Operating room scheduling in the Erasmus Medical Center

Thesis Operations Research and Quantitative Logistics

Erasmus University Rotterdam

Tim Grevelink 348967

Supervised by: Prof. dr. Albert P.M. Wagelmans

Co-reader: Dr. Wilco van den Heuvel

May 23, 2016

Abstract

In this research we constructed different operating room schedules that fit the needs of the Erasmus Medical Center. The schedules differ in scheduling strategy, working hours, overtime and unused capacity. The operating rooms of the Erasmus Medical Center will be moved and centralized in a new building, which results in a whole new set of restrictions that should be incorporated. Surgical procedures are grouped per specialism and operation codes, from which the durations are estimated. The open scheduling strategy and block scheduling strategy are used to construct schedules. A first-fit heuristic is used to schedule the procedures for both strategies, whereafter a simulated annealing heuristic improves those schedules. The overtime and unused capacity are the lowest under the open scheduling strategy. For both strategies we found schedules that perform better in terms of overtime and unused capacity compared to the current situation. An iterative process is needed to come up with a block schedule that fits all the needs of the Erasmus Medical Center, which should be continued. The constructed model can be used to check the performance of proposed block schedules, which may be useful in the coming years before the migration to the new building.

1 Table of contents

Contents

2 Summary

The operating rooms (ORs) of the Erasmus Medical Center will be centralized at a new location. The capacity of the new operating theatre is unknown, partly because the number of ORs will be decreased. The arriving procedures of the coming years have to be handled, but in which working hours? The specialisms have restrictions on the ORs they can use and the surgical procedures have stochastic durations and occurrence rates. The OR theatre is expensive and has to be used most efficiently. The unused capacity has to be minimized, while the overtime should also be within the boundaries of the Erasmus Medical Center.

To give a substantiated advise to the Erasmus Medical Center we have investigated two of the most commonly used scheduling strategies, namely the block scheduling strategy(BSS) and the open scheduling strategy(OSS). All surgical procedures are grouped by specialism and operation codes to get equally distributed procedures. The distribution of the duration of each type of procedure is estimated. Each schedule we have constructed, consists of preassigned time per specialism where specified procedures are handled. A schedule is constructed with a first-fit heuristic, whereafter the constructed order of procedures is improved by a simulated annealing heuristic. In case of OSS the objective is to minimize the total overtime. In the BSS there are predefined blocks. The goal here is to minimize the overtime, because those blocks should be exceeded as little as possible. All schedules are constructed with data in a training set, whereafter they are tested with data from the test set. The durations are simulated, from where the measures (overtime and unused capacity) have been calculated.

The results, of average overtime and unused capacity, are used to compare the schedules. The schedules that are constructed, differ in terms of input of working hours, number of ORs, block times and the probability that a group of procedures exceeds the scheduled time. All OSS schedules have a better performance than the BSS schedules, just as expected. Comparing the measures with the current situation is difficult, because of the assumptions we made. In the current situation there are elective cases that are intentionally scheduled outside working hours. We assume that all elective cases should be inside working hours and emergency cases do not take place in 'elective ORs'. Under

these assumptions, we improve the schedules definitely. Despite the challenging comparison we emphasize that the schedules, we have used, meet the questions of the Erasmus Medical Center. A BSS schedule is advised, because of its usability for the employees. The schedules with only 7 working hours per OR-day will lead to increasing overtime, which is not favored. The schedules of 11 working hours have nice results in overtime, but are not realistic on a daily basis. Somewhere around 9 hours per OR-day should be realistic for the Erasmus Medical Center.

Some extensions can be made in further research, like the growth in the number of surgical procedures. Besides an iterative process can be used with different insights from the Erasmus Medical Center to improve the schedules we found with BSS. The evaluation of the model can be improved and investigated in an earlier stage with a smaller problem. The outcomes can be improved when using more iterations.

The surgical procedures will fit in the new building. Dependent on the schedule the measures will differ. The Erasmus Medical Center can choose their own schedule based on their own insights. The OSS leads to better outcomes compared to BSS, notwithstanding the BSS is advised. A schedule with OR-days of 7.5 working hours, that are used in the current situation, is not recommended. With less ORs available, it is expected that the working hours should increase to come up with comparable outcomes. The Erasmus Medical Center can try different block times for this strategy, whereafter they can use the schedule that fits their demand best.

3 Introduction

The Erasmus Medical Center (Erasmus MC) started a renovation in 2009. To let all changes run smoothly, there are a few supporting programs for the staff of the Erasmus MC. One of them is 'Werken In de Nieuwbouw (WIN)', which started to get all medical and supporting services ready to move in the new buildings. One of the major changes in the new situation is the centralization of the Operating Rooms (ORs).

Improving the strategy that is used in the OR department may lead to less capacity problems. The high number of stochastic factors in ORs, causes an uneven stream of patients through the hospital. When the knowledge about those factors increases, we are capable to better deal with those fluctuations. This research focuses on the capacity of the ORs. It is unknown what working hours should be used to fulfill all surgical procedures in the upcoming years. Because of the changed restrictions of the new situation we use several methods to construct multiple schedules for those procedures. A schedule consists of multiple blocks from which each block is assigned to a certain specialism and a certain OR on a specific day. The timeframes for the blocks and number(s) of ORs to include in the schedule can be chosen. Multiple restrictions are needed to compose the procedures sets. Each set can include procedures that differ in duration, the cause to operate, the requirements to operate and so on. Because of the initial stage of this project and the varying interests that occur in the health sector, there are multiple research questions and restrictions that are highly relevant to be answered.

Obviously, this research deals with a part of the questions that appear during the WIN project. The capacity of the new building has to be identified. The surgical procedures that are present have to be fulfilled without too much overtime. A number of decisive factors are of main importance for the results of this research. We focus on a few parts. First, we have to deal with the stochastic duration of the surgical procedures. What should we take into account when we want to come to a realistic duration of a procedure at the moment of scheduling? The functionality and number of ORs that should be build have already been decided. Not every surgical procedure can take place in any OR. Therefore we assume the non-identicallity of ORs, while most OR researches assume multifunctional ORs.

Another decision that has to be made, is the way of scheduling. The two strategies, that are often used in OR scheduling, are the 'open scheduling strategy' and the 'block scheduling strategy', which are explained in Section 5. Those two strategies are compared in terms of capacity that is needed, to fulfill all surgical procedures. Changes in the healthcare sector occur frequently. Therefore the methods we are going to use should be easy adaptable to slightly different assumptions.

The main question we want to answer during this research is the following: which capacity of the operating rooms (in terms of working hours) is needed in the new situation to fulfill all surgical procedures without too much overtime?

Before we can answer this, we describe some questions that are directly related. These questions can be seen as a roadmap of this research. We start with questions about the data. How are surgical procedures distinguished, while the operations have stochastic durations? What surgical procedures should be incorporated in the scheduling and what capacity has to be reserved for those operations?

When we look more into the methods, we want to know the reliability and robustness. Do the outcomes differ much when using slightly different input? Finally, we want to get results and give a recommendation to the Erasmus MC. For each constructed schedule we want to know the surgical procedures that are incorporated, the working hours, the amount of overtime, unused OR capacity and the planned slack (scheduled delay). What difference in schedule do we get when using different strategies? Is the gain of a block schedule worth the possible loss in flexibility? How do the constructed schedules perform compared to the current situation in terms of overtime and unused capacity? The questions are based on the interest of the Erasmus MC that wants to know how the surgical procedures fit in the new situation. Besides, literature is used to highlight the topics that are already known.

"If we knew what it was we were doing, it would not be called research, would it?" (Albert Einstein). However, to stay organized, the outline of this research is given here:

in the next section the relevance of this research is explained and for whom the results are of interest. In Section 5 there is some knowledge discussed from previous researches, whereafter the motivation to start this research is addressed. The formal problem and the methods we will use during this research are explained in Section 7 and 8. Section 9 describes the data that are used to come to the results. Thereafter the results are shown and we discuss the research in Section 10 and 11. The last part we conclude this research and give a recommendation to the Erasmus MC, which can be seen in Section 12.

4 Relevance

One of the most expensive departments of the hospital is the operating theatre. The Erasmus MC wants to know, which capacity is needed for their new operating theatre to be able to fulfill all surgical procedures without too much overtime. The ORs have been built and the specialisms know in which ORs they are able to operate. Therefore non-identical ORs are taken into account. The model that will be used to construct the final schedules, should give the opportunity to make changes when needed. This makes the model practical to use for further research when the knowledge and questions about this subject change slightly.

