

Linköping University Post Print

Automating the self-scheduling process of nurses in Swedish healthcare: a pilot study

Elina Rönnerberg and Torbjörn Larsson

N.B.: When citing this work, cite the original article.

The original publication is available at www.springerlink.com:

Elina Rönnerberg and Torbjörn Larsson, Automating the self-scheduling process of nurses in Swedish healthcare: a pilot study, 2010, HEALTH CARE MANAGEMENT SCIENCE, (13), 1, 35-53.

<http://dx.doi.org/10.1007/s10729-009-9107-x>

Copyright: Springer Science Business Media

<http://www.springerlink.com/>

Postprint available at: Linköping University Electronic Press

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-59720>

Automating the self-scheduling process of nurses in Swedish healthcare: a pilot study

Elina Rönnerberg* and Torbjörn Larsson*

November 9, 2009

Abstract: Hospital wards need to be staffed by nurses round the clock, resulting in irregular working hours for many nurses. Over the years, the nurses' influence on the scheduling has been increased in order to improve their working conditions. In Sweden it is common to apply a kind of self-scheduling where each nurse individually proposes a schedule, and then the final schedule is determined through informal negotiations between the nurses. This kind of self-scheduling is very time-consuming and does often lead to conflicts.

We present a pilot study which aims at determining if it is possible to create an optimisation tool that automatically delivers a usable schedule based on the schedules proposed by the nurses. The study is performed at a typical Swedish nursing ward, for which we have developed a mathematical model and delivered schedules. The results of this study are very promising.

Keywords: Nurse scheduling, nurse rostering, self-scheduling, preference scheduling, operations research, integer linear programming.

*Department of Mathematics, Linköping University, SE-581 83 Linköping, Sweden.

1 Introduction

Traditionally, hospital wards in most countries, among them, the authors' home country Sweden, have been using cyclical schedules for their nursing staff. These schedules are fixed and repeated after a predetermined number of weeks, for example every second month. From an administrative point of view, these schedules are easy to manage, but they offer neither the nurses, nor the managers of the ward, much flexibility.

Hospital wards are often served by nurses round the clock, and because of that, nurses often work irregular times and have schedules that highly affect their personal lives. In many countries today, there is a shortage of nurses, and in order to make the nursing profession more popular, it is urgent to try to improve the working conditions for the nurses. One possible improvement is to introduce better scheduling processes, especially with regards to flexibility and adaptation to personal requests.

Staff scheduling in general is suitable for addressing with operations research, since the number of possibilities for how to construct a schedule is huge, the restrictions are quite precise, and it is often possible to formulate a useful mathematical model. The scheduling problems become even more complex and challenging from an operational research point of view if staff is needed around the clock, as is the case for nurses. Ever since the fifties, operations research techniques have been applied to nurse scheduling problems. In their article *The state of the art of nurse rostering*, Section 4, Burke et al. [6] state that even though a lot of research has been devoted to applying operations research to nurse scheduling, there is still a gap between the mathematical models and solution methods found in articles, and the flexibility that is needed to tackle the difficult real life nurse scheduling problems.

In Sweden there is a very popular kind of self-scheduling, to be thoroughly described later on, that we believe to be especially well suited for addressing with operations research. The key characteristic of this self-scheduling is that all nurses individually and independently propose their own schedules, and these proposed schedules are then used as guidelines when creating the complete schedule, that should comply with the staffing demand.

The intentions with our work is to describe and document this particular kind of self-scheduling, to develop a mathematical model to be used for automating this self-scheduling process, and to perform a pilot implementation of the automated process at a Swedish nursing ward. To the best of our knowledge, we are the first to apply operations research techniques to this particular kind of self-scheduling.

Our focus has been on creating an optimisation tool that adapts to the existing self-scheduling process as well as possible, and we have striven to adapt the mathematical model closely to the real-life conditions. During our work, we have collaborated with Swedish healthcare representatives with solid knowledge in nurse scheduling, in order to learn about Swedish nurse scheduling in general. To carry out the pilot implementation, we have worked in close collaboration with the head nurse of a typical Swedish nursing

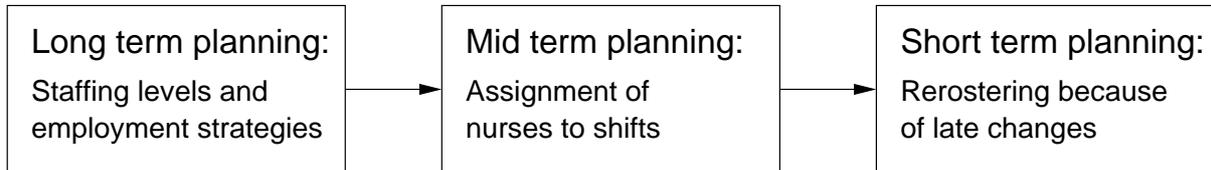


Figure 1: The three phases of nurse scheduling.

ward.

This introduction continues with a brief overview of nurse scheduling in general, and then the second section presents the Swedish self-scheduling process that we focus on. The third section presents the main concept for how to automate this self-scheduling process, and the fourth describes how the pilot implementation was carried out. Conclusions and ideas for further development are presented in the last section. Since this paper does not focus on the details of the modelling, the mathematical model is not presented in the body of the paper, but in Appendix A.

1.1 The nurse scheduling process

Most hospital wards need nurses round the clock every day of the year. The process of ensuring that there are enough nurses present at all times can naturally be divided into three planning phases, as illustrated in Figure 1.

The long term planning is a part of the overall strategic planning for the ward. First the managers of the ward must estimate how many nurses with each of the necessary skills are needed during all possible time periods of the day. Given these needs for nurses, the managers can determine how the days shall be divided into working shifts and the staffing demand for each of these shifts. Based on this staffing demand, it can be decided what kind of nurses need to be employed and the conditions for their employment (for example what kind of shifts a nurse should work). The long term planning is made for the first time when a new ward is opened, and the plan is then updated regularly, for example yearly, and whenever major changes occur. More information about this phase of the planning is found in Arthur and James [1], Ghosh and Cruz [13], Venkataraman and Brusco [19], and Fagerström et al. [12].

When the staffing demand is known and there is a given workforce of nurses, then each nurse should be assigned to a *schedule* specifying which shifts she should work, usually for a scheduling period of four to eight weeks. This phase in the planning process can be referred to as the *the mid term planning*, or *nurse rostering*. From an operational research point of view, this is the most well studied of the three phases of planning. Examples of surveys of staff rostering in general are Ernst et al. [10] and Ernst et al. [11], and surveys specialised on nurse rostering are Burke et al. [6] and Cheang et al. [7]. Mid term planning

is further described in Section 1.2. Our work concerns the mid term planning, and this is henceforth what we refer to when the term *scheduling* is used.

The schedule produced during the mid term planning should be seen as a viable plan, but the conditions for the staffing can change over time and necessitate some rescheduling to be made. For example, a nurse can call in sick just before the shift starts, or the number of nurses needed can be different from the original estimate. Whenever there is a shortage of nurses for a shift, *the short-term planning* consists of deciding whether to use overtime, to call in a nurse on her day off, to call in a substitute nurse, or to try to manage despite some shortage. How to incorporate an optimisation tool in the short term planning is for example dealt with in Bard and Purnomo [4], and in Moz and Pato [18].

1.2 The mid term planning

There are two main approaches when creating a schedule for nurses; either the schedule can be unique for each scheduling period, or it can be cyclic, that is, the same schedule is used repeatedly, period after period. In recent years the non-cyclic scheduling has gained a lot of popularity due to the flexibility that follows from creating a unique schedule each period.

Whether the schedules are cyclic or not, there are some fundamentals that need to be taken into account when creating them. At the ward, there is a given staff of nurses that should be assigned to shifts. The nurses have different skills and individual contracts stating which shifts they can work and for how many hours per week they should work. For each shift, for example a day, evening, or night shift, there are given demands for nurses with certain skills, and these demands should be fulfilled.

Besides that the nurses should be assigned to shifts in order to fulfil the staffing demands, there are *scheduling rules* that prescribe necessary properties of a schedule. Firstly, the schedules must be consistent with the prevailing laws and regulations regarding staff scheduling. Secondly, there are always some quality aspects to consider if the schedule should be acceptable to the nurses, and thereby possible to use. A typical example of such a quality aspect is the even distribution of unpopular shifts.

Given detailed information regarding the requirements described above, it is a complex, but doable, task to create a schedule for a nursing ward. At many hospital wards this work is still carried out manually, which is time-consuming and often results in schedules with undesired properties. The properties and structure of the mid term planning problem makes it well suited for addressing with operations research methodology, and a lot of research has been carried out within this field during the last five decades. Some examples can be found in Dowsland [8], Dowsland and Thompson [9], Bard and Purnomo [3], Berrada et al. [5], and Jaumard et al. [16].

As mentioned earlier, working a mixture of both weekdays and weekends, and both day- and night-time, has large consequences on the personal life of the nurses. The so called *preference scheduling* is a popular way of improving the working conditions for the

nurses through taking their requests and opinions into consideration when creating the schedule. However, the ambition to also take the nurses' opinions into account further complicates the scheduling process.

A rather extreme form of preference scheduling is *self-scheduling*, which is a process where the nurses themselves are responsible for creating a schedule for each scheduling period through informal negotiations, where they in fact *trade shifts*. There are several ways of carrying out this process, for examples see Bailyn et al. [2] and Karlsson [17]. In the next section we will present the self-scheduling process that we have studied.

2 The self-scheduling process studied

The self-scheduling process to be presented here is common and popular in Sweden, see Karlsson [17], and its characteristics makes it very interesting and challenging from an operational research and optimisation point of view. The five steps of the process of creating a schedule are presented in Figure 2. All the steps are essentially carried out manually, although the schedules are often documented by using a computer software.

Usually about six to eight weeks before the upcoming scheduling period begins, each nurse proposes a schedule for herself, independently of the other nurses. This *proposed schedule* shall mirror which shifts the nurse would like to work if she only considers her own preferences, and not when the ward needs her to work. It shall also be consistent with the scheduling rules.

