadow shift is included and 0.07% to 0.36% when an officer works a shadow shift. This is a considerable difference. According to managers and planners at the Clinic of Surgery, rerostering is a challenge in practice, and avoiding rerostering would be very useful to them. Furthermore, in practice the scheduling process is simplified since it is clearly defined who will fill in, in case of a sudden absence. For the employees, the predictability of their schedule is improved, as they are unlikely to be asked to fill in unless they are scheduled for a shadow shift.

The objective function values related to consecutive and congested work are slightly worse in the cases with shadow shifts for some instances. In instance S18 and S19, some occurrences of working consecutive weekends exist, as shadow shifts are included when counting weekend work. In instance S17, no occurrences of working seven consecutive days exist, while in the case without shadow shifts, one occurrence exists. Otherwise there are no differences between the shadow and no-shadow cases for consecutive and congested work.

For objective function values related to time differences among surgeons, both the cases that include shadow shifts and those that do not are penalized somewhat for deviations from ideal work hours. However, no cases stand out as particularly good or bad, implying that the introduction of shadow shifts does not change these roster qualities dramatically.

The last three rows of Table 3.8 shows the number of shadow shifts worked in total for each instance. For all instances, it seems that using the coursework shifts as shadow shifts is favorable as compared with the section shifts. Also in practise, coursework shifts are considered a good "buffer" in case of absence of emergency shifts, but are not planned as strategically as in the matheuristic. The shadow off-shift is the most frequently used shadow shift in all instances. This is expected, as they can both be assigned during weekends and they can be utilized to reduce total work time for surgeons that would otherwise work too much, while contributing to flexibility in rosters.

3.6 Conclusions

We have presented and formalized the novel Semi-Cyclic Ranked Physician Rostering Problem derived from a real-life case study. A two-step matheuristic was developed and used to solve the problem for five real-life instances.

The matheuristic produced five rosters of high-quality without introducing shadow shifts. Notably, the variations in staffing levels at section shifts, that have proven very problematic in real-life, were reduced. There were no cases of understaffing in any roster produced by our matheuristic, as opposed to the frequent understaffing in real-life. The superior quality of the produced rosters was further supported by a detailed comparison with a real-life roster created at the Clinic of Surgery.

Furthermore, we introduced shadow shifts in the matheuristic, meaning to have staff available in case of absence from emergency night shifts among colleagues. Results showed that we are able to assign a surgeon to a shadow shift on every day throughout the planning period with only marginal deterioration of other roster qualities. Our matheuristic vastly outperforms the current way of creating rosters at the Clinic of Surgery and greatly improves robustness to absence through resilient staffing levels and flexibility due to shadow shifts.

In real-life, when a surgeon becomes unavailable for an emergency night shift, this has a significant effect on the schedule of the surgeon that steps in to cover the emergency night shift. Because surgeons wish to work many hours, and are scheduled to do so, rosters are filled with shifts to the point where additional shifts rarely can be added into a surgeon's schedule without some rule or regulation being violated. This makes it very hard for surgeons to cover for each other in cases of absence, and implies that surgeons either break rules or that they must be absent from a subsequent shift, which often transfers the problem of surgeon absence to the next day. At the Clinic of Surgery, this has typically culminated in sporadic cases of reduced staffing at sections. Understaffing at sections normally lead to reducing the level of priority for patient visits, including patient discharging. This has adverse effects. As patient beds are a limited resource, late patient discharging can ramify and affect patient admission and care, while other resources like operating rooms and surgical nurses are waiting in an idle state. The robustness measures included in our matheuristic, in the form of even overstaffing (stability) and shadow shifts (flexibility), reduce the adverse effects of absence among staff considerably.

While our work has focused specifically on increasing the robustness of rosters, our model could easily be adjusted to focus on costs. If the Clinic of Surgery and the surgeons agreed that they would relax the restriction of minimum number of work hours, as long as no understaffing occurred, preliminary testing demonstrates that the total number of work hours would be significantly lower. There is also an argument to be made, that with more time off, schedules would have more free space allowing surgeons to cover for absentees. Rosters would thus likely be more flexible, but also less stable due to lower overstaffing.

Acknowledgements

We want to thank Dr. Torbjørn Dahl, Sonja Hoff, Mariann Aasberg Thingstad, and Marthe Steien Løseth for insights on planning at the Clinic of Surgery at St. Olav's Hospital, Trondheim University Hospital. This work was partly supported by the Research Council of Norway through the AXIOM project. This support is gratefully acknowledged. We are grateful to the reviewers, whose comments helped us improve the paper.

Bibliography

T. Becker, P. M. Steenweg, and B. Werners. Cyclic shift scheduling with on-call duties for emergency medical services. *Health Care Management Science*, 22:676 – 690, 2019.

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

P. D. Bruecker, J. V. den Bergh, J. Beliën, and E. Demeulemeester. Workforce planning incorporating skills: State of the art. *European Journal of Operational Research*, 243 (1):1 – 16, 2015.

E. K. Burke, P. De Causmaecker, S. Petrovic, and G. V. Berghe. Fitness evaluation for nurse scheduling problems. In *Evolutionary Computation, 2001. Proceedings of the 2001 Congress on*, volume 2, pages 1139 – 1146. IEEE, 2001.

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

S. Ceschia, N. Dang, P. D. Causmaecker, S. Haspeslagh, and A. Schaerf. The second international nurse rostering competition. *Annals of Operations Research*, 274:171 – 186, 2019.

P. Chan and G. Weil. Cyclical staff scheduling using constraint logic programming. In *Practice and Theory of Automated Timetabling III*, pages 159 – 175. Springer Berlin Heidelberg, 2001.

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

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

V. Dück, L. Ionescu, N. Kliewer, and L. Suhl. Increasing stability of crew and aircraft schedules. *Transportation Research Part C: Emerging Technologies*, 20(1):47 – 61, 2012.

O. EL-Rifai, T. Garaix, V. Augusto, and X. Xie. A stochastic optimization model for shift scheduling in emergency departments. *Health Care Management Science*, 18:289 – 302, 2015.

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

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

A. Fügener and J. O. Brunner. Planning for overtime: The value of shift extensions in physician scheduling. *INFORMS Journal on Computing*, 31(4):732–744, 2019.

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

J. Ingels and B. Maenhout. Employee substitutability as a tool to improve the robustness in personnel scheduling. *OR Spectrum*, 39(3):623 – 658, 2017.

J. Ingels and B. Maenhout. The impact of overtime as a time-based proactive scheduling and reactive allocation strategy on the robustness of a personnel shift roster. *Journal of Scheduling*, 21(2):143–165, Apr 2018. ISSN 1099-1425.

L. Ionescu and N. Kliewer. Increasing flexibility of airline crew schedules. *Procedia - Social and Behavioral Sciences*, 20:1019 – 1028, 2011.

OECD. *Health at a Glance 2017*. OECD, 2017. doi: https://doi.org/https://doi.org/10.1787/health_glance-2017-en.

L. D. Smith. The application of an interactive algorithm to develop cyclical rotational schedules for nursing personnel. *INFOR*, 14(1):53 – 70, 1976.

D. M. Warner. Scheduling nursing personnel according to nursing preference: A mathematical programming approach. *Operations Research*, 24(5):842 – 856, 1976.

H. Wolfe and J. P. Young. Staffing the nursing unit: Part i. controlled variable staffing. *Nursing Research*, 14(3):236–242, 1965a.

H. Wolfe and J. P. Young. Staffing the nursing unit part ii. the multiple assignment technique. *Nursing Research*, 14(4):299–303, 1965b.

L. Xie and L. Suhl. Cyclic and non-cyclic crew rostering problems in public bus transit. *OR Spectrum*, 37:99 – 136, 2015.

Appendices

A Symbol directory

A.1 Indices

In some cases, the same index symbol is used multiple times, and subscripts are used to differentiate them (e.g. s_1 and s_2.)

Symbol	Description
e	Employee
s	Shift
t, τ	Day
w	Week
r	Rank
l	Lag value
a	Section

A.2 Sets

Symbol	Description
$\mathcal{E}$	Employees
$\mathcal{E}_r^R$	Employees of rank r
$\mathcal{E}_a^A$	Employees affiliated to section a
$\mathcal{S}$	Shifts
$\mathcal{S}^E$	Emergency shifts
$\mathcal{S}^{Off}$	Off-shifts
$\mathcal{S}^N$	Night shifts
$\mathcal{S}^{SS}$	Shadow shifts
$\mathcal{T}$	Days in planning period
$\mathcal{T}^{Sat}$	Saturdays in planning period
$\mathcal{T}^{Sun}$	Sundays in planning period
$\mathcal{T}'$	Days in planning period excluding Saturdays and Sundays
$\mathcal{T}_w^W$	Days in week w
$\mathcal{W}$	Weeks in planning period
$\mathcal{R}$	Ranks
$\mathcal{L}_r$	Lag numbers of rank r
$\mathcal{A}$	Sections
$\mathcal{A}^L$	Large sections

A.3 Key values in sets

Symbol	Description
s^{ED}	Emergency day shift
s^{EN}	Emergency night shift
s^{O}	Other shift (O-shift)
s^{NOff}	Normal off-shift
s^{POff}	Protected off-shift
s^{Sec}	Section shift
s^{CW}	Coursework shift
s^{SO}	Shadow off-shift
s^{SSec}	Shadow section shift
s^{SCW}	Shadow coursework shift
r^{I}	Intern
r^{R}	Resident
r^{O}	Officer

A.4 Parameters and case study values

Table 11: Parameters are listed with values and the constraints in which they appear.

Parameter	Index values	Value	Constraints
D_s^{ET}	$s = s^{ED}$	3	(3.1)
D_s^{ET}	$s = s^{EN}$	2	(3.1)
D_{rs}^{ER}	$r = r^I, s = s^{ED}$	3	(3.2)
D_{rs}^{ER}	$r = r^I, s = s^{EN}$	2	(3.2)
D_{rs}^{ER}	$r = r^R, s = s^{ED}$	2	(3.2)
D_{rs}^{ER}	$r = r^R, s = s^{EN}$	1	(3.2)
D_{rs}^{ER}	$r = r^O, s = s^{ED}$	1	(3.2)
D_{rs}^{ER}	$r = r^O, s = s^{EN}$	1	(3.2)
D_a^{A}	$a \in \mathcal{A} \backslash \mathcal{A}^L$	1	(3.9)(3.39)
D_a^{A}	$a \in \mathcal{A}^L$	2	(3.9)(3.39)
$\overline{P}^{N}$		1	(3.22)(3.23)
$\overline{P}^{CW}$		2	(3.26)
$\overline{P}^{EDS}$		1	(3.27)
$\overline{P}^{WE}$		2	(3.31)
$\overline{P}^{NES}$		1	(3.34)
X_{est}^{S1}	$e \in \mathcal{E}, s \in \mathcal{S}^E, t \in \mathcal{T}$	$\{0, 1\}$	(3.38)
$\overline{P}^{SR}$		1	(3.41)
$\overline{P}^{NS}$		2	(3.51)
P_{rst}^{H}	$e \in \mathcal{E}, s \in \mathcal{S}, t \in \mathcal{T}$	See Table 12	(3.57)(3.58)(3.60)(3.61)
$\overline{H}^{W}$		60	(3.57)
P^{TH}		41	(3.58)(3.59)

Table 12: P^H_{rst} is the time spent working different shifts on different days of the week. No hours are spent on off-shifts. For shadow shifts, the time of the unrealized shift is used.

Ranks	Shifts	Mon	Tue	Wed	Thu	Fri	Sat	Sun
$r = r^I$	$s = s^{ED}$	13.25	13.25	13.25	13.25	13.25	13	12
	$s = s^{EN}$	12	12	12	12	12	13	12
	$s = s^{Sec}$	8.25	8.25	8.25	8.25	7.75		
	$s = s^{CW}$	8.25	8.25	8.25	8.25	7.75		
$r = r^R$	$s = s^{ED}$	13.25	13.25	13.25	13.25	13.25	13	12
	$s = s^{EN}$	12	12	12	12	12	13	12
	$s = s^{Sec}$	8.25	8.25	8.25	8.25	7.75		
	$s = s^{CW}$	8.25	8.25	8.25	8.25	7.75		
$r = r^O$	$s = s^{ED}$	8.25	8.25	8.25	8.25	7.75	13	12
	$s = s^{EN}$	17	17	17	17	17.5	13	12
	$s = s^{Sec}$	8.25	8.25	8.25	8.25	7.75		
	$s = s^{CW}$	8.25	8.25	8.25	8.25	7.75		

Table 13: Parameters are listed with values and the constraints in which they appear.

Parameter	Index values	Value	Constraints
W^{O1}		2	(3.37)(3.66)
W^{O2}		1	(3.37)(3.66)
W^{U}		100	(3.37)(3.66)
W^{SS}_r	$r = r^I$	$5\frac{1}{3}$	(3.66)
W^{SS}_r	$r = r^R$	$5\frac{2}{3}$	(3.66)
W^{SS}_r	$r = r^O$	5	(3.66)
W^{CN}		2	(3.37)(3.66)
W^{CE}		20	(3.37)(3.66)
W^{CW}		2	(3.37)(3.66)
W^{CD}		5	(3.66)
W^{E}		20	(3.37)(3.66)
W		100	(3.66)
W^{RH}		20	(3.66)
W^{R_2H}		30	(3.66)

A.5 RSA-matrices (case study instances)

Table 14: RSA-matrix for S17. Urology is a small section, while the other sections are large.

		Ranks			
		Interns	Residents	Officers	Sum
Sections	Urology	1	1	2	4
	Vascular-endocrine-and-pediatric	2	2	3	7
	Upper gastric	1	3	4	8
	Lower gastric	4	2	1	7

Table 15: RSA-matrix for A17. Urology is a small section, while the other sections are large.

		Ranks			
		Interns	Residents	Officers	Sum
Sections	Urology	1	1	2	4
	Vascular-endocrine-and-pediatric	1	2	4	7
	Upper gastric	3	2	2	7
	Lower gastric	3	3	2	8

Table 16: RSA-matrix for S18. Urology is a small section, while the other sections are large.

		Ranks			
		Interns	Residents	Officers	Sum
Sections	Urology	1	3	1	5
	Vascular-endocrine-and-pediatric	1	1	5	7
	Upper gastric	4	1	1	6
	Lower gastric	2	3	3	8

Table 17: RSA-matrix for A18. Urology is a small section, while the other sections are large.

		Ranks			
		Interns	Residents	Officers	Sum
Sections	Urology	1	2	2	5
	Vascular-endocrine-and-pediatric	1	4	3	8
	Upper gastric	3	0	3	6
	Lower gastric	3	2	2	7

Table 18: RSA-matrix for S19. Urology is a small section, while the other sections are large.

		Ranks			
		Interns	Residents	Officers	Sum
Sections	Urology	1	2	2	5
	Vascular-endocrine-and-pediatric	1	2	3	6
	Upper gastric	1	2	3	6
	Lower gastric	5	2	2	9

A.6 Decision variables

Table 19: Decision variables are listed with domains and the constraints in which they appear.

Parameter	Index values	Domain
x_{est}	$e \in \mathcal{E}, s \in \mathcal{S}, t \in \mathcal{T}$	$\{0,1\}$
λ_{el}	$r \in \mathcal{R}, e \in \mathcal{E}_r^R, l \in \mathcal{L}_r$	$\{0,1\}$
y_{at}^{O1}	$a \in \mathcal{A}, t \in \mathcal{T}'$	$\{0,1\}$
y_{at}^{O2}	$a \in \mathcal{A}, t \in \mathcal{T}'$	$\{0,1\}$
y_{at}^{U}	$a \in \mathcal{A}, t \in \mathcal{T}'$	$\{0,1\}$
u_{et}^{Sec}	$e \in \mathcal{E}_a^A, t \in \mathcal{T}'$	$\{0,1\}$
u_{et}^{NC}	$e \in \mathcal{E}, t \in \mathcal{T}'$	$\{0,1\}$
y_{rt}^{SS}	$r \in \mathcal{R}, t \in \mathcal{T}$	$\{0,1\}$
δ_{ew}	$e \in \mathcal{E}, w \in \mathcal{W}$	$\{0,1\}$
z_{et}^{CN}	$r \in \mathcal{R}, e \in \mathcal{E}_r^R, t \in \mathcal{T} \vert r \notin \mathcal{R}^O$	$\{0,1\}$
z_{et}^{CW}	$e \in \mathcal{E}, t \in \mathcal{T}^{Sat}$	$\{0,1\}$
z_{rt}^{CE}	$r \in \mathcal{R}, t \in \mathcal{T} \vert r \notin \mathcal{R}^O$	$\{0,1\}$
z_{et}^{CD}	$e \in \mathcal{E}, t \in \mathcal{T}$	$\{0,1\}$
$\overline{v}_r^E, \underline{v}_r^E, v_r^E$	$r \in \mathcal{R}$	≥ 0
$\overline{v}, \underline{v}$		≥ 0
$\overline{v}_r^{RH}, \underline{v}_r^{RH}, v_r^{RH}, v_r^{R_2H}$	$r \in \mathcal{R}$	≥ 0

B The Semi-Cyclic Ranked Physician Rostering Problem

We present the formulation of the full problem in groups of constraint similar to the groups presented in Steps 1 and 2 in Section 3.4.

Cyclic emergency shifts

$$\sum_{e \in \mathcal{E}} x_{est} = D_s^{ET}, \qquad s \in \mathcal{S}^E, t \in \mathcal{T} \quad (67)$$

$$\sum_{r \in \mathcal{R} | r \geq r_2} \sum_{e \in \mathcal{E}_r^R} x_{est} \geq D_{r_2 s}^{ER}, \qquad r_2 \in \mathcal{R}, s \in \mathcal{S}^E, t \in \mathcal{T} \quad (68)$$

$$x_{est} \in \{0,1\}, \qquad e \in \mathcal{E}, s \in \mathcal{S}, t \in \mathcal{T} \quad (69)$$

$$\lambda_{el}(x_{(\epsilon_r)st} - x_{es(t+7l)}) = 0, \quad r \in \mathcal{R}, e \in \mathcal{E}_r^R, s \in \mathcal{S}^E, t \in \mathcal{T}, l \in \mathcal{L}_r \quad |e \neq \epsilon_r, t \leq 7|\mathcal{E}_r^R| \quad (70)$$

$$x_{est} - x_{es(t+7|\mathcal{E}_r^R|)} = 0, \qquad r \in \mathcal{R}, e \in \mathcal{E}_r^R, s \in \mathcal{S}^E, t \in \mathcal{T} \quad |t \leq |\mathcal{T}| - 7|\mathcal{E}_r^R| \quad (71)$$

$$\sum_{e \in \mathcal{E}_r^R} \lambda_{el} = 1, \qquad r \in \mathcal{R}, l \in \mathcal{L}_r \quad (72)$$

$$\sum_{l \in \mathcal{L}_r} \lambda_{el} = 1, \qquad r \in \mathcal{R}, e \in \mathcal{E}_r^R \quad (73)$$

$$\lambda_{el} \in \{0,1\}, \qquad r \in \mathcal{R}, e \in \mathcal{E}_r^R, l \in \mathcal{L}_r \quad (74)$$

Section shifts

$$\sum_{e \in \mathcal{E}_a^A} x_{es^{Sec}t} - y_{at}^{O1} - y_{at}^{O2} + y_{at}^{U} - u_{et}^{SecA} \geq D_a^A, \qquad a \in \mathcal{A}, t \in \mathcal{T}' \quad (75)$$

$$\sum_{s \in \mathcal{S}^{SS}} x_{es(t-1)} + x_{es^{Sec}t} - u_{et}^{SecA} \leq 1, \qquad e \in \mathcal{E}, t \in \mathcal{T}' \quad (76)$$

$$\sum_{e \in \mathcal{E}_r^R} \sum_{s \in \mathcal{S}^{SS}} x_{est} - y_{rt}^{SS} \geq 0, \qquad r \in \mathcal{R}, t \in \mathcal{T} \quad (77)$$

$$\sum_{r \in \mathcal{R}} y_{rt}^{SS} \leq \overline{P}^{SR}, \qquad t \in \mathcal{T}' \quad (78)$$

$$\sum_{t \in \mathcal{T}} (x_{es^{CW}t} + x_{es^{SCW}t}) = |\mathcal{W}|/2, \qquad e \in \mathcal{E} \quad (79)$$

$$y_{at}^{O1} \in \{0,1\} \qquad a \in \mathcal{A}, t \in \mathcal{T}' \quad (80)$$

$$y_{at}^{O2} \in \{0,1\}, \qquad a \in \mathcal{A}^L, t \in \mathcal{T}' \quad (81)$$

$$y_{at}^{U} \in \{0,1\}, \qquad a \in \mathcal{A}, t \in \mathcal{T}^{'} \quad (82)$$

$$u_{et}^{SecA} \in \{0,1\}, \qquad e \in \mathcal{E}, t \in \mathcal{T}^{'} \quad (83)$$

$$y_{rt}^{SS} \in \{0,1\}, \qquad r \in \mathcal{R}, t \in \mathcal{T}^{'} \quad (84)$$

Create space for off-weeks

$$\sum_{s \in \mathcal{S}^{Off}} \sum_{t \in \mathcal{T}_w^W} x_{est} \geq 7\delta_{ew}, \qquad r \in \mathcal{R}, e \in \mathcal{E}_r^R, w \in \mathcal{W} \quad (85)$$

$$\sum_{e \in \mathcal{E}_r^R} (\delta_{e(w-1)} + \delta_{ew}) = 1, \qquad r \in \mathcal{R}, w \in \mathcal{W} \quad |mod(w,2) = 0 \quad (86)$$

$$\delta_{e(w-1)} + \delta_{ew} = \delta_{e(w-1+2|\mathcal{E}_r^R|)} + \delta_{e(w+2|\mathcal{E}_r^R|)}, \quad r \in \mathcal{R}, e \in \mathcal{E}_r^R, w \in \mathcal{W} \quad |mod(w,2) = 0 \quad (87)$$

$$\delta_{ew} \in \{0,1\}, \qquad e \in \mathcal{E}, w \in \mathcal{W} \quad (88)$$

Regulations and norms

$$\sum_{s \in \mathcal{S}} x_{est} = 1, \qquad e \in \mathcal{E}, t \in \mathcal{T} \quad (89)$$

$$x_{es^{EN}(t-2)} + \sum_{s_2 \in \mathcal{S}^{Off}} x_{es_2(t-1)} + x_{es^{EN}t} \leq 2, \qquad e \in \mathcal{E}, t \in \mathcal{T} \quad (90)$$

$$\sum_{\tau = t - \overline{P}^N}^{t} x_{es^{EN}\tau} - z_{et}^{CN} \leq \overline{P}^N, \qquad r \in \mathcal{R}, e \in \mathcal{E}_r^R, t \in \mathcal{T} \quad |t \notin \mathcal{T}^{Sat}, t > \overline{P}^N \quad (91)$$

$$\sum_{\tau = t - \overline{P}^N - 1}^{t} x_{es^{EN}\tau} \leq \overline{P}^N + 1, \qquad e \in \mathcal{E}, t \in \mathcal{T} \quad |t \notin \mathcal{T}^{Sat}, t > \overline{P}^N + 1 \quad (92)$$

$$x_{es^{ED}t} - x_{es^{ED}(t-1)} = 0, \qquad e \in \mathcal{E}, t \in \mathcal{T}^{Sun} \quad (93)$$

$$x_{es^{EN}t} - x_{es^{EN}(t-1)} = 0, \qquad e \in \mathcal{E}, t \in \mathcal{T}^{Sat} \quad (94)$$

$$\sum_{t \in \mathcal{T}_w^W} x_{es^{ED}t} - z_{rt}^{CE} \leq P^{EDS}, \qquad r \in \mathcal{R}, e \in \mathcal{E}_r^R, w \in \mathcal{W} \quad |r \neq r^O \quad (95)$$

$$\sum_{s \in \mathcal{S}^E} \sum_{t \in \mathcal{T}_w^W | t \notin \mathcal{T}^{Sat}} x_{est} \leq \overline{P}^{WE}, \qquad e \in \mathcal{E}, w \in \mathcal{W} \quad (96)$$

$$\sum_{s \in \mathcal{S}^E} \sum_{t \in \mathcal{T}} x_{est} - \overline{v}_r^E \leq 0, \qquad r \in \mathcal{R}, e \in \mathcal{E}_r^R \quad (97)$$

$$\sum_{s \in \mathcal{S}^E} \sum_{t \in \mathcal{T}} x_{est} - \underline{v}_r^E \geq 0, \qquad r \in \mathcal{R}, e \in \mathcal{E}_r^R \quad (98)$$

$$\overline{v}_r^E - \underline{v}_r^E - v_r^E \leq \overline{P}^{NES}, \qquad r \in \mathcal{R} \quad (99)$$

$$x_{es^{EN}(t-1)} - \sum_{s_2 \in \{s^{EN}\} \bigcup \{s^{SO}\} \bigcup \mathcal{S}^{Off}} x_{es_2t} \leq 0, \qquad e \in \mathcal{E}, t \in \mathcal{T} \quad (100)$$

$$x_{es^{EN}(t-2)} + \sum_{s_2 \in \mathcal{S}^{Off}} x_{es_2(t-1)} + x_{es^{EN}t} \leq 2, \qquad e \in \mathcal{E}, t \in \mathcal{T} \quad (101)$$

$$x_{es_1(t-1)} + x_{es^{POff}t} \leq 1, \qquad e \in \mathcal{E}, s_1 \in \mathcal{S}^N, t \in \mathcal{T} \quad (102)$$

$$\sum_{t \in \mathcal{T}_w^W} x_{es^{POff}t} = 1, \qquad e \in \mathcal{E}, w \in \mathcal{W} \quad (103)$$

$$\sum_{s \in \mathcal{S}^N} \sum_{\tau = t - \overline{P}^{NS}}^{t} x_{es\tau} \leq \overline{P}^{NS}, \qquad r \in \mathcal{R}, e \in \mathcal{E}_r^R, t \in \mathcal{T} | t \notin \mathcal{T}^{Sat}, t > \overline{P}^{NS} \quad (104)$$

$$x_{est} - x_{es(t-1)} = 0, \qquad e \in \mathcal{E}, s \in \mathcal{S}^{SS}, t \in \mathcal{T}^{Sat} \quad (105)$$

$$\sum_{s \in \mathcal{S}^N} \sum_{\tau=1}^{\overline{P}^{CW}+1} x_{es(t-7(\tau-1))} - z_{et}^{CW} \leq \overline{P}^{CW}, \qquad e \in \mathcal{E}, t \in \mathcal{T}^{Sat} \quad (106)$$

$$\sum_{s \in \mathcal{S}^{Off}} \sum_{\tau = t - P^{CD}}^{t} x_{est} \geq 1 - z_{et}^{CD}, \qquad e \in \mathcal{E}, t \in \mathcal{T} \quad | t > P^{CD} \quad (107)$$

$$\sum_{s_1 \in \mathcal{S}^{SS}} x_{es_1(t-1)} + x_{es^{ED}t} \leq 1, \qquad e \in \mathcal{E}, t \in \mathcal{T} \quad (108)$$

$$z_{et}^{CN} \in \{0,1\}, \qquad r \in \mathcal{R}, e \in \mathcal{E}_r^R, t \in \mathcal{T} \quad | r \neq r^O \quad (109)$$

$$z_{et}^{CW} \in \{0,1\}, \qquad e \in \mathcal{E}, t \in \mathcal{T}^{Sat} \quad (110)$$

$$z_{rt}^{CE} \in \{0,1\}, \qquad r \in \mathcal{R}, t \in \mathcal{T} \quad | r \neq r^O \quad (111)$$

$$\overline{v}_r^E, \underline{v}_r^E, v_r^E \geq 0, \qquad r \in \mathcal{R} \quad (112)$$