In this research it is interesting to see what the non-identical ORs can do to the scheduling. The scheduling becomes much more complex than it already is. Because of the comprehensive problem that rises, the challenge is to use a quick method that is well applicable in this situation. Exact methods are probably not capable to find a solution in reasonable time. In the near future, the Erasmus MC wants to construct an integrated capacity model where not only the physical capacity, but also the personal availability and other factors are dependent of each other in all cases. This integrated capacity model possibly becomes an additional strategy decision making tool. Therefore not only the direct outcomes are highly important, but also the possibility to change the assumptions of the model and the applicability of the model in other or extended situations are of interest.

5 Literature

OR planning and scheduling is a widely examined subject. To get an overview of the problems that appear, Cardoen et al. (2010) made a literature review of operational research on OR planning and scheduling. This paper evaluates the relation to either the problem setting or the solution techniques and uncertainty incorporation that are used in other researches. This review examines not only researches that only focus on the OR department, but also researches that take more links of the chain of a patient in the hospital into account. The solution techniques vary from simulation, mathematical programming and improvement heuristics. Moreover, the uncertainty that is incorporated in the models differs from no uncertainty, to uncertainty in arrival, duration or other elements.

The performance measures that are taken into account lead to a variety of outcomes. In a medical center the priority does not only depend of the best possible care, the satisfaction of the patients or the utility of the staff, but also in the financial aspects. The main performance measures that are found in Cardoen et al. (2010) are waiting time, throughput, utilization, leveling, makespan, patient deferrals, financial measures and preferences.

Scheduling ORs is possible in two ways: The 'open scheduling' and the 'block scheduling' strategy (Fei et al. (2009)). In the first strategy a procedure can be handled at any workday, regardless of the specialism. To optimize the scheduling, the open scheduling strategy is more flexible with less restrictions. This is different in the block scheduling strategy. First, the specialisms are scheduled in blocks during the day. Those blocks are time frames that are assigned before the actual procedures are scheduled. The advantages of this block schedule is that the employees of the specialisms know their working hours for a longer period (the period for which the block schedule is built). Secondly, the specialisms assign surgeries to specific blocks that are part of their block schedule. The block scheduling strategy is a more often used method in hospitals, because of fixed working days of the employees. In this way they can work in one specialism for multiple surgical procedures.

One of the most obvious assumptions is the use of identical ORs for the whole scheduling.

Fei et al. (2009), Hans et al. (2008), Van Oostrum et al. (2008), Denton et al. (2010) all use identical/multifunctional ORs in their research. However, some specialisms need unique ORs that are differently equipped than others. The use of a heart pump requires special connections in the OR. Ophthalmology requires a fixed microscope on the ceiling. These are examples of requirements that are not possible to build in every OR. Although some specialisms need specific equipment it is possible in an emergency to operate elsewhere, which is probably not in favor of the patient.

Also Dexter et al. (1999) assumed identical ORs where they used a simulation, based on a next-fit, first-fit, best-fit or worst-fit algorithms which are used in a bin-packing problem. The ORs can be seen as the bins and the surgical procedures as the packages that have to be collected in the bins. The duration of the procedure determines the size of the package, just as the capacity of the OR determines the bin size. The objective in Dexter et al. (1999) is to minimize the waiting time together with high OR utilization. It can be seen that the use of these simple approaches may lead to accurate outcomes and easy to use models. Dexter et al. also used a block scheduling strategy. In the simulation two of the parameters were the number of hours in one block and the number of blocks in a week. A comparable research is Hans et al. (2008), who describe a surgery loading problem, which is a generalization of the so-called "general bin packing problem with unequal bins." In this problem the unequal sized bins can be extended by planning overtime. Both the bin-packing problem and surgery loading problem are strongly $\mathcal{NP}$-hard. The objective of Hans et al. (2008) is to maximize the OR utilization while minimizing the risk of overtime without cancelled surgeries. In this case they start with simple constructive heuristics, but extend the model with some local search heuristics. Those lead to significant improvements of the base solution.

An exact approach as column generation is also used in researches about OR scheduling. An example is Van Oostrum et al. (2008) who introduces a master surgical schedule(MSS) for the elective procedure that occur frequently. This MSS is a cyclic block schedule that repeats over time. The research directly involved the leveling of requirements of hospital beds into the problem. Another approach is stochastic programming which is used by Denton et al. (2010). A stochastic OR allocation problem is solved to optimize the costs of opening an OR when stochastic OR durations are taken into account. The model is especially useful to decide the number of ORs that have to be open on a specific day.

One of the shortcomings of most methods is the computational time that is needed to come to an appropriate solution. Pham and Klinkert (2008) uses a MILP formulation to solve the multi-mode blocking job shop problem they formulated. In the general job shop problem, jobs/operations have to be handled by machines/ORs. In this case, each job consists of a sequence of operations, where each operation needs a set of resources. This set is called a mode. When their is more than one mode available we deal with a multi-mode problem. A mode is available for a time interval. If this mode is chosen for an operation, the processing consists of the duration of the procedure and possibly the setup time and cleanup time. Pham and Klinkert conclude that (good) feasible solutions are only found in small to medium instances and other approaches are needed in larger cases.

Not only the methods to construct a model for the scheduling, but also the way of incorporating uncertainty is of high importance for the usability of the outcomes. A widely used research, Strum et al. (2000), models the uncertainty in surgical procedure times. They conclude to use lognormal duration times of surgical cases. Because the sum of lognormal distributed variables has no closed form, often the approximation of Fenton (1960) is used to model the total duration time. The goodness of fit is often tested with the Kolmogorov-Smirnov test (Massey Jr (1951)) or the Anderson-Darling test (Anderson and Darling (1954)). The null hypothesis of both tests is that the data come from a certain proposed distribution. If the distance with the empirical (sample) distribution is too large, the null is rejected. When the parameters of the proposed distribution are unknown/estimated, the critical values of the original Anderson-Darling test are not valid. The paper Stephens (1976) shows how the critical values of the Anderson-Darling test are calculated when the parameters are unknown.

In this thesis the OR scheduling problem is modeled with the use the bin-packing problem, which is described above. Each specialism, which is a block of operations/objects, has a set of ORs (bins) where it can be handled. The goal is to minimize the total number of ORs/bins used. In this way we know in which working hours all operations can be handled in the new situation. Non-identicallity of the ORs complicates the (bin-packing) problem drastically. It is however not known how this non-identicallity of the ORs/bins can be handled in a good way. Therefore we use a simple constructive heuristic from which the solution is improved later on.

In Melouk et al. (2004) and Eglese (1990) a simulated annealing(SA) approach is described to use as a tool for optimization problems. SA is a local search algorithm where improvements are made from an initial solution by exploring the neighborhood solutions. The solutions in the neighborhood are calculated, whereafter the possible new objective value is calculated. If the new objective is an improvement of the old objective, this new solution is chosen. If the new solution leads to a worse solution, this solution can also be chosen with some probability, which is done to make it possible to escape from local optima.

6 Motivation for this research

This research is important for the EMC to make sure that all surgical procedures will fit in the new situation within reasonable working hours. The difference compared to other researches is the centralization of the ORs at one location. A research needs a initial situation where it can be compared to. The change of the centralization of the ORs is new in literature.

In the health sector, changes are commonplace. Therefore, the method we use should be easy adaptable to slightly different assumptions. When the knowledge about the data changes or for example the surgeon should be taken into account in the scheduling procedure, it should be doable to make such small changes. Besides, the idea of the method stays the same, but with an extra restriction. The main interest of this research is the scheduling of the specialisms in the non-identical ORs.

In comparison with previous research, this paper considers the use of non-identical ORs in scheduling. The ideal situation for hospitals is to deal with identical ORs that can be used by every specialism. Unfortunately, this is not always possible. An example is the specialism Thorax where they make use of a heart pump. This requires special connections in the OR, which are very expensive.

The block scheduling is a widely used strategy in hospitals. In the current situation of the EMC, the block scheduling is also used. Therefore it is also interesting to compare the outcomes when using an open scheduling strategy. The tradeoff between the flexibility of the open scheduling and the usability of the block scheduling is clarified. Comparing the outcomes of the two different strategies leads to insight into the (dis)advantages of those strategies.

7 The problem

In this section we describe the problem, which is handled during this research. The schedule we construct, should be functional after a period of time. The main problem consists of multiple smaller problems that we discuss in this section. First the problem is described in a more general way, whereafter we go into more detail.