To give the nurses further influence, they usually have the possibility to put *veto*s *against*, or *for*, working a shift; the former means that the nurse is then promised not to work the shift, and the latter means that she is promised to work the shift. The vetoes against working a shift are often allowed to be extended to concern a whole day. The rules for how to apply vetoes differ from ward to ward; commonly each nurse is allowed to use one veto per week.

At a closing date, all the proposed schedules are compiled and printed out on a sheet of paper or written on a whiteboard. The individually proposed schedules are, of course, very unlikely to be consistent with the staffing demand of the ward, and the next step of the self-scheduling process is to identify for which shifts there are shortages or excesses, and the result is then summarized together with the proposed schedules.

The key idea behind self-scheduling is that the nurses should trade shifts on their own initiative in order to fulfil the staffing demand. In Sweden this trading of shifts is typically made through informal negotiations during the coffee breaks. All nurses are aware of the fact that in order to create a schedule, they need to compromise with their initial requests, but at the same time they consider their own personal interests. The trading is not overseen and each nurse is left out to her own ability to negotiate. The initial requests are not documented and therefore there is no way to measure the outcome of the trading with respect to the fulfilment of the initial requests. The nurses themselves

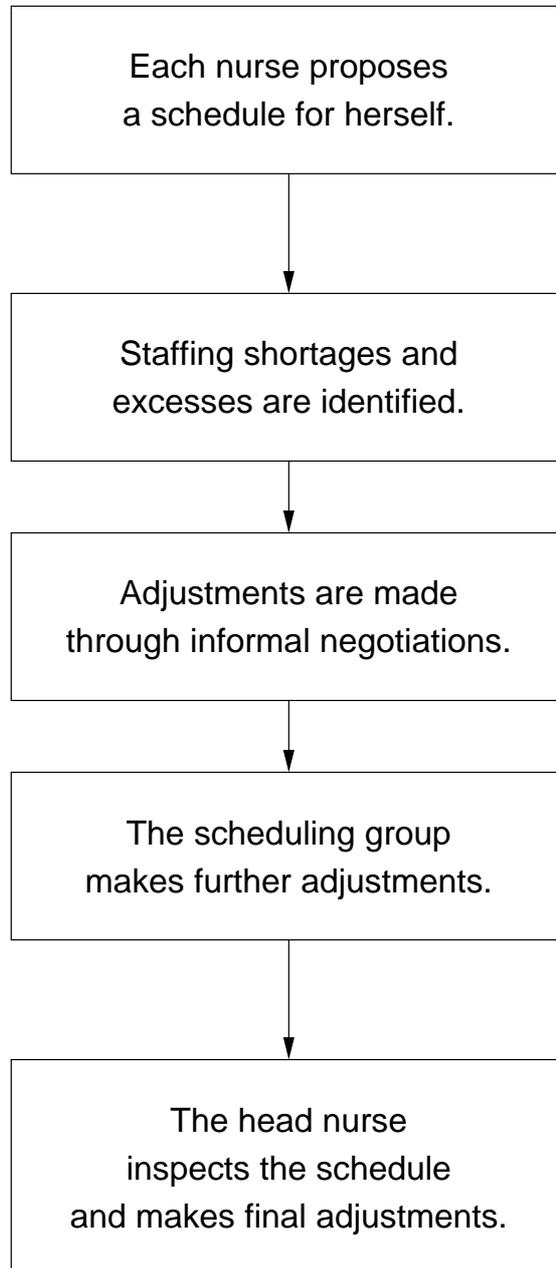


Figure 2: How the studied self-scheduling process is carried out manually.

are often able to improve the fulfilment of the staffing demand considerably, but after a while the trading stagnates and it is no longer meaningful to let it continue.

The responsibility for making the remaining adjustments is now handed over to a *scheduling group*, which consists of a few of the nurses at the ward. There are two aspects for the group to consider while making their adjustments. Firstly, during the trading it can be difficult for the nurses to fulfil all scheduling rules, and the scheduling group should identify, and correct, all violations that are of significance. Secondly, they should try to eliminate the remaining staffing shortages and excesses. At the beginning of its work, the scheduling group asks the nurses for agreement before each change. When it is no longer possible to make the nurses agree to the proposed changes, the group is mandated to do them anyway. The scheduling group usually spends many hours adjusting the schedule for each period. To give an example, the scheduling group at the pilot ward used around 20 hours each period.

When the scheduling group has finished its work, it is supposed to have a usable schedule to present to the head nurse, who is ultimately responsible for the schedule. Depending on how thorough the scheduling group has been, the schedule can be more or less consistent with the staffing demand and the scheduling rules. To some extent, violations of the rules and staffing shortages can be permitted by the head nurse, but for each scheduling period she needs to decide which of the violations in the schedule that are unacceptable, and have them corrected. How often there are remaining defects in the schedule after the scheduling group has finished its work depends on the skills of the group, the traditions, and the attitude of the nurses working at the ward.

Finally, this self-scheduling process ends with the head nurse approving the schedule, at the latest two weeks before it shall be used. For the nurses, it is then time to propose a schedule for the next scheduling period.

Through contacts with several Swedish nursing wards, we have learned that this kind of self-scheduling is much appreciated by the nurses, but also associated with considerable complications. These complications are however not unique for this kind of self-scheduling, but are instead rather common for self-scheduling in general, see Baily et al. [2]. The great benefit of this self-scheduling is that the scheduling is individualised and that the nurses have an influence on when to work. The value of this influence is the reason why many Swedish wards use this self-scheduling, despite the drawbacks to be presented.

As mentioned before, this self-scheduling process is almost constantly active, which makes it strenuous for the nurses in their everyday work. It is a very difficult task to manually create a schedule, and it is therefore likely that some staffing shortages and excesses will be left unadjusted, and because of that, more substitute nurses than should be necessary may be needed. When the nurses trade shifts simultaneously it is virtually impossible for them to comply with all the scheduling rules, even if they intend to do so. What makes the situation even more uncontrolled is that some nurses also tend to fulfil their own requests for shifts at the expense of fulfilling the scheduling rules. As an overall result, the schedules created by the nurses themselves commonly violate the scheduling

rules. These violations should be observed and put right by the scheduling group, but since this group consists of nurses from the ward it might ignore some of the violations, and instead making them a problem for the head nurse.

Another drawback with this manual self-scheduling process is the fact that the outcome relies completely on the nurses' ability to cooperate and negotiate. Since no record is kept of how many of the requested shifts that a nurse finally gets fulfilled and since the trading of shifts is made in a social context, the outcome of the self-scheduling will certainly become unfair, both on short and long terms. Unfairnesses have a tendency to lead to conflicts, and we have even heard about nursing wards that have ceased using self-scheduling because of the conflicts it has caused within the nursing staff.

To summarise, the primary goal of self-scheduling is to give the nurses influence on the scheduling process. The self-scheduling process described above is however very time-consuming and has several weaknesses due to the negotiations and the fact that the whole process is carried out manually. The proposed schedules represent how the nurses want their schedules to be, and since they are all independent and completed simultaneously, they are very well suited to be used as guidelines when creating a schedule in an automated scheduling process.

3 Automating the self-scheduling process studied

The purpose of this section is to present the main principles for how to automate the self-scheduling process described above. These principles have been developed through discussions with some Swedish healthcare representatives with solid knowledge in nurse scheduling.

An important goal of our work was to create an automated process that adapts to the existing manual self-scheduling process as well as possible, and at the same time preserves its advantages and reduces the disadvantages. The reason for not changing the current scheduling process more than necessary is that it should be easier for the nurses to accept a new process if it is similar to what they are already used to. In the automated process, each nurse proposes a schedule for herself, and then the optimisation tool suggests a schedule that is possible for the ward to use.

The described self-scheduling is appreciated because of the influence it offers the nurses, both initially when they propose a schedule, and also later when they negotiate during the trading phase. When an optimisation tool replaces the trading, some of this influence and control is lost; to compensate for this we have added some new types of requests, which make it possible for the nurses to prioritise when proposing a schedule.

In the manual self-scheduling, five kinds of requests can be used; out of these, two must be approved beforehand by the head nurse, namely holidays and *individual tasks*. An individual task is when a nurse is on duty without contributing to the fulfilment of the staffing demand at the ward, for example when she attends a meeting or special training.

The remaining three types of requests represent the nurses' personal opinions on when they prefer to work, that is, a request for working a shift, and a veto either against, or for, working a certain shift.

It is reasonable to assume that when a nurse proposes a schedule, some of her requests are made because they are important to her, while some of them are made simply because she needs to request enough shifts to fill out the right number of hours each week.

Example: A nurse requests a day shift on Tuesday, because on Tuesday evening she is meeting with a friend and does not want to work then. It suits her to work on Wednesday also, but it does not really matter if she works a day or an evening shift, but she decides, quite arbitrarily, to request a day shift. From her proposed schedule it is of course not possible to distinguish her motives behind the requests made for Tuesday and Wednesday, respectively, but they will become clear during the trading.

To compensate for the loss of information when the trading phase is excluded, we have widened the range of possible requests by introducing *a strong request for working a shift* and *a strong request against working a shift*.

Example, continued: Given these new possibilities, the nurse has two ways to distinguish between her requests on Tuesday and Wednesday. Firstly, by using a strong request against working Tuesday evening she can make it more likely that she will be free then. Secondly, she can indirectly make it more likely to be free on Tuesday evening by using a strong request for working the day shift on Tuesday. Furthermore, it is possible for her to make it even more likely to be free on Tuesday evening by using both of these strong requests.

To simplify for the nurses to make their requests, each of the possible requests are associated with a colour, according to the following list:

- purple:** individual task
- black:** holiday
- blue:** veto for working a shift
- red:** veto against working a shift
- green:** strong request for working a shift
- white:** request for working a shift
- yellow:** strong request against working a shift.

Because of the high priority to fulfil the purple, black, blue, and red requests, they are considered to be *hard* requests that lead to a predetermined assignment of the shifts in question. The *soft* requests, that is, the green, white, and yellow ones, should be fulfilled if possible. A more detailed description of how the requests are handled in the mathematical model is given in Section 4.1.4.