Work time constraints

$$\sum_{s \in \mathcal{S}} \sum_{t \in \mathcal{T}_w^W} P_{rst}^H x_{est} \leq \overline{H}^W \qquad e \in \mathcal{E}_r^R, w \in \mathcal{W} \quad (113)$$

$$\sum_{s \in \mathcal{S}} \sum_{t \in \mathcal{T}} P_{rst}^H x_{est} - |\mathcal{W}| \overline{v} \leq |\mathcal{W}| P^{TH}, \qquad r \in \mathcal{R}, e \in \mathcal{E}_r^R \quad (114)$$

$$\sum_{s \in \mathcal{S}} \sum_{t \in \mathcal{T}} P_{rst}^H x_{est} + |\mathcal{W}| \underline{v} \geq |\mathcal{W}| P^{TH}, \qquad r \in \mathcal{R}, e \in \mathcal{E}_r^R \quad (115)$$

$$\sum_{s \in \mathcal{S}} \sum_{t \in \mathcal{T}} P_{rst}^H x_{est} - \overline{v}_r^{RH} \leq 0, \qquad r \in \mathcal{R}, e \in \mathcal{E}_r^R \quad (116)$$

$$\sum_{s \in \mathcal{S}} \sum_{t \in \mathcal{T}} P^{H}_{rst} x_{est} - \underline{v}^{RH}_{r} \geq 0, \qquad r \in \mathcal{R}, e \in \mathcal{E}^{R}_{r} \quad (117)$$

$$\overline{v}^{RH}_{r} - \underline{v}^{RH}_{r} - v^{RH}_{r} \leq 0, \qquad r \in \mathcal{R} \quad (118)$$

$$\overline{v}^{RH}_{r_1} - \overline{v}^{RH}_{r_2} - v^{R_2H}_{r_1} \leq 0, \qquad r_1, r_2 \in \mathcal{R} | r_1 < r_2 \quad (119)$$

$$\overline{v}, \underline{v} \geq 0 \quad (120)$$

$$\overline{v}^{RH}_{r}, \underline{v}^{RH}_{r}, v^{RH}_{r}, v^{R_2H}_{r} \geq 0, \qquad r \in \mathcal{R} \quad (121)$$

Symmetry breaking constraints

$$\sum_{l_2=1}^{l_1} \lambda_{e_1 l_2} - \sum_{l_2=1}^{l_1} \lambda_{e_2 l_2} \leq 0,$$
$$a \in \mathcal{A}, r \in \mathcal{R}, e_1, e_2 \in (\mathcal{E}^{A}_{a} \cap \mathcal{E}^{R}_{r}), t \in \mathcal{T}, l_1 \in \mathcal{L}_r \quad | e_1 \neq e_2, e_1 \neq \epsilon_r, e_2 \neq \epsilon_r \quad (122)$$

Objective function

$$Max\ Z = \sum_{a \in \mathcal{A}} \sum_{t \in \mathcal{T}'} W^{O1} y^{O1}_{at} + \sum_{a \in \mathcal{A}} \sum_{t \in \mathcal{T}'} W^{O2} y^{O2}_{at} - \sum_{a \in \mathcal{A}} \sum_{t \in \mathcal{T}'} W^{U} y^{U}_{at} + \sum_{r \in \mathcal{R}} \sum_{t \in \mathcal{T}} W^{SS}_{r} y^{SS}_{rt}$$
$$- \sum_{e \in \mathcal{E}} \sum_{t \in \mathcal{T}} W^{CN} z^{CN}_{et} - \sum_{r \in \mathcal{R}} \sum_{t \in \mathcal{T}} W^{CE} z^{CE}_{rt} - \sum_{e \in \mathcal{E}} \sum_{t \in \mathcal{T}^{Sat}} W^{CW} z^{CW}_{et} - \sum_{e \in \mathcal{E}} \sum_{t \in \mathcal{T}^{Sat}} W^{CD} z^{CD}_{et}$$
$$- \sum_{r \in \mathcal{R}} W^{E} v^{E}_{r} - W(\overline{v} + \underline{v}) - \sum_{r \in \mathcal{R}} W^{RH} v^{RH}_{r} - \sum_{r \in \mathcal{R} | r \neq r^{O}} W^{R_2H} v^{R_2H}_{r} \quad (123)$$

C Linearization

Constraints (124) and (125) replace Constraints (3.4) in the matheuristic in Section 3.4.

$$x_{(\epsilon_r)s\tau} - x_{est} - \lambda_{el} \geq -1, \quad e \in \mathcal{E}^{R}_{r}, s \in \mathcal{S}^{E},$$
$$r \in \mathcal{R}, \tau \in \mathcal{T}, l \in \mathcal{L}_r \quad | e \neq \epsilon_r, t = mod(\tau + 7 \cdot l, |T|) \quad (124)$$

$$x_{(\epsilon_r)s\tau} - x_{est} + \lambda_{el} \leq 1, \quad e \in \mathcal{E}^{R}_{r}, s \in \mathcal{S}^{E},$$
$$r \in \mathcal{R}, \tau \in \mathcal{T}, l \in \mathcal{L}_r \quad | e \neq \epsilon_r, t = mod(\tau + 7 \cdot l, |T|) \quad (125)$$

D Full SCRPRP run with commercial solver

In Table 20, we present results of running the full SCRPRP for all 5 real-life instances, for cases with and without shadow shifts, for 2 and 10 hours. Results obtained after 2 and 10 hours are identified by the subscripts 2 and 10 respectively. Z-values and BB-values represent the best solutions and best bound values found at the time indicated by

Table 21: Results after running Step 1 and Step 2 of the five real instances with shadow shifts. For each instance we present a column of weighted (W) a column of non-weighted (NW) sums of variable types in the objective function. The second column indicates if a positive value of a variable has a positive (+) or negative (-) contribution to the objective function value.

	+/-	S17		A17		S18		A18		S19	
		W	NW	W	NW	W	NW	W	NW	W	NW
Step 1											
Z		1390.0		1388.0		1328.0		1382.0		1321.0	
BB		1430.0		1430.0		1430.0		1430.0		1430.0	
LP-bound		1430.0		1430.0		1430.0		1430.0		1430.0	
Step 2											
Z		2111.0		2103.0		2013.9		2048.8		1949.6	
$\hat{Z}$		1202.7		1194.7		1103.9		1138.8		1039.6	
BB		2161.0		2157.8		2048.6		2083.1		1992.1	
LP-bound		2161.6		2160.0		2049.1		2083.4		1995.4	
y_{at}^{O1}	$+$	1018.0	509.000	1016.0	508.000	1032.0	516.000	1022.0	511.000	998.0	499.000
y_{at}^{O2}	$+$	216.0	216.000	220.0	220.000	166.0	166.000	178.0	178.000	133.0	133.000
y_{at}^{U}	$-$	0.0	0.000	0.0	0.000	0.0	0.000	0.0	0.000	0.0	0.000
y_{rt}^{SS}	$+$	908.3	182.000	908.3	182.000	910.0	182.000	910.0	182.000	910.0	182.000
z_{et}^{CN}	$-$	0.0	0.000	0.0	0.000	0.0	0.000	0.0	0.000	0.0	0.000
z_{rt}^{CE}	$-$	0.0	0.000	0.0	0.000	0.0	0.000	0.0	0.000	0.0	0.000
z_{et}^{CW}	$-$	0.0	0.000	0.0	0.000	-2.0	1.000	0.0	0.000	-6.0	3.000
z_{et}^{CD}	$-$	0.0	0.000	-5.0	1.000	0.0	0.000	0.0	0.000	-5.0	1.000
v_{r}^{E}	$-$	0.0	0.000	0.0	0.000	-60.0	3.000	-20.0	1.000	-40.0	2.000
$\overline{v}$	$-$	-9.6	0.097	-12.5	0.125	-14.4	0.144	-22.1	0.221	-23.1	0.231
$\underline{v}$	$-$	-11.5	0.115	-11.5	0.115	-6.7	0.067	-4.8	0.048	-3.8	0.038
v_{r}^{RH}	$-$	-10.1	0.509	-12.3	0.615	-11.0	0.548	-14.2	0.712	-13.5	0.673
$v_{r}^{R_2H}$	$-$	0.0	0.000	0.0	0.000	0.0	0.000	0.0	0.000	0.0	0.000
ShadowSec			0.000		0.000		22.000		30.000		42.000
ShadowCourse			109.000		98.000		88.000		89.000		72.000
ShadowOff			145.000		162.000		155.000		154.000		124.000

subscripts, respectively. The values of best bound obtained after running the model for 10 hours, $BB_{e_{10}}$, are the basis for calculating the Gap in Section 3.5.

Table 20: Results after running the full SCRPRP with a commercial solver, as a benchmark to compare with the matheuristic. For each instance we present a column where shadow shifts are not included in the problem and one where shadow shifts are included.

Instance	S17		A17		S18		A18		S19	
	No Shadow	Shadow	No Shadow	Shadow	No Shadow	Shadow	No Shadow	Shadow	No Shadow	Shadow
LP-bound$_e$	1260.6	2170.6	1260.6	2170.6	1260.6	2170.6	1260.6	2170.6	1248.8	2158.7
Z_{e_2}	-10682.1	NA	NA	NA	NA	NA	NA	NA	NA	NA
BB$_{e_2}$	1260.6	2170.6	1260.6	2170.6	1260.6	2170.6	1260.6	2170.6	1247.9	2157.8
$Z_{e_{10}}$	1081.5	-3687.3	1088.7	-616.82	1130.2	-3327.9	-6359.8	-7509.2	977.54	NA
BB$_{e_{10}}$	1260.6	2170.6	1260.6	2170.6	1260.6	2170.6	1260.6	2170.6	1247.9	2157.6

E Full table of shadow shifts

To improve the understanding of the rosters we create at the Clinic of Surgery, we present an excerpt of a full roster in Table 22. Emergency shifts are visible in red cells in the table. Surgeons are presented in the order of the lag variables, so that the semi-cyclic structure of the roster is clear. Note how Intern1 works an emergency shift on Wednesday of week 1, Intern2 works an emergency shift Wednesday of week 2, etc. The shared cyclic structure of night shifts between interns and residents is also visible, as Intern7 works an emergency night shift on day 7 and Residents 1 works it on day 14.

Shadow shifts are presented in blue cells in Table 22. Every day, there is at least one officer working a night shift. In some cases, multiple surgeons work shadow shifts, but the additional shadow shifts do not improve the objective function value.

The stability of the roster is clear from the number of surgeons affiliated to each section working section shifts. These are counted in the rightmost columns. Lastly, it is visible in Table 22 how some surgeons have some weeks off, see e.g. Officer1 during week 1.

F Excerpt of full roster

Table 22: Excerpt of the first three weeks of the roster produced for the Spring19-instance. Notice that interns and residents have an aggregated demand for emergency night shifts (exemplified in red highlighted cells), implying the cyclic structure of these shifts are shared for the two ranks. The cyclic structure of the emergency night shifts worked by officers is not aggregated with other ranks, and thus cycle independently from interns and residents. The cyclic structure of emergency day shifts are exemplified in the yellow highlighted cells for interns. There is no aggregate demand for emergency day shifts, and thus all ranks have separate cyclic structures for emergency day shifts. The bottom four rows contain sums of surgeons affiliated to each section working section shifts.

	Day	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Intern1	Upper gastric	s^{SSec}	s^{SO}	s^{ED}	s^{SSec}	s^{SSec}	s^{POff}	s^{SO}	s^{ED}	s^{SSec}	s^{SSec}	s^{SSec}	s^{CW}	s^{POff}	s^{SO}	s^{SSec}	s^{CW}	s^{EN}	s^{SO}	s^{ED}	s^{POff}	s^{SO}
Intern2	Lower gastric	s^{SO}	s^{CW}	s^{POff}	s^{CW}	s^{SSec}	s^{ED}	s^{ED}	s^{SO}	s^{CW}	s^{ED}	s^{SSec}	s^{SSec}	s^{POff}	s^{SO}	s^{ED}	s^{SSec}	s^{SSec}	s^{CW}	s^{SSec}	s^{POff}	s^{SO}
Intern3	Lower gastric	s^{SSec}	s^{SSec}	s^{SCW}	s^{CW}	s^{SSec}	s^{POff}	s^{SO}	s^{POff}	s^{CW}	s^{SSec}	s^{SSec}	s^{SSec}	s^{ED}	s^{ED}	s^{SO}	s^{SSec}	s^{ED}	s^{SSec}	s^{SSec}	s^{POff}	s^{SO}
Intern8	VEaP	s^{POff}	s^{SSec}	s^{SSec}	s^{ED}	s^{EN}	s^{EN}	s^{SO}	s^{SSec}	s^{SSec}	s^{SSec}	s^{CW}	s^{SSec}	s^{POff}	s^{SO}	s^{POff}	s^{SSec}	s^{SSec}	s^{SO}	s^{SSec}	s^{ED}	s^{ED}
Intern4	Lower gastric	s^{SO}	s^{CW}	s^{SSec}	s^{SSec}	s^{SSec}	s^{POff}	s^{SO}	s^{POff}	s^{SSec}	s^{SSec}	s^{ED}	s^{EN}	s^{EN}	s^{SO}	s^{SO}	s^{SSec}	s^{SO}	s^{SSec}	s^{CW}	s^{POff}	s^{SO}
Intern5	Lower gastric	s^{SSec}	s^{ED}	s^{SSec}	s^{SSec}	s^{SO}	s^{POff}	s^{SO}	s^{NOff}	s^{NOff}	s^{NOff}	s^{NOff}	s^{NOff}	s^{POff}	s^{NOff}	s^{SSec}	s^{POff}	s^{SSec}	s^{ED}	s^{EN}	s^{EN}	s^{SO}
Intern6	Lower gastric	s^{SSec}	s^{SSec}	s^{SSec}	s^{SSec}	s^{ED}	s^{POff}	s^{SO}	s^{SSec}	s^{ED}	s^{CW}	s^{SSec}	s^{SSec}	s^{POff}	s^{SO}	s^{NOff}	s^{NOff}	s^{NOff}	s^{NOff}	s^{NOff}	s^{POff}	s^{NOff}
Intern7	Urology	s^{ED}	s^{SSec}	s^{SSec}	s^{SSec}	s^{SSec}	s^{POff}	s^{EN}	s^{SO}	s^{SSec}	s^{SSec}	s^{SSec}	s^{ED}	s^{POff}	s^{SO}	s^{SSec}	s^{ED}	s^{SSec}	s^{CW}	s^{CW}	s^{POff}	s^{SO}
Resident1	Upper gastric	s^{EN}	s^{SO}	s^{ED}	s^{SSec}	s^{SSec}	s^{POff}	s^{SO}	s^{ED}	s^{SSec}	s^{SSec}	s^{SSec}	s^{SSec}	s^{POff}	s^{EN}	s^{SO}	s^{SSec}	s^{SSec}	s^{SSec}	s^{ED}	s^{POff}	s^{SO}
Resident5	Urology	s^{CW}	s^{SSec}	s^{SSec}	s^{POff}	s^{SSec}	s^{ED}	s^{ED}	s^{EN}	s^{SO}	s^{ED}	s^{SSec}	s^{SO}	s^{POff}	s^{SO}	s^{ED}	s^{SSec}	s^{SSec}	s^{SSec}	s^{SSec}	s^{POff}	s^{EN}
Resident3	Lower gastric	s^{NOff}	s^{NOff}	s^{NOff}	s^{NOff}	s^{NOff}	s^{POff}	s^{NOff}	s^{SSec}	s^{POff}	s^{SSec}	s^{SSec}	s^{SSec}	s^{ED}	s^{ED}	s^{EN}	s^{SO}	s^{ED}	s^{SSec}	s^{SSec}	s^{POff}	s^{SO}
Resident7	VEaP	s^{SSec}	s^{SSec}	s^{SSec}	s^{ED}	s^{SSec}	s^{POff}	s^{SO}	s^{CW}	s^{SSec}	s^{CW}	s^{SSec}	s^{SSec}	s^{POff}	s^{SO}	s^{POff}	s^{SSec}	s^{SSec}	s^{SO}	s^{CW}	s^{ED}	s^{ED}
Resident6	Urology	s^{SSec}	s^{EN}	s^{SO}	s^{SSec}	s^{SO}	s^{POff}	s^{SO}	s^{SSec}	s^{SSec}	s^{SSec}	s^{ED}	s^{SSec}	s^{POff}	s^{SO}	s^{SSec}	s^{SSec}	s^{CW}	s^{CW}	s^{SSec}	s^{POff}	s^{SO}
Resident2	Upper gastric	s^{SSec}	s^{ED}	s^{SSec}	s^{EN}	s^{SO}	s^{POff}	s^{SO}	s^{SSec}	s^{EN}	s^{SO}	s^{SSec}	s^{SSec}	s^{POff}	s^{SO}	s^{SSec}	s^{SSec}	s^{CW}	s^{ED}	s^{SSec}	s^{POff}	s^{SO}
Resident8	VEaP	s^{CW}	s^{SSec}	s^{EN}	s^{SO}	s^{ED}	s^{POff}	s^{SO}	s^{SSec}	s^{ED}	s^{CW}	s^{EN}	s^{SO}	s^{POff}	s^{SO}	s^{SSec}	s^{EN}	s^{SO}	s^{SSec}	s^{CW}	s^{POff}	s^{SO}
Resident4	Lower gastric	s^{ED}	s^{SSec}	s^{SSec}	s^{SSec}	s^{SSec}	s^{POff}	s^{SO}	s^{CW}	s^{SSec}	s^{EN}	s^{SO}	s^{ED}	s^{POff}	s^{SO}	s^{SSec}	s^{ED}	s^{SSec}	s^{EN}	s^{SO}	s^{POff}	s^{SO}
Officer1	Upper gastric	s^{NOff}	s^{NOff}	s^{NOff}	s^{NOff}	s^{NOff}	s^{POff}	s^{NOff}	s^{SSec}	s^{CW}	s^{SSec}	s^{CW}	s^{POff}	s^{ED}	s^{ED}	s^{SSec}	s^{SO}	s^{ED}	s^{ED}	s^{SSec}	s^{POff}	s^{SO}
Officer8	VEaP	s^{EN}	s^{SO}	s^{SSec}	s^{SSec}	s^{ED}	s^{POff}	s^{SO}	s^{SSec}	s^{SSec}	s^{SSec}	s^{SSec}	s^{SSec}	s^{POff}	s^{SO}	s^{SSec}	s^{SSec}	s^{SSec}	s^{CW}	s^{POff}	s^{ED}	s^{ED}
Officer9	VEaP	s^{SSec}	s^{EN}	s^{SO}	s^{SSec}	s^{SSec}	s^{POff}	s^{SO}	s^{EN}	s^{SO}	s^{SSec}	s^{SSec}	s^{ED}	s^{POff}	s^{SO}	s^{CW}	s^{CW}	s^{CW}	s^{SSec}	s^{SSec}	s^{POff}	s^{SO}
Officer6	Urology	s^{SSec}	s^{ED}	s^{SCW}	s^{EN}	s^{SO}	s^{POff}	s^{SO}	s^{SSec}	s^{EN}	s^{SO}	s^{CW}	s^{SSec}	s^{POff}	s^{SO}	s^{EN}	s^{SO}	s^{CW}	s^{SSec}	s^{ED}	s^{POff}	s^{SO}
Officer10	VEaP	s^{SSec}	s^{SCW}	s^{EN}	s^{SO}	s^{SSec}	s^{POff}	s^{SO}	s^{POff}	s^{ED}	s^{CW}	s^{EN}	s^{SO}	s^{SO}	s^{SO}	s^{SSec}	s^{EN}	s^{SO}	s^{SSec}	s^{SSec}	s^{POff}	s^{SO}
Officer2	Upper gastric	s^{ED}	s^{SSec}	s^{SSec}	s^{SSec}	s^{SSec}	s^{POff}	s^{EN}	s^{SO}	s^{CW}	s^{EN}	s^{SO}	s^{CW}	s^{POff}	s^{SO}	s^{POff}	s^{ED}	s^{SSec}	s^{EN}	s^{SO}	s^{SO}	s^{SO}
Officer7	Urology	s^{SCW}	s^{SSec}	s^{SCW}	s^{SSec}	s^{CW}	s^{POff}	s^{SO}	s^{ED}	s^{SCW}	s^{SSec}	s^{SSec}	s^{CW}	s^{POff}	s^{EN}	s^{SO}	s^{CW}	s^{EN}	s^{SO}	s^{CW}	s^{POff}	s^{SO}
Officer3	Upper gastric	s^{SSec}	s^{SSec}	s^{SSec}	s^{POff}	s^{EN}	s^{EN}	s^{SO}	s^{SSec}	s^{SSec}	s^{CW}	s^{SSec}	s^{SSec}	s^{POff}	s^{SO}	s^{ED}	s^{SSec}	s^{SSec}	s^{SSec}	s^{SSec}	s^{POff}	s^{EN}
Officer4	Lower gastric	s^{POff}	s^{SSec}	s^{ED}	s^{ED}	s^{SSec}	s^{SO}	s^{SO}	s^{SSec}	s^{SSec}	s^{SSec}	s^{POff}	s^{EN}	s^{EN}	s^{SO}	s^{SSec}	s^{SSec}	s^{SSec}	s^{SSec}	s^{SSec}	s^{POff}	s^{SO}
Officer5	Lower gastric	s^{SSec}	s^{SSec}	s^{SSec}	s^{SSec}	s^{POff}	s^{ED}	s^{ED}	s^{SSec}	s^{SSec}	s^{ED}	s^{ED}	s^{SO}	s^{POff}	s^{SO}	s^{SSec}	s^{POff}	s^{SCW}	s^{SSec}	s^{EN}	s^{EN}	s^{SO}
	Upper gastric	3	2	3	3	3			3	3	3	4	3			3	3	3	2	3		
	Lower gastric	4	5	6	5	5			4	4	4	4	4			4	4	5	5	4		
	VEaP	3	4	3	2	3			3	3	3	3	3			3	3	3	3	3		
	Urology	3	3	4	3	2			2	3	3	3	2			2	2	2	2	2		

Paper III

K. K. Klyve, I. Senthooran, M. Wallace:

Nurse Rostering with Fatigue Modelling - Incorporating a Validated Sleep Model with Biological Variations in Nurse Rostering

In a reviewing process for publication in an international scientific journal. The version presented here has not yet been revised.

Nurse Rostering with Fatigue Modelling - Incorporating a Validated Sleep Model with Biological Variations in Nurse Rostering

Abstract

We use a real nurse rostering problem and a validated model of human sleep to formulate the Nurse Rostering Problem with Fatigue. The fatigue modelling includes individual biologies, thus enabling personalised schedules for every nurse. We create an approximation of the sleep model in the form of a look-up table, enabling its incorporation into nurse rostering. The problem is solved using an algorithm that combines Mixed Integer Programming and Constraint Programming with a Large Neighborhood Search. A postprocessing algorithm deals with errors, to produce feasible rosters minimising global fatigue. The results demonstrate the realism of protecting nurses from highly fatiguing schedules and ensuring the alertness of staff. We further demonstrate how minimally increased staffing levels enable lower fatigue, and find evidence to suggest biological complimentarity among staff can be used to reduce fatigue. We also demonstrate how tailoring shifts to nurses' biology reduces the overall fatigue of the team, which means managers must grapple with the meaning of fairness in rostering.

4.1 Introduction

Adverse psychological and physiological effects of night rotations on nurses are well documented Muecke (2005). Impaired vigilance and performance occurs as a result of increased sleepiness and can seriously compromise workers' health and safety Boivin and Boudreau (2014), as well as patient safety Hughes and Rogers (2004). This underscores the importance of avoiding nurse rosters that cause fatigue. Shift work regulations have been established to hinder employee exhaustion. Such rules and regulations are a key part of the constraints in the Nurse Rostering Problem (NRP), see for example Burke et al. (2004). Because such rules over-simplify the conditions underlying fatigue, sleep deprivation and different kinds of fatigue continue to have adverse effects on nurses.

In this work we expand from the typical NRPs to incorporate modeling of fatigue using a validated sleep model. We include individual biology in the fatigue modelling and minimise fatigue to enhance nurse health and reduce the risk of human errors due to impaired vigilance. This is formalised in the Nurse Rostering Problem with Fatigue (NR-PwF). We incorporate an approximation of the Phillips et al. (2010) fatigue model in the form of a lookup-table. The NRPwF is solved using an algorithm combining Constraint Programming and Mixed Integer Programming with a Large Neighborhood Search to solve realistic problem instances based on real-life data. Our research demonstrates that the worst cases of fatigue can be significantly reduced. It serves as a proof of concept for incorporating a general sleep model in NRP, and is generalisable to other rostering and workforce planning problems. The NRPwF implementation produces rosters minimising the global maximum fatigue, and demonstrates how biology is an important factor when creating fatigue minimising rosters. We further demonstrate how minimally increasing the number of staff makes it possible to significantly reduce the fatigue experienced by nurses.

Our main contributions are listed below:

- Creating an approximation of an advanced sleep model, and demonstrating it can be integrated into the novel Nurse Rostering Problem with Fatigue (NRPwF)
- Introducing realistic biological profiles enabling personalised schedules
- Creating a new algorithm combining Mixed Integer Programming and Constraint Programming to facilitate a Large Neighborhood Search solving the NRPwF, with a postprocessing procedure to handle cases where the approximation is erroneous
- Demonstrating that to minimise the global maximum fatigue of nurses for realistic instances, they must be assigned different numbers of tiring shifts and shift combinations, i.e. be treated differently depending on their biology
- Demonstrating how minimally increased staff levels enable reduction of nurse fatigue in realistic instances

The outline of this paper is as follows. In Sections 4.2.2 and 4.2.1 we present relevant literature in sleep research and nurse rostering. In Section 4.3, the fatigue model at the core of our project is presented, and preliminary analyses of its effects are performed. We go on to create a typical NRPwF in Section 4.4, and demonstrate how a fatigue model approximation can be utilized despite cases of imprecision. This is done by implementing an algorithm using a Constraint Programming (CP) solver in a Large Neighbourhood Search (LNS) to find high-quality solutions based on the approximation in Section 4.5. In Section 4.6 the use of our algorithm is demonstrated, verified, and in some cases postprocessed. We further perform analyses on the effects of rosters in light of biological profiles and staffing levels. In Section 4.7, we make concluding remarks and give suggestions for future research.

4.2 Related literature

This section introduces the literature on typical nurse rostering problems and the most prominently used solution methods. It briefly reviews how fatigue is included in Operations Research literature, as well as presenting relevant fatigue models from the realm of sleep research. Lastly it summarises the identified gaps in related literature.

4.2.1 Nurse Rostering literature

The Nurse Rostering Problem (NRP) is a scheduling problem which assigns a number of shifts with predefined start and end times to a set of nurses in a given planning period. NRPs typically include coverage constraints, i.e. some constraints ensuring a minimum number of nurses on duty; time related constraints e.g. a number of hours to be worked during the planning period; and a set of work regulations Burke et al. (2004). A range of different rules and regulations exist in the Nurse Rostering literature. The many different variations are too many to mention explicitly, but for additional details, we refer readers to Burke et al. (2004) and Haspeslagh et al. (2014). As no widely accepted standard NRP exists, we create an NRP based on guidelines from Safe Work Australia Safe Work Australia (2013).

NRPs are solved in numerous ways, e.g. Artificial Intelligence (AI) approaches, Constraint Programming, metaheuristics and mathematical programming approaches Ernst et al. (2004). In the realm of CP, Downing (2016) tackle Nurse Rostering, among other problems, with lazy clause generation. Other examples of CP include Pizarro et al. (2011) and Métivier et al. (2009). Examples of AI methods include Meyer auf'm Hofe (2001), which builds on CP and integrates fuzzy constraints with branch and bound. The hybrid artificial bee colony algorithm presented in Awadallah et al. (2015) is another AI method used, where the bee operator is replaced with the hill climbing optimizer. According to Burke et al. (2008), metaheuristic methods seem to be the dominant technique when solving real-world problems. Examples are the tabu search based metaheuristic of Rönnberg and Larsson (2010) and the case-based reasoning approach of Beddoe et al. (2009).

There are three main drawbacks to these meta heuristic approaches. Firstly they have parameters which require tuning for each application. Secondly, once the parameters have been tuned for a certain set of example inputs, it is unclear for which other inputs the same tuning works. Thirdly the user has no feedback if the parameter tuning is incorrect. The results may be good, or poor, but unless there is another approach to compare them with the user cannot know.

In the realm of mathematical programming approaches, the standard mixed integer programming (MIP) models are among the most explored ones, see e.g. Ásgeirsson and Sigurðardóttir (2016) and Mischek and Musliu (2019). Different decomposition methods have also been explored, with variants of column generation being popular modeling choices Dohn and Mason (2013), Beliën and Demeulemeester (2007). These approaches typically provide optimality gaps, which provides some confidence about the solution quality. Unfortunately for complex rostering problems the measure is typically too large to be useful.