The surgical procedures that occur in a hospital have a varying duration. Therefore they can not all be handled in the same way. Maybe the large operations will occur more frequently, while the occurrence of the shorter operations is more varying. A type of procedure can be seen as surgical procedures that are similar in specialism and type of operation. The duration of each type surgical procedure has a certain distribution and each type occurs multiple times. After all procedures have been assigned to a type, we know the basic information that is needed for the scheduling. In the scheduling procedure we have to keep in mind the restriction that not all types of procedures can be handled in any OR. The ORs are non-identical and most ORs have already been assigned to a certain specialism, which results in an important restriction.

When we start to schedule we have to incorporate the possible delay of surgical procedures. Time to avoid that staff has to work in overtime, when there is a net delay of multiple operations, is called planned slack. If the amount of planned slack is taken minimal, the probability that operations are handled outside working hours increases. Besides, the planned slack is not only dependent on the surgical procedure itself, but also on the other operations in the same OR. Surgical procedures within the same specialism can exchange their hours. The most important aspect is that surgical procedures of equal specialism need less planned slack when they are scheduled together compared to the case where they are scheduled apart, this is called the portfolio effect (Hans et al. (2008)). Therefore we need to schedule procedures of the same specialism together. Besides we want to schedule the procedures with a commonly specialism together, because these can interchange time, which is not directly possible with procedures of different specialisms.

If we know what we want to schedule, we have to decide how we are going to schedule. The scheduling strategy describes the way of scheduling. In the 'open scheduling strategy' (OSS) the operations can take place at 'any' moment in time, while the 'block scheduling strategy' (BSS) has assigned blocks per specialism of a certain size that occur frequently over time. 'Any' is highlighted, because there are definitely restrictions. In both strategies there should be defined times in which the procedures may be handled. In BSS we need to specify the number of hours in each block along with the number of blocks in each week per specialism (block times). Obviously the assignment of those numbers per specialism is done before the actual scheduling is executed. The main difference between the strategies is that the block times are fixed and preassigned for a longer period in the BSS.

Selecting the block times together with the actual scheduling of the procedures is an iterative process. How should we choose the block times and how should the surgical procedures be scheduled in those blocks? The actual scheduling of the procedures is done for a smaller horizon than the construction of the block times. Those block times and restrictions of the BSS should be known before the actual scheduling starts. In the OSS the surgical procedures can be scheduled without considering the preassigned time that is given to a specialism.

7.1 Formal problem

First we have to deal with the uncertainty of this problem, which is the duration of the surgical procedures. We focus on the elective cases that can be scheduled ahead. These procedures will be separated into groups with the same characteristics. Each surgical procedure will be classified in a specialism $s = 1, \ldots, S$ with corresponding type $t = 1, \ldots, t_1 - 1, t_1, t_1 + 1, \ldots t_s - 1, t_s, t_s + 1 \ldots T_S$. The specialisms have one or multiple types of procedures, type $t = 1, \ldots t_1$ belong to specialism $s = 1$, $t = t_1 + 1, \ldots t_2$ belong to specialism $s = 2$ and so on. Each type surgical procedure has a certain distribution, from which we know the time that should be scheduled for each type t procedure.

To schedule the operations we denote an OR i on day j as OR-day(i, j). The total duration for all procedures of specialism s in OR-day(i, j) equals μ_{ijs}. This duration is dependent of the procedures that are scheduled in the same OR-day. The quality of the block schedule is also very dependent on the number of a type surgical procedure that has to be scheduled. This number can be varying during the year for each type of procedure. The time to schedule not only depends on the expected duration and the number of procedures, but also on the planned slack that is taken into account to avoid that surgical procedures are handled in overtime.

The amount of planned slack depends on the number of surgical procedures of each specialism in one OR-day. Specialisms themselves get a certain amount of OR time in a year that they can spend. If a specialism works in overtime, this is possibly the assigned time of another specialism. The exchange of time is in this case not directly possible. If operations of the same specialism are handled in one block, they can exchange time. If one of the two exceeds the average and the other is finished early, the OR may finish within working hours. The case when the surgical procedures are of the same specialism, but with different (average) lengths, complicates the problem. The exchange of time is possible, but they do not compensate each other in an equal way. So the planned slack may be less when the procedures are scheduled in the same OR, but not as much as the case of identical distributed procedures.

Each specialism s in one OR-day(i, j) has got planned slack δ_{ijs}. Each group of operations consists out of several types of procedures of one single specialism. If all procedures are scheduled into separate OR-days, there is much more planned slack built in the schedule and the total scheduled time is increased, because of the portfolio effect that was mentioned earlier. When all procedures of the same specialism are handled in one OR-day, the loss of time in terms of planned slack is minimized. This leads to the lowest amount of planned slack and thus the least amount of time that has to be scheduled. In this case also the minimum amount of capacity is needed. An OR-day is however limited with its capacity. In this problem, we assume certain capacities as given.

Besides the unknown capacity of the ORs, they are also non-identical. This means that not all surgical procedures can be handled in any OR, which complicates the problem. The next problem we have to deal with are the different strategies. For the BSS, we need to assign for each OR-day(i, j) and per specialism s a block of B_{ijs} hours. This can be varying from zero to the capacity of one OR-day. The block times are fixed over a period of time and should be known upfront. For OSS we need to know in which OR-days the specialisms can operate, whereafter the actual scheduling can start.

For both strategies we need a total scheduling horizon of length Q, which consists of n independent schedules of q days. If the whole scheduling horizon Q is scheduled, we can compare the outcomes. The actual scheduling is done shortly before the schedule is taken into production. We want the least scheduled overtime compared to the capacity of the schedule. Overtime is defined as the duration of procedures outside the capacity of an OR-day compared to the total duration in a schedule. With the actual scheduling, the problem arises how procedures should be scheduled when the capacity is limited. What if the total capacity is insufficient? The methods to construct an OSS or BSS schedule should incorporate these possible problems. Logically, the actual decision of the size of the blocks, that are most desirable, is an iterative process. When the total duration of procedures increases from one schedule to another and the block times stay the same, this is not in favor of the actual overtime.

After the scheduling is done, we need to compare the schedules with the realized durations. How can we compare these measures? In the next subsection we describe the problem of OSS and BSS in formulas, whereafter we introduce the methods we use to solve these problem.

7.2 Mathematical problem

7.2.1 Open scheduling strategy

The objective is to minimize the total amount of overtime under a given OR-day capacity. The possible OR-day capacity (c_{ij}, the capacity of OR i on day j) is unknown upfront. The total capacity c_s, in which specialism s is assigned to operate, is given.

We introduce a decision variable X_{ijt} to indicate the number of type t procedures that are assigned to OR-day(i, j) and Z_{ijs} as the maximum amount of time of specialism s that is assigned to OR-day(i, j). Constraint (2) is introduced to make sure that every surgical procedure is assigned to an OR-day, where a_t is the number of procedures of type t, that have to be scheduled.

As mentioned in Section 7.1: μ_{ijs} is the scheduled duration of specialism s in OR-day(i, j) and δ_{ijs} the planned slack of specialism s in OR-day(i, j). Both variables are dependent on the procedures of specialism s that are scheduled within one OR-day, $\mu_{ijs} = F^{-1}(0.5)$ and $\delta_{ijs} = F^{-1}(p) - F^{-1}(0.5)$, where F is the cumulative distribution function of the sum distribution of all procedure durations of specialism s in OR-day(i, j) and p is the probability that the duration of the procedure(s) stay(s) within working hours, with $p > 0.5$. This means that the planned slack equals the difference between the median duration and the duration that will not be exceeded with probability p.

The sum of the assigned times over all OR-days should be less than the total amount of time available for a specialism, see Equation (4). Equation (3) is that the planned time of a specialism s on a specified OR-day(i, j) is not exceeded, in which we consider surgical procedure duration after regular time as overtime O_{ijs} (on OR-day(i, j) for specialism s). Constraint (5) is that the total capacity of each OR-day is not exceeded by the sum of time assigned to the separate specialisms. The last two constraints are for the decision variables.

$$\min \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{s=1}^{S} O_{ijs} \tag{1}$$

$$\text{s.t.} \sum_{i=1}^{I}\sum_{j=1}^{J} X_{ijt} = a_t \qquad \forall t = 1,...,T_S \tag{2}$$

$$\mu_{ijs} + \delta_{ijs} \leq Z_{ijs} + O_{ijs} \qquad \forall i = 1,...,I, j = 1,...,J, s = 1,...,S \tag{3}$$

$$\sum_{i=1}^{I}\sum_{j=1}^{J} Z_{ijs} \leq c_s \qquad \forall s = 1,...,S \tag{4}$$

$$\sum_{s=1}^{S} Z_{ijs} \leq c_{ij} \qquad \forall i = 1,...,I, j = 1,...,J \tag{5}$$

$$X_{ijt} \in \mathbb{N} \qquad \forall i = 1,...,I, j = 1,...,J, t = 1,...,T_S \tag{6}$$

$$Z_{ijs} \in \mathbb{R} \qquad \forall i = 1,...,I, j = 1,...,J, s = 1,...,S \tag{7}$$

The number of variables become very large when we have multiple OR-days and multiple types per specialism. Therefore we have to use methods that can deal with this challenge.