When all the nurses have proposed a schedule, the optimisation tool shall produce a schedule that is consistent with the staffing demand and all the scheduling rules. The schedule shall also fulfil as many of the green, white, and yellow requests as possible, and this fulfilment should be fair between the nurses. Even though the resulting schedule is consistent with the explicitly stated scheduling rules, there will always be some aspects, for example minor exceptions from the stated rules, that have not been considered in the model. These exceptions will probably be difficult to incorporate in any optimisation tool, and because of that, the schedule produced should be regarded as a qualified suggestion, and the nurses should be encouraged to trade shifts themselves if there is something they want to adjust.

4 Pilot implementation at a Swedish nursing ward

The work of developing a prototype to an optimisation tool, to be incorporated in the studied self-scheduling process, has been carried out in close collaboration with a typical Swedish nursing ward. At this *pilot ward*, the head nurse had experienced some difficulties with the self-scheduling and was well motivated to participate in a study that could help improve the self-scheduling. The pilot ward is specialized in treating patients with infectious diseases and has the capacity to care for 30 patients at a time, and it is staffed round the clock. The staff usually consists of 17 trained nurses and 14 assistant nurses (in Swedish: *sjuksköterskor* and *undersköterskor*, respectively), with the defining difference between the two categories being that the trained nurses have a university degree in nursing, while the assistant nurses have only an upper secondary nursing education. The trained nurses have more responsibilities and are permitted to perform more qualified tasks than the assistant nurses. The nursing ward is integrated with a clinic, also specialised in infectious diseases, and some of the trained nurses work both at the ward and the clinic.

In the next section we will present what is taken into account in our optimisation model. All scheduling rules and quality aspects are formulated as constraints, and the objective is to adapt the final schedules to the requests in the proposed ones.

4.1 Model components

The optimisation model is formulated as an integer linear program, implemented in AMPL 10.0, see [14], and solved by CPLEX 10.0, see [15], on a Sun Sunblade 1000 computer (750 MHz UltraSparc-III processor and 3 GB RAM). Because the mathematical description of the model is lengthy, it is found in Appendix A only, and the presentation in this section is, though with some exceptions, verbal. For each introduced constraint we give a reference to the mathematical description in Appendix A.

The key decision variables are

$$x_{ijl} = \begin{cases} 1 & \text{if nurse } i \text{ works shift } j \text{ on day } l \\ 0 & \text{otherwise.} \end{cases}$$

We begin by presenting all the constraints, which are categorised according to the subheadings below, and then conclude this section by presenting the objective function.

4.1.1 Staffing demand

At the pilot ward, there are three types of shifts, day (ca 6.45 a.m. – 3.15 p.m.), evening (ca 1.30 p.m. – 9.45 p.m.), and night (ca 8.45 p.m. – 7.00 a.m.). The given times are only approximate because there is a flexibility in when the shifts start and end (± 30 minutes), depending on the nurses' contracts. There are two staffing demands specified for each shift during a week; one stating the least number of trained nurses that are required, and the other one stating the total number of nurses, trained and assistant, that are needed, see (1)–(2). The figures stating the demands at the pilot ward are found in Tables 1 and 2.

Trained nurses	Mon	Tue	Wed	Thu	Fri	Sat	Sun
Day	3	3	3	3	3	2	2
Evening	2	2	2	2	2	2	2
Night	1	1	1	1	1	1	1

Table 1: The staffing demand for trained nurses.

Nurses	Mon	Tue	Wed	Thu	Fri	Sat	Sun
Day	7	7	8	7	7	5	5
Evening	5	5	5	5	5	4	4
Night	3	3	3	3	3	3	3

Table 2: The total staffing demand.

There is also a staffing demand for the nearby clinic, and usually they need one trained nurse each day shift during the weekdays. This demand can be fulfilled by a few of the trained nurses only, and the shifts at the clinic shall be evenly distributed among these nurses, see (10)–(11).

The number of nurses employed at the ward can vary somewhat, but during our pilot study there were 17 trained nurses and 14 assistant nurses working there. Most of these nurses, namely 12 trained nurses and 12 assistant nurses, work all the three types of shifts, while the other nurses work day and evening shifts only, see (3). Not all of the nurses are working full-time, and how large percentage of full-time they work can be seen in Table 3.

For trained nurses, full-time means 37 hours per week, if they work day and evening shifts only, and 36.2 hours if they work all kinds of shifts. For the assistant nurses, the corresponding numbers of hours are 38.15 and 36.2, respectively.

Part of full-time	50%	75%	80%	85%	90%	100%
Trained nurses	0	2	3	1	1	10
Assistant nurses	1	5	2	4	0	2

Table 3: The number of nurses working a certain percentage of full-time.

It might not always be possible to fulfil the staffing demand by only using the nurses employed at the ward, and therefore substitute nurses are sometimes needed. In the mathematical model, this is described by the variables

y_{jl} = the number of substitute nurses that work shift j on day l .

Because of the extra cost that arises when substitute nurses are hired, it is not desirable to use them, and the sum of the number of substitute nurses is therefore penalised in the objective function, see Section 4.1.4.

When a nurse is carrying out administrative tasks, or attending meetings or courses, she is on duty without fulfilling the staffing demand at the ward or the clinic. These occasions are usually known in advance, and they should be taken into account when the schedule is created, see (4)–(9). It is also known in advance when a nurse is going on holiday and this is marked in the proposed schedule. For further information see *individual task* and *holiday* in Sections 3 and 4.1.4.

4.1.2 Scheduling rules

A schedule presented by an optimisation tool must of course be consistent with prevailing laws and regulations regarding staff scheduling. This section gives our interpretation of how these rules are applied at the pilot ward.

A nurse can work, at the most, one shift per day, see (12), (22), (28)–(30). Working a day or an evening shift will not affect which shifts the nurse can work the following days. If a nurse works a night shift, then it is neither possible for her to work a day or an evening shift the next day, nor a day shift the day after that, see (13)–(17), (23)–(26).

At the pilot ward, a nurse is allowed to work at most five days in a row, see (20), (29), because then she is assumed to need at least one day off to rest. For each nurse, there is a *nominal number of weekly working hours*; it would be very limiting to try to schedule each nurse this exact number of hours each week, and in order to achieve more flexibility when creating the schedule, we have introduced the limits that a nurse shall work between 40% and 150% of her nominal number of hours per week, see (18).

Since each nurse is employed for working a certain number of hours per week, on average, it is necessary to keep track of how many hours she has actually been at work each specific week, see (27), (31). The pilot ward uses a *time bank* which registers the total number of hours that a nurse has been working. This number of hours is compared with the number of hours that the nurse should have worked, yielding a *time bank status*. The rule at the ward is that this status can vary between -20 and 20 hours. The time bank implemented in our model allows this same interval. The current time bank status is given as input data to the model, and then the status is constrained to be within the given interval each week, see (21). To further ensure that each nurse works about the right number of hours, the total number of hours worked during a whole scheduling period is allowed to deviate at the most $\pm 10\%$ from the nominal total number of hours, see (19).

Together with the time bank, the rules described by the constraints (18), (20), and (29) distribute the scheduled shifts evenly over the weeks. The time bounds given in percentage of full-time, see constraint (18), have large impact for those nurses that work a low percentage of full-time. For the nurses working full-time, or almost full-time (ca $> 80\%$), the rule of working at most five days in a row, see constraints (20) and (29), has more impact for evening out the workload between weeks.

4.1.3 Quality aspects

In order to create a schedule that complies with laws and regulations, it is sufficient to take into account the aspects described previously in this section, but considering these aspects only will not be sufficient for producing a schedule that can actually be used by a nursing ward. We will here present the remaining scheduling rules that we have identified to be necessary when creating a schedule; these will be referred to as quality aspects. At a first glance one might believe that the quality aspects are less important than the rules described earlier, but this is actually not the case. The quality aspects have, in practice, shown to be just as important as the other scheduling rules.

Most nurses prefer not to work during weekends, which necessitates some special care when scheduling the weekend shifts. A principle at the ward is that if a nurse works a day or an evening shift during a weekend, then she shall be on duty all that weekend. Being *on duty during a weekend* means that the nurse shall work both Saturday and Sunday (a day or an evening shift), and at least either the evening shift on Friday or the day shift on Monday, or both, see (34)–(38).

It is also required that the weekend shifts are evenly distributed among the nurses. The weekend shifts are defined to be all shifts from Friday evening to Sunday night, and the nominal number of weekend shifts for a nurse to work is the average value when the weekend shifts are evenly distributed among the nurses. The number of weekend shifts that a nurse actually works is allowed to deviate at the most by ± 2 shifts from the nominal number of shifts. To ensure complete fairness in the long run, this deviation is kept in a *weekend bank* to the next scheduling period, see (33). If the weekend bank is omitted,

then there is of course a risk that the same nurse could be the least favoured one for several consecutive periods.

The night shifts are often unpopular and they should therefore also be evenly shared. Here the same line of reasoning as for the weekend shifts is applied. The nominal number of night shifts for a nurse is the average value when all these shifts are evenly distributed among those nurses that work night shifts. The deviation from the nominal number of shifts is kept in a *night bank*, and the value of this deviation is constrained to be at the most ± 3 shifts.

4.1.4 Fulfilment of requests in the proposed schedules

As emphasized earlier, we aim at developing an optimisation tool that is practically useful and has properties that makes it easy for a nursing ward that is using the studied self-scheduling process to incorporate it into the existing scheduling process. Because the manual process is centered around how to fulfil the requests in the proposed schedules, a crucial component of the mathematical modelling is how to take these requests into account when creating the schedule.

When making the pilot implementation, all the types of requests presented in Section 3 were used. As described there, the requests can be categorised into two groups, hard requests, that must be fulfilled, and soft requests that should be fulfilled whenever possible.

Four of the requests are considered to be strong, and those are vetoes for and against working, individual tasks, and holidays. At the pilot ward, the nurses are allowed to use at the most one veto per week, but they are also allowed to save up to three vetoes and use all of them the same week to form a short holiday. A veto against working can be used for a single shift or be extended to count for the whole day.