Notably, literature on Nurse Rostering often focuses on solution techniques Petrovic and Berghe (2012). We argue there should be an increased focus on creating models that are useful in practice and that provide insights for real-life decision makers. We aim to do so by designing a Nurse Rostering Problem that focuses on minimising fatigue to reduce risks of accidents and improve nurse health, and create managerial insights for decision makers based on our computational results. This implies less focus on proof of technical concepts such as the optimality gap, but rather finding high-quality solutions in terms of reducing fatigue more than the standard scheduling rules do within reasonable run times for realistic instances, and identifying managerial insights.

4.2.2 Fatigue modeling literature

When models in Operations Research (OR) deal with subjects such as tiredness, stress and work strain, they often present some version of a fatigue model. The fatigue term is used ambiguously; often loosely defined, if defined at all. In Michalos et al. (2010) a job rotation tool designed to provide less monotonous and repetitive tasks for employees is presented. Authors define fatigue as "the physical stress that each process induces on the operators", and it was shown that job rotation plans could reduce the total accumulated physical fatigue per operator Michalos et al. (2013). According to Jamshidi (2019), "Fatigue is a stochastic factor that changes according to other factors such as environmental conditions, work type, and work duration", which they handle using chance constraints. In Goel and Vidal (2014), fatigue is not defined explicitly, but rather linked to road transport crashes and falling asleep while driving, in an effort to evaluate regulations.

While the different approaches to modelling fatigue in the examples mentioned above are useful, literature in medical sciences and biology often distinguishes between acute fatigue and chronic fatigue, further differentiated into muscular fatigue, mental fatigue, psychomotor fatigue and chronic fatigue associated with post-viral syndromes Dawson et al. (2011). We argue that literature within sleep research best fits the fatigue experienced by shift workers. This literature deals with fatigue fitting the definition provided in Dawson et al. (2011): "the drive to sleep". This is the sense in which we use the term *fatigue* in this work (but note that the term sleep drive can be used interchangeably).

There are several models of human sleep that can be utilized either directly or as part of quantitative tools to evaluate the fatigue of shift workers, e.g. Borbély (1982), Åkerstedt et al. (2004), Mallis et al. (2004), Hursh et al. (2004), McCauley et al. (2009), Rajdev et al. (2013), St. Hilaire et al. (2016), Postnova et al. (2016). These models tend to be used as tools of retrospective evaluation. It is rare for such tools to be deployed prospectively, i.e. explicitly incorporating them into models that perform planning. However, we have identified some few examples of this.

The fatigue model in Tvaryanas and Miller (2010) is based on the Hursh et al. (2004) "Sleep, Activity, Fatigue, and Task Effectiveness (SAFTE)" model, and incorporates it in a staff scheduling tool, where the SAFTE-model has been used to simulate the fatigue score of all possible schedules and ranking them in four categories depending on degree of fatigue. In Wang and Ke (2013), the authors were inspired by the Fatigue Audit Inter Dyne (FAID) system Roach et al. (2004). Wang and Ke (2013) simplified the FAID model, resulting in a linearisation of an exponential function suitable for a MIP framework. The problem considers minimising fatigue in work shift scheduling for air traffic controllers. The linear fatigue model is improved in Wang and Liu (2014), by the addition of a dampening parameter in cases of extreme fatigue. This is shown to fit the results of the FAID model better. A similar model and technique is used for shift scheduling of aircraft maintenance crews in Liu and Wang (2013). Lin et al. (2013) present a MIP model for nurse scheduling taking into account fatigue using two different approaches. The first is survey-based and the second uses a sinusoidal function that includes a parameter implying individual nurses' chronotype (propensity to sleep at different times), based on work presented in Dawson and Fletcher (2001) to approximate fatigue at the end of a week. Bowden (2016) proposes the TDSPFM, a Truck Driver Scheduling Model where fatigue is modeled using the non-linear fatigue model proposed in Ingre et al. (2014) model, which is itself based on the three process model of Åkerstedt et al. (2004). The non-linear TDSPFM was solved using the evolutionary algorithm of the built-in Excel 2013 solver.

Our fatigue modelling is based on the Phillips et al. (2010), where a validated model of human sleep and circadian rhythms is presented. It is discussed in more detail in Section 4.3. We aim to minimise the changes and adjustments to it, both to conserve the realism provided by the sleep model itself and to ensure the continued relevance of our approach as sleep models are improved and extended.

4.2.3 Gaps in related literature

We identify four interesting gaps in this literature. First, we have found no published research on nurse rostering integrating a validated sleep model. This gap includes both nurse rostering models and also the techniques and algorithms needed to solve them. We present the novel Nurse Rostering Problem with Fatigue and develop an algorithm to find high-quality solutions. Second, we utilise an approximation technique unusedin related literature; namely a lookup table. This technique is conceptually simple, but the generality of the approximation technique makes it relevant when sleep research progresses and new models are produced, as long as the implicit assumption holds. Third, the referenced works based on validated sleep models have made significant adjustments to models to fit the OR-framework We demonstrate that our general approximation technique combined with postprocessing finds solutions that truly match the validated sleep model. Fourth, all works where sleep models are used in prospective planning, except Lin et al. (2013), assume homogeneous biology among staff. We include individual nurse biology to the prospective planning of fatigue minimising rosters, leading to interesting managerial insights.

4.3 The fatigue model

In this section we present the fatigue model based on the sleep model of Phillips et al. (2010). This sleep model has been subject to testing and parameter-tuning, and similar models have been based on it since. It combines the Phillips and Robinson (2007) model of the ascending arousal system with the Forger et al. (1999) human circadian pacemaker. In the fatigue model, a sleep/wake switch is included, which models how a human falls asleep and wakes up as a result of internal processes in the brain and light conditions. The impact of shift work on the model is that it precludes sleep. The times a person is at work, the fatigue model is restricted from entering a sleeping state. This functionality of forced wakefulness has been used in other works, such as Phillips and Robinson (2007) and Fulcher et al. (2010) to model total sleep deprivation, Postnova et al. (2012) to model shift work, and Skeldon et al. (2017) and Swaminathan et al. (2017) to model work schedules. The Phillips et al. (2010) fatigue model is written in Matlab Dorf and Bishop (1998) and solved using a built-in ordinary differential equation solver.

A notable characteristic of typical rostering problems, as opposed to more general scheduling problems, is that a set of possible shifts is defined. Our NRPwF model admits four shifts in accordance with Safe Work Australia (2013) guidelines for managing the risk of fatigue at work to represent realistic and advisable shift times:

- Day shift “D” 07:00 - 15:00
- Evening shift “E” 14:30 - 22:30
- Night shift “N” 22:00 - 07:30(+1 day)

- Off-shift “O”

To add to the realism, we have chosen to include 45 minutes of forced wake-time before and after work, to represent commuting. Depending on the roster a nurse works, the fatigue model calculates fatigue based on his or her shifts.

The initial values of the fatigue model variables reflect a well-rested individual where the circadian rhythm has been given time to stabilize in the individual’s preferred phase. To ensure this, we simply let the sleep model run for long periods without any work, thus obtaining the default initial fatigue model state. The individual has typical biological parameter values, meaning default parameter values from the Phillips et al. (2010) model are used. These have been validated in previous works. For details see Phillips et al. (2010), Phillips and Robinson (2007), and Forger et al. (1999).

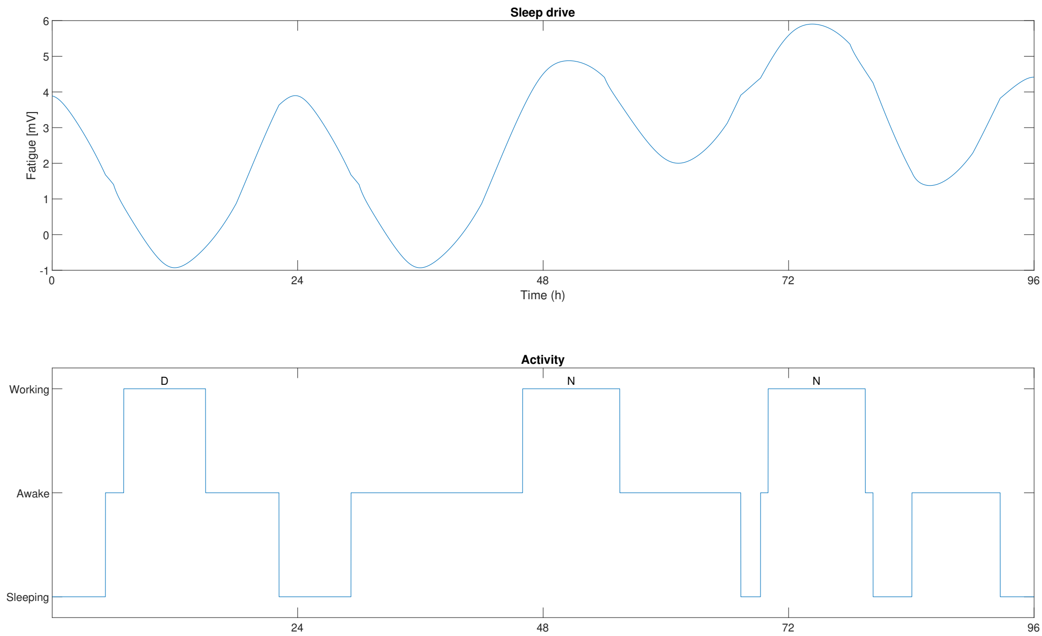

Figure 4.1: Plots of how fatigue and activity of a nurse with typical biological parameters change as time passes. The nurse is scheduled to work the 4-day roster {D,N,N,O}.

In Figure 4.1, two plots of a four day example roster {’D’,’N’,’N’,’O’} is presented. The fatigue is illustrated in the top plot and visibly oscillates according to the time of the day (hours 0, 24, etc. represent midnight). As a result of the two night shifts, the fatigue level is notably higher during the third and fourth day, compared to the two previous days. The sleep drive typically exists in an interval [-2mV,8mV] depending on biological parameters and other factors that affect sleep. In this work, it is sufficient to compare fatigue values knowing that a lower fatigue is always beneficiary. However, for a more intuitive understanding of how different periods of sleep deprivation correspond to different values of fatigue, see e.g. Fulcher et al. (2010).

In the bottom plot, the activities at all times in the form of sleep, wakefulness and work are presented. Note that night shifts are defined to begin during the end of a day, so the night shifts during days 2 and 3 begin at hours 46 and 70. From the activity plot it is clear that the nurse only got a short period of sleep between night shifts, falling asleep around hour 67, before the nurse was forced to awaken 45 minutes prior to the second night shift beginning at hour 70.

In this work, the fatigue model is taken to be the best representation available of a nurse's fatigue at any time, and the fatigue scores provided by the model are thus sometimes referred to as the *true fatigue* of a nurse. In Section 4.3.2 the fatigue model is approximated for incorporation in NRPs. It is referred to as the *rolling horizon approximation* or simply the approximation.

4.3.1 Biological variations

In this work we wish to take into account that fatigue develops differently for different individuals, as "knowledge of individual circadian phase in shift workers could identify times of impaired alertness and thereby inform individualized countermeasures for improving workplace safety, overall health, and wellbeing." Stone et al. (2019) The model presented in Phillips et al. (2010) has previously been used to gain insights into the physiological basis for interindividual differences in circadian timing (see e.g. Phillips et al. (2010) Swaminathan et al. (2017) Skeldon et al. (2017)) and the circadian response to simulated shift work (see e.g. Postnova et al. (2013) Stone et al. (2020)). It is thus a good fit for introducing individual biological differences. We present our approach to modelling biological variations as a set of 9 biological profiles in Section 4.6.1.

4.3.2 Approximating the fatigue model

The fatigue model is inherently non-linear, and incorporating it in an NRP is not trivial. When creating a parameter to represent fatigue in our NRP, time is discretisised into days. We believe the most relevant value to represent a nurse's fatigue throughout a day, both in terms of patient safety and nurse health, is the highest fatigue experienced during the 24 hours of that day. Evaluating the fatigue created by all possible rosters of realistic sizes is not realistic. The number of possible rosters is simply too large. For example, for a roster with 4 shifts and a planning period of 42 days an upper bound for the number of possible rosters would be approximately 1.93×10^{25}. The number of practically feasible rosters would be lower, but for realistic sets of constraints, the number of rosters is still huge.

To deal with this issue, we develop a rolling horizon approximation of the fatigue model, using explicit enumeration of all possible rosters of a given number of days T^{hor}. The rosters of length T^{hor} are stored in a lookup table. This approach implicitly assumes there exists a finite number of days (T^{hor}) shorter than the planning horizon of the full roster, that provides a useful approximation of the fatigue. When evaluating a time period $[t - T^{hor} + 1, t]$, this period is referred to as the *evaluation horizon*. The sequence of shifts worked during the evaluation horizon is referred to as the *evaluation pattern*. For every evaluation pattern we elicit the nurse's maximum fatigue on day t. Clearly the longer the horizon, the better the estimate. The best estimate from the model is, of course, when the complete work history of the nurse is entered into it: in effect this is an infinite horizon.

The action of performing an evaluation of a full individual roster, thus obtaining the model's best possible prediction of the fatigue scores (the true fatigue), is referred to as a *Full Roster Evaluation* (FRE). The action of performing an evaluation of an individual roster using the rolling horizon approximation is referred to as a Rolling Horizon Evaluation (RHE). $RHEs$ can be performed for different evaluation horizons t, indicated through the notation RHE_t.

Day	-1	0	1	2	3	4	5	6	7
indrost			N	N	O	D	D	E	O
indpat1	O	O	**N**						
indpat2		O	N	**N**					
indpat3			N	N	**O**				
indpat4				N	O	**D**			
indpat5					O	D	**D**		
indpat6						D	D	**E**	
indpat7							D	E	**O**

Table 4.1: Demonstration of how the rolling horizon approximation evaluates the different 3-day patterns that exist as parts of indrost. The rolling horizon approximation uses the information from the last day of the evaluation patterns, and save them to comprise the approximated fatigue scores for all days. Shift codes in bold represent the scores stored for each evaluation pattern.

In Table 4.1, we present an example of a RHE_3, i.e. the rolling horizon evaluation given a three-day evaluation horizon. The 7-day individual roster $indrost = \{N, N, O, D, D, E, O\}$ begins on day 1 and that it is approximated using a 3-day rolling horizon approximation. The true fatigue is found by evaluating the full $indrost$ and storing the fatigue scores each day, while the approximation evaluates the 7 different 3-day individual rosters $indpat1 \dots indpat7$ and store the fatigue score obtained on the last day, as demonstrated in Table 4.1. For days 1 and 2, we assume off-days before the beginning of $indpat3$, which do not affect the initial values of the fatigue model. As a consequence, the 3-day RHE results match the FRE results for days $1 \dots 3$, but from thereon differences may arise.

We noted that the fatigue model's initial values reflect a well-rested individual with a stable circadian rhythm before introducing the RHE. However, in the case of $RHEs$, some evaluation patterns can follow a night shift (see e.g. $indpat5$ in Table 4.1). Because a night shift stretches into the following day and forces a state of wakefulness in the beginning of that day, we introduce an additional initial fatigue model state for all evaluation patterns that succeed a night shift - essentially this is the RHE approximation extended to include the initial night-shift.

4.3.3 Testing the rolling horizon approximation

We run our FRE and our RHEs for different evaluation horizons on a collection of 30 real-life rosters of 42 days worked by anonymous nurses at the Austin hospital in Melbourne to evaluate the quality of our approximation. We perform our analysis with $T^{hor} \in [3, \dots, 7]$. This is because preliminary testing implies $T^{hor} \leq 2$ is insufficient, and $T^{hor} \geq 8$ would imply generating a very large lookup-table. The large look-up table would be time consuming to generate and potentially increase the complexity of our NRP, depending on implementation. The first T^{hor} days of the RHEs will naturally be identical to the FRE.[1] Thus, we disregard the data for the first 7 days. This gives us 30 rosters of 35 days for 9 biological profiles. Every day, in each roster, for all biological profiles, we identify the fatigue scores, and thus get 9450 data points to compare the FRE with each of the $RHEs$.

A difference in sleep drive of 0.10 mV is regarded as irrelevant by the developer of the model in Phillips et al. (2010), thus there is a good match if in most cases the errors between FRE and $RHEs$ are less than this. To evaluate the full model (FRE) the algorithm solves a differential equation using the MATLAB ordinary differential equation solver "ode23", see Shampine and Reichelt (1997). The 0.10 mV benchmark for magnitudes of errors necessitated a significant reduction in the tolerances of the differential equation solver, as compared to the default values. Our tests included several rosters in violation of the Safe Work Australia guidelines, which should thus be considered relatively tough.

To evaluate the quality of the rolling horizon approximations of different evaluation horizons, we want to compare each data point in the FRE with each data point in the $RHEs$, by quantifying the errors of the approximations. For every RHE of a given evaluation horizon, for each data point, we subtract the value provided by the RHE from the value provided by the FRE for the same data point. E.g., for a 3-day rolling horizon approximation we find the value of $FRE - RHE_3$ for all 9450 data points. We then sort the errors, and obtain percentiles to get an overview of how large the errors are. Negative values in a given percentile would imply the $RHEs$ are larger than the FRE and vice versa.

Evaluation Horizon	1st perc.	5th perc.	10th perc.	90th perc.	95th perc.	99th perc.
$FRE - RHE_3$	-1.5024	-0.3074	-0.1169	0.0224	0.0895	0.7230
$FRE - RHE_4$	-1.3593	-0.3014	-0.0969	0.0144	0.0631	0.6078
$FRE - RHE_5$	-1.1767	-0.2859	-0.0879	0.0088	0.0454	0.5438
$FRE - RHE_6$	-1.0473	-0.2606	-0.0743	0.0058	0.0411	0.4636
$FRE - RHE_7$	-1.0355	-0.2019	-0.0551	0.0049	0.0389	0.4968

Table 4.2: Results of subtracting values of $RHEs$ of different evaluation horizons from the FRE of 30 real rosters. Values for all biological profiles are used. (The units in the fatigue model are millivolts.)

Results of the analysis are presented in Table 4.2. Firstly, we note that errors decrease for longer evaluation horizons, as expected. For all percentiles, the longer evaluation horizons have errors closer to 0 in Table 4.2. Secondly, it is notable how the magnitude of the errors are larger than the irrelevant magnitude 0.1 (less than -0.1 or more than 0.1) in roughly 7% of cases. Furthermore, it is notable that approximations both over- and underestimate the fatigue of nurses regularly.

To understand how errors occur, we consider one of the evaluated 42-day rosters, denoted $realroster_1$. We present up to 21 shifts on each row below:

[1] except for possible errors due to the use of numerical methods in the differential equation solver

$realroster_1 =$
$\{O, O, E, E, D, O, O, E, E, D, O, O, E, D,$
$D, E, D, D, O, O, E, E, D, D, O, O, O, O,$
$O, N, N, N, O, O, O, N, N, N, O, O, O, O\}$

During the first 30 days, the FRE and RHE are close to identical, with only irrelevant differences. On day 30, lasting into day 31, the nurse works the first of three consecutive night shifts, ending with a shift from late hours on day 32 until the morning on day 33. On days 34 and 35, the nurse is still recovering from this shift sequence, and fatigue is above rested levels. This leads to a small but visible error appearing few days later as presented in Figure 4.2.

On day 36, the FRE has evaluated a long roster and has slightly different parameter values than in the fully rested state. If the nurse had some continuous off-days from day 36, the FRE would eventually fall back to fit the RHE_4 again, but as consecutive night shifts occur days 36, 37, and 38, the errors rather increase. As a result, errors marginally larger than the 0.1 threshold occur in several of the following days, although the shapes of the two graphs are very similar and intuitively imply high precision in the approximation.

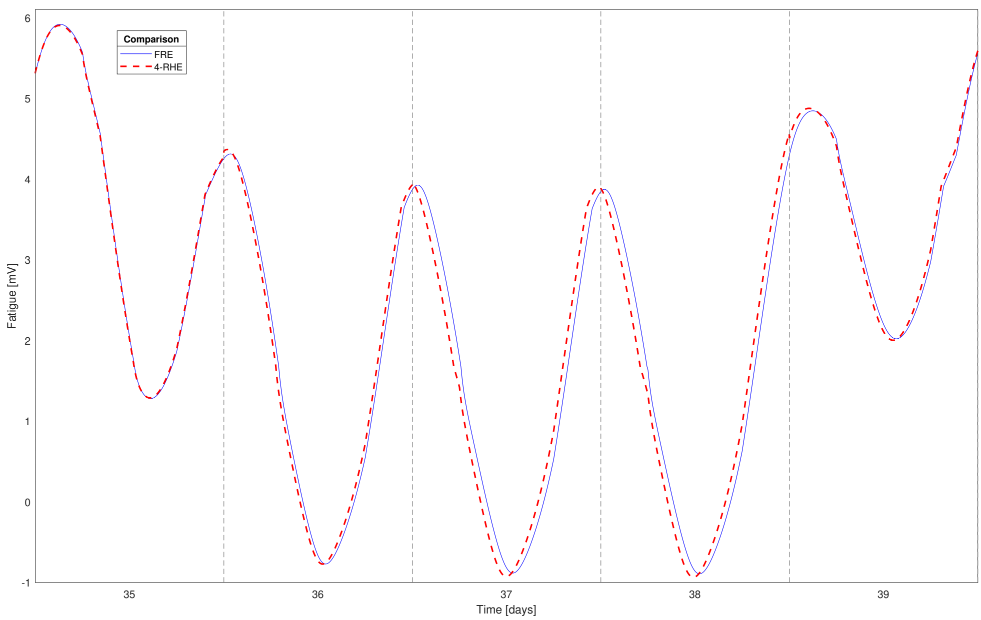

Figure 4.2: Excerpt of the fatigue on days 35 to 39 as $roster_1$ is evaluated through FRE and RHE_4. Notice a small error from the beginning of day 36, lasting throughout the days in the plot.

However, in Table 4.2, some errors are far larger than those illustrated in Figure 4.2. This is due to an additional effect, and we illustrate a particularly tough roster from the collection of real rosters to demonstrate it, $realroster_2$:

$$realroster_2 = \{O, O, O, N, N, N, N, O, O, O, N, N, N, N, O, D, O, N, N, N, O, O, O, O, N, N, N, N, O, O, O, N, N, N, N, O, O, O, N, N, N, O\}$$

In $realroster_2$, the consecutive night shifts on days 4 to 7 lead to a similar shift in the circadian rhythm as we observed in Figure 4.2. However, when the second sequence of four consecutive night shifts occur, one can observe an interesting difference in the activity plot in the Figure 4.3. On day 13, the RHE_4 calculates that the nurse will have a short nap before going to work (notice the dip in the red dotted activity plot), while the FRE does not. From that day and onwards, the two graphs diverge consistently both in terms of maximum daily fatigue values and in the shapes of the two graphs.

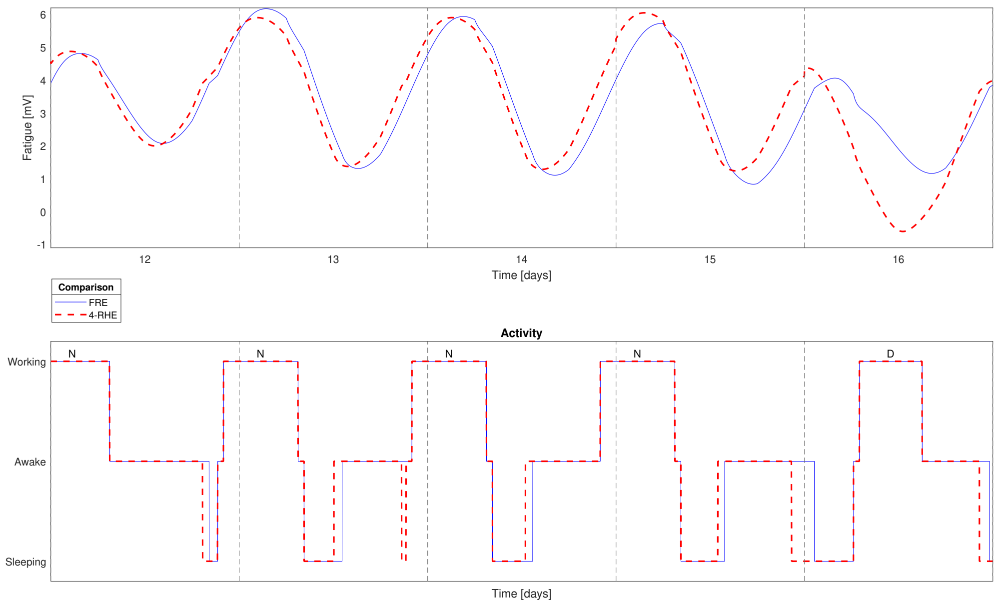

Figure 4.3: Excerpt of both the fatigue and the corresponding activity on days 12 to 16 as $roster_2$ is evaluated through FRE and RHE_4.

The RHE_4 calculates a nap on day 13 because it just exceeds a threshold for specific values of brain activity inherent in the fatigue model (not just the sleep drive/fatigue value), while the FRE only comes close to that same threshold. In real life it is unclear whether the nurse would, in fact, sleep at this time or not. However the divergence between FRE and RHE shows that the two scenarios - staying awake or sleeping at this time - has significant knock-on effects for the nurse's alertness.

The notion that minor errors at any point in time can lead to different activities and thus escalate into large differences, brings up an important consideration. In reality, the nurse might or might not have a nap prior to the night shift, depending on a variety of external factors such as noise, light, telephone interruptions etc. Consequently it could be FRE that diverges from reality and not RHE. In this work, however, we regard the FRE is the best model available for predicting real-life sleep patterns and treat it as a prediction of true fatigue values.

With this discussion in mind, for RHE we choose a 4-day horizon in our computational experiments. This gives a practical size of the NRPwF, accepting that whatever the horizon, some differences from FRE are likely to occur.

4.4 The Nurse Rostering Problem with Fatigue

In this Section, we present a brief formal problem description in Section 4.4.1 and a model formulation in Section 4.4.2. As we use a mix of different techniques for implementation of our model, we provide a short explanation of key concepts in Constraint Programming before presenting the model according to Mixed Integer Programming-tradition.

4.4.1 Problem description

The following hard constraints are based on Safe Work Australia's guide for managing the risk of fatigue at work Safe Work Australia (2013). Every day in the planning period, each nurse should be allocated either one work shift or an off-day. At least a minimum number of nurses must be assigned to work on each day, evening and night shift. The number of successive night shifts is restricted. After ending a night shift, or a sequence of consecutive night shifts, every nurse should have two consecutive nights without work. Nurses should not be assigned backward rotation, meaning that on the day after a shift, the next shift should be the same shift type, or a shift starting later. Restricting backward rotation thus ensures minimum rest times between shifts. There is also an upper limit to the number of consecutive days of work a nurse can have.

Nurses have a maximum number of hours they cannot exceed on average throughout the planning horizon, and for realism we also constrain the average minimum number of hours. Furthermore, there exists a maximum number of hours a nurse can work in any week. In our case, this can be regarded as a maximum number of weekly shifts, as all work shifts last 8.5 hours (see Section 4.3). Nurses are guaranteed to have a weekend off with a given frequency. A weekend off is defined as not working the night shift Friday, any shift Saturday, nor the day or evening shift Sunday. Two consecutive off-days should be ensured for each nurse with a reasonable frequency.

Until this point, we have abstained from specifying what expression of fatigue we will minimise. While it is clear that a lower fatigue level is generally preferable to a higher one, it is not obvious which objective function best represents a combined effort to ensure nurse health and patient safety. While the relation between fatigue due to sleep deprivation and performance is subject to ongoing research, models combining homeostatic and circadian drives (such as our fatigue model) can be used to predict a variety of performance and sleepiness measures Fulcher et al. (2010) Postnova et al. (2018). The effects of increased fatigue levels are disproportional to the reduction in performanceRaslear et al. (2011), and as a result we formulate an objective function reflecting that the highest fatigue scores are especially disadvantageous to both health and safety. We thus minimise the highest fatigue experienced by any nurse at any time in the planning period; the global maximum fatigue (GMF).