7.2.2 Block schedule strategy

In BSS the specialisms get preassigned blocks. Therefore the flexibility of scheduling decreases. The main idea of scheduling stays the same, except for the restriction that surgical procedures can only be assigned to the blocks where they can be handled instead of the ORs where the specialism are allowed to operate.

At this time we assume that the number of hours of the preassigned blocks are given for each specialism. The time of all operations in OR-day(i, j) of specialism s may not exceed B_{ijs}. In the formulation above, the variable Z_{ijs} is replaced by parameter B_{ijs}, what restricts the solution area:

$$\min \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{s=1}^{S} O_{ijs} \tag{8}$$

$$\text{s.t.} \sum_{i=1}^{I}\sum_{j=1}^{J} X_{ijt} = a_t \qquad \forall t = 1,...,T_S \tag{9}$$

$$\mu_{ijs} + \delta_{ijs} \leq B_{ijs} + O_{ijs} \qquad \forall i = 1,...,I, j = 1,...,J, s = 1,...,S \tag{10}$$

$$X_{ijt} \in \mathbb{N} \qquad \forall i = 1,...,I, j = 1,...,J, t = 1,...,T_S \tag{11}$$

In the construction of B_{ijs} the capacity restrictions of the OR-days and specialisms have to be taken into account. The difference of BSS with OSS is that B_{ijs} does not change in the different weeks, where Z_{ijs} can differ much.

7.3 Number of hours of block time B_{ijs}

To allocate capacity to a certain specialism we need information about the factors that are of influence. B_{ijs} should be found for each specialism s and OR-day(i, j). If we construct a set A of all possible allocations of surgical procedures in a week, we have to construct the block time such that all surgical procedures in a week can be handled within block-time. This has to hold for *most* weeks of a year.

To make sure that all surgical procedures can be fulfilled in time, there should exist a subset A_b of all surgical procedure allocations that comply to the block-time restrictions B_{ijs}. As can be seen, the BSS introduces a two-stage procedure. First the blocks should be known, whereafter the procedures can be scheduled in those blocks. Both stages are very dependent of each other, which is called interdependence. As said earlier, it may require many adjustments in block times before this approach leads to the desired result in practice. Since we are searching for a reasonable comparison, the problem can be simplified by choosing B_{ijs}. This is discussed more extensively in the next section.

8 Methods

In this section we describe which steps are taken to come to a desirable result. The steps will be explained extensively in the remainder of this section. First the duration that has to be scheduled will be discussed together with the use of planned slack. Thereafter the construction of block-times is explained. Thirdly, we explain the scheduling procedure, where we need some input information, like the sets of ORs that each specialism can use, the possible working hours and the amount of planned slack that is incorporated in scheduling. After these basics are known, the initial solution will be found with the use of a constructive heuristic. This initial solution will be improved with a local search heuristic, whereafter we try different inputs for our solution approach. The outcomes, based on a simulation, are compared with respect to used/unused capacity and overtime.

The initial solution for the scheduling is constructed with a first-fit heuristic and improved with the use of Simulated Annealing, which are also used by Dexter et al. (1999) and Hans et al. (2008). These two methods are used in the actual scheduling, but also in the construction of block times. First we explain how the duration of the surgical procedures is incorporated.

8.1 Duration of procedures

For each type t there are a_t surgical procedures found in the used data that we take into account with a stochastic duration η_t, where η_t is a certain distribution per type. Besides, a_t should be large enough to get a proper distribution. As mentioned in Section 7.1, we want to know the duration of each type of procedure t. The procedures are divided in a training set and a test set, which is explained in more detail in the next section. The procedures from the training set are grouped on characteristics. Procedures with the same characteristics belong to one type t, from which the corresponding distribution η_t is calculated. This results in an expected duration μ_t and the standard deviation σ_t of each type t. With the use of the distribution we should be able to calculate the scheduled duration and planned slack of a group of procedures in an OR-day.

We sum all types of procedures that belong to specialism s in a specific OR-day and calculate the planned slack for that specific group of operations. Strum et al. (2000) indicates that surgical procedures are lognormal distributed. The sum of lognormal variables is again lognormal distributed, but has no closed form. Fenton (1960) constructed a widely used approximation of the parameters of this lognormal distribution, which we need to get the planned slack. These parameters are calculated using the following formulas:

$$\sigma_z^2 = \ln\left(\frac{\sum_{t=t_{s-1}+1}^{t_s}\left(e^{2\mu_t+\sigma_t^2}(e^{\sigma_t^2}-1)\right)}{\sum_{t=t_{s-1}+1}^{t_s}(e^{\mu_t+\sigma_t^2/2})}+1\right) \tag{12}$$

$$\mu_z = \ln\left(\sum_{t=t_{s-1}+1}^{t_s}\left(e^{\mu_t+\sigma_t^2/2}\right)\right)-\frac{\sigma_z^2}{2} \tag{13}$$

The scheduled duration for specialism s in OR-day(i,j) equals $\mu_{ijs} = F^{-1}(0.5|\mu_z,\sigma_z)$, where F^{-1} is the lognormal inverse distribution function. The planned slack for specialism

s in OR-day(i,j) is calculated as follows:

$$\delta_{ijs} = F^{-1}(p|\mu_z, \sigma_z) - F^{-1}(0.5|\mu_z, \sigma_z) \tag{14}$$

where p is the probability that the duration of the procedure(s) stay(s) within working hours and $p > 0.5$ The planned slack equals the difference between the median duration and the duration that is not exceeded with probability p.

To test whether the fitted lognormal distributions are acceptable to use, we evaluate the distribution with the Anderson-Darling(AD) goodness of fit test, stated in Stephens (1976). The parameters of the lognormal distributions will be estimated, which results in different critical values compared to the case where the parameters are known up-front. The null hypothesis of the Anderson-Darling is that the data come from a certain distribution. Because there are multiple types of procedures, it is acceptable that the null hypothesis will be rejected sometimes. Favorable is that all types of procedures fit the lognormal distribution. Because of the sum distributions above, it is difficult to use different distributions for different types of procedures. This is discussed further in Section 11.

The test statistic measures the distance between the hypothesized distribution $F(x)$ and $F_n(x)$, which is the empirical CDF. The general statistic equals

$$n \int_{-\infty}^{\infty} (F_n(x) - F(x))^2 w(x) dF(x) \tag{15}$$

where weight function $w(x) = [F(x)\,(1 - F(x))]^{-1}$ and the statistic is calculated over the ordered data sample $x_1 < x_2 < \cdots < x_n$ with n data points. We use the AD statistic:

$$A_n^2 = -n - \sum_{i=1}^{n} \frac{2i-1}{n} \left[\ln(F(x_i)) + \ln(1 - F(x_{n+1-i}))\right] \tag{16}$$

The critical values are calculated with use of a Monte-Carlo simulation. First, there is drawn a random dataset from the proposed distribution with estimated parameters. With the use of these datapoint, the AD test statistic is calculated. These steps are repeated multiple times, such that the Monte Carlo standard error for the p-value is small enough, see Equation (17), with m is the number of Monte-Carlo simulations. This standard error is the error due to simulating the p-value. The critical value equals the average of all simulated test statistics that are larger than the actual test statistic for the proposed distribution.

$$SE_{MC} = \sqrt{\frac{\hat{p}(1-\hat{p})}{m}} \tag{17}$$

8.2 Construction of block times

As explained in Section 7.2 the allocation of the number of hours in one block and the number of blocks per specialism can be seen as given in the actual scheduling. As mentioned earlier, the allocation of block-time is an iterative process. The construction of block times is not of main interest of this research. However, the outcomes of the block scheduling strategy are to a large extent dependent on the allocation of the block times.