In the model, all the hard requests are fulfilled through predetermined assignments of the shifts in question, before the remaining schedule is created, see (39)–(40). There is of course a risk that it is impossible to make all these predetermined assignments simultaneously without violating any scheduling rule. Our policy is that if it is not possible to fulfil all these hard requests, then the head nurse is responsible for deciding which of them should be fulfilled or not, since this decision requires complete knowledge of the motives for the requests.

Modelling the soft requests, that is, the requests for working and the strong requests for and against working, is not as immediate as modelling of hard requests. When designing this part of the model we focused on two, in our opinion, vital properties. Firstly, in order to facilitate acceptance from the nurses, the way of modelling requests cannot introduce a lot of complicated rules for the nurses to consider when proposing a schedule. Secondly, the model should be such that it is difficult, or at best, impossible for the nurses to outwit and take advantage of its properties when proposing a schedule.

When using the manual self-scheduling, the nurses request to work as many shifts as they are supposed to work during the scheduling period, and this is a rule that we have

chosen to keep. What needs some consideration is the use of strong requests, both for and against working shifts. We decided to let the nurses use as many strong requests as they wish, and to construct the model such that the more strong requests they use, the less each request will be worth. For example, if a nurse uses no strong requests against working, then it shall be equivalent if all her requests for working are either ordinary ones only, or strong ones only, given that she uses the same number of requests in both cases.

It is further desirable to make it possible to compare the fulfilment of requests, irrespective of the number of requests made and the percentage of full-time. We therefore introduce a predetermined reference value, p , that will for each week represent that all of a nurse's requests are fulfilled, on average. The fulfilment of requests for each nurse could then be represented as a portion of this reference value. The value of p can be chosen arbitrarily since it serves as a baseline only. We consider all requests to be fulfilled if a nurse works all the shifts that she has requested to work and is off work all the shifts that she has requested not to work. The value $-p$ represents that a nurse got none of her requests fulfilled, that is, when she has to work all the shifts that she has requested not to work and she does not work any of the shifts that she has requested to work.

The soft requests are quantified by associating each kind of request with a *desirability grade* representing how important it is that the request is fulfilled. A request for working a shift is associated with a positive value of the grade, while a request against working is associated with a negative value. The absolute value of the desirability grade represents the strength of the request.

Which requests that become fulfilled by the optimisation tool depend on the choice of values for the desirability grades (see below), and these values must be chosen to reflect how to prioritise between the requests. The choice of value for the grade representing a strong request for working compared to the grade for an ordinary request for working has mainly been decided empirically through a dialogue with the head nurse.

An inherent property of the model is that it is easier to fulfil a request for not working than to fulfil a request for working, since the number of shifts that a nurse shall not work is larger than the number of shifts that she shall work. Because of this, the grade representing a strong request against working has a lower absolute value than the strong request for working.

In a proposed schedule, there are numerous shifts without any requests, and those shifts are associated with the desirability grade 0. We use the desirability grade 1 for the request for working a shift, the grade 5 for the strong request for working a shift, and the grade -3 for the strong request against working a shift. In the model, the desirability grades are represented by the coefficients

$$c'_{ijl} = \text{desirability grade for nurse } i, \text{ shift } j, \text{ and day } l.$$

The fulfilment of requests is quantified by introducing a *score* for each nurse. This score is an aggregated measure of the fulfilment of requests. Since the nurses might have

different numbers of requests, the desirability grades for each nurse are scaled by dividing them by the sum of the absolute values of all desirability grades. The grades are multiplied by p and the number of weeks w , in order to normalise and establish a maximal possible score of pw per period, and a minimal possible score of $-pw$ per period. (See further Section 4.3.) Together this yields the *scaled desirability grades*

$$c_{ijl} = \frac{pw}{\sum_l \sum_j |c'_{ijl}|} c'_{ijl}, \quad \forall i, j, l.$$

The score for nurse i

$$P_i = \sum_l \sum_j c_{ijl} x_{ijl} - \sum_l \sum_j c_{ijl} (1 - x_{ijl}), \quad \forall i,$$

consists of two sums, where the first one contributes when a nurse works a shift, and the second one when she does not, see (41). The absolute value of the contribution to the score for one shift will always be the absolute value of the scaled desirability grade for that shift, and what differs is the sign of the contribution; when a request is fulfilled, the contribution is positive, and otherwise it is negative.

If the sum of the scores for all nurses is maximised, then the total fulfilment of requests is maximised without regarding any fairness among the nurses. For example, all nurses but one can have a high fulfilment of requests at the expense of a really poor fulfilment for the last nurse. To prevent this situation from occurring we have introduced an auxiliary variable,

$$Q = \min_i P_i.$$

By maximising the value of this variable, it becomes important to fulfil requests for the least favoured nurse, and thereby some degree of fairness is enforced, see (42)–(43). As a compromise, we consider both of these goals described.

The overall objective function is a linear combination of the three goals: maximise the sum of the scores for all nurses, maximise the lowest score, and minimise the number of substitute nurses used, that is

$$z = \sum_i P_i + \alpha Q - \beta \sum_l \sum_j y_{jl}.$$

Because it is always prioritised to minimise the number of shifts staffed by substitute nurses, the value of the coefficient β is chosen large. When creating schedules for the pilot ward, we used $\beta = 1000$ and no substitute nurse was used.

When assigning a value for α , one needs to consider which is most important, to fulfil many requests or to fulfil them in a fair way. Tests with different values of α revealed

that the result obtained is not sensitive to small changes with respect to this, probably because there is no conflict between the two first terms in the objective function, instead they are striving in the same direction. After some testing and discussions with the head nurse of the pilot ward, we chose to use $\alpha = 2$.

4.2 The optimisation model

To give an overview of the optimisation model described above, its components are summarised as follows:

Minimise	The number of shifts staffed by substitute nurses.
Maximise	The fulfilment of requests in the proposed schedules in a fair way.
Subject to	Staffing demand: <i>Demand for trained nurses, assistant nurses, and nurses at the clinic; individual tasks; holidays.</i>
	Scheduling rules: <i>Number of hours worked per week and total; maximum five days worked in a row; feasible combinations of consecutive shifts; special restrictions for nights; the time bank.</i>
	Quality aspects: <i>The night bank; the weekend bank; sharing of the shifts at the clinic; rules for working weekends.</i>
	Auxiliary constraints: <i>Veto; definition of least score; etc.</i>

The number of constraints and variables used in the mathematical model depends on how it is actually formulated and on the number of auxiliary variables used. In our formulation there are a few types of auxiliary variables and about 25 different kinds of constraints, some of them being very straightforward and some of them being more complicated. When scheduling 31 nurses for 8 weeks and after presolving in CPLEX, our model contains approximately 3000–5000 variables, mainly binary ones, and 5000–7000 constraints.

When using the optimisation tool, only one scheduling period is considered at a time. Nevertheless, the schedules must follow the same rules in the transition from one period to the next as during a scheduling period. To enforce this, all the information needed is saved from one scheduling period to the next, where it is used as input data. Examples of this kind of information are how many days in a row each nurse has worked at the end of the scheduling period, which nurses worked the last weekend, and the statuses of the time, weekend, and night banks.

As described earlier, variables representing the use of substitute nurses have been introduced in the model to ensure the staffing demand constraint to become fulfilled.

Despite this, there is a risk that the problem will lack a feasible solution, for example if it is not possible to make all the predetermined assignments. For all our instances of data, it has however been possible to create a feasible schedule.

An inherent property of the optimisation tool is that its performance relies heavily on the quality of the proposed schedules. If all the proposed schedules would comply with the scheduling rules and together fulfil the staffing demand, then the result produced by the optimisation tool would be that the nurses should work according to their proposed schedules. On the other hand, if the requests in the proposed schedules would deviate very much from the staffing demand, then most of them would be impossible to fulfil. These issues will be further discussed in Section 4.3, where the results from the pilot implementation are presented and evaluated.

Our work with developing the model can be described as an iterative process involving meetings with the head nurse and preliminary tests and evaluations. The scheduling rules used at the pilot ward were poorly documented, which made them cumbersome to identify and formalise. After developing a preliminary model, which was believed to be sufficiently complete and accurate to be useful, we began tests on real data, that is, schedules proposed by the nurses at the ward. The first three attempts to create a schedule for the pilot ward did however reveal both weaknesses and defects in the model, and also some misunderstandings in the communication with the head nurse. Because of the short time window, a few days only, for delivering a schedule, we did not succeed to correct the model on time and therefore the nurses used the manual process instead. During this work we learned that the head nurse found it much easier to convey their scheduling rules when studying a suggestion for a schedule, than by trying to list all rules and restrictions that they consider. After finishing these three initial tests, we delivered schedules for the pilot ward for three scheduling periods, and these schedules were used after some minor adjustments only.

In the next section we present some results from these scheduling periods. We evaluate the results with respect to both the outcome from the nurses' perspective, and the optimisation tool itself.

4.3 Result

Within the scope of the pilot implementation, our optimisation tool was used to deliver schedules to the pilot ward for three scheduling periods. During these periods, the model remained almost unchanged, except for adaptations to small changes in the conditions at the ward. An example of such an adaptation is that for the third period, the ward employed two newly graduated trained nurses and they were not allowed to work the evening or night shifts together, because of the low staffing levels at these shifts. The head nurse reported that they made some minor adjustments of the delivered schedules, something that probably is, and always will be, inevitable; but compared to the work of creating the schedules manually from scratch, this work is negligible.

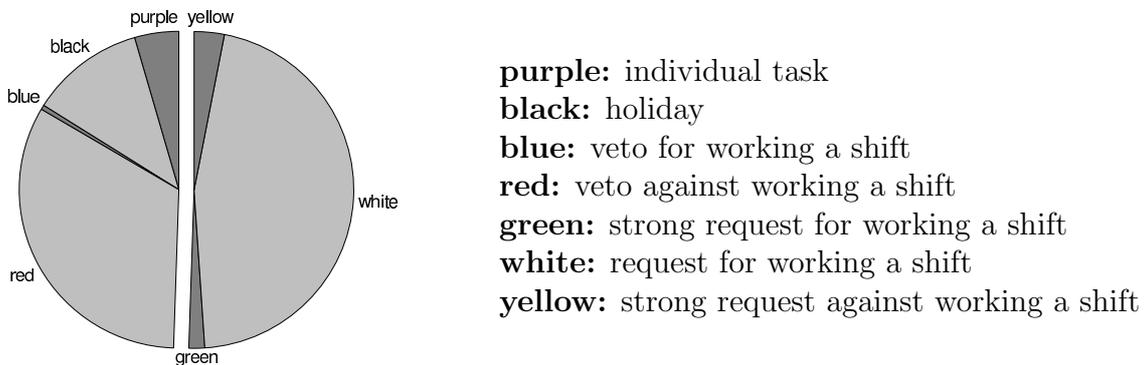


Figure 3: Distribution of requests in the proposed schedules.