This is a relatively coarse objective function in the sense that it does not take into account any other fatigue scores than the very worst one. As a result, rosters could have arbitrarily high fatigue scores for other nurses and on other days as long as they are below the GMF without deteriorating the objective function value. While it would be ideal to minimise the fatigue scores at every minute of every day, we must choose how to balance increased fatigue at one time against lower fatigue at another. This motivates our decision to minimise the maximum fatigue score. We base our modelling on approximated RHE fatigue scores, which may differ from scores returned by the full FRE model, as discussed previously. However, we perform a full FRE evaluation on the rosters our system computes, and asses the effects of such errors afterwards.

Another issue affected by our objective function is fairness. It is quite typical in nurse rostering to model fairness as treating all nurses in the same way, e.g. restricting the difference in working tiring or unpopular shifts. However, we envisage an alternative perspective on fairness, where avoiding the highest fatigue levels for every nurse is more fair than treating everyone the same. We also argue it is more fair to patients to minimise fatigue levels of staff and to avoid huge differences in alertness among nurses. We thus believe minimising the maximum fatigue level is interesting and arguably can result in more fair rosters. It should be noted that the traditional scheduling rules are in place, which treat every nurse the same regardless of their biology. This limits how differently nurses can be scheduled, as the rules treat every nurse the same.

4.4.2 Modelling the problem

The NRPwF is modelled in the MiniZinc language Nethercote et al. (2007), which can map the model onto either MIP or CP solvers, or hybrids. While we choose to formulate the model according to MIP tradition in this work, some key CP concepts utilised in the algorithm should be explained briefly.

Unlike MIP, where constraints must be in linear form, a CP specification (model) can use more expressive built-in constraints (e.g. not equal, append, alldifferent, etc.), and even new constraints defined within the specification. This flexibility helps to simplify the specification, reflect the original problem definition, and is well suited for the problem presented in this paper.

In CP, a problem is defined by a set of variables, representing the choices to be made in reaching a solution; constraints, representing properties/requirements of the problem which must be satisfied in any solution; and the objective, whose value is to be optimised. Each variable can take a set of values, known as its *domain*, and each constraint involves a subset of these variables. In CP, a process called *filtering* is performed first where an appropriate resolution method is applied on each constraint to reduce the domains of its variables; i.e. the values of variables that violate the constraint are removed. When a domain of the variable is changed, it is beneficial to run through all constraints that contain this variable and see whether this change leads to new domain reductions. This process is called *propagation*. Iteratively, a variable is chosen and a value from its domain is assigned to it. The filtering and propagation process is triggered on each assignment. This sometimes leads to the removal of all the values of a variable resulting in a failed value assignment. In the event of a failure, the latest value assignment is reconsidered, called *backtracking*, and a new value is tried. The iterative value assignment, and backtracking process is called *search*. So, as defined in Régin (2011), CP is based on three strategies: filtering, propagation and search.

The choice of variables and constraints in a CP model can impact its efficiency. In particular, the use of sophisticated *global* constraints, which have specialised filtering algorithms, can enhance its performance. Therefore, when modelling our problem, we have used the 'regular' global constraint to capture requirements that apply to every sequence of rostered days or shifts; and the 'cardinality' constraint to enforce coverage for each shift in a roster.

Before presenting our model, we note that a symbol directory and a full model formulation are available in Appendix A. In the NRPwF we assign nurses $n \in \mathcal{N}$ to shifts $s \in \mathcal{S}$. $\mathcal{S}$ consists of all the work shifts $\mathcal{S}^W$ and the off-shift s^O. The work shifts consist of day, evening and night shifts ($s \in \mathcal{S}^W = \{s^D, s^E, s^N\}$). They are allocated during all days in a defined planning period $t \in \mathcal{T}$. In this problem we assume all nurses have had a long period of off days before the beginning of this roster. For constraints stretching back in time to days prior to the defined set of days, we thus assume all nurses were assigned off-shifts s^O. As some restrictions apply specifically for weekends, a set of Sundays $\mathcal{T}^S$ is defined such that $\mathcal{T}^S = \{t \in \mathcal{T} | mod(t, 7) = 0\}$.

Coverage

Nurses are assigned through binary variables $y_{nst} \in \{0, 1\}$. $y_{nst} = 1$ if nurse n works shift s on day t, 0 else.

$$\sum_{n \in \mathcal{N}} y_{nst} \geq \underline{P}_s^C, \qquad s \in \mathcal{S}^W, t \in \mathcal{T} \tag{4.1}$$

Constraints (4.1) ensure coverage, by enforcing that required staffing levels $\underline{P}_s^C$ must be respected for all work shifts s on all days. Constraints (4.1) are implemented as global constraints in our algorithm.

Short-term rest

To ensure sufficient rest, different shift transitions and limitations on work patterns are not allowed. Constraints in this section are modelled using global constraints in the CP implementation.

$$\sum_{s \in \mathcal{S}} y_{nst} = 1, \qquad n \in \mathcal{N}, t \in \mathcal{T} \tag{4.2}$$

$$\sum_{\tau = t - \overline{P}^{CN}}^{t} y_{ns^N\tau} \leq \overline{P}^{CN}, \qquad n \in \mathcal{N}, t \in \mathcal{T} \tag{4.3}$$

$$y_{ns^N(t-1)} + y_{ns^Dt} + y_{ns^Et} \leq 1, \qquad n \in \mathcal{N}, t \in \mathcal{T} \tag{4.4}$$

$$y_{ns^E(t-1)} + y_{ns^Dt} \leq 1, \qquad n \in \mathcal{N}, t \in \mathcal{T} \tag{4.5}$$

$$y_{ns^N(t-2)} + \sum_{s \in \{s^D, s^E, s^O\}} y_{ns(t-1)} + y_{ns^Nt} \leq 2, \qquad n \in \mathcal{N}, t \in \mathcal{T} \tag{4.6}$$

$$\sum_{s \in \mathcal{S}^W} \sum_{\tau = t - \overline{P}^{CD}}^{t} y_{ns\tau} \leq \overline{P}^{CD}, \qquad n \in \mathcal{N}, t \in \mathcal{T} \tag{4.7}$$

Constraints (4.2) enforce that exactly one shift is assigned per day. Constraints (4.3) ensure no nurse works more than $\overline{P}^{CN}$ consecutive nights. Constraints (4.4) and (4.5) make sure that backward rotation is not possible, thus securing a minimum period of rest between shifts for all nurses. Constraints (4.6) state that it is not possible to work a night shift followed by an off day and then another night shift. This implies that after ending a sequence of consecutive night shifts, the nurse will not be working during the night in any of the two following days. Constraints (4.7) set the maximum number of consecutive work days to $\overline{P}^{CD}$.

Long-term rest

To ensure two consecutive days of rest, a new variable is introduced. z_{nt} is a binary auxiliary variable used to indicate if a nurse is allocated any work shifts during a period of two consecutive days ending on day t. Due to nurse preferences, the two-day period considered includes the night shift on day $t-2$ rather than the night shift on day t. This especially affects weekends, as $z_{nt} = 0,\ t \in \mathcal{T}^S$, implies nurse n has the nights off on Friday and Saturday when they have the weekend off.

$$\underline{H} \le \sum_{s \in \mathcal{S}} \sum_{t \in \mathcal{T}} P^H_s y_{nst} \le \overline{H}, \qquad n \in \mathcal{N} \tag{4.8}$$

$$\sum_{s \in \mathcal{S}^W} \sum_{\tau = t-6}^{t} P^H_s y_{ns\tau} \le \overline{H}^W, \qquad n \in \mathcal{N}, t \in \mathcal{T}^S \tag{4.9}$$

$$\begin{aligned} & 2z_{nt} - y_{ns^N(t-2)} - \sum_{s \in \mathcal{S}^W} y_{ns(t-1)} \\ & - y_{ns^D t} - y_{ns^E t} \ge 0, \qquad n \in \mathcal{N}, t \in \mathcal{T} \end{aligned} \tag{4.10}$$

$$\sum_{\tau=0}^{\overline{P}^{CW}} z_{n(t-7\tau)} \le \overline{P}^{CW}, \qquad n \in \mathcal{N}, t \in \mathcal{T}^S \tag{4.11}$$

$$\sum_{\tau = t - \overline{P}^Z}^{t} z_{n\tau} \le \overline{P}^Z, \qquad n \in \mathcal{N}, t \in \mathcal{T} \tag{4.12}$$

Constraints (4.8) restrict the hours worked by each nurse to be in the interval $[\underline{H}, \overline{H}]$. The length of shift s is denoted P^H_s. Constraints (4.9) restrict working more than $\overline{H}^W$ hours every week. Constraints (4.10) ensure z_{nt} indicates work during a two-day period. Due to the short-term rest constraints in Section 4.4.2, the big M-value 2 is sufficient in constraints (4.10). Constraints (4.11) ensure that no nurse works $\overline{P}^{CW}$ consecutive weekends. Furthermore, every nurse should have two consecutive days off at least once every $\overline{P}^Z$ days, as instructed through constraints (4.12).

Objective function

The variable f_{nt} represents the fatigue score of nurse $n \in \mathcal{N}$ on day $t \in \mathcal{T}$. The value of f_{nt} is retrieved from a lookup table where there are multiple additional inputs. The biological profile $b \in \mathcal{B}$ of nurse n, information on whether the nurse worked a night shift the day prior to the evaluation pattern, and the evaluation pattern itself are all relevant inputs for the value of f_{nt}. However, we use the simple f_{nt} syntax here, and present a more detailed version in Appendix A.6. Here, auxiliary variable f^{GM} is introduced to represent the global maximum fatigue score, and is assigned the correct value due to constraints (4.13). The objective function is presented in constraint (4.14).

$$f^{GM} - f_{nt} \geq 0 \qquad n \in \mathcal{N}, t \in \mathcal{T} \tag{4.13}$$

$$\text{Minimise } f^{GM} \tag{4.14}$$

4.5 The solution method

The main idea of our solution method is presented here. For a more detailed review of the algorithm, pseudo-code is provided in Algorithms 1 and 2 in Appendices B and C.

We first use a MIP solver to find a feasible solution; our current best solution. From this point on, in every iteration we have a current best solution available, with a current roster parameter denoted y^*_{nst} and a current global maximum fatigue parameter denoted f^{GM*}. We also denote the current 4-day rolling horizon approximated fatigue parameter of nurse n on day t f^*_{nt}, and introduce the current individual maximum fatigue parameter of nurse n f^{IM*}_n, defined as the highest fatigue experienced by nurse n in the planning period $f^{IM*}_n = max_{t \in \mathcal{T}}(f^*_{nt})$.

To reduce complexity when performing iterations, we fix the roster y_{nst} to be identical to the roster in the current best solution y^*_{nst}, except for some specifically chosen combinations of nurses n and days t denoted by the neighborhood parameter N_{st}. It takes the value 1 if y_{nst} is not fixed to the value of y^*_{nst}, and 0 else.

In every iteration of our algorithm, we create a new roster y_{nst} with new approximated fatigue scores f_{nt} resulting in a new GMF f^{GM}. In every iteration, the algorithm either reduces the GMF ($f^{GM} < f^{GM*}$), finds a new solution with unchanged GMF and fewer occurrences of the GMF ($f^{GM} = f^{GM*} \bigcap (sum_{n \in \mathcal{N}, t \in \mathcal{T}}(f_{nt} = f^{GM*}) < sum_{n \in \mathcal{N}, t \in \mathcal{T}}(f^*_{nt} = f^{GM*})))$, or is not able to find a better solution and keeps the current best ($y_{nst} = y^*_{nst}$). Attempting to reduce the GMF is the standard approach, while attempting to reduce occurrences of the GMF is done when symmetry or unsuccessful previous attempts indicate this is more promising.

When we are no longer able to improve the solution, we perform a Full Roster Evaluation of it, and if errors have relevant magnitude, we repeat the process. We provide a conceptual illustration of the algorithm in Figure 4.4 with a brief description of each step.

As seen from Figure 4.4, the algorithm firstly creates an initial solution, identifying y^*_{nst} and f^{GM*}. Assume there are *not* many occurrences of the f^{GM*} and that a new solution *has* been found recently, the standard optimisation approach of *minimising the GMF* is implemented. This means we implement the objective function presented in constraint (4.14) in Section 4.4. All combinations of nurses and days described below constitute the neighborhood when performing the standard optimisation approach:

1. The full individual schedules of a set of nurses $\mathcal{N}^F$ experiencing the GMF ($N_{nt} = 1,\ n \in \mathcal{N}^F, t \in \mathcal{T}$, provided that $n \in \mathcal{N}^F | f^{IM*}_n = f^{GM*}, n \in \mathcal{N}$)

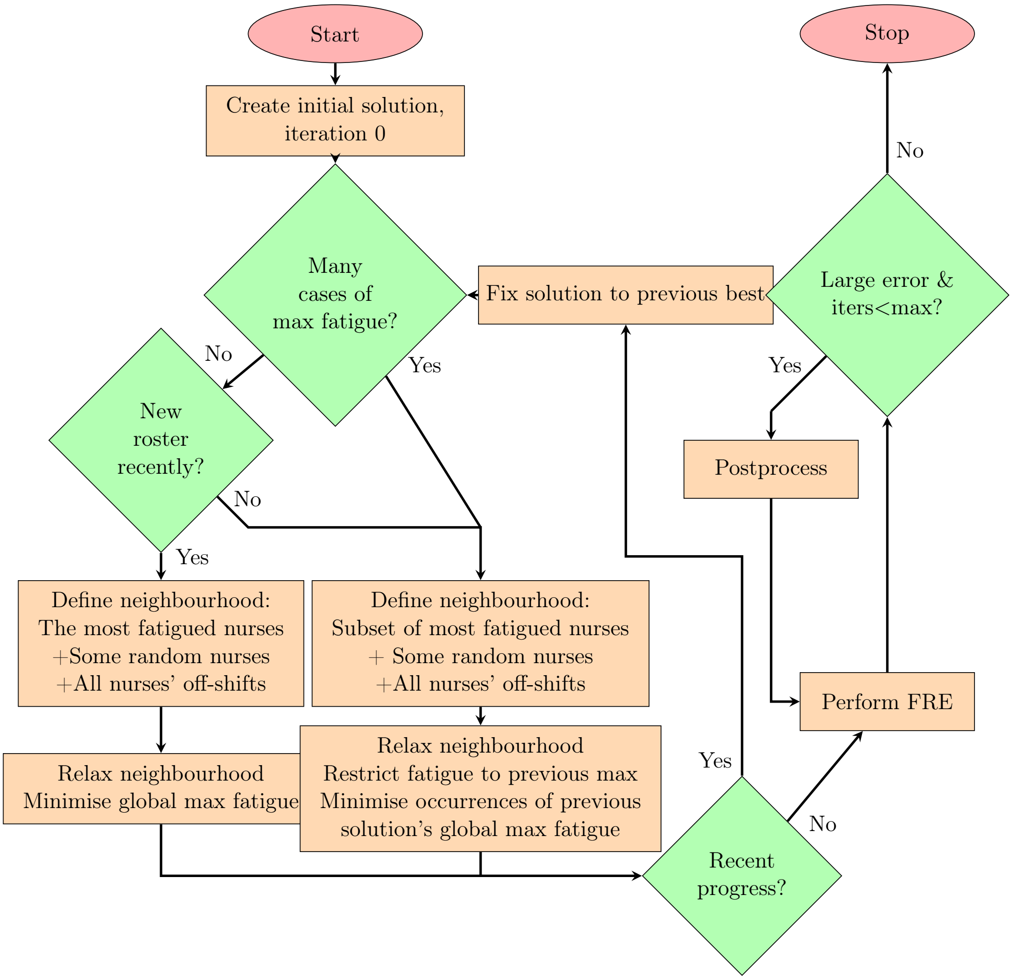

Figure 4.4: Flow chart of algorithm.

2. The full individual schedules of a set of n^R random nurses $\mathcal{N}^R$ ($N_{nt} = 1$, $n \in \mathcal{N}^R, t \in \mathcal{T}$, provided that $\mathcal{N}^R = rand(\mathcal{N} - \mathcal{N}^F, n^R)$)

3. The off-shifts of all nurses in the roster ($N_{nt} = 1$, $n \in \mathcal{N}, t \in \mathcal{T} | y_{ns^O t} = 1$)

We fix the roster $y_{nst}$of any iteration to be equal to the current best roster y^*_{nst}, except for the defined neighborhood where $N_{nt} = 1$, as below:

$$y_{nst} = y^*_{nst}, \qquad n \in \mathcal{N}, s \in \mathcal{S}, t \in \mathcal{T} | N_{nt} = 0 \tag{4.15}$$

Assume the response to the question of recent progress in Figure 4.4 is "yes", and another iteration is performed. If there *are* many occurrences of nurses experiencing the GMF ($f_{nt} = f^{GM*}$ for more than some few $n \in \mathcal{N}, t \in \mathcal{T}$) and/or our algorithm has *not* been able to produce a new solution in recent iterations (we consider ourselves stuck in a local optima). In this case, the neighborhood is defined in the same way as for the standard optimisation approach, with the notable exception that if there are more than n^F nurses in $\mathcal{N}^F$, the set is redefined as a randomly drawn subset of maximum n^F of the nurses experiencing the GMF. Furthermore, we *minimise the occurrences of the GMF.* This also entails restricting the maximum fatigue score of all nurses to the current best BMF. While the nonlinearities are handled by the CP solver in our algorithm, we still provide a MIP linearisation in for an intuitive understanding. Let the binary variable $f_{nt}^{Occ} \in \{0,1\}$ be equal to 1 if nurse n experiences the GMF on day t, 0 else.

$$f_{nt} \leq f^{GM*}, \qquad n \in \mathcal{N}, t \in \mathcal{T} \tag{4.16}$$

$$f_{nt}^{Occ} - f_{nt} + f^{GM*} > 0, \qquad n \in \mathcal{N}, t \in \mathcal{T} \tag{4.17}$$

$$\text{Minimise} \sum_{n \in \mathcal{N}} \sum_{t \in \mathcal{T}} f_{nt}^{Occ} \tag{4.18}$$

Constraints (4.16) ensure that no nurse is assigned a fatigue score that is higher than the GMF of the previous iteration, while constraints (4.17) state that f_{nt}^{Occ} must take a value higher than 0 if $f_{nt} = f^{GM*}$, but can be 0 otherwise. Minimising f_{nt}^{Occ} in constraint 4.18 thus entails minimising the number of occurrences of the GMF.

Assume, after some iterations of the algorithm in Figure 4.4, that there has been no *recent progress* (the evaluation of this question is discussed in more detail in Section 4.6.2)A FRE is performed to unveil the true fatigue of the best solution, before the GMF of the FRE is compared to f^{GM*}. If the true GMF turns out higher than the f^{GM*}, with a margin larger than the 0.10mV threshold of relevance, we perform postprocessing and repeat the FRE on the new roster produced in postprocessing. We assume extras (casuals) can step in on some limited number of shifts when necessary, as is common in real-life. In our solution method, that means we can substitute a tiring shift in our roster with an off-day. Post-processing is thus done by removing the shift prior to the true GMF identified in the FRE. This is repeated until the true GMF is below $f^{GM*} + 0.10$mV or a maximum number of postprocessing iterations is reached. The postprocessing is discussed further in Section 4.6.2, and pseudocode is available in Algorithm 2 in Appendix C.

4.6 Computational Study

The algorithm is run using Python3.6.8 to define neighbourhoods and call MiniZinc2.3.2 using built-in MIP-solver gurobi8.1.1 to find the initial solution and the CP-solver Chuffed0.10.4 Chu (2011) to search within the given neighbourhoods. Matlab R2018 is called to create the lookup-table of approximated fatigue values and to perform the FRE. Computational experiments are run using an HP EliteBook 820 G3 with the specifications below:

CPU:	Intel Core i7-6500U CPU @ 2.50GHz - 2 cores
RAM:	16Gb

4.6.1 Instances

When performing computational studies, we would ideally use real instances. However, as collecting information about individual biology would be both complicated and controversial, we create biological profiles $b \in \mathcal{B}$ by making changes to two parameters in the fatigue model that represent common differences in biology related to sleep Phillips et al. (2010). These two parameters represent the average sleep-time and the chronotype of a human.

The average-sleep-times parameter is by default calibrated for a person with a normal chronotype, which is $\approx$ 7 hours sleep when fully rested beforehand. According to the developer of the Phillips et al. (2010) model, 5 and 9 hours are realistic variations within the adult population in real-life. To identify the right parameter values for the sleep times, we thus vary the one parameter typically reflecting this in the adult population to get the right hours of sleep (see constant offset D_0 in Phillips et al. (2010)), while all other parameters are left at their default values. To get a meaningful number of these nurses represented, we draw biological profiles for nurses with a 10% chance of having a short sleep time ($\approx$ 5 hours) or a long sleep time ($\approx$ 9 hours), leaving an 80% probability of the most common value ($\approx$ 7 hours). The chronotype parameter is by default set to its standard value representing the most common "day-time chronotype". We similarly provide chronotype parameters that are somewhat uncommon, but within realistic variations in the adult population, to create "morning-type" and "evening-type" biological profiles (see intrinsic period τ_c in Phillips et al. (2010)). We assume these two parameters are not correlated, and thus produce the 9 different biological profiles in Table 4.3.

Table 4.3: Illustration of the probability of drawing different biological profiles for the nurses in our instances. The index-value of each of the biological profiles is given in parenthesis.

		Average sleep time		
		Short $\approx$ 5	Normal $\approx$ 7	Long $\approx$ 9
Chronotypes	Probabilities	0.1	0.8	0.1
Morning-type	0.1	0.01 (5)	0.08 (4)	0.01 (6)
Day-type	0.8	0.08 (2)	0.64 (1)	0.08 (3)
Evening-type	0.1	0.01 (8)	0.08 (7)	0.01 (9)

In Table 4.3, the probability of drawing each profile is given depending on the average sleep time and the chronotype of any given nurse. Clearly this very coarse grouping of biological profiles does not come close to capturing the real-world variations of biology affecting sleep. However, the profiles facilitate analyses of the effects of some common differences in biology among nurses. Furthermore, as these parameters represent two aspects of sleep that would be possible to unveil using e.g. a survey, our profiles represent a realistic and pragmatic approach to including biology in real-life rostering.

Furthermore, deciding on the number of nurses relative to minimum staffing levels, as ensured by coverage constraints, is not trivial. We have used real-life 12-week rosters from the Intensive Care Unit (ICU) at the Alfred in Melbourne, Australia as a basis for the minimum staffing parameter. This ICU is an ideal starting point for creating instances that are realistic while ensuring that biological profiles are the only differences between nurses. Firstly, the ICU is the largest in the state of Victoria in Australia State Government of Victoria (2019), making it scalable despite the skill-mix variations in real-life rostering problems. Secondly, the activity at the ICU is inherently interminable, and as a result shift work is planned around the clock.

In the real-life data, several nurses work part-time. This counteracts the desired homogeneousness of our nurses, but by calculating the ratios between full-time equivalents and day, evening and night shifts, we calculate conservative estimates of the minimum coverage requirements. These are 7, 5, and 5 nurses for day shifts, evening shifts, and night shifts respectively, given a total staff of 30 full-time nurses, and correspond to the parameter $\underline{P}_s^C, s \in \mathcal{S}^W$ in the model in Section 4.4. Other data and parameter values are retrieved from the Safe Work Australia guidelines Safe Work Australia (2013), and can be found in Appendix A. We generate 20 instances and analyse them in the following sections.

4.6.2 Minimising the global maximum fatigue scores

In Figure 4.5, the minimisation of the global maximum fatigue score of Instance 1 is illustrated. Black circles with the black lines striking through represent the GMF (f^{MG*}) in each iteration. Blue circles represent one or more nurse's individual maximum fatigue score (f_n^{IM*}) in each iteration (some circles are hidden behind each other). The green line represents the number of occurrences of the current GMF in each iteration ($\sum_{n\in\mathcal{N}}\sum_{t\in\mathcal{T}} f^{Occ}$).

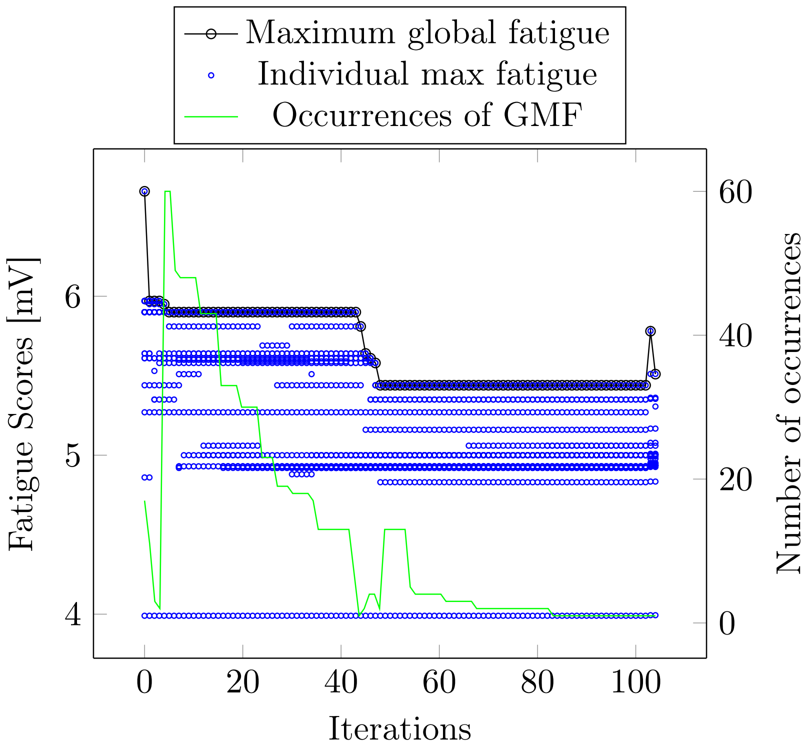

Figure 4.5: Figure demonstrating change in the value, and the number of occurrences, of the global max fatigue for Instance 1. Iteration 0 is the initial feasible solution, while Iterations 1 to 101 are based on the approximated fatigue scores. Iteration 102 is the true global maximum fatigue score and Iteration 103 is the result of postprocessing.

In Figure 4.5, Iteration 0 simply produces any feasible solution to the NRPwF. During the first iterations, the global maximum fatigue clearly decreases, before stabilising at 5.90. Had we only performed 40 iterations on this instance, and not noted the change in number of occurrences of the GMF, we would likely conclude that the algorithm seemed to converge. However, from the green line in Figure 4.5 we can see that while the global maximum fatigue score remains 5.90 during iterations 5 to 43, the number of occurrences of the 5.90-score gradually decreases from 60.

From this point, the global maximum fatigue continues to decrease from iteration 43 to 48, converging to 5.44. For this value of the GMF, however, the number of occurrences also remains unchanged, and we can thus be more confident we have found a high-quality solution. These results illustrate that rather than determining an exact number of iterations for all instances, we should terminate our algorithm when we have reason to believe the maximum global fatigue score has converged, i.e. when our algorithm has no progress in a reasonable number of iterations. We set this limit to 20 iterations without either finding a new GMF or reducing the number of occurrences of the GMF.

Errors and postprocessing

In Figure 4.5, there is a spike in the fatigue score in iteration 103. Iteration 103 represents the FRE performed after termination of our algorithm. From the large difference in fatigue score in iteration 102 and iteration 103, there is clearly at least one case of significantly higher true GMF than estimated in iteration 102. To counteract this, we use extras, as mentioned briefly in Section 4.5.

In reality, it is common in most hospitals to have access to some number of extras that can cover for staff when necessary. We assume that a ward of 30 full-time nurses has access to extras that can cover one shift per week on average, i.e. up to six shifts can be covered by extras in our planning period. When the true GMF is 0.10mV, or more, higher than the approximated GMF, the last work day prior to the true GMF is replaced with an off-shift for the nurse experiencing the GMF. If the process is performed six times and the global maximum fatigue is still more than 0.10mV larger than the approximated global maximum fatigue score, we assume that we have exhausted the ward's budget for extras on single shifts, and accept that the fatigue is higher than implied by the approximation. Pseudocode for the postprocessing procedure is presented in Algorithm 2 in Appendix C.