To come up with likely outcomes for the chosen block time we choose block times from the OSS strategy for the training set. The OSS schedules can also be constructed for a part of the training set, from which the block times can be used as input for the BSS. Ones we chose block times, we can improve those with some iterations. Dexter et al. (1999) made use of a simulation where the block time was an input value that was changed during the research. We construct a model where the blocks can be inserted, whereafter the performance can be measured.

We choose B_{ijs} as the blocks where the planned slack is maximal from the OSS schedule for the last year of the training set. These blocks are used to schedule the procedures of the test set as described in the following subsections. Possibly we extend this with some iterations where a few block times are adjusted.

8.3 Scheduling

In the schedule, all the operations of the past years should be fulfilled. Before scheduling the procedures we identify where they can be scheduled, how many OR-days are available, what working hours are used and how much planned slack is taken into account. In both strategies it is possible that the total capacity is insufficient and not all procedures fit in the OR-days. Time is assigned to a specialism, whereafter the procedures are scheduled with the objective of minimal overtime. The question is how to assign time to a specialism where (most of) the procedures can be handled with minimal overtime and under some capacity restrictions? Both the allocation of time and the actual scheduling is done with the methods described below.

From the operations that have to be scheduled we need to know the number of surgical procedures in a week and the total duration. This differs a lot in a year, which has to be taken into account. The scheduling is done before the schedule is used. We should know how many procedures should be scheduled beforehand. We use the occurrence from the data, as described in Section 9.

When the duration per specialism has been calculated, the schedule can be constructed. We know how much surgical procedures per type have to be scheduled. The planned slack is minimized when all operations are scheduled in one OR-day, see Equation 14. Because this is not possible, we have to find a schedule where the similar types of procedures of the specialisms are combined as much as possible.

8.3.1 Initial solution / constructive heuristic

We start with the specialism that has the least OR-days where it can operate. Dependent on the number of different types of procedures we construct a schedule for only this specialism. If there is only one type, this is scheduled in the least possible OR-days.

If there are multiple types of procedures we want to use the least possible OR-days for this specialism. This will decrease the planned slack and therefore the total needed capacity. The optimal solution of this subproblem will be found by scheduling the used OR-days as full as possible. In this way we have the least planned slack. This problem looks similar to the bin packing problem, which is a special case of the cutting stock problem. The cutting stock problem is $\mathcal{NP}$-hard problem and therefore it is not likely to be polynomial solvable. Where the bin packing problem deals with bins with the same characteristics, the OR scheduling problem we discuss has to deal with non-identical ORs in addition to the similar bin packing problem. These non-identical ORs complicate the problem even more, from which we conclude that this problem will also be $\mathcal{NP}$-hard.

The heuristic we use to solve this $\mathcal{NP}$-hard problem is the first-fit heuristic. In Figure 1 the procedure of this initial heuristic is described. First, we should construct a procedure list. This is done for the period that has to be scheduled. We know which procedures have to be scheduled and we will use the training set to construct the parameters of the duration of each type of procedure. This way we can calculate the duration and planned slack, dependent on the OR-day where the procedure is scheduled.

After the procedure list is constructed, we select the specialism which has the least OR capacity available. The priority specialisms are the ones with certain ORs where they can operate. Those are scheduled first. If some specialisms have the same remaining capacity, we choose the specialism with the largest (in duration) remaining operation. We know the specialism from which we schedule the procedure with the longest duration. The ORs that are allowed the chosen specialism x to operate are ordered. The ORs that are allowed for specialism x are denoted with OR_x. Priority ORs are favored for a certain specialism above other ORs and are tried first. The (priority) ORs are checked for enough capacity. The (priority) ORs with the most capacity left, are tried earlier. If multiple ORs have the same capacity left, the ORs where the least other specialisms can operate are chosen first. $\mathrm{OR}_x max$ is the last allowed OR for specialism x that can be tried, to fit the procedure.

If we find an OR where the procedure fits, we update the OR capacity and the planned slacks (of the corresponding specialism(s)). Otherwise the procedure is put in a separate list, which has to be scheduled in the updating procedure. The procedure is than deleted from the list whereafter all steps are repeated until the procedure list is empty.

In the basic version of this algorithm, the procedures are sorted in decreasing order. In the case of non-identical ORs we have to take the allowable ORs into account. Otherwise it is possible that an OR where a certain specialism can operate is already out of capacity when those procedures have to be scheduled. After all surgical procedures are scheduled in an OR-day or in an imaginary OR, we want to improve the found solution. Especially because the worst case scenario of this first-fit heuristic can be pretty bad. The improvement of the found solution is done with a local search heuristic, which is described hereafter.

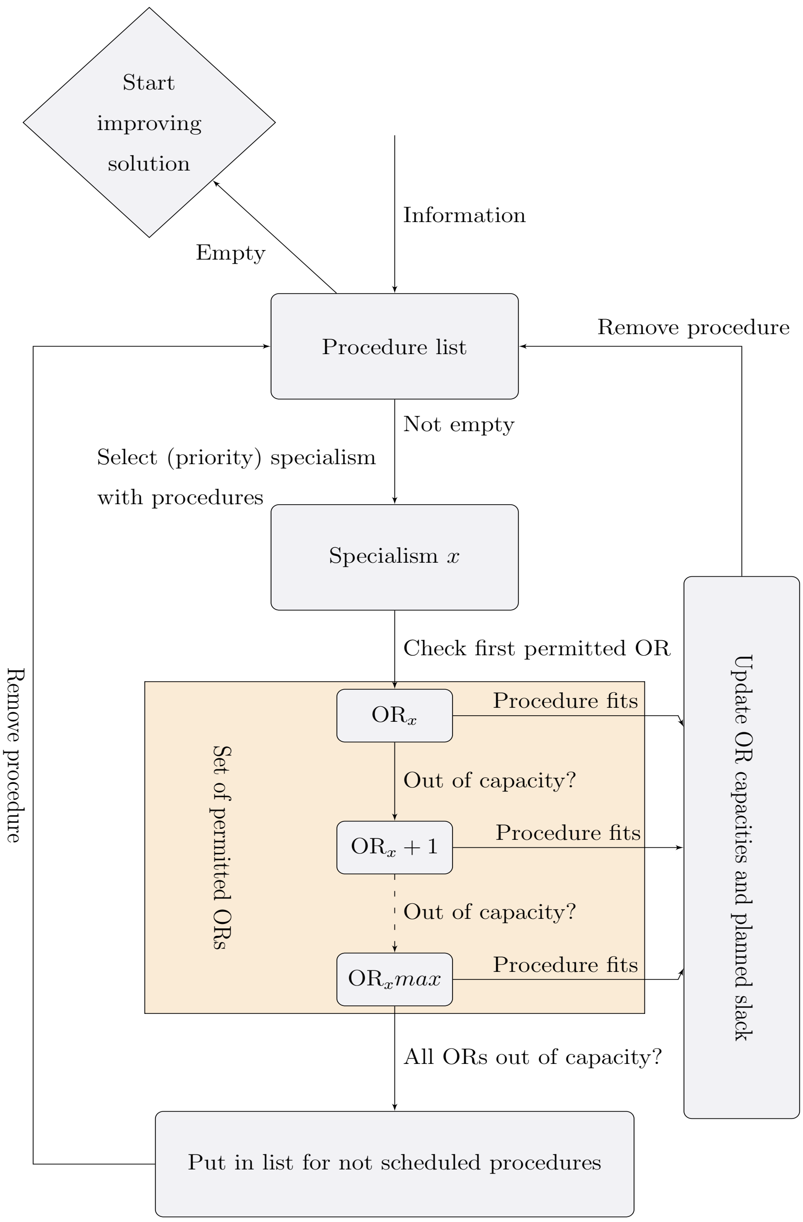


Figure 1: Steps of the initial heuristic

8.3.2 Updating procedure / improvement heuristic

For updating this initial schedule we use a simulated annealing (SA) heuristic, as described in Section 5. The initial solution we find, is stated in the previous subsection. This will be improved to a solution where the objective is minimized. The difficulty of improving the current solution is again that not all surgical procedures can be handled in any OR-day. Besides, the planned slack is dependent on the specialism that are currently scheduled in an OR. The neighborhood of this SA heuristic are all allowed swaps of procedures. A swap is allowed when the procedures that will be swapped, may be handled in each others ORs with respect to the non-identical ORs. In Figure 2 we see the steps that are taken in the improvement heuristic. The pseudo code of the SA heuristic is described in Algorithm 1.