The three periods, for which we produced usable schedules, will henceforth be referred to as Periods 1, 2, and 3. Each scheduling period was eight weeks long, and the staff was almost the same in all three periods, with the above mentioned exception of the two new nurses hired for Period 3.

During the pilot study, the nurses proposed schedules as they are used to do, but we encouraged them to also use the strong requests for, and against, working, which was the new kind of requests introduced when automating the process. The nurses were not so willing to try this new opportunity, something that hopefully would improve over time if the optimisation tool was to be used permanently.

For each period, about 1500 requests were given in the proposed schedules, and the statistics for all three periods is presented together in Figure 3. As can be seen from the figure, about 50% of the requests are predetermined assignments of the shifts in question. The number of purple requests represents how many individual tasks that need to be carried out, and this can vary both over time and between the nurses. Some of the nurses have recurring individual tasks, while other have such tasks only seldomly. The use of black requests depends on how many nurses who will be on holiday, and during the pilot study, some of the nurses had a few weeks of holiday. Almost all nurses used the maximum number of vetoes (the blue ones for working and the red ones against working, respectively) with the ones against working being the most popular. A veto against working can be used for a single shift only or for a whole day, and the nurses prefer to use these vetoes for a whole day. In the statistics presented here, we count a veto for a whole day as three vetoes, one for the day, evening, and night shift, respectively. Among the requests that are fulfilled when possible, the ordinary request for working (white) is dominating, but some of the nurses did use the possibility of making the strong requests, both for and against working (green and yellow, respectively).

The problem is solved by CPLEX, that uses branch and bound, see Wolsey [20]. It was neither possible, nor of interest, to wait for a solution that is guaranteed to be optimal. For the solution to be practically useful it is sufficient to find a solution where the difference

	Execution time	Gap	Substitute nurses	Average score	Lowest score
Period 1	6 minutes	5%	0	461	253
Period 2	4 minutes	2%	0	580	308
Period 3	< 1 minute	2%	0	477	94

Table 4: Results from the three scheduling periods.

between the dual bound and best integral solution found is a few per cent. This difference will henceforth be called *the gap*.

Out of the three periods, the first one turned out to be the most difficult one to schedule; six minutes were required to find an integral solution within a gap of 5%. For the two following periods, less than four minutes were sufficient to find a solution within a 2% gap. Data representing the outcome from the pilot study is presented in Table 4, with the first two columns presenting the execution time and the gap, respectively, and the three last columns containing the three components of the objective function.

No substitute nurses were used in any of the periods, which is a welcome result since the head nurse reported that the ward often has shortages in their manually created schedules. The average maximum score per week, p , introduced in Section 4.1.4, was here set to 100, yielding the score interval -800 to 800 as the possible outcome for each nurse. For the first two periods, the score of the least favoured nurse was quite high, but in the third period one of the nurses had a score that was significantly lower than the other nurses' scores. The reason for the low fulfilment of her requests was easy to reveal; this nurse had requested to carry out some individual tasks, combined with some strong requests for working, and the combination of these requests was impossible to comply with. Since the requests for individual tasks must be fulfilled, the conflicting strong requests for working could not be fulfilled, which corresponded to a loss of at least 230 points.

To more thoroughly present the outcome of the scheduling from the nurses' perspective, Figure 4 presents an histogram over how many nurses who received scores within given intervals. The outcomes from all the periods are presented in the same histogram, since they were quite similar. As shown in Figure 4, 90% of the nurses had a score of 300 or more, which corresponds to 68% or more of the requests being fulfilled. The average score is 507, which corresponds to 81% of the requests being fulfilled. We believe these results from the pilot study to be very promising, but since the possibility to fulfil requests highly depends on the properties of the proposed schedules, it would require a much more extensive study to draw any certain conclusions.

The kind of results presented this far are interesting for evaluating the outcome of the pilot study, but we also want to further evaluate and analyse *the concept* of automating the self-scheduling process. A cornerstone in the presented self-scheduling is that the nurses shall request as many shifts as they should work. The deviation between the proposed

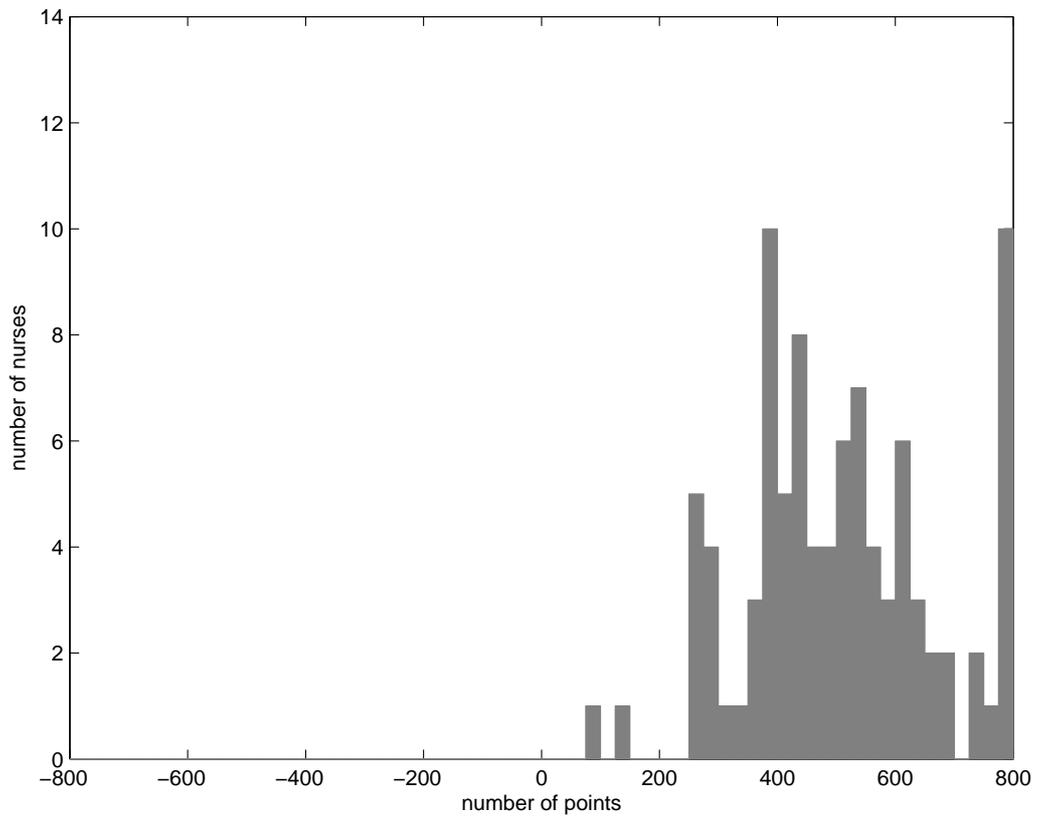


Figure 4: The total result of how many nurses that received scores within the given intervals.

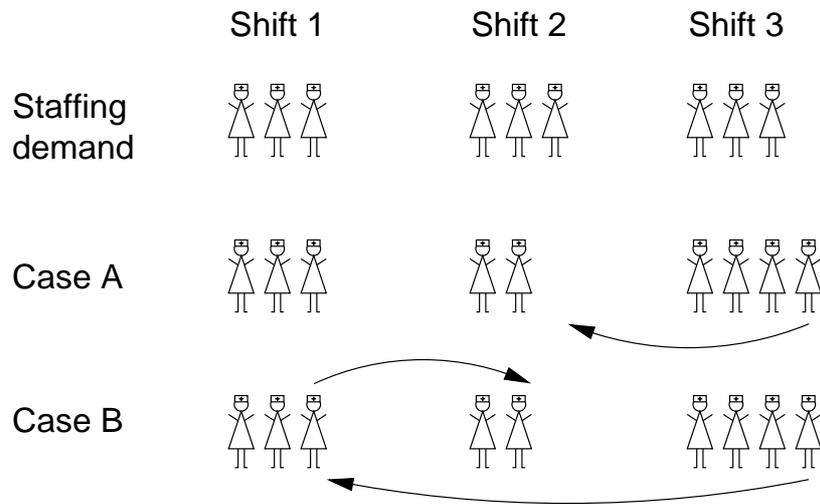


Figure 5: Illustration of the two kinds of swaps; in Case A only a primary swap is carried out, and in Case B both a primary and a secondary swap is carried out

schedules and the final ones can therefore be represented by *swaps* of shift, that is, if a nurse does not work a requested shift, then she has to work some other shift. The swaps can be divided into two types, *primary swaps* and *secondary swaps*. By primary swaps are meant those that are directly caused by the difference between the staffing demand and the requests in the proposed schedules. The primary swaps are sufficient to redistribute nurses from over-staffed shifts to the ones with a shortage, but in order to create a schedule that is also consistent with the scheduling rules, further swaps might be required, and those are called secondary. The two types of swaps are illustrated in Figure 5. In this example, there are three shifts that need to be staffed, these shifts can represent any three shift in a scheduling period, and they do not need to be consecutive or ordered. The staffing demand is three nurses for each shift. In their proposed schedules, there are three nurses requesting the first shift, two requesting the second, and four requesting the third one. In Case A it is possible to let one of the nurses that requested to work the third shift to work the second shift instead, requiring a primary swap only. In Case B, it is not possible for any of the nurses requesting the third shift to work the second one, for example because they do not have the right skills. Instead one of these nurses is moved to shift one, which she is allowed to work, and then one of the nurses from shift one, who can work the second shift, is moved to the second shift. In this case both a primary swap and a secondary swap is required.