In Figure 4.5, the last iteration, iteration 104, is the result of using an extra worker for one single shift. In the example of Instance 1, the approximated global maximum fatigue value was 5.44, the true evaluation was 5.78, and after postprocessing (adding one off-shift) the true fatigue value became 5.51.

In the column of Iteration 0 in Table 4.5, the global maximum fatigue scores of the initial feasible roster is provided. The subsequent columns up to the column denoted "Last" present the global maximum fatigue based on the approximated fatigue scores. Later columns all include the true fatigue scores provided through $FREs$ and errors found by comparing the $FREs$ with the approximated fatigue score in the latest iteration of our algorithm. If errors are greater or equal to the threshold of relevance, 0.10mV, extras are used to cover single shifts, and new FRE-values are provided in subsequent columns.

Immediately we notice the stark difference between GMF in rosters in Iteration 0 and the column denoted "Last". As all feasible solutions to the NRPwF must respect the Safe Work Australia guidelines, one might not expect the potential for reducing the GMF was very large to begin with. However, results in Table 4.4 underscore that the algorithm

Table 4.4: The table presents global maximum fatigue scores for 20 instances. All values are given in milliVolts.

Instance	Initial	Values based on RHE_4							Values based on FRE				
Iterations	0	10	20	40	60	80	100	Last	True	Error	1 extra	2 extras	3 extras
1	6.66	5.90	5.90	5.90	5.44	5.44	5.44	5.44	5.78	0.34	5.51		
2	6.66	5.90	5.90	5.44	5.35	5.35		5.35	5.35	0.00			
3	6.66	5.90	5.90	5.90	5.90	5.90		5.90	5.92	0.02			
4	6.66	5.90	5.90	5.44	5.35	5.35		5.35	5.35	0.00			
5	6.66	5.90	5.90	5.44	5.44	5.44		5.44	5.47	0.03			
6	6.66	5.90	5.90	5.90	5.90			5.90	6.19	0.29	5.91		
7	6.66	5.90	5.44	5.44	5.35	5.35	5.35	5.35	6.13	0.78	5.36		
8	6.66	5.90	5.90	5.90	5.90	5.66	5.44	5.44	5.50	0.06			
9	6.66	5.90	5.90	5.90	5.44	5.44	5.35	5.35	5.36	0.01			
10	6.66	5.90	5.90	5.90	5.53	5.44	5.44	5.44	5.54	0.10	5.44		
11	6.66	5.90	5.90	5.90	5.90	5.90		5.90	5.94	0.04			
12	6.66	5.90	5.90	5.59	5.35	5.35	5.35	5.35	5.36	0.01			
13	6.66	5.90	5.90	5.90				5.90	6.06	0.16	6.05	6.01	5.96
14	6.66	5.97	5.90	5.90	5.90	5.90	5.90	5.90	5.97	0.07			
15	6.66	5.90	5.81	5.44	5.35	5.35		5.35	5.89	0.54	5.7	5.69	5.36
16	6.66	5.90	5.90					5.90	6.06	0.16	6.01	5.96	
17	6.56	5.90	5.90	5.44	5.44	5.44	5.44	5.44	5.51	0.07			
18	6.66	5.90	5.90	5.90	5.90	5.35	5.35	5.35	5.36	0.01			
19	6.66	5.90	5.90	5.90	5.90	5.90		5.90	6.21	0.31	6.06	5.95	
20	6.66	5.90	5.90					5.90	6.08	0.18	5.98		

is able to reduce the approximated GMF vastly by explicitly minimising it (1.06mV on average for all instances).

From results in Table 4.4, it is clear that postprocessing is necessary in 9 of the 20 instances to reduce the error to an irrelevant magnitude. In five cases one extra shift is sufficient, in two cases we need two extras, and in two cases we need three extras. The postprocessing technique seems effective, as evidenced by the results in Table 4.4. Notably, all errors are positive (FRE-scores are larger than RHE_4-scores), which can seem surprising given results in Table 4.2. However, the global maximum fatigue scores obtained after FRE are not necessarily the same nurses and shifts that are estimated to have the maximum global fatigue by the RHE_4. With 30 nurses and 42 days in a roster, there could potentially be large true fatigue scores in seemingly arbitrary parts of the final roster, as became clear when analyzing errors in Section 4.3.3. With this in mind, it makes sense that errors tend to be positive.

Results in Table 4.5 demonstrate that focusing only on the approximated values from our lookup-table (in practise this means focusing on shorter shift patterns) does not guarantee against producing tiring schedules for some nurses (see e.g. Instance 7), but it proves to be a highly useful proxy. The longer shift patterns are considered, the better the proxy, as implied by results in Section 4.3.3. Furthermore, managers can ensure a high-quality roster if they combine this approach with a full evaluation of rosters after they are created and also use extras for single shifts. Our results suggest this number can be small, and in many cases 0.

Roster insights

The large plateaus in Figure 4.5 stand out as an interesting characteristic, which provides some insights to the structure of the NRPwF when minimising the global maximum fatigue. Note that 5.90 is the approximated fatigue score for a nurse of biological profile 1 working the shift pattern {O,N,N,N} or the pattern {D,N,N,N}. In other words, to reduce the objective function value, all such patterns for nurses of biological profile 1 must be removed without introducing an even higher fatigue score somewhere in the roster. This is an example of a pattern that can occur frequently, as the probability of a nurse having biological profile 1 is 64%, and three consecutive nights is the maximum number of nights.

However, as the algorithm has removed high fatigue scores from the roster, this has affected the number of different shifts worked by nurses of different biological parameters. In Table 4.5 we present some key information on the average number of different shifts worked by nurses of each biological profile. 30 nurses working 42 days and a minimum of 7 day shifts, 5 evening shifts, and 5 night shifts per day, implies each nurse should work a minimum of 9.8 day shifts and 7 evening and night shifts each, as seen in the last row in Table 4.5.

Table 4.5: Roster statistics for each of the 9 biological profiles in the 20 instances.

Biological profile	Sleep time	Chrono-type	Avg. Day	Avg. Evening	Avg. Night	Avg. Nr. triple night
1	Normal	Day	10.41	8.21	7.10	0.10
2	Short	Day	6.94	6.10	12.51	0.55
3	Long	Day	13.33	9.82	2.59	0.00
4	Normal	Morning	10.11	7.18	8.44	0.49
5	Short	Morning	5.00	6.00	14.75	2.50
6	Long	Morning	12.83	1.17	11.83	1.33
7	Normal	Evening	10.79	6.49	8.51	0.59
8	Short	Evening	14.56	2.22	9.22	1.67
9	Long	Evening	16.40	8.20	1.40	0.00
Avg. Min.			9.8	7	7	

It is clear from Table 4.5 that nurses of different biological profiles are assigned very different numbers of different shifts. Nurses of the most typical biological profile 1 work on average 10.41 day shifts, 8.21 evening shifts, and 7.10 night shifts. They thus work approximately their share of each shift type. Nurses of biological profile 1 work on average 0.10 of the tiring triple night-patterns. The low number of triple night-patterns is notable, and make sense seeing that most instances in Table 4.4 have no approximated fatigue scores of 5.90.

Nurses of all the nine biological profiles have little negative impact on their sleep from working day shifts, and it is thus natural to compare profiles by looking at the number of night shifts and evening shifts they are able to perform without producing high fatigue scores. If we analyse the values in Table 4.5, we can see that the sleep times tend to affect the number of shifts worked during hours the nurse would otherwise sleep. Nurses that have short sleep times, i.e. nurses of biological profiles 2 (12.51 nights), 5 (14.75 nights), and 8 (9.22 nights), work more night shifts than the minimum requirement per nurse.

On the other hand, nurses with long sleep times work few night shifts on average (profile 3 works 2.59 and profile 9 works 1.40) with the exception of the morning chronotype profile 6 (11.83 nights). The case of profile 6 is interesting, because it contradicts a notion of a straight-forward relation between length of sleep times and the frequency of night shifts. However, we can see that the number of evening shifts worked by nurses of biological profile 6 is 1.17. This could mean that for nurses with a morning chronotype, it is more problematic to work evening shifts and easier to work night shifts than for nurses of other chronotypes. Intuitively this makes sense, as morning type sleepers go to bed early and as a result this can make evening shifts more challenging than for other chronotypes. Also, with most other biological profiles clearly favouring evening shifts over night shifts, it is practical not to assign evening shifts to nurses of profile 6 from a combinatorial perspective.

It seems that when minimising the global maximum fatigue, we must take special notice of the needs of nurses with long sleep times and customize the rosters for them. This entails treating nurses differently in order for them to be similarily fatigue when they are the most tired, possibly denouncing the simplified fairness rules typically presented in rostering literature, where e.g. nurses are assigned the same number of the unpopular night shifts. The customization could both entail ensuring sufficient off-days and rest times, but it is also notable that the chronotype of a nurse seems to decide which shifts are most disadvantageous to the nurses.

4.6.3 The effects of increased staff levels on maximum fatigue levels

To analyze the effect of staff size on the global maximum fatigue score in a roster, we take Instances 1-20 solved in Section 4.6.2 and add one or two full-time nurses of biological profile 1 to evaluate the effects. The instances with extra full-time nurses are simply referred to with +1 or +2 in subscripts, e.g. Instance 1_{+1}, RHE_{4+1} or FRE_{+1}. Results based on approximated fatigue scores are presented in Table 4.6, while values based on FREs are presented in Table 4.7.

Results in Table 4.6 can provide insights to the magnitude in the decrease of fatigue scores from increasing staff numbers. Most notably, there are large improvements for some instances, see e.g. RHE_4-RHE_{4+1} for instances 6 and 19, or RHE_{4+1}-RHE_{4+2} for instance 11. However, to be certain of the practical impacts of increasing the staff levels, we must evaluate the true fatigue scores for each case as well. Results using true fatigue scores are presented in Table 4.7.

In Table 4.7 we present true fatigue scores for all 20 instances after postprocessing.[2]

Despite errors occurring when performing FREs, there is only one case of increased objective function values when calculating the FRE_3^{PP}-FRE^{PP} (Instance 15), which is lower than the 0.10 threshold of relevance. Results in Table 4.7 thus seem realistic.

As in Table 4.6, there are some instances where the global maximum fatigue scores are reduced greatly in Table 4.7, while there are others that have no or very little improvement. From the average values in the last row, we can see a slightly larger average improvement from adding the first nurse in column FRE_{+1}^{pp} compared to FRE^{pp}, than comparing the two later columns with their priors (FRE_{+1}^{pp} to FRE_{+2}^{pp} and FRE_{+2}^{pp} to FRE_{+3}^{pp}), but average differences in GMF from adding a nurse are generally small. It

[2] To evaluate the effect of adding full-time nurses to instances, we could look at both the FRE-values without post-processing and the FRE-values with post-processing. However, ignoring post-processing entails assuming managers would ignore surprisingly high fatigue scores, and we believe this is unrealistic. We thus compare the cases with post-processing, and use pp in superscript to mark this (e.g. FRE_{+1}^{pp}), but note that the number of extra shifts introduced in post-processing varies, from 0 (most common) up to six (the maximum number of added off-shifts).

Table 4.6: Approximated values of objective functions after running our algorithm on Instances 1-20 with zero, one and two full-time extra nurses added to the staff. The decrease in global maximum fatigue score is provided in the last columns. Average values are provided in the last row.

Instance	RHE_4	RHE_{4+1}	RHE_{4+2}	RHE_4-RHE_{4+1}	RHE_{4+1}-RHE_{4+2}
1	5.44	5.35	5.27	0.09	0.08
2	5.35	5.35	5.35	0.00	0.00
3	5.90	5.69	5.35	0.21	0.34
4	5.35	5.35	5.27	0.00	0.08
5	5.44	5.35	5.35	0.09	0.00
6	5.90	5.35	5.27	0.55	0.08
7	5.35	5.35	5.35	0.00	0.00
8	5.44	5.44	5.35	0.00	0.09
9	5.35	5.35	5.27	0.00	0.08
10	5.44	5.35	5.27	0.09	0.08
11	5.90	5.90	5.35	0.00	0.55
12	5.35	5.35	5.35	0.00	0.00
13	5.90	5.53	5.35	0.37	0.18
14	5.90	5.69	5.69	0.21	0.00
15	5.35	5.35	5.35	0.00	0.00
16	5.90	5.44	5.44	0.46	0.00
17	5.44	5.44	5.35	0.00	0.09
18	5.35	5.35	5.35	0.00	0.00
19	5.90	5.37	5.35	0.53	0.02
20	5.90	5.90	5.53	0.00	0.37
Average	5.59	5.46	5.36	0.13	0.10

is interesting to compare the second column FRE^{pp} containing true fatigue scores after post-processing for the original instances with the last column, as we do in FRE^{pp} -FRE^{pp}_{+3}, where the total improvement in GMF from adding three nurses to the staff is presented.

It is clear that instances with the highest GMF in the original instances FRE^{pp} have the largest improvements. That is, Instances 3, 6, 11, 13, 14, 16, 19, and 20 all have improvements of 0.50 or more, and they all had RHE_4-values of 5.90 in Table 4.6 and FRE^{pp}-values in the region of 5.90-6.00 in Table 4.7. These improvements have quite clearly come as a direct result of being able to remove triple night-patterns for the nurses of a normal biological profile and in some cases adding extra off-shifts to compensate for errors in our 4-day rolling horizon approximation. On the other hand, of the 7 instances that had FRE^{pp}-values in the range 5.30-5.40 (Instance 2, 4, 7, 9, 12, 15, and 18) in the original instance, only three had an improvement of relevant magnitude.

The above mentioned results highlight two interesting insights. Firstly, and perhaps unsurprising to practicing nurses, increased staff levels enable less tiring rosters. Managers should note that avoiding the most undesirable shorter patterns tend to reduce nurse fatigue. Secondly, when biological profiles are as coarsely divided as in our experiments, the effects of removing every occurrence of a short and tiring pattern becomes important. When adding an additional nurse means the last nurse of biological profile 1 no longer has to work any triple night shifts, the GMF is typically reduced by a lot. If adding the additional nurse only reduces the occurrences of triple night shifts among nurses with biological profile 1, the GMF is unchanged or changed within the threshold

Table 4.7: True fatigue scores of running our algorithm on Instances 1-20 with none, one and two full-time extra nurses added to the staff. Average values in the last row.

Instance	FRE^{pp}	FRE^{pp}_{+1}	FRE^{pp}_{+2}	FRE^{pp}_{+3}	FRE^{pp}-FRE^{pp}_{+3}
1	5.51	5.36	5.30	5.19	0.32
2	5.35	5.35	5.35	4.94	0.41
3	5.92	5.80	5.36	5.35	0.57
4	5.35	5.35	5.31	5.22	0.13
5	5.47	5.36	5.35	5.35	0.12
6	5.91	5.38	5.19	5.15	0.76
7	5.36	5.36	5.36	5.36	0.00
8	5.50	5.46	5.47	5.34	0.16
9	5.36	5.37	5.24	4.9	0.46
10	5.44	5.44	5.30	4.92	0.52
11	5.94	5.90	5.36	5.34	0.60
12	5.36	5.35	5.36	5.35	0.01
13	5.96	5.36	5.36	5.35	0.61
14	5.97	5.82	5.80	5.35	0.62
15	5.36	5.40	5.36	5.39	-0.03
16	5.96	5.49	5.47	5.44	0.52
17	5.51	5.43	5.47	5.35	0.16
18	5.36	5.35	5.36	5.36	0.00
19	5.95	5.37	5.35	5.35	0.60
20	5.98	5.97	5.59	5.43	0.55
Average	5.63	5.48	5.39	5.27	0.35

of relevance. However, in reality, every individual's biology will differ to some extent, and if this information was truly available and incorporated in the fatigue model, there would likely be small reductions in the GMF in cases where our model only shows a reduction in occurrences of the GMF. It is therefore reasonable to look at the average values of reduction of the GMF in Table 4.7 to estimate the effect of adding one additional nurse to the staff. Thus, the average values of 0.14, 0.11 and 0.10 mV decrease in GMF per added staff are likely reasonably close to the actual decrease one can expect from adding a full-time nurse to a ward of 30 full-time nurses. Simply put, the reduction in fatigue by adding an additional nurse is small, but not irrelevant.

4.6.4 The value of knowing each individual's biotype

While utilizing sleep models in rostering is in itself uncommon at most real-life hospital wards, the introduction of different biological profiles is especially novel. To evaluate the impact of it, we run our algorithm minimising the global fatigue score for a set of 30 nurses that all have the normal biological profile 1. When the final roster is produced, we perform new FREs on the final roster, this time applying other biological profiles to all nurses. That is, all nurses are evaluated assuming biological profile 2, before all nurses are evaluated assuming biological profile 3, etc. Essentially we test how well we can minimise global fatigue scores without taking into account individual biology.

In Table 4.8 the global maximum fatigue scores are provided. Comparing the values in this column across all biological profiles, we see that the standard profile 1 has a global maximum fatigue score of 6.00, which is near the average maximum global fatigue score across all biological profiles of 5.96.

Table 4.8: Key fatigue score statistics for the same roster where all nurses are assumed to have the biological profile in the leftmost column. The roster was produced by our algorithm minimising the global maximum fatigue score for 30 nurses, all with biological profile 1.

Biological profiles	Sleep time	Chrono-type	Global maximum fatigue score
1	Normal	Day	6.00
2	Short	Day	5.64
3	Long	Day	6.65
4	Normal	Morning	5.62
5	Short	Morning	5.05
6	Long	Morning	6.38
7	Normal	Evening	6.09
8	Short	Evening	5.26
9	Long	Evening	7.00
Average			5.96

Table 4.8 provides some pointers to which profiles contribute to increasing and reducing the fatigue. The highest profiles with global maximum fatigue scores higher than biological profile 1 are profiles 3 (6.65), 6 (6.38) and most notably 9 (7.00). Nurses with these three biological profiles have in common their long sleep time. This indicates that the long sleep time is a key characteristic of the nurses that are easily fatigued, which corresponds to results in Section 4.6.2. In those experiments, long time sleepers were spared the most tiring shifts. In this case, such individual customization was not made, and the global maximum fatigue scores of long time sleepers is far higher than in any of the rosters produced in Section 4.6.2 as a result. Similarly, short time sleepers tend to have lower fatigue scores than biological profile 1 had in Table 4.8, with scores 5.64 for profile 2, 5.05 for profile 5, and 5.26 for profile 8.

We note that the morning chronotype nurses all have a lower global maximum fatigue score than the day and evening chronotype nurses with the same sleep times (profile 4 has lower global maximum fatigue score than profiles 1 and 7, etc.) This is interesting, especially knowing that rosters were created to minimise the global maximum fatigue of nurses with a day chronotype. This indicates that it is advantageous for a nurse to have a morning chronotype over a day chronotype, although results would be dependant on shift definitions and commuting. The difference between day and evening chronotypes is more unclear, as global maximum fatigue scores are 6.00, 5.64, and 5.82 for day chronotypes and 6.09, 5.26, and 7.00 for evening chronotypes in Table 4.8.

By revisiting results in Section 4.6.2 of running our 20 instances in Table 4.4, an interesting realisation occurs, that the fatigue score of the normal biological profile 1 in Table 4.8 is in fact quite poor. The approximated GMF for the case of 30 nurses with biological profile 1 is 5.90 (triple night pattern) and the true GMF is 6.00.[3] The poor results for 30 nurses of the normal biological profile 1 imply that variations in biological profiles are beneficial when minimising the GMF. With the insights acquired in Section

[3]The error is just below 0.10 in this roster, and thus no postprocessing occurred. Some rosters were postprocessed in Table 4.4, making a direct comparison of the GMF between rosters imperfect.

4.6.2 in mind, it seems nurses with different chronotypes act as complimentary resources at the ward, an interesting notion for those looking to recruit new shift workers.[4]

4.7 Conclusions and future work

We have presented and formalised the Nurse Rostering Problem with Fatigue by approximating and incorporating an advanced sleep model in a general Nurse Rostering Problem. An algorithm combining Mixed Integer Programming and Constraint Programming to form a Large Neighborhood Search was introduced. The algorithm created high-quality rosters minimising the global maximum fatigue for all nurses. Instances were created as 6-week rosters using real-life data. We further demonstrated the use of a post-processing technique in cases where approximation errors are larger than a threshold of relevance.

4.7.1 Technical Outcomes

Nurse rostering is a time-consuming task, and poor rostering choices can easily result in fatigue and medical errors. This paper described how a fatigue model can be successfully integrated with a nurse rostering model and solved to a practical scale (30 nurses over 6 weeks) using a hybrid algorithm. Current systems implement rules, such as the Safe Work Australia (2013) guidelines, to avoid fatiguing rosters. The results show that, compared with a roster which merely implements such a generic set of rules, levels of fatigue can be reduced by more than $1.00mV$ with an integrated fatigue model in the rostering system.

4.7.2 Managerial insights

This work demonstrates that prospective use of advanced sleep models in Nurse Rostering is realistic. In practice, managers should be aware of the potential benefits for nurse health and patient safety, and innovative health care institutions should consider pilot projects with real-life implementation.

Among other results, our research demonstrates two closely linked insights:

1. Minimising the global maximum fatigue for nurses of different biological profiles entails assigning them different numbers of shifts during evening and night time
2. Without customisation to individual nurses' biology, we cannot expect to create rosters that limit the highest fatigue levels considerably

In a practical setting this means that managers must treat nurses differently in order to minimise the global maximum fatigue. This entails grappling with an idea of what fairness is in rostering. While it is easy and tempting to treat every nurse exactly the same irrespective of their reaction to working round the clock, this does not suffice if managers wish to create rosters that focus on nurse health and patient safety.

Furthermore, our results support the intuitive notion that biological parameters linked to sleep time affect the fatigue experienced from shift work. That is, nurses who are

[4] An interesting parallel to the notion of using complementary chronotypes in scheduling exists in Walker (2017), where the reason for differences in chronotypes are theorised from an evolutionary perspective. Walker (2017) argue that humans likely evolved to co-sleep as families or even tribes, and that different chronotypes would reduce the time they were collectively vulnerable, thus enhancing safety.

rested and uninterrupted and sleep approximately 5 hours are more resilient to shift work than those who sleep 7 or 9 hours. Furthermore, the fatigue experienced from working at different hours seem to be affected by the chronotype of a nurse, and results from minimising the global maximum fatigue of nurses demonstrate that while day and evening chronotype nurses should not be assigned a lot of night shifts, the evening chronotype nurses should not be assigned too many evening shifts. Furthermore, our results indicate that nurses' different chronotypes can prove complimentary when creating rosters.

Our research demonstrates how minimally increasing the staff levels makes it possible to decrease the global maximum fatigue levels. Our results indicate that the average of global maximum fatigue scores decrease by a small but relevant magnitude for each additional nurse, assuming the additional nurse has a normal biology in terms of sleep.

4.7.3 Future work

The general approach for incorporating the sleep model, where the approximation is created through a look-up table, is likely useful in Operations Research within other areas of application. Furthermore, as sleep models are improved, research on rostering using sleep models should be updated and improved. Noting the vast impact of variations in two of the most common biological profiles, it would be very interesting to see the impact of more refined biology in nurses. From a rostering perspective, it would be particularly interesting if future sleep models were able to take into account how other factors such as individuals' social life etc. affects sleep patterns. There are examples of attempts of this, see e.g. Postnova et al. (2012). For many nurses, work-life balance includes a preference towards following the circadian rhythm of the rest of society, to enable daily chores and meeting others with a more standard work schedule.

Real-life implementation and evaluation of the impact on nurse health and patient safety would be very useful. Furthermore, the NRPwF can be expanded to include other aspects of rostering, e.g. individual preferences, different fairness measures or personnel costs. Alternative objective functions to the minimisation of the global maximum fatigue would also be interesting to introduce and analyse in detail. The current algorithm works well for problems of the size presented in this work, but run times are typically somewhat below 1 hour for solving one instance. Thus, it would be interesting to see work on alternative algorithms reducing the computation time. There could be potential in exploring more comprehensive use of regular constraints to cut away all shift patterns that imply fatigue scores above the current global maximum fatigue score.

Introducing other objective functions to the NRPwF would likely imply the need for a new solution method, so these topics of future research are closely intertwined. The current post-processing method is realistic, but simple. It would be interesting to see work where alternative strategies to simply removing shifts and assigning an off-day are considered.

We want to thank Dr. Andrew Phillips for providing us with full access and insights to his sleep model. We also wish to thank Dr. Pierre le Bodic for useful discussions and feedback.

Conflict of interest

The authors declare that they have no conflict of interest.

Bibliography

T. Åkerstedt, S. Folkard, and C. Portin. Predictions from the three-process model of alertness. *Aviation, space, and environmental medicine*, 75(3):A75–A83, 2004.

E. I. Ásgeirsson and G. L. Sigurðardóttir. Near-optimal MIP solutions for preference based self-scheduling. *Annals of Operations Research*, 239(1):273–293, Apr 2016.

M. A. Awadallah, A. L. Bolaji, and M. A. Al-Betar. A hybrid artificial bee colony for a nurse rostering problem. *Applied Soft Computing*, 35:726 – 739, 2015.

G. Beddoe, S. Petrovic, and J. Li. A hybrid metaheuristic case-based reasoning system for nurse rostering. *Journal of Scheduling*, 12(2):99, 2009.

J. Beliën and E. Demeulemeester. On the trade-off between staff-decomposed and activity-decomposed column generation for a staff scheduling problem. *Annals of Operations Research*, 155(1):143–166, 2007.

D. Boivin and P. Boudreau. Impacts of shift work on sleep and circadian rhythms. *Pathologie Biologie*, 62(5):292 – 301, 2014.

A. A. Borbély. A two process model of sleep regulation. *Hum neurobiol*, 1(3):195–204, 1982.

Z. E. Bowden. *Behavioral Logistics and Fatigue Management in Vehicle Routing and Scheduling Problems.* PhD thesis, Virginia Tech, 2016.

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

E. K. Burke, T. Curtois, G. Post, R. Qu, and B. Veltman. A hybrid heuristic ordering and variable neighbourhood search for the nurse rostering problem. *European journal of operational research*, 188(2):330–341, 2008.

G. G. Chu. *Improving combinatorial optimization.* PhD thesis, University of Melbourne, 2011.

D. Dawson and A. Fletcher. A quantitative model of work-related fatigue: background and definition. *Ergonomics*, 44(2):144–163, 2001.

D. Dawson, Y. I. Noy, M. Härmä, T. Åkerstedt, and G. Belenky. Modelling fatigue and the use of fatigue models in work settings. *Accident Analysis & Prevention*, 43(2):549 – 564, 2011.

A. Dohn and A. Mason. Branch-and-price for staff rostering: An efficient implementation using generic programming and nested column generation. *European Journal of Operational Research*, 230(1):157–169, 2013.

R. C. Dorf and R. H. Bishop. *Modern control systems.* Pearson (Addison-Wesley), 1998.

N. R. Downing. *Scheduling and rostering with learning constraint solvers.* PhD thesis, The University of Melbourne, 2016.

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

D. B. Forger, M. E. Jewett, and R. E. Kronauer. A simpler model of the human circadian pacemaker. *Journal of Biological Rhythms*, 14(6):533–538, 1999.

B. Fulcher, A. Phillips, and P. Robinson. Quantitative physiologically based modeling of subjective fatigue during sleep deprivation. *Journal of Theoretical Biology*, 264(2):407 – 419, 2010.

A. Goel and T. Vidal. Hours of service regulations in road freight transport: An optimization-based international assessment. *Transportation Science*, 48(3):391–412, 2014.

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

R. G. Hughes and A. E. Rogers. Are you tired?: Sleep deprivation compromises nurses' health—and jeopardizes patients. *AJN The American Journal of Nursing*, 104(3):36–38, 2004.

S. R. Hursh, D. P. Redmond, M. L. Johnson, D. R. Thorne, G. Belenky, T. J. Balkin, W. F. Storm, J. C. Miller, and D. R. Eddy. Fatigue models for applied research in warfighting. *Aviation, space, and environmental medicine*, 75(3):A44–A53, 2004.

M. Ingre, W. Van Leeuwen, T. Klemets, C. Ullvetter, S. Hough, G. Kecklund, D. Karlsson, and T. Åkerstedt. Validating and extending the three process model of alertness in airline operations. *PLOS ONE*, 9(10):1–15, 10 2014.