Before we start with the improvement heuristic, we calculate the initial objective and the 'temperature', which is needed in the SA algorithm. The probability p of accepting a non-improving solution is dependent of the Temperature T: $p = \exp\left\{\frac{Obj_{max} - Obj'_{max}}{T}\right\}$ with Obj_{max} incumbent solution, Obj'_{max} the proposed (neighboring) solution. The begin Temperature $T_{\text{begin}} = Obj_{worst} - Obj_{best}$, where the best objective is chosen as the planned slack/overtime where all procedures of one specialism are scheduled together and the worst objective is chosen as the planned slack/overtime where all procedures are scheduled apart. Besides, then planned slack/overtime is minimal when the procedures of one specialism are scheduled together and maximal when they are scheduled apart. We start with a 'high' temperature, because in that way most non-improving solutions will be accepted. Later in the process we want less of those non-improving solutions to be accepted and that is why we use a cooling rate r to decrease the temperature. This rate is mostly somewhere between 0.8 and 0.99, see Eglese (1990). The parameter θ is chosen such that only the improvement of a single procedure should be executed. The probability of non-improving should be close to zero.

From the algorithm, it can be seen that we need some information in the current schedule before we can use this algorithm. For each procedure we need to know the duration, the specialism, the current OR-day where it is handled, the distribution parameters and the planned slack that belongs to the operation. In the schedule there is for every OR-day a dummy procedure which has a duration of the remaining capacity in that OR-day. The advantage of these dummy procedures is that it is possible to get more/less number of procedures than the amount in the initial situation in an specific OR-day after updating.

Algorithm 1 SA heuristic

```
procedure UPDATING(scheduleCurrent, specialismORPermission, coolRate, θ)
    while Temperature > θ do
        noImprovement ← 0
        while noImprovement < N do
            procedure1 ← random procedure from scheduleCurrent
            possibilitySwaps ←procedures allowed to swap with
                                    procedure1, specialismORPermission
            fits ← no
            while fits ≠ yes do
                procedure2 ← random procedure from possibilitySwaps
                if procedure1 fits in OR2 & procedure2 fits in OR1 then
                                    ▷ Fit with respect to the capacity restrictions
                    fits ← yes
                    scheduleNew ← new schedule with swap of procedure1, procedure2
                else
                    possibilitySwaps ← possibilitySwaps without procedure2
            objectiveNew ← sum of plannedSlack of scheduleNew
            if objectiveNew < objectiveCurrent then
                scheduleCurrent ← scheduleNew
                objectiveCurrent ← objectiveNew
                if objectiveNew < bestFound then
                    bestFound ← objectiveNew
                    bestSchedule ← scheduleNew
                    noImprovement ← 0
                else
                    noImprovement ← noImprovement + 1
            else
                if random(0,1) < exp (objectiveCurrent−objectiveNew)/Temperature then
                    noImprovement ← noImprovement + 1
                    scheduleCurrent ← scheduleNew
                    objectiveCurrent ← objectiveNew
        Temperature ← Temperature ∗ coolRate
    updatedSchedule ← bestSchedule
```

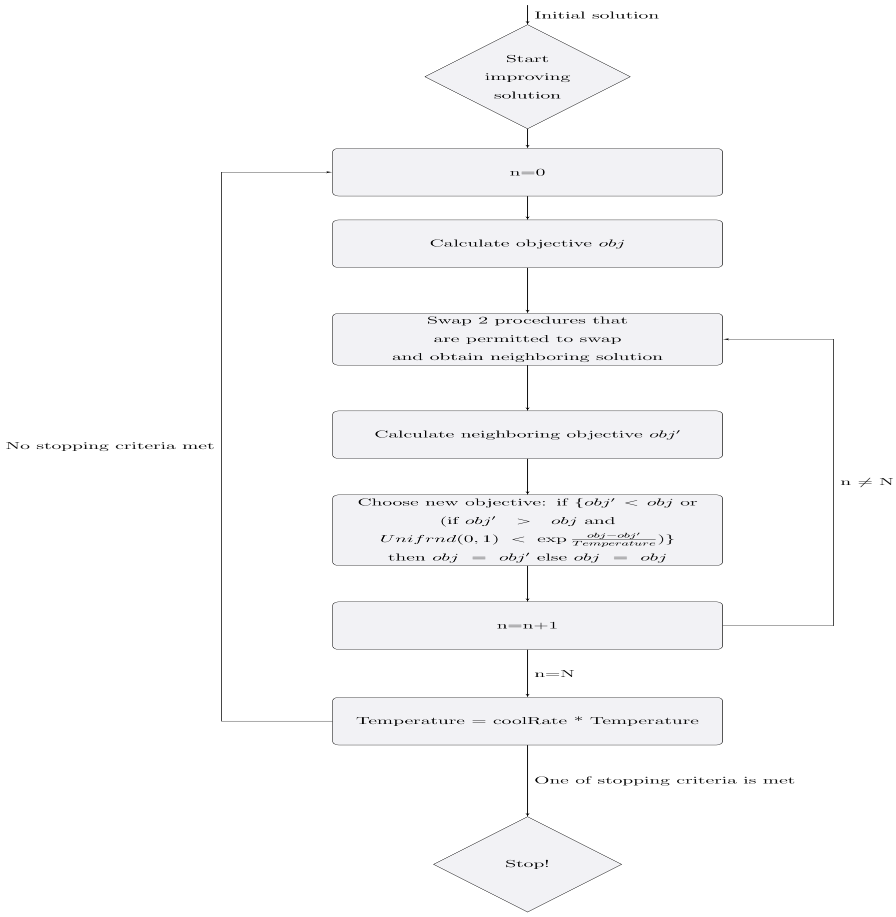


Figure 2: Steps of the improvement heuristic

8.4 Construction OSS and BSS schedules compared

In case of the OSS we construct a schedule of length q (as mentioned in Section 7.2) with the use of the methods above, where the overtime will be minimized. When the capacity that is chosen per OR-day is too limited or the chosen uncertainty factor is chosen too large, it is possible that we can find a solution where we need more than the given OR-days. Probably not all procedures are scheduled directly in this case. The procedures of each specialism should be scheduled together as much as possible, which can be seen as groups or 'blocks'. With the same methods we schedule the procedures of the week with minimal scheduled overtime. This strategy is repeated for schedule in the scheduling horizon Q. The 'blocks' can change drastically over the schedules. If the whole scheduling horizon is constructed, we could test the schedules with the simulation described below.

The BSS uses one of the 'blocks' that are constructed with the OSS schedules. The schedule with the most scheduled duration is used in the BSS, because this way we use blocks that should be relatively large. After we know these block times, we also use the methods above to schedule the procedures. However, the block times are not adjusted over the whole scheduling horizon Q. The block times stay the same, but can also be tested with the same simulation of the procedures.

8.5 Simulation

After each schedule is constructed we simulate the procedures with the parameters that were not used during the construction of the schedule. The procedures that were scheduled with the methods above are replaced with the duration from the simulation. All procedures from an OR-day are summed, from which we can calculate the overtime or unused capacity. This is done multiple times to come up with reasonable outcomes. We calculate these measures for each schedule in the scheduling horizon Q. Those measures can be compared for both strategies and all other input that can be given before the scheduling starts, which hopefully leads to valuable insights. Next we discuss the data that are used to come up with those results.

9 Data description

The data are provided by the Erasmus MC over a period from 1 January 2005 until 31 December 2015. The first 10 years are used to train the model, whereafter the last year is used to construct the measures of this research and test the model. Q is one year and we construct schedules of 5 working days, which is also done in the current situation. The year is chosen such that all seasonal effects are incorporated. The BSS will suffer from this, but can be improved at a later stage.

The information we get are all surgical procedures that took place in an OR during the given timeframe. For each surgical procedure we know the date, starting time, end time, the responsible specialism, the producing specialism, the location where the procedure took place and we know the operation codes for each procedure. All surgical procedures are classified in either an elective case or an emergency, this information is also known. It happens that some of the information is missing, which is not desired. How we deal with missing information, will be explained in the remainder of this section. The inclusion of the emergency cases is beyond the scope of this research. Therefore we only use the data of elective cases in the scheduling strategies.