In the analysis of our approach for automating the self-scheduling process, we only consider requests for working shifts. In Figure 6 there are three bars representing the result for each scheduling period. The first, black, bar shows the total staffing demand for

the period. The second, grey, bar shows the number of requests in the proposed schedules that coincide with the staffing demand, that is, the difference between the first and the second bar is the number of primary swaps. The last, white, bar represents how many of the requests, of any strength, for working that were fulfilled by the optimisation tool. Hence, the difference between the second and the third bar is the number of secondary swaps, plus the number of requests not fulfilled because the schedule does not correspond to an optimal solution.

In our example, the total staffing demand is nine nurses, and for each of the shifts, the number of requests that coincide with the staffing demand are three, two, and three, respectively. In the kind of diagram described, the first and the second bar would be nine and eight units high, respectively, with the difference corresponding to the single primary swap required. In Case A, the third bar would be of the same height as the second one, since only a primary swap is needed, while in Case B, the third bar would be one unit lower, because of the secondary swap.

The number of primary swaps required can clearly not be affected by any optimisation tool, or any other approach for creating the schedule, because it depends solely on the difference between the staffing demand and the requests in the proposed schedules. If this difference was significantly large, there would be no reason to use self-scheduling at all, because it would then not be possible to meet many of the requests anyway. However, as can be seen in Figure 6, about 85% of the requests made during the pilot study coincides with the staffing demand, which is a promising result.

Beforehand, we could not foresee whether the number of secondary swaps would be large or not, and it was a welcome result that they were as few as shown in Figure 6. Although the secondary swaps are induced by the scheduling rules, the number of such swaps is highly dependent on how the ward is staffed. If all nurses work a high percentage of full-time, then it is more difficult to create a schedule than if there are more nurses, all working a lower percentage of full-time. It is also of importance how many nurses who have the same skills, and therefore can fulfil the same demand, since it increases the likelihood that a primary swap is sufficient.

4.4 Qualitative feedback

This far we have only evaluated the results quantitatively, but the qualitative feedback from the head nurse and the nurses working at the ward is also of interest; we conclude this section by presenting this feedback. Since we performed the pilot study at this ward, its nurses might not be the best to ask when evaluating the idea of incorporating an optimisation tool into self-scheduling, because they have been exposed to our mistakes during the study, and it must be difficult for them to distinguish between factors that depend on our mistakes and those that do not.

The head nurse felt that it took some time to understand what we were trying to achieve, how difficult it was, and what information that was needed to carry it out. Dur-

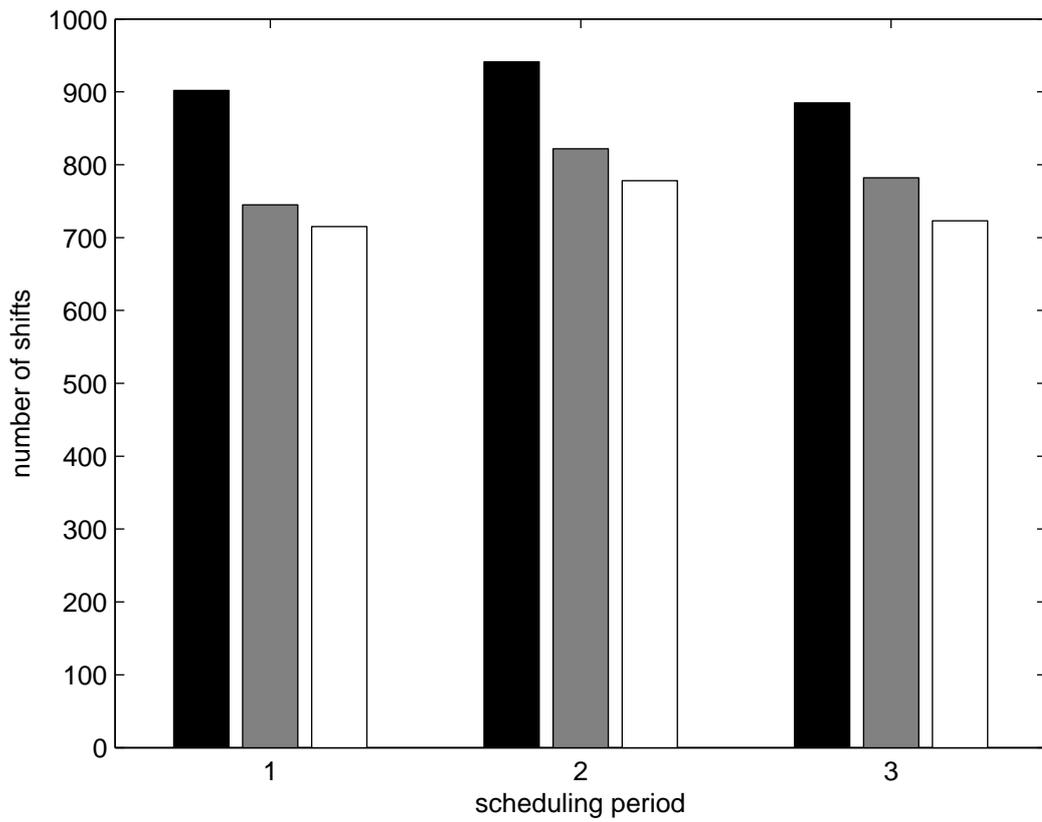


Figure 6: Fulfilment of requests for working a shift. For each scheduling period the first bar shows the number of requests used, the second one how many of these that coincide with the staffing demand, while the third one shows to the number of requests fulfilled.

ing the pilot study the communication between ourselves and the head nurse improved significantly, but it took some time and some misunderstandings along the way. Throughout the whole study, the head nurse has been very encouraging and interested in our work, which has simplified a lot.

We have only had direct contact with the head nurse of the pilot ward and the opinions from the other nurses there have been passed on to us by her. According to the head nurse some of the nurses were very sceptical to the idea that a computer, that is, our optimisation tool, should be involved in their scheduling process, mainly because they were afraid to lose their influence on the scheduling. More encouraging is that there were also nurses that were curious about the idea and thought it would be a relief if we could develop a computerised tool for their scheduling.

The head nurse seems convinced that it should be possible to incorporate an optimisation tool in the kind of self-scheduling that they use, but that it is crucial how it is implemented and presented to the nurses. She also agrees that it is important to consider the schedule created by this tool as a suggestion that can be adjusted by the nurses themselves, whenever needed.

5 Conclusions and further development

The long-term goal of our work is to develop an optimisation tool that is practically useful for automating the kind of self-scheduling described in this paper, and we find it very promising that our prototype tool managed to deliver usable schedules of high quality to the pilot ward. We consider this pilot implementation to be only the first step in our work, and in order to be able to draw further conclusions, the model needs to be generalised and tested at more wards.

During our work, the responses from the nurses have been both expectant and sceptical; expectant because of the time-consuming work and difficulties associated with the manual process, and sceptical mainly because of the nurses' loss of influence on the outcome of the scheduling. Because of the nurses' scepticism it is important to emphasize that the optimisation tool only provides a qualified suggestion for a schedule, and encourage the nurses to make minor adjustments themselves if beneficial. The great benefit of using our approach should be the time and effort saved if the head nurse is handed a schedule that complies with the scheduling rules, takes the requests in the proposed schedules into account, and, if possible, fulfils the staffing demand.

The principles for the scoring are quite satisfying, but some improvements are needed. In the model presented, we maximise the sum of the scores and also the least score, to achieve fairness for each single scheduling period. During our development work we have learned that the level of fairness is crucial to the nurses, and to improve the level of fairness during a single scheduling period it is better to also maximise the score for, say, the three least favoured nurses. It could also be of interest to study the impact of using

other values for the desirability grades.

Furthermore, we think that a memory for passing on the score between the scheduling periods is needed. In particular, if a nurse has a low fulfilment of requests in one scheduling period, then she must be favoured the next. Preferably, the average value of the scores should tend to the same value for all nurses after many scheduling periods. This long term fairness could for example be achieved by introducing a score bank, similar to the night bank.

The pilot ward usually needs some substitute nurses to cover shortages in the fulfilment of the staffing demand. In retrospect, it is difficult to conclude which of these shortages were known in advance and occurred because of scheduling difficulties, and which occurred in the short term planning. A welcome result is that the schedules we delivered did not include any planned use of substitute nurses. It is further a promising result that the number of secondary swaps needed can be as small as indicated by the result in the pilot study.

Our experience from the Swedish healthcare is that the composition of the staff is determined empirically. We believe it to be of great interest to use operations research methods to study how the number of secondary swaps, and the possibility to schedule without the use of substitute nurses, depend on how a ward is staffed, for example if there are many or few nurses working part time.

The next, and ongoing, step of our work is to generalise the model, and especially make it flexible enough to be used at most Swedish nursing wards. When generalising, our goal is to create a mathematical model that can be adapted to the varying conditions at different wards by mainly changing the input data, and not the model itself. A key feature would be that the managers of the ward are provided with a set of possible scheduling rules and then they choose which ones to use. We believe it to be important to not only offer the most basic scheduling rules, but to also offer quality aspects that can be used as an instrument for the managers to control the properties of the schedules.

In this pilot study, we have used the modelling language AMPL and the solver CPLEX, because of the simplicity and flexibility the combination of these two tools offer when developing a mathematical model for a complex problem. It could also be advantageous to use a meta-heuristic for solving this nurse scheduling problem, and we have the ambition to develop a tabu search strategy for this purpose. Because of the inherent properties of the self-scheduling problem under consideration, a tabu search method could be designed to start from the proposed schedules and to then strive to fulfil the scheduling rules and the staffing demand, while deviating from the proposed schedules as little as possible.

There are, of course, a lot of additional aspects to consider when designing an optimisation tool to be used in real-life at nursing wards. How well it will succeed is probably highly dependent on how the tool is introduced to the nurses, and if the schedules delivered are considered to have the desired properties and enough fulfilment of requests. Also, if the optimisation tool shall be used at many nursing wards, a graphical user interface for proposing and delivering schedules needs to be developed.