R. Jamshidi. Stochastic human fatigue modeling in production systems. *Journal of Industrial and Systems Engineering*, 12(1):270–283, 2019.

R.-C. Lin, M. Y. Sir, E. Sisikoglu, K. Pasupathy, and L. M. Steege. Optimal nurse scheduling based on quantitative models of work-related fatigue. *IIE Transactions on Healthcare Systems Engineering*, 3(1):23–38, 2013.

C. C. Liu and T. C. Wang. Optimal aircraft maintenance crews work shifts with integer programming. In *Applied Mechanics and Materials*, volume 319, pages 479–484. Trans Tech Publ, 2013.

M. Mallis, S. Mejdal, T. T Nguyen, and D. Dinges. Summary of the key features of seven mathematical models of human fatigue and performance. *Aviation, space, and environmental medicine*, 75:A4–14, 04 2004.

P. McCauley, L. Kalachev, A. Smith, G. Belenky, D. Dinges, and H. Van Dongen. A new mathematical model for the homeostatic effects of sleep loss on neurobehavioral performance. *Journal of theoretical biology*, 256:227–239, 2 2009.

J.-P. Métivier, P. Boizumault, and S. Loudni. Solving nurse rostering problems using soft global constraints. In *International Conference on Principles and Practice of Constraint Programming*, pages 73–87. Springer, 2009.

H. Meyer auf'm Hofe. Solving rostering tasks as constraint optimization. In E. Burke and W. Erben, editors, *Practice and Theory of Automated Timetabling III*, pages 191–212, Berlin, Heidelberg, 2001. Springer Berlin Heidelberg.

G. Michalos, S. Makris, L. Rentzos, and G. Chryssolouris. Dynamic job rotation for workload balancing in human based assembly systems. *CIRP Journal of Manufacturing Science and Technology*, 2(3):153 – 160, 2010.

G. Michalos, S. Makris, and G. Chryssolouris. The effect of job rotation during assembly on the quality of final product. *CIRP Journal of Manufacturing Science and Technology*, 6(3):187–197, 2013.

F. Mischek and N. Musliu. Integer programming model extensions for a multi-stage nurse rostering problem. *Annals of Operations Research*, 275(1):123–143, Apr 2019.

S. Muecke. Effects of rotating night shifts: literature review. *Journal of Advanced Nursing*, 50(4):433–439, 2005.

N. Nethercote, P. J. Stuckey, R. Becket, S. Brand, G. J. Duck, and G. Tack. Minizinc: Towards a standard cp modelling language. In C. Bessière, editor, *Principles and Practice of Constraint Programming – CP 2007*, pages 529–543, Berlin, Heidelberg, 2007. Springer Berlin Heidelberg.

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

A. Phillips and P. Robinson. A quantitative model of sleep-wake dynamics based on the physiology of the brainstem ascending arousal system. *Journal of Biological Rhythms*, 22(2):167–179, 2007.

A. J. K. Phillips, P. Y. Chen, and P. A. Robinson. Probing the mechanisms of chronotype using quantitative modeling. *Journal of Biological Rhythms*, 25(3):217–227, 2010.

R. Pizarro, G. Rivera, R. Soto, B. Crawford, C. Castro, and E. Monfroy. Constraint-based nurse rostering for the valparaíso clinic center in chile. In *International Conference on Human-Computer Interaction*, pages 448–452. Springer, 2011.

S. Postnova, A. Layden, P. A. Robinson, A. J. Phillips, and R. G. Abeysuriya. Exploring sleepiness and entrainment on permanent shift schedules in a physiologically based model. *Journal of Biological Rhythms*, 27(1):91–102, 2012.

S. Postnova, P. A. Robinson, and D. D. Postnov. Adaptation to shift work: Physiologically based modeling of the effects of lighting and shifts' start time. *PloS one*, 8(1):e53379, 2013.

S. Postnova, S. W. Lockley, and P. A. Robinson. Sleep propensity under forced desynchrony in a model of arousal state dynamics. *Journal of Biological Rhythms*, 31(5): 498–508, 2016.

S. Postnova, S. W. Lockley, and P. A. Robinson. Prediction of cognitive performance and subjective sleepiness using a model of arousal dynamics. *Journal of Biological Rhythms*, 33(2):203–218, 2018.

P. Rajdev, D. Thorsley, S. Rajaraman, T. L. Rupp, N. J. Wesensten, T. J. Balkin, and J. Reifman. A unified mathematical model to quantify performance impairment for both chronic sleep restriction and total sleep deprivation. *Journal of Theoretical Biology*, 331:66 – 77, 2013.

T. G. Raslear, S. R. Hursh, and H. P. Van Dongen. Predicting cognitive impairment and accident risk. In *Progress in brain research*, volume 190, pages 155–167. Elsevier, 2011.

J.-C. Régin. Global constraints: A survey. In *Hybrid optimization*, pages 63–134. Springer, 2011.

G. Roach, A. Fletcher, and D. Dawson. A model to predict work-related fatigue based on hours of work. *Aviation, space, and environmental medicine*, 75:A61–9; discussion A70, 04 2004.

E. Rönnberg and T. Larsson. Automating the self-scheduling process of nurses in swedish healthcare: a pilot study. *Health Care Management Science*, 13(1):35–53, Mar 2010.

Safe Work Australia. Guide for managing the risk of fatigue at work. `https://www.safeworkaustralia.gov.au/system/files/documents/1702/managing-the-risk-of-fatigue.pdf`, 2013. Accessed 22 August 2019.

L. F. Shampine and M. W. Reichelt. The matlab ode suite. *SIAM journal on scientific computing*, 18(1):1–22, 1997.

A. C. Skeldon, A. J. Phillips, and D.-J. Dijk. The effects of self-selected light-dark cycles and social constraints on human sleep and circadian timing: a modeling approach. *Scientific reports*, 7:45158, 2017.

M. A. St. Hilaire, M. Rüger, F. Fratelli, J. T. Hull, A. J. K. Phillips, and S. W. Lockley. Modeling Neurocognitive Decline and Recovery During Repeated Cycles of Extended Sleep and Chronic Sleep Deficiency. *Sleep*, 40(1), 12 2016.

State Government of Victoria. Alfred ICU expands victorian government health information. `http://www.health.vic.gov.au/healthvictoria/nov19/icu.htm`, 2019. Accessed Nov. 11 2020.

J. E. Stone, X. L. Aubert, H. Maass, A. J. Phillips, M. Magee, M. E. Howard, S. W. Lockley, S. M. Rajaratnam, and T. L. Sletten. Application of a limit-cycle oscillator model for prediction of circadian phase in rotating night shift workers. *Scientific reports*, 9(1):1–12, 2019.

J. E. Stone, S. Postnova, T. L. Sletten, S. M. Rajaratnam, and A. J. Phillips. Computational approaches for individual circadian phase prediction in field settings. *Current Opinion in Systems Biology*, 22:39 – 51, 2020.

K. Swaminathan, E. B. Klerman, and A. J. K. Phillips. Are individual differences in sleep and circadian timing amplified by use of artificial light sources? *Journal of Biological Rhythms*, 32(2):165–176, 2017.

A. P. Tvaryanas and N. L. Miller. Human systems integration domain trade-offs in optimized manning-the task effectiveness scheduling tool. Technical report, Naval Postgraduate School Monterey CA, 2010.

M. Walker. *Why we sleep: The new science of sleep and dreams.* Penguin UK, 2017.

T.-C. Wang and G.-C. Ke. Fatigue minimization work shift scheduling for air traffic controllers. *International Journal of Automation and Smart Technology*, 3(2):91–99, 2013.

T.-C. Wang and C.-C. Liu. Optimal work shift scheduling with fatigue minimization and day off preferences. *Mathematical Problems in Engineering*, 2014, 2014.

Appendices

A Symbol directory and model

A.1 Indices

Symbol	Description
n	Nurse
s	Shift
s^D	Day shift
s^E	Evening shift
s^N	Night shift
s^O	Off-shift
t, τ	Day
b	Biological profile
w	Indication of night shift one day prior

A.2 Sets

Symbol	Description	Range
$\mathcal{N}$	Nurses	$1 \dots 30$
$\mathcal{T}$	Days	$1 \dots 42$
$\mathcal{T}^S$	Sundays	$7, 14 \dots 42$
$\mathcal{S}$	Shifts	$\{s^D, s^E, s^N, s^O\}$
$\mathcal{S}^W$	Work shifts	$\{s^D, s^E, s^N\}$

A.3 Parameters

Symbol	Description	Value
$\underline{P}^C_{s^D}$	Minimum staff coverage of work shift day	7
$\underline{P}^C_{s^E}$	Minimum staff coverage of work shift day	5
$\underline{P}^C_{s^N}$	Minimum staff coverage of work shift day	5
$\underline{H}$	Minimum total work hours	210
$\overline{H}$	Maximum total work hours	228
$P^H_{s^D}$	Length of day shift [hours]	8.5
$P^H_{s^E}$	Length of evening shift [hours]	8.5
$P^H_{s^N}$	Length of night shift [hours]	8.5
$\overline{P}^{CN}$	Maximum number of consecutive night shifts	3
$\overline{P}^{CD}$	Maximum number of consecutive work days	6
$\overline{H}$	Maximum number of weekly work hours	50
$\overline{P}^{CW}$	Maximum number of consecutive work weekends	2
$\overline{P}^{CW}$	Maximum number of days between two off-days	10

A.4 Decision variables

$y_{nst} \in \{0,1\}$ is a binary variable determining if shift $s \in \mathcal{S}^W$ is worked by nurse $n \in \mathcal{N}$ on day $t \in \mathcal{T}$. y_{nst} constitutes the roster in the NRPwF.
$z_{nt} \in \{0,1\}$ is a binary auxiliary variable indicating if nurse $n \in \mathcal{N}$ works during a two-day period ending on a day $t \in \mathcal{T}$. f_n^{Max} is a continuous variable equal to the maximum fatigue score value of nurse $n \in \mathcal{N}$ for the entire planning period.

A.5 Model

Coverage

$$\sum_{n\in\mathcal{N}} y_{nst} \geq \underline{P}_s^C, \qquad t \in \mathcal{T}, s \in \mathcal{S}^W \tag{19}$$

Short-term rest

$$\sum_{s\in\mathcal{S}} y_{nst} = 1, \qquad n \in \mathcal{N}, t \in \mathcal{T} \tag{20}$$

$$\sum_{\tau=t-\overline{P}^{CN}}^{t} y_{ns^N\tau} \leq \overline{P}^{CN}, \qquad n \in \mathcal{N}, t \in \mathcal{T} \tag{21}$$

$$y_{ns^N(t-1)} + y_{ns^Dt} + y_{ns^Et} \leq 1, \qquad n \in \mathcal{N}, t \in \mathcal{T} \tag{22}$$

$$y_{ns^E(t-1)} + y_{ns^Dt} \leq 1, \qquad n \in \mathcal{N}, t \in \mathcal{T} \tag{23}$$

$$y_{ns^N(t-2)} + y_{ns^O(t-1)} + y_{ns^Nt} \leq 2, \qquad n \in \mathcal{N}, t \in \mathcal{T} \tag{24}$$

$$\sum_{s\in\mathcal{S}^W} \sum_{\tau=t-\overline{P}^{CD}}^{t} y_{ns\tau} \leq \overline{P}^{CD}, \qquad n \in \mathcal{N}, t \in \mathcal{T} \tag{25}$$

$$\tag{26}$$

Long-term rest

$$\underline{H} \leq \sum_{s\in\mathcal{S}} \sum_{t\in\mathcal{T}} P_s^H y_{nst} \leq \overline{H}, \qquad n \in \mathcal{N} \tag{27}$$

$$\sum_{s\in\mathcal{S}^W} \sum_{\tau=t-6}^{t} P_s^H y_{ns\tau} \leq \overline{H}^W, \qquad n \in \mathcal{N}, t \in \mathcal{T}^S \tag{28}$$

$$2z_{nt} - y_{ns^N(t-2)} - \sum_{s\in\mathcal{S}^W} y_{ns(t-1)}$$
$$- y_{ns^Dt} - y_{ns^Et} \geq 0, \qquad n \in \mathcal{N}, t \in \mathcal{T} \tag{29}$$

$$\sum_{\tau=0}^{\overline{P}^{CW}} z_{n(t-7\tau)} \leq \overline{P}^{CW}, \qquad n \in \mathcal{N}, t \in \mathcal{T}^S \tag{30}$$

$$\sum_{\tau=t-\overline{P}^Z}^{t} z_{n\tau} \leq \overline{P}^Z, \qquad n \in \mathcal{N}, t \in \mathcal{T} \tag{31}$$

Objective function

$$f^{GM} - f_{nt} \geq 0 \qquad n \in \mathcal{N}, t \in \mathcal{T} \tag{32}$$

$$\text{Minimise } f^{GM} \tag{33}$$

A.6 The relation between the fatigue score variable and the lookup-table

Assume the lookup-table is described by parameter $P^{Score}_{bs_1s_2s_3s_4w}$, where index b is the biological profile, indices s_1, s_2, s_3, and s_4 are the shifts assigned on days $t-3$, $t-2$, $t-1$, and t of the evaluation pattern, and the index w is 1 if nurse n worked a night shift prior to the evaluation pattern, 0 else. The constraints below ensure a linear relation between variables f_{nt} and the lookup-table of fatigue scores. We also introduce the binary auxiliary variable $p_{ns_1s_2s_3s_4wt}$, indicating if nurse n works the pattern indicated by the s-indices prior to a night shift or not on day t. Furthermore, we introduce the sets $\mathcal{B}$ consisting of all biological profiles and the sets $\mathcal{N}^B_b$ consisting of all nurses of biological profile b.

$$p_{ns_1s_2s_3s_41t} - y_{ns^Nt-4} - y_{ns_1t-3} - y_{ns_2t-2} - y_{ns_3t-1} - y_{ns_4t} \leq -4,$$
$$n \in \mathcal{N}, s_1, s_2, s_3, s_4 \in \mathcal{S}, t \in \mathcal{T} \tag{34}$$

$$p_{ns_1s_2s_3s_40t} - \sum_{s \in \mathcal{S} \setminus s^N} y_{nst-4} - y_{ns_1t-3} - y_{ns_2t-2} - y_{ns_3t-1} - y_{ns_4t} \leq -4,$$
$$n \in \mathcal{N}, s_1, s_2, s_3, s_4 \in \mathcal{S}, t \in \mathcal{T} \tag{35}$$

$$\sum_{s_1 \in \mathcal{S}} \sum_{s_2 \in \mathcal{S}} \sum_{s_3 \in \mathcal{S}} \sum_{s_4 \in \mathcal{S}} \sum_{w=0}^{1} p_{ns_1s_2s_3s_40t} = 1, \qquad n \in \mathcal{N}, t \in \mathcal{T} \tag{36}$$

$$f_{nt} - P^{Score}_{bs_1s_2s_3s_4w} p_{ns_1s_2s_3s_4wt} \geq 0, b \in \mathcal{B}, n \in \mathcal{N}^B_b, s_1, s_2, s_3, s_4 \in \mathcal{S}, w \in \{0,1\}, t \in \mathcal{T} \tag{37}$$

Constraints (34) ensure that the variable $p_{ns_1s_2s_3s_41t}$ can only have the value 1 if nurse n works the evaluation pattern defined by the s-indices if it occurs subsequent to a night shift, while constraints (35) ensure the same relation if the evaluation pattern occurs subsequent to a different shift than the night shift. Constraints (36) ensure that the variable $p_{ns_1s_2s_3s_4wt}$ can only be equal to 1 once for every combination of nurses and days, ensuring that the variable $p_{ns_1s_2s_3s_4wt}$ becomes an indicator of the evaluation pattern and prior night shift for nurse n on day t. Constraints (37) ensure the variable f_{nt} cannot be lower than the approximated fatigue score of nurse n in the lookup-table if the nurse works the evaluation pattern and prior night shift defined by s-indices and w on any day t.

B Pseudocode for algorithm

Algorithm 1 Solving the Nurse Rostering Problem with Fatigue

```
Draw set $\mathcal{N}$ with realistic biological profiles according to Table 4.3.
Initialise sets of fatigued nurses $\mathcal{N}^F$, random nurses $\mathcal{N}^R$
Initialise set of days $\mathcal{T}$ in planning period
Initialise empty roster $y$ of size $|\mathcal{N}| \times |\mathcal{T}|$, empty list of rosters $Y$
Initialise empty array of fatigue scores $f$ of size $|\mathcal{N}| \times |\mathcal{T}|$
Initialise variable indicating minimisation of occurrences of GMF $p^{Occ}$
$y, f \leftarrow solve\_feasibility\_problem\_using\_gurobi(y)$
while $!(|Y| \geq 20 \ \wedge \ y = Y(end - 19))$ do
    $\mathcal{N}^F, \mathcal{N}^R \leftarrow \{\}$
    $f^{GM} \leftarrow max(\mathrm{f})$
    for each $n \in \mathcal{N}$ do
        if $max(f(n)) = f^{GM}$ then
            $\mathcal{N}^F \leftarrow \mathcal{N}^F \bigcup n$
        end if
    end for
    if $|\mathcal{N}^F| > 5$ then
        $\mathcal{N}^F \leftarrow rand(\mathcal{N}^F, 5)$
        $p^{Occ} \leftarrow 1$
    else if $|Y| > 2 \ \wedge \ y = Y(end - 2)$ then
        $p^{Occ} \leftarrow 1$
    else
        $p^{Occ} \leftarrow 0$
    end if
    $\mathcal{N}^R \leftarrow rand(\mathcal{N}\text{-}\mathcal{N}^F, 5)$
    $N \leftarrow$ array of zeros of size $|\mathcal{N}| \times |\mathcal{T}|$
    for each $n \in \mathcal{N}$, $t \in \mathcal{T}$ do
        if $r(n,t) = s^O \vee n \in \mathcal{N}^{CF} \bigcup \mathcal{N}^R$ then
            $N(n,t) \leftarrow 1$
        end if
    end for
    if $p^{Occ} = 0$ then
        $(r, f) \leftarrow minimise\_global\_fatigue\_using\_chuffed(f, N)$
    else if $p^{Occ} = 1$ then
        $(r, f) \leftarrow minimise\_occurences\_of\_global\_fatigue\_using\_chuffed(f, N)$
    end if
    $R \leftarrow R + r$
end while
True fatigue scores $f^{FRE} \leftarrow perform\_FRE(r)$
$y, f^{FRE} \leftarrow postprocess(y, f, f^{FRE})$
```

C Pseudocode for postprocessing function

Algorithm 2 Postprocessing function

procedure POSTPROCESS(y,f,f^{FRE})
  Initialise set $\mathcal{N}$ of nurses in y
  Initialise set $\mathcal{T}$ of days in y
  $i \leftarrow 1$
  while $max(f^{FRE}) \geq max(f) + 0.10 \ \wedge \ i \leq 6$ **do**
    for each $n \in \mathcal{N}$, $t \in \mathcal{T}$ **do**
      if $f^{FRE}(n,t) = max(f^{FRE})$ **then**
        if $t > 1 \wedge \ y(n,t-1)! = s^O$ **then**
          $y(n,t-1) \leftarrow s^O$
        else if $t > 2 \wedge \ y(n,t-2)! = s^O$ **then**
          $y(n,t-2) \leftarrow s^O$
        else if $t > 3 \wedge \ y(n,t-3)! = s^O$ **then**
          $y(n,t-3) \leftarrow s^O$
        else$t > 4 \wedge \ y(n,t-4)! = s^O$
          $y(n,t-4) \leftarrow s^O$
        end if
        $i \leftarrow i + 1$
      end if
    end for
    True fatigue scores $f^{FRE} \leftarrow perform\_FRE(y)$
  end while
  return y, f^{FRE}
end procedure

Paper IV

K. K. Klyve and A. N. Gullhav:

Future hospital rostering – experiences with new technology

Accepted for the scientific anthology Norsk arbeidsliv mot 2050 [Norwegian work life towards 2050]

Future hospital rostering – experiences with new technology

Abstract

We describe the introduction of a new decision support tool that exemplifies how technology development and digitalisation will affect the future working life of shift workers in health care, at Norwegian hospitals in particular. The tool was used in two pilot projects at Maternity Ward West at St. Olav's Hospital, and provides valuable insight into the potential that can be realised as well as the challenges to improving roster planning. We interviewed two other actors who work with similar initiatives elsewhere in Scandinavia and drew on their experiences. The results indicate that if such a decision support tool establishes itself as the norm for roster planning, it will mean increased standardisation of both the planning process and of the considerations that are taken into account for different stakeholders. Decisions will be made at a higher level in the health organisations than they are today, which will strengthen the trade union's role in clarification of the premisses for roster planning. Trade unions will have to develop an understanding of how the decision support tools work and communicate this effectively to the individual employees. The tug-of-war between employer, trade union and the individual employee will largely be as before, but the employer and trade union will make more detailed clarifications when adopting rostering agreements. This will increase fairness, but while the individual customisation will be better for the employees, the types of customisation will be more standardised. The opportunities for individual customisation will therefore be limited to the needs that the trade union understands and demands on the employees' behalf.

5.1 Introduction

According to the OECD report Health at a Glance from 2019, Norway differs from other OECD countries in terms of the number of employees in the health service. Norway has the most nurses employed per capita (1.77%) and is ranked between number two and four in the number of doctors per capita (0.47%) (OECD, 2019). Norway is also ranked second and third for public expenditure in the health sector per person and expenditure in the health sector per person among OECD countries respectively (OECD, 2019). The large use of resources carries expectations of the health service's results. In Table 5.1, we list four key indicators for the quality of health services from the OECD (2019), as well as the percentage of employees in health and social services and the total costs per person for health in the columns on the far right. We present the numbers for Norway and some selected countries we believe are comparable; Sweden, Denmark and Iceland.

Table 5.1: Key indicators for the health services, and the percentage of employees and costs, in Norway, Sweden, Denmark and Iceland. For details and definitions, see OECD (2019).

	Life expectancy (years)	Avoidable mortality (per 100 000 people)	Chronic disease morbidity (% of adults)	Self-rated health as poor (% of population aged 15+)	Share of employees in health and social services (% of workforce)	Health expenditure per person (USD PPP)
Norway	82.7	145	5.3	7.2	20.9	6187
Sweden	82.5	144	4.8	5.7	17.3	5447
Denmark	81.2	184	6.4	7.5	17.5	5299
Iceland	82.7	140	5.3	6.4	10.9	4349

Table 5.1 shows that Denmark generally has the weakest results for all the key indicators, while it is more unclear how Norway, Sweden and Iceland should be ranked in comparison with each other. This very limited analysis is not sufficient to determine the effectiveness of the resource utilisation in the different countries. Nevertheless, it is interesting to note that Norway stands out with its extensive staff resources and high costs in the health sector, without this necessarily being reflected in the results. It is therefore appropriate to increase the focus on the use of staff resources, both because of a political desire to reduce costs and because the supply of staff with certain skills is limited regardless of budgets.

In this work we discuss new technology for planning staff resources at hospitals. We refer to this as optimisation-based rostering (OBR). OBR is a type of decision support software for roster planning. In practice, this involves giving a computer with OBR installed the relevant information to create a roster, and the OBR software then solving the major planning problem and presenting a proposed roster. Computers are particularly suited to this task because rostering entails solving complicated combinatorial problems. In the OBR system, so-called *constraints* are defined, which are requirements that can never be breached. Typical constraints include provisions in the Working Environment Act, collective agreements and rostering agreements. The rostering agreement is a mutually binding agreement between union representatives and managers that stipulates local provisions and clarifications for rota work in relation to, for example, rest breaks and weekend work. Provided that there are not too many constraints, or that they are not too extensive, there will typically be an abundance of valid rosters. These rosters are likely to have vastly different qualities, which will also depend on who is reading them. For example, extra staffing on a particular shift is one quality, and granting a request for a day off is another quality. In OBR, this is done by assigning different qualities with weights that are measured against each other in cases where they are mutually exclusive. We refer to these as *weighted qualities*. In order for the OBR system to have a real opportunity to utilise the weighted qualities and make priorities, the constraints must not be too extensive. Otherwise, the system's flexibility to choose between the weighted qualities will be lost.

In 2017 and 2018, we conducted a case study of Maternity Ward West at St. Olav's

Hospital, where we developed a customised OBR solution. The rosters at Maternity Ward West were drawn up for periods of six months (26 weeks). We carried out two *pilot projects* where OBR replaced the traditional, manual planning of both rosters in 2018. We have also conducted tests to evaluate how OBR can be used as a tool for understanding the effect of various more long-term measures. Finally, we interviewed two other actors who work with similar initiatives in Norway and Denmark. This provides a good basis for discussing how staff planning at hospitals, and to some extent other parts of the health service, will develop in the period up to 2050.

OBR is an example of the use of optimisation or operations research to solve rostering problems. Since Warner (1976) presented the opportunity to frame nurse roster planning as an optimisation problem, considerable work has been done in the optimisation of rosters for this occupational group (Burke et al., 2004, den Bergh et al., 2013). Some of the efforts have also been aimed at doctors (Erhard et al., 2018).

Literature on the optimisation of rosters can roughly be divided into two categories; one that focuses on mathematics and solution methods, and one that focuses on solving real-life planning problems. The latter is much less common. In the 64 studies of nurse roster planning in the literature review by den Bergh et al. (2013), 11 were categorised as 'Applied', but only a few described any kind of case study. In contrast, a lot of attention has been given to seeking effective solution methods. Some popular methods are integer programming (Bard and Purnomo, 2005a, Burke et al., 2010), mixed integer programming (Fügener et al., 2015, Yilmaz, 2012), column generation (Bard and Purnomo, 2005b), heuristic methods (Bellanti et al., 2004, Guessoum et al., 2020, Puente et al., 2009) and the use of artificial intelligence (Kumar et al., 2019). In this literature, however, the focus is on stylistic and general rostering problems (Burke et al., 2004). This means that much of the literature does not appear to be connected to real-life roster planning (Kellogg and Walczak, 2007). However, there are some examples of inspiring case studies (Bester et al., 2007, Rönnberg and Larsson, 2010) and systems that have been implemented in several hospitals (Burke et al., 2006). The lack of research into the practical use of tools such as OBR emphasises the need to test the decision support tools in practice, learn from the challenges and clarify future use.

In the next section, we present background information on the case study and explain the method used. We then present and discuss the results from the pilot projects under Results and evaluation. Finally, we present Development up to 2050, where we discuss how the use of OBR will change working life at hospitals and to some extent other health organisations over the next 30 years.

5.2 Pilot testing OBR

This work is based on a case study, and the method for developing and pilot testing OBR is strongly influenced by the case at Maternity Ward West. We will therefore begin by describing some important background information about the case before presenting the methodological approach.

5.2.1 Background

When we started working on the development of OBR at Maternity Ward West, we were largely driven by two motivating factors. One was the aforementioned knowledge gap in the research literature related to the implementation of OBR. The second was that we saw significant potential to make more use of the existing staff resources at hospital departments that practise manual roster planning. We will discuss this potential later in the text. Manual roster planning entails one or more people making the decisions about who will work which shifts in a roster without the aid of a decision support tool. At St. Olav's Hospital, a software called the resource management system (RMS) is used to register desired rotas and comments, as well as to check that the Working Environment Act, collective agreements and rostering agreements are being complied with. The RMS does not, however, provide support for generating rosters, and the rostering process is therefore regarded as manual.

Different variants of two approaches to manual nurse rostering at Norwegian hospitals are mainly used: planning with *basic rotas* and *desired rotas*. When drawing up a basic rota, a manager or planner sets up a roster, and then the employees have the opportunity to swap shifts internally. This represents a top down approach to roster planning, unlike desired rotas. A desired rota is when all employees draw up their ideal rota, i.e. they only plan their own shifts for the entire period. Each person is typically required to include an agreed number of different types of shift, such as night and weekend shifts. Once everyone has submitted their personal desired rota, there will be some shifts that many employees want to work and some shifts that are not popular. A potentially long period of negotiation follows, where the employees have to compromise and change their original rota in order to accommodate operational factors.

Maternity Ward West was using the desired rota method when we introduced OBR. We were therefore aware of the importance of OBR reflecting the employees' wishes to the greatest extent possible. We set ourselves the ambitious goal of our OBR being able to take all the same considerations into account as the manual planning, provided that they could be formulated explicitly.