In Table 1 we show some basic information from the elective cases. The training set is used to construct the distribution of the types of procedures. The number of procedures of 2014 are also used to construct block times for the BSS. As the name suggests, the information of the test set is used to test the data. The scheduled procedures are simulated with the parameters, calculated from the test set. For the scheduling we assumed that the number of procedures was known beforehand. In the current situation, the scheduling is also done a few days before the production of the next week starts.

	training set (2005-2014)	test set(2015)
Total number of elective procedures	167,978	16,061
Total duration of elective procedures (minutes)	$2.0 \cdot 10^7$	$2.1 \cdot 10^6$
Average duration (minutes)	118	133
Standard deviation (minutes)	105	112
Maximum duration (minutes)	2830	1136
Minimum duration(minutes)	1	1
Number of procedures with multiple registrations	28	4
Number of procedures without time registration	854	17
Number of procedures with multiple entries	7622	995
Number of registered specialisms	22	19
Number of registered operation codes	2546	1491

Table 1: Information about the used data

Here we mention some problems we encountered. The first problem deals with the procedures with multiple operation codes or multiple specialisms. This can happen when there is a complex case, where multiple surgeons are needed. Because of the data that are not complete enough we chose to spread the duration of the procedure evenly over the different possibilities. Besides, we have the specialisms in the Erasmus MC that do not work uniformly, which results in a variety of data. The shortcomings are discussed briefly. The specialism Thoracic surgery does not work with operation codes. Therefore, the distinction between different types of procedures is not available. Other specialisms register a procedure multiple times when there occurs a complication after a procedure, which can be caused by the different information systems they have to use. There are also procedures that are registered without time information.

In Figure 3 we can see what the relation is between the data that are used. First the procedures are divided in either elective or emergency. Thereafter the data are clustered per responsible specialism. Each specialism has operation codes that occur. There is also a node without operation code, which is used if the operation code is unknown for a surgical procedure, as in the case of Thoracic surgery. The last deepening are the week numbers. Some procedures will occur more often in a certain time of the year, which can be seen in the deepest nodes.

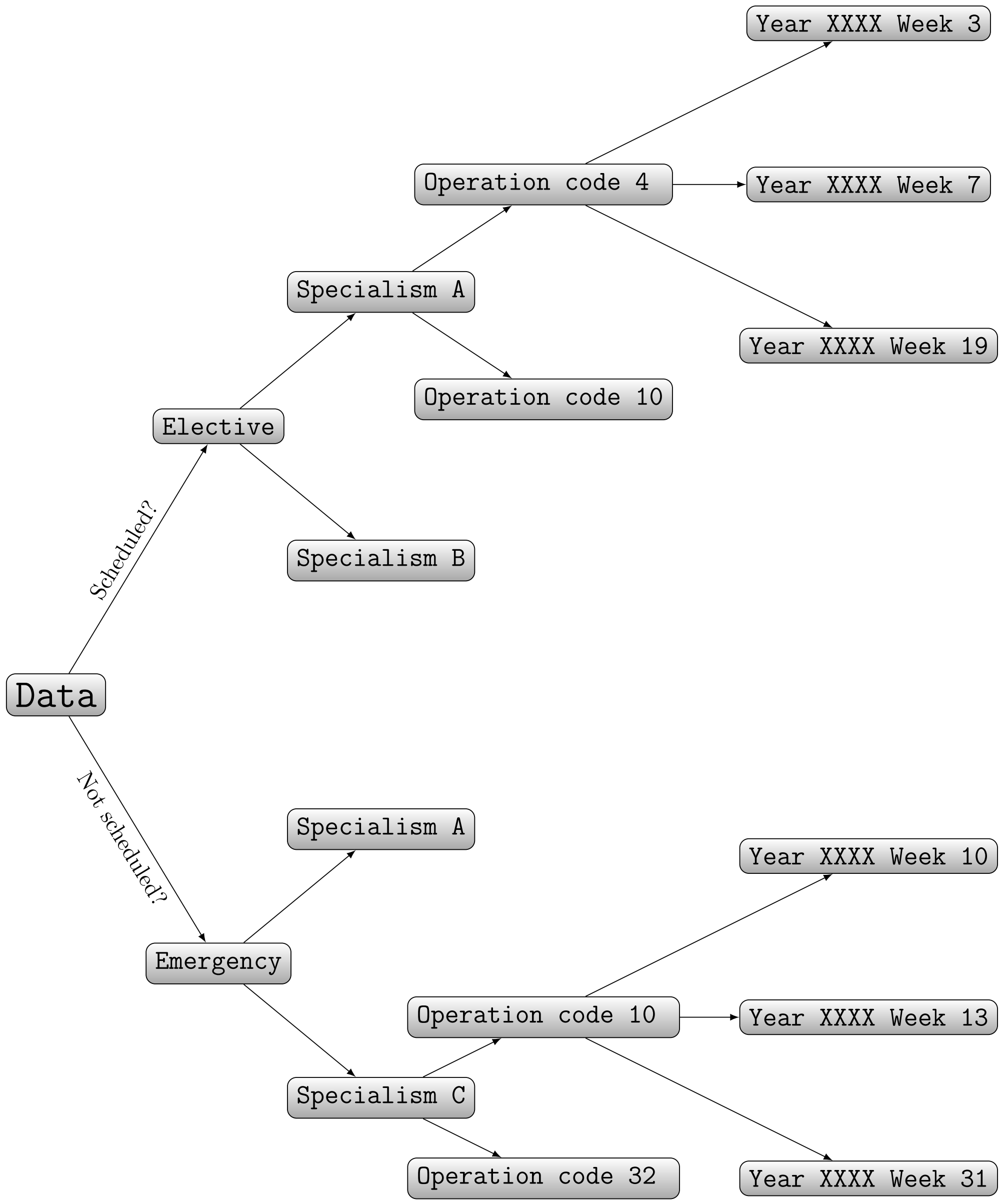


Figure 3: Example of how the data are ordered

In each 'operation code' node we have a set of multiple procedures that belong to the same type. Those durations are fitted to a lognormal distribution, which is used when scheduling such type of procedure. From all surgical procedures there are cases that do not occur frequently. Normally, the procedures where the patients are children take place at a different location (Sophia). It happened that some of those procedures occurred at one of the involved locations. There are also types of procedures for which the most recent operation occurred more than two years ago. We do not take those types of procedures and specialisms into account during this research. To construct an acceptable distribution, we chose a minimum of 10 procedures per type before we used the type. When there were less procedures that took place, we used the distribution of the whole specialism to schedule this procedure. It occurred that some operation codes from the test set were unknown in the training set. We use the parameters of the specialism in the training set for this type of procedures.

Especially the data from the past years are of higher importance because of strategy and procedural changes during the years. It is known that not every specialism can operate in any OR, therefore we made a list of possible ORs per specialism. In Appendix A we put a list of the ORs where each specialism can operate. For example Orthopedic surgery has only one OR per day where it can do all its procedures (so five OR-days in a week). This is the same with Ophthalmology. As mentioned in Section 8, those specialisms will therefore be scheduled first so that these ORs are not already filled with other 'less-restricted' specialisms.

The information of the emergency cases is given in Appendix B. The duration of the emergency cases ($3.6 \cdot 10^5$ minutes) if we divide them over 365 days per year with 2 emergency ORs, we use them around 8 hours per day for each OR. Those emergency ORs are always available to operate. However, some emergency procedure do also take place in elective ORs, when there is a room available. In the new situation the idea is that there will no longer be preassigned ORs for emergencies. Operations that are scheduled, can be postponed or canceled when there occurs an emergency. To find the capacity of the OR department, we can not disregard the other surgical procedures that take place. Therefore, we also try schedules where there are OR-days reserved for emergency procedures. It is investigated by the Erasmus MC itself, that reserving 2 ORs per day for emergency cases should be enough to interrupt the elective schedules not too severely.

In Figure 4 we can see the plan of the new building where the ORs are built. The room numbers 1-21 and 26 are ORs that are used to schedule procedures. In total that are 22 ORs, which results in 110 OR-days if all are used for elective cases. For every specialism we know in which ORs they can operate, this is shown in Appendix A.

In addition to the assumptions that are already made in the model, we assume that the amount of ICU beds and staff available, are not a bottleneck in this case. Besides, the amount of planned slack is chosen to be such that less than half of the groups of procedures could be in overtime. In the current situation there is also a certain margin, that we want to try. To answer the research questions we try different working hours (and capacities). In the current situation, they use a schedule from 5 working days with in total 28 ORs where each OR-day is available for 8 hours. In one year there is $28 \cdot 7.5 \cdot 5 \cdot 52 = 54,600$ hours of OR capacity available.

The current working hours are not always used for elective cases. The total duration of emergency cases versus elective cases in the training set is 16.5%. A part of these emergency cases are handled within working hours, which influences the available time. Because the current situation and the new situation will not only differ in the method of scheduling, but also in the number of ORs, the comparison with this old schedule is only used as an indication of the measures.