6 Acknowledgment

Thanks to all healthcare representatives that we have been in contact with, for sharing your knowledge about nurse scheduling. Special thanks to head nurse Elisabet Shimekaw and all the nurses working at the pilot ward.

We also thank the reviewers for careful reading and constructive suggestions for improvements.

References

- [1] T. Arthur and N. James. Determining nurse staffing levels: A critical review of the literature. *Journal of Advanced Nursing*, 19:558–565, 1994.
- [2] L. Bailyn, R. Collins, and Y. Song. Self-scheduling for hospital nurses: An attempt and its difficulties. *Journal of Nursing Management*, 15:72–77, 2007.
- [3] J. F. Bard and H. W. Purnomo. Preference scheduling for nurses using column generation. *European Journal of Operational Research*, 164:510–534, 2005.
- [4] J. F. Bard and H. W. Purnomo. Hospital-wide reactive scheduling of nurses with preference considerations. *IIE Transactions*, 37:589–608, 2005.
- [5] I. Berrada, J. A. Ferland, and P. Michelon. A multi-objective approach to nurse scheduling with both hard and soft constraints. *Socio-Economic Planning Sciences*, 30:183–193, 1996.
- [6] E. K. Burke, P. de Causmaecker, G. vanden Berghe, and H. van Landeghem. The state of the art of nurse rostering. *Journal of Scheduling*, 7:441–499, 2004.
- [7] B. Cheang, H. Li, A. Lim, and B. Rodrigues. Nurse rostering problems — a bibliographic survey. *European Journal of Operational Research*, 151:447–460, 2003.
- [8] K. A. Dowsland. Nurse scheduling with tabu search and strategic oscillation. *European Journal of Operational Research*, 106:393–407, 1998.
- [9] K. A. Dowsland and J. M. Thompson. Solving a nurse scheduling problem with knapsacks, networks and tabu search. *Journal of the Operational Research Society*, 51:825–833, 2000.
- [10] A. T. Ernst, H. Jiang, M. Krishnamoorthy, B. Owens, and D. Sier. An annotated bibliography of personnel scheduling and rostering. *Annals of Operations Research*, 127:21–144, 2004.

- [11] A. T. 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.
- [12] L. Fagerström, A-K. Rainio, A. Rauhala, and K. Nojonen. Professional assessment of optimal nursing care intensity level: A new method for resource allocation as an alternative to classical time studies. *Scandinavian Journal of Caring Science*, 14: 97–104, 2000.
- [13] B. Ghosh and G. Cruz. Nurse requirement planning: A computer-based model. *Journal of Nursing Management*, 13:363–371, 2005.
- [14] *ILOG AMPL CPLEX System, Version 10.0, User's Guide*. ILOG, 2006.
- [15] *ILOG CPLEX 10.0, User's manual*. ILOG, 2006.
- [16] B. Jaumard, F. Semet, and T. Vovor. Case study: A generalized linear programming model for nurse scheduling. *European Journal of Operational Research*, 107:1–18, 1998.
- [17] P. Karlsson. Rapport om kartläggning av arbetstidslösningar inom landstinget. Stockholms läns landsting, 2005.
- [18] M. Moz and M. V. Pato. Solving the problem of rostering nurse schedules with hard constraints: New multicommodity flow models. *Annals of Operations Research*, 128:179–197, 2004.
- [19] R. Venkataraman and M. J. Brusco. An integrated analysis of nurse staffing and scheduling policies. *Omega: The International Journal of Management Science*, 24: 57–71, 1996.
- [20] L. A. Wolsey. *Integer programming*. John Wiley & Sons, 1998.

A The mathematical model

This appendix contains a mathematical description of the model presented in Section 4. First, the parameters, variables, and sets are introduced, and then follows the model, with the constraints categorised as in Section 4.2. In A.3 to A.7, the constraints are numbered as referred to in the text, while formulas for calculating parameter values are not numbered.

The most fundamental sets and indices used are:

- The sets of assistant and trained nurses are denoted by I^a and I^t , respectively, and $I = I^a \cup I^t$ will denote the set of all nurses.
- The three types of shifts, day, evening, and night, are indexed by $j = 1, 2, 3$.
- The days of a scheduling period are indexed by $l = k + 7(w - 1)$, where $k = 1, \dots, 7$ are the seven weekdays and $w = 1, \dots, sw$ are the weeks of the scheduling period.

In order to simplify the description in Section 4, we used a basic version of the variables. For instance, x_{ijl} used there, is in this appendix represented by both X_{ijl} and AP_{ijl} . Furthermore, the variables Y_{jl}^a and Y_{jl}^t in the appendix are in the text only represented by y_{jl} , not specifying the kind of substitute nurse, and the same simplification is made for Q^a and Q^t , called Q in the text.

A.1 Sets and parameters

notation	description
b_{kw}^{rec}	staffing demand at the clinic on weekday $k = 1, \dots, 7$ in week $w = 1, \dots, sw$
b_{jl}^t	the minimum number of trained nurses required on shift $j = 1, 2, 3$ on day $l = 1, \dots, 7sw$
b_{jl}^{tot}	the exact number of nurses required on shift $j = 1, 2, 3$ on day $l = 1, \dots, 7sw$
bn_{mn}	the nurse $i \in I$ the predetermined assignment $m \in mb$ considers, $n = 1$ the week $w = 1, \dots, sw$ the predetermined assignment $m \in mb$ considers, $n = 2$ the weekday $k = 1, \dots, 7$ the predetermined assignment $m \in mb$ considers, $n = 3$ the shift $j = 1, 2$ the predetermined assignment $m \in mb$ considers, $n = 4$ the number of hours that the predetermined assignment $m \in mb$ considers, $n = 5$
c_{ijl}	scaled desirability grade for nurse $i \in I$ at shift $j = 1, 2, 3$ on day $l = 1, \dots, 7sw$
c'_{ijl}	desirability grade for nurse $i \in I$ at shift $j = 1, 2, 3$ on day $l = 1, \dots, 7sw$
en	the average number of nights that a nurse shall work during the period
et_i	the average number of hours that nurse $i \in I$ shall work per week
ew	the average number of weekends that a nurse shall work during the period
I	the set of nurses
I^a, I^t	the sets of assistant and trained nurses, respectively
lbt^w	lower bound on the number of hours to be worked in one week, in per cent
lbt^{tot}	lower bound on the total number of hours to be worked, in per cent
$lsan$	the subset of nurses that worked the last Saturday night in the previous period
$lsun$	the subset of nurses that worked the last Sunday night in the previous period
lw	the set of nurses with $NM_i = 1$ in the previous period
mb	the set of predetermined assignments
$me1$	a set of exceptions from the set of predetermined assignments mb
$me2$	a set of exceptions from the set of predetermined assignments mb
n_i	$= 1$ if nurse $i \in I$ can work night shifts, and $= 0$ otherwise
ns_i	night bank status for nurse $i \in I$ when the period starts
p	the maximum score per week
pt_i	percentage of full-time for nurse $i \in I$
rn	the set of nurses that work at the clinic, $rn \subseteq I^t$
sw	number of weeks in the scheduling period
t_{ijk}	the length of shift $j = 1, 2, 3$ on weekday $k = 1, \dots, 7$ for nurse $i \in I$, in hours
t^{rec}	the length of a shift at the clinic, in hours
ts_i	time bank status for nurse $i \in I$ when the period starts
ubt^w	upper bound on the number of hours to be worked in one week, in per cent
ubt^{tot}	upper bound on the total number of hours to be worked, in per cent
ws_i	weekend bank status for nurse $i \in I$ when the period starts

A.2 Variables

Three types of variables are used, binary, integer, and continuous ones, this is specified by respectively using the marks (B), (I), and (C) in the definition.

notation	description
AP_{ijl}	= 1 if nurse $i \in I$ is working shift $j = 1, 2, 3$ on day $l = 1, \dots, 7sw$ without contributing to the staffing demand b_{jl}^{tot} , and = 0 otherwise, (B)
AT_{iw}	the number of hours nurse $i \in I$ works without contributing to the staffing demand b_{jl}^{tot} in week $w = 1, \dots, sw$, (C)
MP_{ijl}	= 1 if nurse $i \in I$ is on holiday or performs an individual task on shift $j = 1, 2, 3$ on day $l = 1, \dots, 7sw$, and = 0 otherwise, (B)
MT_{iw}	the number of hours that nurse $i \in I$ spends on holiday or individual tasks in week $w = 1, \dots, sw$, (C)
NM_i	= 1 if nurse $i \in I$ shall work the first Monday the next period, and = 0 otherwise, (B)
NNS_i	night bank status for nurse $i \in I$, (C)
P_i	score for nurse $i \in I$, (C)
Q^a, Q^t	lowest score for the nurses in I^a and I^t , respectively, (C)
R_{ikw}	= 1 if nurse $i \in rn \subseteq I^t$ works at the clinic on day $k = 1, \dots, 7$ in week $w = 1, \dots, sw$, and = 0 otherwise, (B)
RP_{ijl}	= 1 if nurse $i \in I$ works at the clinic on shift $j = 1, 2, 3$ on day $l = 1, \dots, 7sw$, and = 0 otherwise, (B)
RT_{iw}	the number of hours nurse $i \in I$ works at the clinic in week $w = 1, \dots, sw$, (C)
TNS_{iw}	time bank status for nurse $i \in I$ and week $w = 1, \dots, sw$, (C)
W_{iw}	= 1 if nurse $i \in I$ works the weekend in week $w = 1, \dots, sw$, and = 0 otherwise, (B)
WNS_i	weekend bank status for nurse $i \in I$, (C)
X_{ijl}	= 1 if nurse $i \in I$ contributes to the staffing demand b_{jl}^{tot} on shift $j = 1, 2, 3$ on day $l = 1, \dots, 7sw$, and = 0 otherwise, (B)
Y_{jl}^a	number of assistant substitute nurses on shift $j = 1, 2, 3$, on day $l = 1, \dots, 7sw$, (I)
Y_{jl}^t	number of trained substitute nurses on shift $j = 1, 2, 3$, on day $l = 1, \dots, 7sw$, (I)