Figure 5.1 illustrates the flow of information in the planning process as it was *before* our pilot projects were implemented. The collective agreement stipulates that the maternity ward must issue rosters for six-month periods at a time. The combination of the desired rota method and a long planning horizon means that the planning process itself becomes extensive. At Maternity Ward West, rosters were drawn up in stages as described below. For each stage, we specify how many weeks it is until the first working day of the roster. We note that before each planning period, a rostering agreement is made, which serves as the basic premiss for the roster planning.

1. Registering pre-determined shifts in the RMS. 15 weeks.
2. Registering desired rotas in the RMS. 15 weeks.
3. The information from the RMS is retrieved by the operations coordinator and checked. 12 weeks.
4. The operations coordinator assembles the desired rotas into a roster and sends it to the roster group; a small group of employees with responsibility for roster planning. 12 weeks.
5. The roster group draws up a first draft and sends it to the employees for negotiation. 11.5 weeks.
6. The employees negotiate on who needs to compromise and work an unpopular shift. 11 weeks.

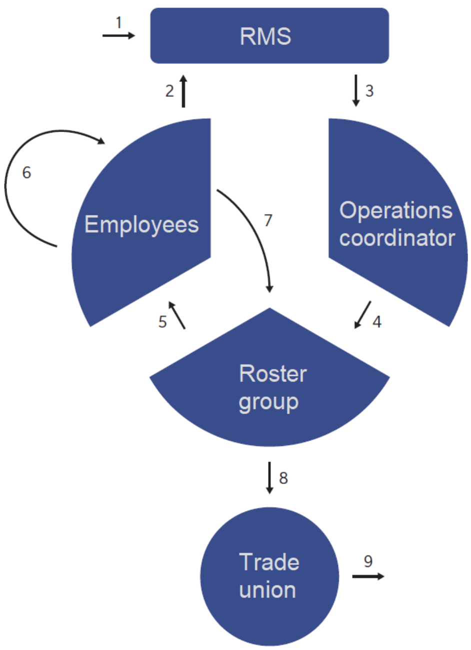

Figure 5.1: The planning process at Maternity Ward West

7. The plan is sent back to the roster group for final changes, and in some cases to force through unpopular shifts. 8 weeks.

8. The final roster is submitted to the trade union, which then provides feedback. 6 weeks.

9. The roster is issued to the employees. 4 weeks. When using OBR, stages 5, 6 and 7 are no longer necessary as the OBR system drafts a roster, which is then sent to the operations coordinator. The operations coordinator can then give feedback, which can be incorporated into the OBR system, before a new roster is produced. If necessary, the feedback and changes process can be repeated several times in the OBR system.

5.2.2 Potential of optimisation-based rostering

The pilot projects were intended to provide insight into whether the potential identified could be realised in practice, and the challenges that arose during implementation. We believed that some of the potential could be realised directly through the use of OBR in the time period of a single roster, while other potential would only be realised by using OBR in a more long-term timeframe. We therefore divided the potential into what we call an operational planning level (the time period covered in a roster) and a tactical-strategic planning level (longer than one roster) and describe them briefly, before discussing the realisation of potential in detail in the section on Results and evaluation.

Table 5.2: Potential identified for key stakeholders upon implementation of OBR, by planning level.

Stakeholders	Tactical-strategic level	Operational level
Employer	•Costs •Operational factors	•Robustness
Trade union	•Health (IA agreement) •Fairness	•Health •Fairness
Individual employee	•Less time spent on planning	•Individual customisation and requests granted

The positioning of some of the potential in the table may be debated, but we believe that Table 5.2 provides a good overview. For the individual employee, OBR's greatest potential lies in its capability to take individual employees into account, whilst also ensuring that other requirements are met, which makes it highly effective in managing the complexity of the rostering problem. The union will see potential in the individual employee's health being prioritised in rosters, and may believe that some employees should be protected against overambitious rotas. As the employees' collective voice, the trade union will be interested in increased fairness, such as how the OBR system can distribute unpopular shifts evenly or in line with explicit norms. At the operational level, most of the staffing costs will be irreversible as employment contracts have already been entered into, but the employer will want robust rosters that reduce the use of temporary staff. This can both help to ensure responsible management and reduce costs associated with temporary staffing.

At the tactical-strategic planning level, the introduction of OBR can simplify the planning process and reduce the individual employee's time spent on planning. Furthermore, OBR provides a unique overview and opportunity to analyse and evaluate rosters, which entails significant potential for the union in terms of health-promoting and fair rosters. For example, if some employees do not work nights for health reasons as part of an inclusive working life agreement (IA agreement), their colleagues will have to work more night shifts, putting them at risk of an overburdensome workload. Such factors can be incorporated into OBR, and the rosters produced can be assessed before any changes are adopted. For the hospital, there is considerable potential in gaining a better overview and control of staffing costs in each department. For example, they can test how rosters will change as a result of reductions and increases in staffing levels, or by changing the competence level of selected employees.

5.2.3 Development of OBR

Our development work was centred around establishing OBR with the stakeholders we considered most relevant; the individual employees, the Norwegian Nurses' Association (NSF) and the hospital management. These stakeholders were also important participants in the process of framing the rostering problem that OBR was to solve. We held regular meetings with the operations coordinator, where we presented the OBR system's latest draft of the roster. The operations coordinator had a good insight into the needs of the individual employees, and was of great assistance in helping us to understand the roster planning problem. Despite the fact that much of the knowledge about the employees' preferences and needs was implicit, we generally managed to unravel the underlying causes and formulate them mathematically. Furthermore, we held meetings with NSF Sør-Trøndelag in order to get their perspectives on the roster planning, and we presented the project for discussion in the working environment committee at St. Olav's Hospital. The employees were also directly involved through three general meetings that were held about the project. The project was rooted in the hospital management through close collaboration with the Medical Director and the head of the Division of Obstetrics and Gynaecology. The project was also presented to the hospital management.

5.2.4 Implementation and pilot projects

The OBR system we developed for Maternity Ward West was trialled in two pilot projects in 2017 and 2018. In the first pilot project, we used OBR to plan for the period 18 December 2017 to 3 June 2018 (hereafter referred to as spring 2018), while the second pilot project covered 4 June to 9 December 2018 (hereafter referred to as autumn 2018). In order to evaluate our success in realising the aforementioned potential of OBR, we conducted a survey among the employees after each pilot project.

5.2.5 Collecting data from similar projects in Scandinavia

In order to support our discussions with further data, we conducted interviews with two other actors working on similar initiatives: Troels Range who works with applied optimisation at Hospital South West Jutland in Denmark, and Jacob Nyman, who works at Visma's department for optimisation in Oslo. They are both developing different variants of OBR. At Hospital South West Jutland, variants of OBR have been implemented as a permanent decision support tool for roster planning for nurses and doctors in various departments, and the solution is gradually being rolled out in more departments. Visma owns and operates computer systems that are used in many Norwegian hospitals, which are reminiscent of the RMS at Maternity Ward West, and they are also developing customised OBR software. Visma plans to conduct pilot projects soon.

5.3 Results and evaluation

In this section, we will present some of our experiences from the pilot projects and a selection of the results from the surveys. Based on these experiences and results, we will discuss whether the potential outlined in Table 2 was realised. The pilot projects provided significant insight into whether we could realise the operational potential, and we summarise the evaluation in Table 3. The tactical-strategic potential is more challenging to evaluate because we only introduced OBR for two planning periods.

5.3.1 Realisation of potential at the operational planning level

The individual employee's potential

Because we developed the OBR system with a clear bottom-up perspective, comprehensive individual customisation was essential for the rosters to be considered in any way relevant to the planning at Maternity Ward West. We believe we achieved this, as feedback and subsequent discussions only related to minor and specific errors or shortcomings, not whether the OBR system is able to produce rosters that epitomise the rostering problem. Furthermore, granting employees' specific requests to work a shift or have time off at stipulated times is probably the most important weighted quality in the rostering problem at Maternity Ward West. In order to communicate their wishes, the employees registered their desired rotas. They could also add comments. Some employees used the comment function extensively, and some even stipulated that they believed they should have time off without taking holiday leave during certain periods, while other employees did not use the function at all. There were no clear guidelines on whether comments should be prioritised over other considerations, but the operations coordinator encouraged us to introduce constraints in OBR that ensured we complied with the wishes expressed in the comments where this was possible.

As we introduced the constraints, we noticed that the OBR system's flexibility to schedule shifts in the roster was gradually reduced. This flexibility could otherwise have been used to secure more of the weighted qualities in the rostering problem. The comment requests were also counterproductive to the objective planning that we had argued would increase fairness compared to manual processes. The procedure for employees putting forward their personal requests for the roster went from avoiding responsibility in the negotiation process to actively asserting their rights in the comments function.

When all the comments were implemented as constraints, it proved impossible to produce a roster that adhered to all the constraints. This changed our understanding of the rostering problem. It was no longer a question of introducing all existing guidelines as constraints and maximising the weighted qualities. Now the rostering was also about navigating which constraints should be removed. The difference may seem small, but because constraints *must be actively deselected by the user of OBR*, while weighted qualities *are optimally prioritised by OBR*, the distinction is important. In the end, we deselected some constraints relating to comment requests in consultation with the operations coordinator, and ended up with a valid roster, but the reduction in flexibility to choose weighted qualities was nevertheless significant.

In contrast to the comment functionality, the desired rota was well suited for use in OBR because the information was well structured. Trying to produce rosters that corresponded as closely as possible to the desired rota was an important weighted quality, and was maximised within the solution space defined by the constraints. OBR achieved 85.8% and 85.4% agreement with the desired rotas for spring 2018 and autumn 2018 respectively. Analyses of an earlier planning period showed comparable results with manual planning of 85.3% (Beckmann and Klyve, 2016). The similarity in the number of requests granted between OBR and manual planning is striking, and emphasises how the lack of flexibility reduced the potential to grant requests.

A high degree of correspondence between the roster and the desired rotas is undoubtedly beneficial, but in practice there are significant weaknesses associated with using each individual day in the desired rota as a measure of employees' preferences.

1. Employees registered one shift request per day, and without the comment functionality, we had no information about which days the employees regarded most important when we were considering their shift requests. The majority of registered shift requests are probably more or less randomly placed in order to fit the number of different shift codes.

2. The planning process was so time-consuming that employees had to submit their requests several months before the rosters were to be realised. They are unlikely

to have had a clear overview of their own preferences at that stage. The frequent changing of shifts after rosters had been drawn up supports this assumption.

3. Requests for shifts and days off may depend on whether shift requests for other days are granted.

If we had ignored the requests in the comments, the aforementioned weaknesses would have prevented us from accommodating the employees' true preferences to any great extent.

Results from the surveys help to shed light on whether the potential for individual customisation was realised. The responses shown in Figure 5.2 paint a varied picture. We believe it is particularly desirable to reduce the number of employees who strongly disagree with the statement, in order for the employees to feel positive about the use of OBR. We note that the proportion who strongly disagree is 18% in the first pilot and 12% in the second pilot. It is important to note that the employees were asked to evaluate the original rosters produced by OBR, not the final rosters that had been further processed by the operations coordinator and the roster group. This means that minor start-up difficulties, such as human error in the input of fixed shifts and misunderstandings related to what information employees had to register, had not been corrected in the rosters evaluated by the employees. We therefore believe that the responses are acceptable, despite it being difficult to know how the employees would have evaluated the manual rosters that were drawn up before the pilot projects.

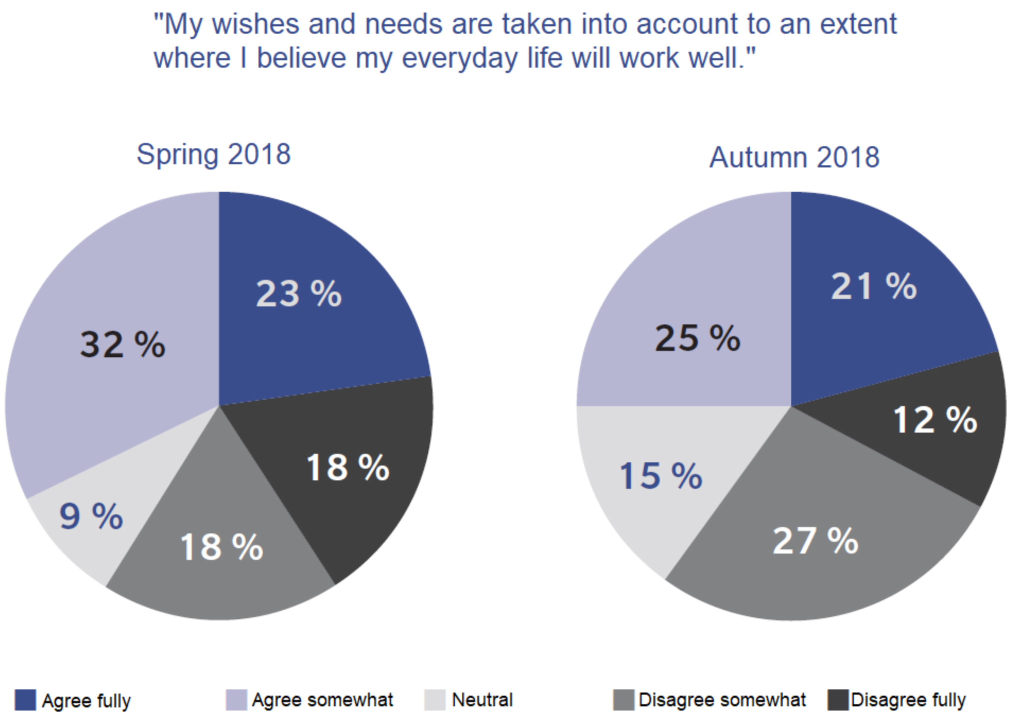

Figure 5.2: Responses to the rosters produced in the two pilot projects

Our overall impression is that we were largely able to accommodate the individual employees' wishes and preferences, but that realising the full potential for individual customisation requires a combination of greater clarity in what can be expected in the way of guidelines (constraints) and further development of the shift request system.

Potential for the trade union

Research shows that working an evening shift immediately followed by a day shift increases the risk of sleep disturbance (Lie et al., 2014), even though this shift combination is not unlawful under the Working Environment Act. This shift combination was effectively ruled out by the OBR system through constraints for most employees. We made exceptions for overnight commuters who requested this. This was because public transport did not cover their needs and they were able to sleep over at the hospital, whereby the overall burden was considered acceptable. Following feedback, evening out the individual employees' workload was given a high priority, but this lead to reduced flexibility for the OBR system to prioritise weighted qualities. The combination of even workloads and the individual employees' requirements, as well as requests for longer periods of continuous leave, significantly reduced the flexibility. Our assessment of the rosters that were produced is that they largely protected employees from unfavourable shift combinations. This was also the case in manual planning, but the constraints in OBR represent extra protection against rosters that are considered detrimental to health.

Evaluating the potential of fairness is a challenge, and we therefore rely on the feedback we received in the surveys after the two pilot projects. It may seem from Figure 5.3 that the potential for fairness has largely been realised. We see that very few of the respondents (4% and 9%) felt that their requests were given a lower priority than those of the other employees. In addition, a large proportion (31% in both periods) disagree that they were down-prioritised. These results are promising, and suggest that the many constraints stemming from comments were not detrimental to the perceived fairness of the rosters.

Potential for the employer

It was already clear to us that robustness in the form of consistency in staffing levels was a sensible weighted quality. However, the focus on this potential diminished as the OBR system was being developed. A weighted quality was actually defined that punished overstaffing, but only on weekends, which were clear bottlenecks. On weekdays, it was not uncommon to have overstaffing on day and evening shifts, despite the desired staffing level being defined as significantly higher than on weekends.

When we presented the results of the pilot projects to the head of the Division of Obstetrics and Gynaecology, a disagreement clearly emerged. We reported that it had been easy to formulate operational factors in the OBR system because they were specific. The head of the division disagreed with the claim, and questioned part of the premiss for the constraints. Simply put, there was no consensus about the minimum staffing levels that were used at Maternity Ward West. This illustrates one of the challenges in not only OBR, but in general rostering.

It is not surprising that there is a tug-of-war between different stakeholders in the balancing act between operational factors, budgets and employee needs. However, what was surprising was that the head of the division apparently disagreed with the premisses

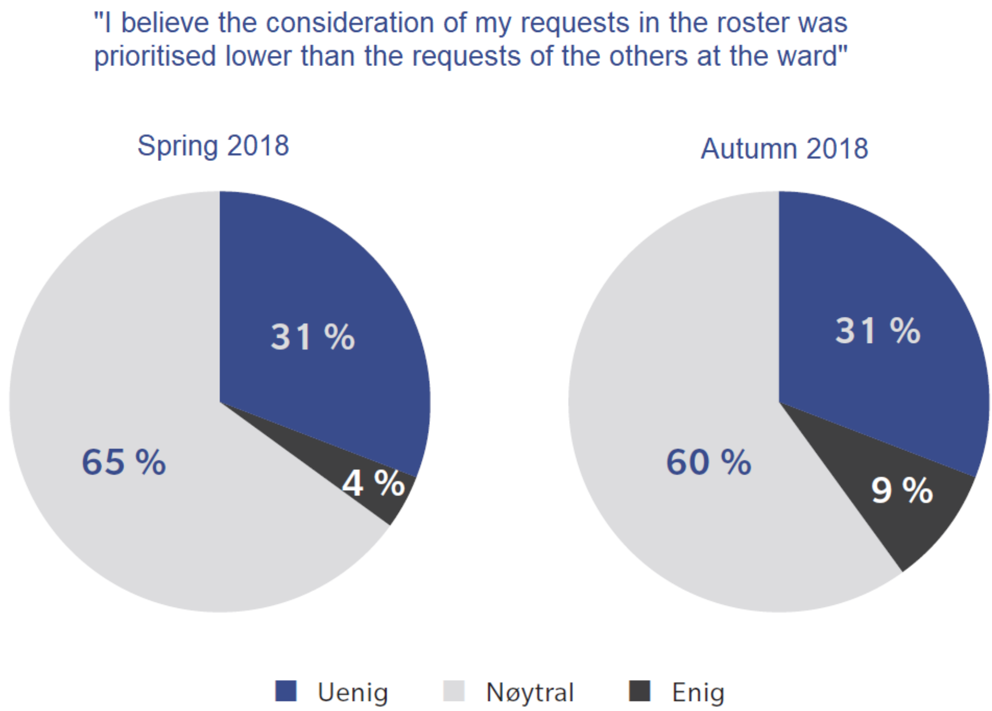

Figure 5.3: Fair distribution of shifts.

on which the roster planning was based at the time OBR was introduced. This may be an indication that we did not include the employer to a sufficient extent in the adaptation of OBR. In hindsight, we believe that this was due to a somewhat naive belief that if OBR was sufficiently adapted to the employees, the flexibility and ability to weight different qualities would enable the rosters to maintain a very high quality. As previously mentioned, there was no flexibility for other considerations if all requirements from individual employees were to be met, as it was impossible to produce such a roster.

Clear guidelines on which considerations should be accommodated would be useful regardless of planning method, and are essential for the success of OBR. Thus, the actual process of developing OBR can be beneficial because disagreements that are not discussed between different stakeholders can surface. This is how we discovered new potential with OBR at the tactical-strategic level, namely that the development and evaluation process itself contributes to resolving existing disagreements about the content of a roster. The disagreements about the basic premisses for constraints inspired us to create the OBR pyramid, presented in Figure 5.4.

Some of the considerations in roster planning are explicitly formulated and there is broad agreement that they should be accommodated among all three of the main stakeholders. These are undoubtedly constraints, and are located at the bottom of the OBR pyramid. The Working Environment Act and the collective agreement, as well as IA agreements and the rostering agreement are good examples of this. At the other end of the scale are the employees' requests. These are placed at the top of the OBR

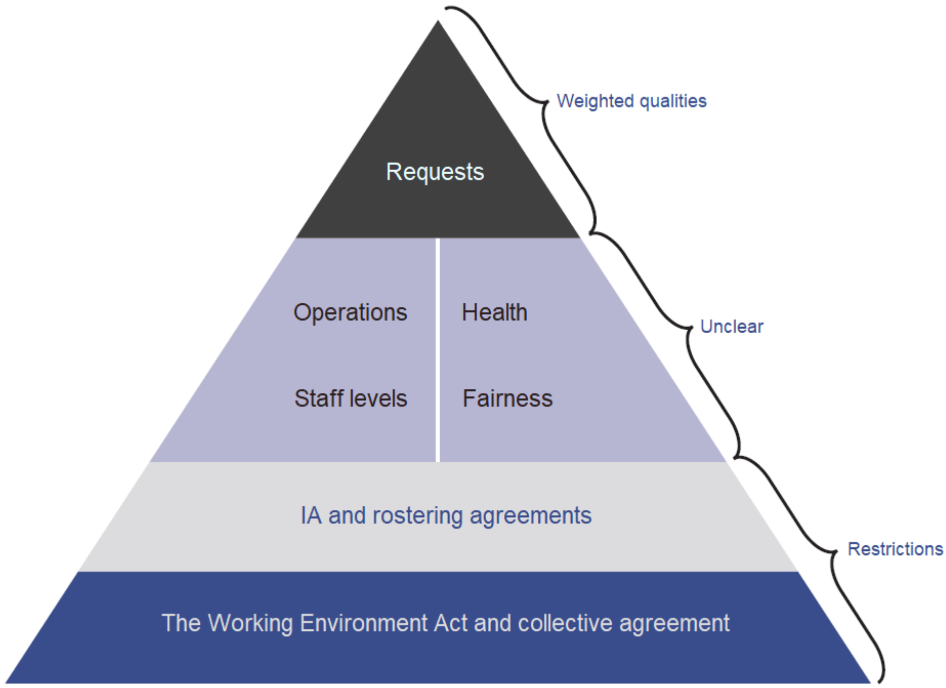

Figure 5.4: The optimisation-based roster planning pyramid (the OBR pyramid).

pyramid, and there is agreement between stakeholders that it is not possible to grant all of these requests. The requests are therefore regarded as weighted qualities. However, there is no consensus on the middle section of the pyramid. Different stakeholders can have different opinions about how to handle these considerations. Consequently, it is also typical that there are no explicit guidelines on how to accommodate the different considerations, making the entire premiss for roster planning unclear. This is a challenge both in manual roster planning and OBR.

5.3.2 Realisation of potential at the tactical-strategic planning level

Our two pilot projects, combined with our limited influence on the basic premisses of roster planning, mean that we cannot claim to have realised the long-term potential for OBR. However, we would like to note some findings that indicate that long-term potential may be realistic in the future.

- OBR reduces the time that employees spend on planning considerably, because it simplifies the actual planning process.

OBR simplifies the planning process, thus saving the employees' time.

- Even for individual rotas, it appeared that employees considered OBR to be fair.

Provided that OBR takes into account information about who has worked the least popular shifts years in previous years (such as Christmas Eve), we see no reason why this cannot continue and remain a major strength of OBR. It is also clear that it is possible to test out different guidelines for IA agreements and see what effect this has on the rosters overall.

- OBR provided a better insight into how measures affect operational factors.

One example that was cited in Beckmann and Klyve (2016), was that an extra day off could be offered to employees who were willing to work an extra weekend. If some employees were willing to do this, the overstaffing on weekdays could be reduced somewhat in exchange for better staffing at the weekend.

Table 3 summarises the results from the pilot projects, where both fully and partially realised potential and indications of realism are marked with a tick, while clearly unrealised potential is indicated with a cross.

Table 5.3: Realisation of potential at the operational level for key stakeholders upon implementation of OBR.

Stakeholders	**Indications of realism in the potential at the tactical-strategic level**	**Realised potential at the operational level**
Employer	•Costs ✓ •Operational factors ✓	•Robustness ✗
Trade union	•Health (IA agreement) ✓ •Fairness ✓	•Health ✓ •Fairness ✓
Individual employee	•Less time spent on planning ✓	•Individual customisation and requests granted ✓

5.4 Development up to 2050

Based on the experiences from our pilot projects and those of Troels Range at Hospital South West Jutland in Denmark and Jacob Nyman at Visma in Oslo, we will discuss some development features that seem likely in roster planning in 24-hour wards, at hospitals in particular, in 30 years' time. These features may also have knock-on effects on other planning in the public health service and activity outside the health services, but this is not our main focus. We point out that we are exploring scenarios we consider to be probable with a clear basic premiss that OBR or a similar decision support tool is used to a large extent in 24-hour wards in hospitals.

5.4.1 Rostering becomes more standardised

A clear development feature of roster planning will be standardisation. The information that is used in OBR must be standardised for the system to be able to interpret the data. This information must then be incorporated into the employer's computer systems in a way that enables the seamless retrieval of data from the OBR system. Desired rotas disappear in favour of systems with more precise information about preferences. However, the choice of data for use in OBR is closely linked to the standardisation of the roster planning process. In order to be able to select what information is relevant, it must be clear to all stakeholders what they can expect from a rota plan. This needs to be clarified between the employer and the employee. We believe the trade union will play a key role here, and discuss this later.

This is supported by experiences from Hospital South West Jutland, which has a structured approach to rolling out OBR in new departments. The guidelines are clarified individually for each department in cooperation with managers, planners and trade unions. The system for shift requests is similar in all departments. Employees can apply to managers for days off under a 'veto' system. Otherwise, they have a limited number of requests they can register in their roster. In exceptional cases, employees still submit comments, which requires the involvement of people with expertise in the development of OBR. Our contact is very confident that a well-functioning standardisation of the roster planning process and system at Hospital South West Jutland will be in place within three years.

Visma in Oslo is currently developing a standardised OBR system that will be used at many different hospitals and departments. It is interesting to note that many hospitals and other organisations have already implemented extensive payroll and staff planning systems from Visma. Their OBR variant will be an add-on module to these systems, and they therefore already have access to significant amounts of structured data.

5.4.2 The introduction of OBR will call into question existing practices

Negotiations on the rostering agreement are an established arena for preparing guidelines for roster planning, and from the perspective of an OBR developer, the rostering agreement is very easy to work with. First, it is explicitly and unambiguously formulated, which makes it easy to formulate it mathematically. Second, it describes constraints, i.e. rules that always need to be complied with. The problem at Maternity Ward West, however, was the issues that the rostering agreement did not address. In the future, there will be a clearer distinction between what employees can expect and what they can only hope for. The unclear considerations placed in the middle of the OBR pyramid must therefore either be constraints that are established in the rostering agreement or weighted qualities. The employees can request the weighted qualities, but cannot demand them from their manager.

Although it is uncertain how representative Maternity Ward West was of other departments in terms of rota work, there will probably be grey areas in terms of what both employers and employees believe they are entitled to in the roster planning. These will surface upon implementation of the OBR system. The grey areas will largely be eliminated when a clearer distinction emerges between what employees are entitled to in the rosters and what they can request. It will be important to ensure that new employees understand this in order to prevent the existing culture from remaining dominant in the workplace.

5.4.3 The employer will demand better utilisation of staff resources

All overstaffing will come to light when the OBR system is implemented, and measures to reduce this can be tested out. The employer will either demand better utilisation of the excess staff resources or a reduction in the overstaffing.

We have previously mentioned that it is conceivable that some employees will be willing to work an extra weekend to get an extra day off. Use of surplus labour as temporary cover in other hospital departments can also be envisaged because OBR will provide an overview of several departments and ensure optimal distribution of the staff resources. However, this capability requires the employee's experience and competence to be suitable for the department they are assigned to. Alternatively, OBR can easily be used to create rosters that require fewer employees. The employer can introduce a temporary recruitment freeze and save on wage expenditure, as shown in Beckmann and Klyve (2016).

The work culture and employment rights at Hospital South West Jutland differ substantially from those in Norway. Nevertheless, it is interesting to see how they have used the system. They have realised large savings after introducing variants of OBR and analysing staffing needs. They found that in 24-hours wards, about half of employees' shifts were day shifts. This was quickly changed, and they were then able to manage with fewer employees. Employment protection for employees appears to be significantly poorer at Hospital South West Jutland than what is typical in Norway, and they frequently cut staff numbers when the OBR system manages to create viable rosters with fewer employees. Hospital South West Jutland represents a scenario where OBR is developed primarily for the benefit of the employer. This scenario is unrealistic for Norway without major changes in the organisation of working life. However, the example highlights how OBR can be used to promote different interests, and emphasises how important it is for all stakeholders to understand how it works.

5.4.4 Employees will demand individual customisation

The standardisation of both processes and data may lead to employees being treated more uniformly, for better or worse. It could be argued that treating all employees or all subgroups of employees uniformly increases fairness and clarity in the roster planning, but it could also entail an unreasonable simplification of the variation in the needs of the individual employees. However, we believe that future variants of OBR will take employees' wishes and needs into account in a positive way. Despite the fact that employees are unique and have different needs, many of their needs and wishes are similar, such as getting time off work at times that cater for their health and leisure time. Furthermore, OBR has the capacity to process large volumes of data and balance many different considerations. Therefore, in the future it will not be a question of whether OBR is able to create rosters that meet the employees' needs, but to what extent these needs are prioritised over weighted qualities that other stakeholders prefer.

In order to have weighted qualities accommodated, they need to be subject to reasonable prioritisation in OBR. Our experiences from Maternity Ward West suggest that the employees will be prioritised. If they meet the employer's wishes for flexibility, they should be able to expect further customisation in return. Our contact at Hospital South West Jutland believes that there will also be a strong emphasis on individual customisation in their variants of OBR in the future, because it will give employers who use OBR wisely a competitive advantage in staff recruitment. In addition, employees can put collective pressure on the employer through the trade union, and thereby demand a reasonable prioritisation of individual considerations in OBR.

5.5 The union must understand the technology so as not to lose its role

When employees want to make the above demands on their employer, this is traditionally done through a trade union. However, concretisation of the guidelines on roster planning presents unions with a challenge. It is easy for the employer to explicitly state its requirements for budget limits, and thereby the number of full-time equivalents they can allow. In contrast, it is difficult for all the employees to convey their preferences this clearly. It is also unrealistic for all employees to have their individual wishes met at the same time. It is therefore the union's task to gather information about the employees' preferences and needs before drawing up guidelines in the form of a very explicit compromise for use in OBR.

Trade unions are used to playing such a role, but their representatives will now also need to have a good understanding of OBR. This insight does not necessarily have to be technical, in the same way as a developer understands OBR, but a thorough understanding of how constraints and weighted qualities work will be important. As mentioned, OBR has something to offer both the employer and the employees. If employees did not feel that their union was able to facilitate compromises that they could benefit from, they would be incentivised to make individual agreements directly with the employer. The union is therefore dependent on keeping up with developments.

As the union gradually manages to draw up good collective guidelines, constraints will probably be used to guarantee that some key preferences will be met. A veto system, where time off on special occasions is guaranteed, is a likely requirement. In addition to these, the individual customisation will largely be formulated using weighted qualities, where everyone is given equal priority. The strongest voices in manual planning will lose their influence, but the collective considerations will instead be well protected. The union's position will be strengthened as it learns to understand the technology, as it will then be in a position to make sensible demands on behalf of the employees and help make the technology more understandable to the employees.

5.5.1 Decisions are taken at a higher level

When more detailed decisions are made in negotiations between the trade union and the employer, and standardised processes serve as a guide, more decisions will be made at a higher level in the health organisations. As a result, we will see a development towards somewhat more vertical management. However, this only applies to the guidelines, i.e. the decisions that serve as a guide for the constraints in OBR. The most detailed decisions, the actual choice of who will work which shifts, are decided by the OBR system. The users of OBR will not necessarily be higher up in the organisation in comparison with manual planning. If you implement OBR in a department with a desired rota, the user of the OBR system will have to be a selected planner, as in our pilot projects at Maternity Ward West. Consequently, there will be no more direct involvement in the negotiation process. In departments that draw up manual basic rotas, the planner can become the user of the OBR system.

At Hospital South West Jutland, the introduction of OBR has led to decisions for roster planning being moved from middle management to the hospital's central finance and planning unit, where the developers of their OBR variants are the users of the system. However, our contact hopes that those who currently work as administrative staff can be trained in OBR, and in the long term he believes that health workers in the departments can be trained as superusers. At Visma, they develop the systems with the aim of the current planners being the users.

Even if the users of OBR are not at the top of the organisations, the perceived distance to the decisions can be greater after the implementation of OBR if there is no understanding of how it works. This may challenge the principle of co-determination, which is a strong component in labour relations in Norway, and emphasises the importance of OBR having the capability to intercept individual employees' considerations and preferences.

5.6 Concluding reflections

Our results bring issues to light where classic lines of conflict appear. OBR can potentially be used both to streamline the operation of health organisations and to accommodate employees' needs, preferences and collective benefits. In addition, a tendency towards standardisation and the perception of a greater distance from decisions can exacerbate conflict and lead to frustration. It is natural, and perhaps a healthy sign, that such clashes come to light when new technology threatens the status quo in planning that affects both operations and employees' leisure time. Nevertheless, OBR represents a development that can lead to Pareto improvements, where both parties benefit from the new technology.

We would like to express our gratitude to the operations coordinator, Vigdis Myhren, who has kindly shared her insight into the operation of Maternity Ward West at St. Olav's Hospital with us, and Francesco Beckmann, who was an essential contributor to the early development of OBR. We would also like to thank the management team at the Center for Health Care Improvement and Professor Henrik Andersson for his guidance and support.

Bibliography

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

J. F. Bard and H. W. Purnomo. Preference scheduling for nurses using column generation. *European Journal of Operational Research*, 164(2):510–534, 2005b.

F. R. Beckmann and K. K. Klyve. Optimisation-based nurse scheduling for real-life instances. Master's thesis, NTNU, 2016.

F. Bellanti, G. Carello, F. Della Croce, and R. Tadei. A greedy-based neighborhood search approach to a nurse rostering problem. *European Journal of Operational Research*, 153 (1):28–40, 2004.

M. Bester, I. Nieuwoudt, and J. H. Van Vuuren. Finding good nurse duty schedules: a case study. *Journal of Scheduling*, 10(6):387–405, 2007.

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

E. K. Burke, P. D. Causmaecker, S. Petrovic, and G. V. Berghe. Metaheuristics for handling time interval coverage constraints in nurse scheduling. *Applied Artificial Intelligence*, 20(9):743–766, 2006.

E. K. Burke, J. Li, and R. Qu. A hybrid model of integer programming and variable neighbourhood search for highly-constrained nurse rostering problems. *European Journal of Operational Research*, 203(2):484–493, 2010.

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

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

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

F. Guessoum, S. Haddadi, and E. Gattal. Simple, yet fast and effective two-phase method for nurse rostering. *American Journal of Mathematical and Management Sciences*, 39 (1):1–19, 2020.

D. L. Kellogg and S. Walczak. Nurse scheduling: From academia to implementation or not? *Interfaces*, 37:355–369, 2007.

M. Kumar, S. Teso, P. De Causmaecker, and L. De Raedt. Automating personnel rostering by learning constraints using tensors. In *2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 697–704. IEEE, 2019.

J. Lie, L. Arneberg, L. Goffeng, H. Gravseth, A. Lie, C. Lojså, and D. Matre. Arbeidstid og helse oppdatering av en systematisk litteraturstudie [work hours and health e an update on a systematic literature study]. *Norway: STAMI*, 2014.

OECD. *Health at a Glance 2019*. OECD, 2019. doi: https://doi.org/https://doi.org/10.1787/4dd50c09-en.

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

E. Rönnberg and T. Larsson. Automating the self-scheduling process of nurses in swedish healthcare: a pilot study. *Health Care Management Science*, 13:35–53, 2010.

D. M. Warner. Scheduling nursing personnel according to nursing preference: A mathematical programming approach. *Operations Research*, 24(5):842–856, 1976.

E. Yilmaz. A mathematical programming model for scheduling of nurses' labor shifts. *Journal of Medical Systems*, 36(2):491–496, 2012.

Appendices

A Mathematical model

This model formalizes the planning problem studied at the Maternity Ward West during the pilot projects performed there. There are some minor differences between the two periods, as the understanding of the planning problem evolved over time. We do not include the large number of constraints, both hard and soft, related to specific periods such as Christmas, Easter, vacations, etc. Some of the logic implemented in balancing these periods, especially the Christmas period, is very detailed and likely to change from year to year. Furthermore, the implementation of these periods in rosters were not very popular, and are likely to remain outside of the scope of future rostering projects at the ward. We thus exclude these constraints from the model, as they provide little insight to the overall planning problem.

Several other problem-specific details are captured by adjusting specific parameters, and thus the mathematical model formulated in this section largely reflects the planning problem observed at the ward. Examples include the number of employees covering shifts across wards, allocating reasonable union representation at the ward, different course days of a kind that should be fixed beforehand, fixed weekly activities at other wards or workplaces, fixing specific requests, making changes according to individuals with specific rights or agreements, etc. The many individual customizations are the reasons why multiple parameters have an employee-specific index, implying there can be individual and group-specific differences in all constraints formulated. However, these do not affect the overall structure of the planning problem, and the comprehensive treatment of parameters are thus simply mentioned briefly here rather than discussing them in detail.

As the development of the model used in the pilot projects was a continuation of the work done in Beckmann and Klyve (2016), the model reuses some constraints and examples from that work.

A.1 Introduction

The rostering problem observed at Maternity Ward West (MWW) is formulated using $\mathcal{N}$ to denote the set of employees, a majority of which are midwives or certified nursing assistants. We denote the set of shifts $\mathcal{S}$. The shift types defined in the problem reflects the real-life shifts in their internal system, and there are thus 22 shift types defined. Important shift types include day, evening, night, normal off-shifts, and protected off-shifts $\{D, E, N, F, \hat{F}\}$. There is a large variation in what the other shift types represent. Some represent different tasks, such as performing supportive services at the ward, attending coursework, coordinating work, etc. Other shift types have similar functions as the most important shift types mentioned, but span a different time period of the day.

We also introduce the set of demand types $\mathcal{U}$. The demand types are key to formulating several different demand coverage constraints at MWW. The distinction between shift types and demand types is important, as the demand for employees working during the night is defined for demand types $u \in \mathcal{U}$, not explicitly for shift types. As a result, midwives and assistants can largely work the same shifts, e.g. the same night shift N, but cover different demand types when working the same shift (except if a midwife ranks down to cover for an assistant). Furthermore, the demand type set $\mathcal{U}$ helps describe several specific demand types that occur sporadically throughout the planning horizon at MWW, e.g. the demand for a certain number of employees to attend a type of course.

The days in the planning period is denoted by $\mathcal{T} = \{1, \dots, T\}$ and we use $t < 1$ to represent days in the previous planning period. We use the set $\mathcal{K}$ to describe weeks in the planning period, and the sets $\mathcal{T}_k$ are the sets of days that exist in each week k.

A large collection of different shift patterns are handled through different sets of shift patterns, the desirable patterns $\mathcal{P}^D$, the undesirable $\mathcal{P}^U$, and the illegal patterns $\mathcal{P}^I$. The binary variable x_{nst} is 1 if nurse n is assigned shift s on day t and 0 otherwise. It is also defined for $t < 1$ to represent the roster that was created in the previous planning period. Remaining sets, parameters and variables used in specific constraints are introduced when relevant.

A.2 Basic constraints

Regulations state that each nurse must have a protected off shift, denoted $\hat{F}$, every week. If the nurse has a weekend off, the weekly protected off shift must be on Sunday. Furthermore, the protected off shift entails a certain number of hours without work, and we introduce $\mathcal{S}_{s_1}^{\hat{F}}$ as shifts such that shift combination $\{s_1, \hat{F}, s_2\}$, $s_2 \in \mathcal{S}_{s_1}^{\hat{F}}$ is illegal with respect to this. The set of days $\mathcal{T}^{SUN}$ contains all Sundays in the planning period.

$$\sum_{s \in \mathcal{S}} x_{nst} = 1, \qquad n \in \mathcal{N}, t \in \mathcal{T} \tag{1}$$

$$x_{ns_1(t-2)} + x_{n\hat{F}(t-1)} + \sum_{s_2 \in \mathcal{S}_{s_1}^{\hat{F}}} x_{ns_2t} \leq 2 \qquad n \in \mathcal{N}, s_1 \in \mathcal{S}, t \in \mathcal{T} \tag{2}$$

$$\sum_{t \in \mathcal{T}_k} x_{n\hat{F}t} = 1 \qquad n \in \mathcal{N}, k \in \mathcal{K} \tag{3}$$

$$\sum_{n \in \mathcal{N}} \sum_{t \in \mathcal{T}^{Sun}} x_{nFt} = 0 \tag{4}$$

Constraints (1) state that every employees must be assigned one shift, including off-shifts, every day of the planning period. Constraints (2) and (3) ensure weekly assignment of protected off shifts with sufficient rest between the shifts before and after. As there are only two types of off-shifts, F and $\hat{F}$, constraint (4) make sure that if a day off is assigned to a Sunday, it must be of type $\hat{F}$. The constraints presented in this subsection are the only ones in the planning problem that are not subject to changes or adjustments based on individual employees, times, etc.

A.3 Covering demand

Parameters $\underline{D}_{tu}$ and $\overline{D}_{tu}$ denote the minimum and maximum demand for coverage of demand type u on day t. The parameters have aggregated values, so that contributions of higher-ranked employees is included in the count, e.g. the demand for midwives and assistants is aggregated in the parameter values for assistants. Similarly, for every combination of day and demand type, there can exist a soft constraint reflecting a desired minimum and maximum level of employees $\underline{D}_{tu}^D$ and $\overline{D}_{tu}^D$. Breaching these levels are penalized by variables $\overline{d}_{tu}$ and $\underline{d}_{tu}$. Parameters $P_{n_1n_2}^{TO}$ indicate if employees n_1 and n_2 can cover certain demand types together. $P_{n_1n_2}^{TO} = 1$ if they can, 0 otherwise. The sets $\mathcal{N}_u^U$ and $\mathcal{S}_u^U$ contain the nurses and shift types that cover demand type u. The set of demand types $\mathcal{U}^{TO}$ are the demand types where some employees cannot work together.

$$\underline{D}_{tu} \leq \sum_{n \in \mathcal{N}_u^U} \sum_{s \in \mathcal{S}_u^U} x_{nst} \leq \overline{D}_{tu}, \qquad t \in \mathcal{T}, u \in \mathcal{U} \tag{5}$$

$$\underline{D}_{tu}^D + \underline{d}_{tu} \leq \sum_{n \in \mathcal{N}_u^U} \sum_{s \in \mathcal{S}_u^U} x_{nst} \leq \overline{D}_{tu}^D - \overline{d}_{tu}, \qquad t \in \mathcal{T}, u \in \mathcal{U} \tag{6}$$

$$\underline{D}_{nu}^I \leq \sum_{s \in \mathcal{S}_u^U} \sum_{t \in \mathcal{T}} x_{nst} \leq \overline{D}_{nu}^I, \qquad u \in \mathcal{U}, n \in \mathcal{N}_u^U \tag{7}$$

$$x_{n_1 st} + x_{n_2 st} \leq 1,$$
$$n_1, n_2 \in \mathcal{N}, u \in \mathcal{U}^{TO}, s \in \mathcal{S}_u^U, t \in \mathcal{T}, \ |P_{n_1 n_2}^{TO} = 0 \tag{8}$$

Constraints (5) ensure that all types of demand u are covered every day t with the possibility of ranking down. Similarly, for every combination of day and demand type, there can exist a soft constraint reflecting a desired minimum and maximum level of employees, formalized in constraints (6). Demand constraints (5) and (6) cover the demand for the typical needs for staff, but also ensure new hires are assigned courses, etc. Constraints (7) ensure each employee is given the right amount of shifts covering each demand type u. For most demand types, the limits $\underline{D}_{nu}^I$ and $\overline{D}_{nu}^I$ will be 0 and $|\mathcal{T}|$ respectively, thus not affecting the problem. However, for other demand types, e.g. coursework that is sporadically planned during the planning horizon, it is important to spread the attendance in a reasonable way between employees. The constraints also ensure that employees live up to contractual agreements that spring from opting in on an around-the-clock-agreement that gives them reduced total work time. Constraints (8) ensure that certain employees n_1 and n_2 never work together.

A.4 Required rest

The parameters $\overline{M}_n^{CW}$ describe the maximum number of consecutive work shifts employee n can work. Similarly, parameters $\overline{M}_{ns}^{CS}$ describe the maximum number of consecutive shifts of type s employee n can work. These parameters represent limits to the consecutive work employees can perform. However, for many employees it is also valuable to even out their workload over time. Thus, for any number of days T_{nt}^{EW} ending on day t, employee n must not work more than a maximum of $\overline{M}_{nt}^{EW}$ days. Similarly, some employees should have specific shifts spread out evenly. For a period of T_{nt}^{NP} days ending on day t, employee n must not work more than a maximum of $\overline{M}_{nt}^{NP}$ work shifts. Parameters H_{nst} reflect the number of hours employee n works during shift s on day t, and $\overline{H}_n^{7D}$ is the maximum limit for work hours during any 7-day work stretch. The set $\mathcal{S}^W$ contains all work shifts and the set $\mathcal{S}^{CS}$ contains all shifts that should be explicitly spread out for any employee.

$$\sum_{s \in \mathcal{S}^W} \sum_{\tau = t - \overline{M}_n^{CW}}^{t} x_{ns\tau} \leq \overline{M}_n^{CW}, \qquad n \in \mathcal{N}, t \in \mathcal{T} \tag{9}$$

$$\sum_{\tau = t - \overline{M}_{ns}^{CS}}^{t} x_{ns\tau} \leq \overline{M}_{ns}^{CS}, \qquad n \in \mathcal{N}, s \in \mathcal{S}^{CS}, t \in \mathcal{T} \tag{10}$$

$$\sum_{s\in\mathcal{S}^W}\sum_{\tau=t-T_n^{EW}}^{t} x_{ns\tau} \leq \overline{M}_n^{EW}, \qquad n\in\mathcal{N}, t\in\mathcal{T} \tag{11}$$

$$\sum_{s\in\mathcal{S}^N}\sum_{\tau=t-T_n^{NP}}^{t} x_{ns\tau} \leq \overline{M}_{nt}^{NP}, \qquad n\in\mathcal{N}, t\in\mathcal{T} \tag{12}$$

$$\sum_{s\in\mathcal{S}^W}\sum_{\tau=t-6}^{t} H_{nst}x_{ns\tau} \leq \overline{H}_n^{7D}, \qquad n\in\mathcal{N}, t\in\mathcal{T} \tag{13}$$

Constraints (9) and (10) limit maximum consecutive work. Constraints (11) and (12) even out the work. There are multiple constraints implemented for different combinations of values for T_{nt}^{EW} and $\overline{M}_{nt}^{EW}$, as well as combinations of values of T_{nt}^{NP} and $\overline{M}_{nt}^{NP}$ with the structure described in constraints (11) and (12). Constraints (13) ensure that no employee works more than $\overline{H}_n^{7D}$ hours in any given 7-day period.

A.5 Work hours

To calculate the correct number of hours to assign to each employee in the planning horizon, we introduce parameters H_{nst}^C. Due to regulations such as the around-the-clock-agreement that gives reduced total work time and hour bonuses provided by working during holidays and Sundays, the calculated work hours during the year differs from the real hours worked. Parameters $\underline{H}_n^{CW}$ and $\overline{H}_n^{CW}$ denote the minimum and maximum number of calculated hours employee n should work throughout the planning horizon. These parameters are also processed significantly depending on contracts and agreements regarding vacations.

$$\underline{H}_n^{CW} \leq \sum_{s\in\mathcal{S}^W}\sum_{t\in\mathcal{T}} H_{nst}^C x_{nst} \leq \overline{H}_n^{CW}, \qquad n\in\mathcal{N} \tag{14}$$

Constraints (14) ensure that each employee n works a correct number of hours throughout the planning horizon. For nurses whom contracts change during the planning horizon, similar restrictions as constraints (14) are implemented for the planning periods before and after the contract change.

A.6 Related to fairness

Some constraints are included to even out the number of certain shifts each employee works. For some shifts, this entails allocating more than a minimum number $\underline{M}_{ns}^S$ and less than a maximum number $\overline{M}_{ns}^S$ of shift s to employee n. Similar bounds for working specific shifts in the set of specific days, e.g. night shifts during weekends, are provided by $\underline{M}_{ns}^{SS}$ and $\overline{M}_{ns}^{SS}$. For other shifts, the difference in number of shifts between employees should be less than M_{ns}^{SD}. The set $\mathcal{S}^E$ contains only shifts that should be evened out, and the set $\mathcal{T}_s^{SS}$ contains days where the number of shifts s worked should be within some threshold.

$$\underline{M}_{ns}^S \leq \sum_{t\in\mathcal{T}} x_{nst} \leq \overline{M}_{ns}^S, \qquad n\in\mathcal{N}, s\in\mathcal{S} \tag{15}$$

$$\underline{M}_{ns}^{SS} \leq \sum_{t \in \mathcal{T}_s^{SS}} x_{nst} \leq \overline{M}_{ns}^{SS}, \qquad n \in \mathcal{N}, s \in \mathcal{S} \tag{16}$$

$$\sum_{t \in \mathcal{T}} (x_{n_1 st} - x_{n_2 st}) \leq M_{ns}^{SD}, \qquad n_1 \in \mathcal{N}, n_2 \in \mathcal{N}, s \in \mathcal{S} \tag{17}$$

Constraints (15) ensure that employee n works no less than $\underline{M}_{ns}^{TS}$ number of shifts s and no more than $\overline{M}_{ns}^{TS}$ during the planning horizon. Similarly, constraints (16) restrict the number of specific shifts s that employee n works during specific days t. Constraints (17) even out the number of popular and unpopular shifts s allocated to employees.

A.7 Weekends

$\underline{K}_n^W$ is the minimum number of weeks between two working weekends for nurse n.

$$\sum_{s \in \mathcal{S}^W} \sum_{\tau=0}^{\underline{K}_{nt}^W - 1} x_{ns(t-7\tau)} \leq 1, \qquad n \in \mathcal{N}, t \in \mathcal{T}^{SUN} \tag{18}$$

Constraints (18) ensure that employees cannot work during weekends more than once every $\underline{K}_{nt}^W$ week. The value of $\underline{K}_{nt}^W$ can vary between employees and during the year.

A.8 Shift patterns

We deal with a wide definition of shift patterns in this model, as there are different constraints related to combinations of short and long sequences of shifts in the implemented model. We use similar notation as presented in Paper 2 to describe how shift patterns work, but include some additional notation.

A shift pattern p is defined by P_{stp} which is 1 if shift s is included in shift pattern p on day t and 0 otherwise and L_p is the number of active shifts, i.e. shifts with $P_{stp} = 1$, in the pattern. The pattern notation is used to formulate a wide collection of different shift pattern constraints. These can be categorized as:

- Desirable patterns, rewarded by soft constraints
- Undesirable patterns, penalized by soft constraints
- Illegal patterns, restricted by hard constraints

These categories of patterns are described using sets of shift patterns. The sets $\mathcal{P}_n^D$, $\mathcal{P}_n^U$, and $\mathcal{P}_n^I$ of shift patterns contain desirable, undesirable and illegal shift patterns for employee n respectively.

$$\sum_{s \in \mathcal{S}} \sum_{t \in \mathcal{T}} P_{stp} x_{nst} - L_p w_{np}^D \geq 0 \qquad n \in \mathcal{N}, p \in \mathcal{P}_n^D \tag{19}$$

$$\sum_{s \in \mathcal{S}} \sum_{t \in \mathcal{T}} P_{stp} x_{nst} - w_{np}^U \leq L_p - 1 \qquad n \in \mathcal{N}, p \in \mathcal{P}_n^U \tag{20}$$

$$\sum_{s \in \mathcal{S}} \sum_{t \in \mathcal{T}} P_{stp} x_{nst} \leq L_p - 1 \qquad n \in \mathcal{N}, p \in \mathcal{P}_n^I \tag{21}$$

Constraints (19) ensure $w^D_{ntp} = 0$ as long as nurse n does not work all shifts in the desirable shift pattern p. Similarly, constraints (20) make sure $w^D_{ntp} = 1$ when nurse n works all shifts in an undesirable shift pattern p. For both desirable and undesirable shift patterns, there can be a number of desirable and undesirable shorter sequences. However, for each pattern lasting the full planning period, the objective function weight of the pattern reflects its aggregated desirability or undesirability. Lastly, constraints (21) hinders allocation of illegal patterns.

Pattern examples

We provide exemplary patterns belonging to each set in Table 4.

Table 4: Examples of patterns belonging to different pattern sets that appear in the problem at MWW.

Pattern	$t-4$	$t-3$	$t-2$	$t-1$	t	L_p
$\mathcal{P}^D$			E	D	F	3
$\mathcal{P}^U$				E	F	2
$\mathcal{P}^I$		F	F	N	F	4
$\mathcal{P}^I$	N	N	N	F	D	5
$\mathcal{P}^I$	N	N	N	F	E	5
$\mathcal{P}^I$	N	N	N	F	N	5

The pattern in the first row is a desirable pattern for employees who are commuters living in an area with reduced access to public transport. Many of them wish to work a $\{E, D\}$ pattern, where they sleep at the hospital in between shifts. The second pattern applies to the same group of commuters. Their reduced access to public transport makes it unpractical to have an off-day after an evening shift, as they cannot get home during the evening using public transport. The third row is a typical illegal pattern, where a series of off-days are interrupted by a night shift. This is undesirable to an extent where it is considered illegal in the model, but several patterns like it could just as well be modelled as a heavily weighted undesirable pattern.

One category of illegal shift patterns might be counterintuitive and needs some explanation. These are the illegal shift patterns that are defined to enforce that one sequence of shifts is mandated to follow another sequence of shifts in the individual roster of an employee n. One such example is that for most employees, a sequence of three night shift, $\{N, N, N\}$ must be followed by a sequence of two off-shifts. In the mathematical model, this is formulated by stating that after the three initial night shifts, all sequences that do not entail allocating two off-shifts immediately after are illegal. This entails that any pattern including the shift sequences in rows four to six in Table 4, among many others, are part of the shift patterns $\mathcal{P}^I$.

A.9 Variable declarations

$$x_{nst} \in \{0, 1\}, \qquad n \in \mathcal{N}, s \in \mathcal{S}, t \in \mathcal{T}^P \tag{22}$$

$$\overline{d}_{tu} \geq 0, \qquad t \in \mathcal{T}, u \in \mathcal{U} \tag{23}$$

$$\underline{d}_{tu} \geq 0, \qquad t \in \mathcal{T}, u \in \mathcal{U} \tag{24}$$

$$w_{np}^D \in \{0,1\}, \qquad n \in \mathcal{N}, p \in \mathcal{P}_n^D \tag{25}$$

$$w_{np}^U \in \{0,1\}, \qquad n \in \mathcal{N}, p \in \mathcal{P}_n^U \tag{26}$$

A.10 Objective function

The objective function is presented below. The parameter P_{nst}^R is the parameter reflecting the daily requests that the employees register prior to the rostering process. P_{nst}^R is 1 if employee n requests shift s on day t, 0 otherwise. Furthermore, we introduce weight parameters with base letter W for each term i the objective function. W_{nst}^R is the objective function reward of assigning shift s to employee n on day t. Weights $\underline{W}_{tu}^C$ and $\overline{W}_{tu}^C$ are the weights penalizing lower and higher staff levels, respectively, than desired on day t for demand type u. Weights W_{np}^D reward the allocation of desired pattern p for employee n, and W_{np}^U penalize the allocation of undesired pattern p for employee n.

$$\begin{aligned} max\ Z = & \sum_{n\in\mathcal{N}} \sum_{s\in\mathcal{S}} \sum_{t\in\mathcal{T}} W_{nst}^R x_{nst} - \sum_{t\in\mathcal{T}} \sum_{u\in\mathcal{U}} (\underline{W}_{tu}^C \underline{d}_{tu} + \overline{W}_{tu}^C \overline{d}_{tu}) \\ & + \sum_{n\in\mathcal{N}} \sum_{p\in\mathcal{P}_n^D} W_{np}^D w_{np}^D - \sum_{n\in\mathcal{N}} \sum_{p\in\mathcal{P}_n^U} W_{np}^U w_{np}^U \end{aligned} \tag{27}$$

ISBN 978-82-326-6030-8 (printed ver.)
ISBN 978-82-326-5745-2 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

NTNU
Norwegian University of
Science and Technology