A.3 The objective function

$$\max \quad z = \sum_{i \in I} P_i + \alpha(Q^a + Q^t) - \beta \sum_{l=1}^{7sw} \sum_{j=1}^3 (Y_{jl}^a + Y_{jl}^t)$$

A.4 Staffing demand

$$\sum_{i \in I^t} X_{ijl} + Y_{jl}^t \geq b_{jl}^t, \quad j = 1, 2, 3, \quad l = 1, \dots, 7sw \quad (1)$$

$$\sum_{i \in I} X_{ijl} + Y_{jl}^t + Y_{jl}^a = b_{jl}^{tot} \quad j = 1, 2, 3, \quad l = 1, \dots, 7sw \quad (2)$$

$$X_{i3l} = 0, \quad i \in I : n_i = 0, \quad l = 1, \dots, 7sw \quad (3)$$

$$MT_{iw} = \sum_{m \in m'} bn_{m5}, \quad m' = \{m : bn_{m1} = i, bn_{m2} = w\}, \quad i \in I, \quad w = 1, \dots, sw \quad (4)$$

$$me1 := \{m : bn_{m2} = 1 \text{ and } bn_{m3} = 1, 2\} \cup \{m : bn_{m4} = 2\}$$

$$me2 := \{m : bn_{m2} = 1 \text{ and } bn_{m3} = 1\}$$

$$X_{bn_{m1}, 3, bn_{m3} - 2 + 7(bn_{m2} - 1)} = 0, \quad m \in mb \setminus me1 \quad (5)$$

$$X_{bn_{m1}, 3, bn_{m3} - 1 + 7(bn_{m2} - 1)} = 0, \quad m \in mb \setminus me2 \quad (6)$$

$$X_{bn_{m1}, 1, bn_{m3} + 7(bn_{m2} - 1)} = 0, \quad m \in mb \quad (7)$$

$$X_{bn_{m1}, 2, bn_{m3} + 7(bn_{m2} - 1)} = 0, \quad m \in mb \quad (8)$$

$$X_{bn_{m1}, 3, bn_{m3} + 7(bn_{m2} - 1)} = 0, \quad m \in mb \quad (9)$$

$$\sum_{i \in rn} R_{ikw} = b_{kw}^{rec}, \quad k = 1, \dots, 5, \quad w = 1, \dots, sw \quad (10)$$

$$\frac{1}{|rn|} \sum_{w=1}^{sw} \sum_{k=1}^5 b_{kw}^{rec} - 3 \leq \sum_{w=1}^{sw} \sum_{k=1}^5 R_{ikw} \leq \frac{1}{|rn|} \sum_{w=1}^{sw} \sum_{k=1}^5 b_{kw}^{rec} + 3, \quad i \in rn \quad (11)$$

A.5 Scheduling rules

$$\sum_{j=1}^3 X_{ijl} \leq 1, \quad i \in I, l = 1, \dots, 7sw \quad (12)$$

$$X_{i3l} + X_{i,1,l+1} \leq 1, \quad i \in I, l = 1, \dots, 7sw - 2 \quad (13)$$

$$X_{i3l} + X_{i,2,l+1} \leq 1, \quad i \in I, l = 1, \dots, 7sw - 2 \quad (14)$$

$$X_{i3l} + X_{i,1,l+2} \leq 1, \quad i \in I, l = 1, \dots, 7sw - 2 \quad (15)$$

$$X_{i11} = 0, \quad i \in lsan \quad (16)$$

$$X_{ijl} = 0, \quad i \in lsun, (j, l) \in \{(1, 1), (2, 1), (1, 2)\} \quad (17)$$

$$et_i := pt_i(36, 2n_i + 37(1 - n_i)), \quad i \in I^a$$

$$et_i := pt_i(36, 2n_i + 38.15(1 - n_i)), \quad i \in I^t$$

$$lbt^w et_i \leq \sum_{k=1}^7 \sum_{j=1}^3 t_{ijk} X_{i,j,k+7(w-1)} + AT_{iw} \leq ubt^w et_i, \quad i \in I, w = 1, \dots, sw \quad (18)$$

$$lbt^{tot} et_i sw \leq \sum_{w=1}^{sw} \left(\sum_{k=1}^7 \sum_{j=1}^3 t_{ijk} X_{i,j,k+7(w-1)} + AT_{iw} \right) \leq ubt^{tot} et_i sw, \quad i \in I \quad (19)$$

$$\sum_{ll=l}^{l+5} \sum_{j=1}^3 (X_{i,j,ll} + AP_{i,j,ll}) \leq 5, \quad i \in I, \quad (20)$$

$$l \in \{1, \dots, 7sw - 5\} \setminus \{l : \exists c'_{i,1,ll} \Leftrightarrow \text{PURPLE}, ll = l, \dots, l + 5\}$$

$$\sum_{k=1}^7 \sum_{j=1}^3 t_{ijk} X_{i,j,k+7(w-1)} + AT_{iw} - et_i + TNS_{i,w-1} = TNS_{iw}, \quad i \in I, w = 1, \dots, sw$$

with $TNS_{i0} = ts_i$ and $TNS_{iw} \in [-20, 20]$ (21)

$$X_{i,j,k+7(w-1)} + R_{ikw} \leq 1, \quad i \in rn, \quad j = 1, \dots, 3, \quad k = 1, \dots, 5, \quad w = 1, \dots, sw \quad (22)$$

$$\begin{aligned} X_{i,3,k-1+7(w-1)} + R_{ikw} &\leq 1, \\ i \in rn, \quad (k, w) &\in \{(k, w) : k = 1, \dots, 5; w = 1, \dots, sw; k = w \neq 1\} \end{aligned} \quad (23)$$

$$\begin{aligned} X_{i,3,k-2+7(w-1)} + R_{ikw} &\leq 1, \quad i \in rn, \\ (k, w) &\in \{(k, w) : k = 1, \dots, 5; w = 1, \dots, sw; k = w \neq 1; k = w + 1 \neq 2\} \end{aligned} \quad (24)$$

$$R_{i11} = 0, \quad i \in lsan \quad (25)$$

$$R_{ik1} = 0, \quad i \in lsun, \quad k = 1, 2 \quad (26)$$

$$RT_{iw} = \begin{cases} t^{rec} \sum_{k=1}^5 R_{ikw}, & i \in rn, \quad w = 1, \dots, sw \\ 0, & i \in I \setminus rn, \quad w = 1, \dots, sw \end{cases} \quad (27)$$

$$RP_{i,j,k+7(w-1)} = \begin{cases} R_{ikw}, & i \in rn, \quad j = 1, \quad k = 1, \dots, 4, \quad w = 1, \dots, sw \\ 0, & \text{otherwise} \end{cases} \quad (28)$$

$$AP_{ijl} = MP_{ijl} + RP_{ijl}, \quad i \in I, \quad j = 1, \dots, 3, \quad l = 1, \dots, 7sw \quad (29)$$

$$RP_{ijl} + MP_{ijl} \leq 1, \quad i \in I, \quad j = 1, \dots, 3, \quad l = 1, \dots, 7sw \quad (30)$$

$$AT_{iw} = RT_{iw} + MT_{iw} \quad i \in I, \quad w = 1, \dots, sw \quad (31)$$

A.6 Quality aspects

$$en := \sum_{l=1}^{7sw} b_{3l}^{tot} / \sum_{i \in I} n_i$$

$$\sum_{l=1}^{7sw} X_{i3l} - en + ns_i = NNS_i, \quad i \in I : n_i = 1, \quad \text{with } NNS_i \in [-2, 2] \quad (32)$$

$$ew := \sum_{w=1}^{sw} \left(\sum_{j=2}^3 b_{j,5+7(w-1)}^{tot} + \sum_{k=6}^7 \sum_{j=1}^3 b_{j,k+7(w-1)}^{tot} \right) / |I|$$

$$\sum_{w=1}^{sw} \left(\sum_{j=2}^3 X_{i,j,5+7(w-1)} + \sum_{k=6}^7 \sum_{j=1}^3 X_{i,j,k+7(w-1)} \right) - ew + ws_i = WNS_i, \quad i \in I$$

with $WNS_i \in [-2, 2]$ (33)

$$W_{iw} = \sum_{j=1}^2 X_{i,j,6+7(w-1)} = \sum_{j=1}^2 X_{i,j,7+7(w-1)}, \quad i \in I, \quad w = 1, \dots, sw \quad (34)$$

$$W_{iw} \leq X_{i,2,5+7(w-1)} + X_{i,1,7w+1}, \quad i \in I, \quad w = 1, \dots, sw - 1 \quad (35)$$

$$\sum_{i \in I^a} NM_i = \sum_{i \in I^t} NM_i = 2 \quad (36)$$

$$W_{i,sw} \leq X_{i,2,5+7(sw-1)} + NM_i, \quad i \in I \quad (37)$$

$$X_{i11} = 1, \quad i \in lw \quad (38)$$

A.7 Auxiliary constraints

$$X_{ijl} + RP_{ijl} = 1, \quad i \in I, \quad j = 1, 2, 3, \quad l = 1, \dots, 7sw \text{ such that } c'_{ijl} \Leftrightarrow \text{BLUE} \quad (39)$$

$$X_{ijl} + RP_{ijl} = 0, \quad i \in I, \quad j = 1, 2, 3, \quad l = 1, \dots, 7sw \text{ such that } c'_{ijl} \Leftrightarrow \text{RED} \quad (40)$$

$$c_{ijl} := (p \cdot sw \cdot c'_{ijl}) / \sum_{l=1}^{7sw} \sum_{j=1}^3 |c'_{ijl}|, \quad i \in I, \quad j = 1, \dots, 3, \quad l = 1, \dots, 7sw,$$

such that $c'_{ijl} \Leftrightarrow \text{GREEN, WHITE, or YELLOW}$

$$P_i = \sum_{l=1}^{7sw} \sum_{j=1}^3 \left(2c_{ijl}(X_{ijl} + AP_{ijl}) - c_{ijl} \right), \quad i \in I \quad (41)$$

$$Q^a \leq P_i, \quad i \in I^a \quad (42)$$

$$Q^t \leq P_i, \quad i \in I^t \quad (43)$$