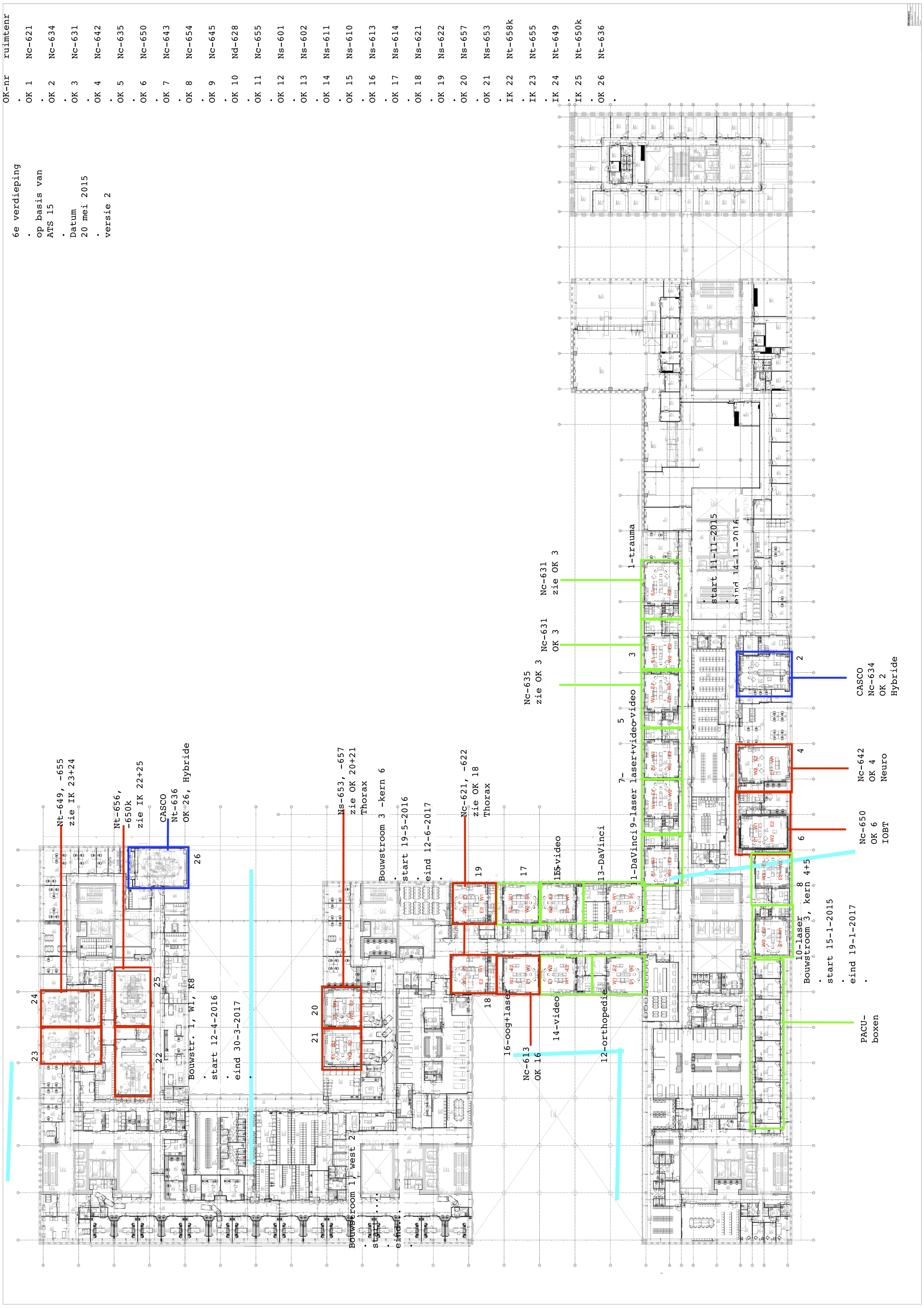


Figure 4: The plan of the 6th floor where the new operation rooms are built.

10 Results

In Section 8 we described the methods we used during this research. The first-fit heuristic is used on the training instance to get an initial schedule. This solution is improved with simulated annealing. The schedules that we found are tested with the test instance, from where we compare the results with the realized measures from the current situation in the Erasmus MC. The data have been made available by the Erasmus MC as described in Section 9.

First the lognormal distributions are tested for the different types of procedures. The general schedules are discussed, whereafter we show the results for both strategies. The schedules are compared with each other and with the current situation. A schedule is denoted with the input factors, for example: 110-8-0.75 is a schedule with 110 OR-days, 8 working hours per OR-day and probability $p = 0.75$.

Matlab 2014b is used to fit the data to a distribution and to get the estimated parameters for the procedure durations. Also the Anderson-Darling goodness of fit test is constructed in Matlab. We used a computer with a 2.6 GHz Intel Core I5 processor with 8 GB 1600 MHz DDR3 RAM.

10.1 Anderson-Darling Goodness of fit test

We test whether the data of the different types of procedures fit a lognormal distribution. For all types of procedures that occur in the training set, this test is repeated (Table 2). If a type has less than four data points, it can not be tested. We only use the parameters of the type if we have more than nine datapoints (as described in Section 9). The null hypothesis, that the data fit the lognormal distribution, is 1422 times not rejected with a significance level of 5% and the Monte Carlo standard error smaller than 0.01. In case of 349 types of procedures the null hypotheses are rejected.

The procedures with less than 10 data points get the estimated parameters with the use of the data of the whole specialism. These are also tested for a lognormal distribution. In this case 17 of the 22 times the lognormal distribution is rejected. Only the types with too little data points will use these parameters.

Specialism	Different types	Not tested	Accepted H_0	Rejected H_0
Anesthesiology	97	29	54	14
Surgery	469	220	196	53
Dermatology	41	22	15	4
Gynecology	148	63	56	29
Hematology	1	0	1	0
Internal Oncology	1	1	0	0
Oral surgery	258	186	67	5
Ear-Nose-Throat (ENT)	414	163	203	48
Children Lung Diseases	1	1	0	0
Lung diseases	13	4	8	1
Gastro logy	10	8	2	0
Neurosurgery	273	133	117	23
Neurology	1	0	0	1
Accident medicine	271	133	117	21
Ophthalmology	74	17	39	18
Orthopedic surgery	350	103	192	55
Pain relief	87	32	45	10
Plastic surgery	440	171	243	26
Radiotherapy	22	9	5	8
X-Ray	5	4	0	1
Thoracic surgery	2	1	0	1
Urology	152	59	62	31

Table 2: Number of operation codes in each specialism, that could not be tested, accepted or rejected the H_0 of a lognormal distribution with an AD-test with 5% significance level

To compare the outcomes of this test, we also fitted the data to a normal distribution with unknown parameters. The null hypothesis is not rejected for 1080 types of procedures. 691 times the null hypothesis is rejected, which is worse compared to the lognormal distribution. This is the same for the data of the specialisms, where 19 of the 22 times the null hypothesis is rejected.

10.2 Differences and similarities in strategies

We compare the outcomes of multiple schedules, which differ in strategy of scheduling, proposed working hours and the allowed OR-days for the specialisms. The BSS has more restrictions than the OSS, which is likely to result in less performing schedules. The schedules with OSS do have block times that can be changed every week, where the BSS schedules have fixed block times for a longer period. The loss in flexibility in BSS is in favor of the staff, that can work with fixed working hours each week.

Before we start to construct an initial schedule we need a procedure list, where all procedures of the week are stated. The data are connected as can be seen in Figure 3. When constructing the procedure list, we check if there is enough data to get an appropriate estimate for this specific operation code. If there are less than 4 such procedures in the data, we use the parameters of the specialism itself.

The methods, as discussed in Section 8, are used in the OSS and BSS strategy. For the first strategy the objective is to minimize the total planned slack whereafter the procedures are scheduled within the constructed time windows with the least amount of overtime. In the second strategy we assume a block schedule where the scheduled overtime will be minimized. To construct all schedules we use the parameters of the training set. All schedules are tested with a simulation of the procedures, where the durations are drawn a 1000 times. The parameters in this simulation are based on the data in the test set. The schedules are compared using the measures overtime and unused capacity. The percentage overtime is the duration that procedures are handled outside working hours compared to the total duration of procedures in a time period. The percentage unused capacity is the time within the working hours that is not used, compared to the total time within working hours in a certain time period.

The schedules are constructed using different inputs. In total there are 22 ORs, which results in 110 OR-days in a schedule. It is however possible that some OR-days will be reserved for emergency cases. This difference can be examined. Also the difference in working hours leads to whole different schedules, which variates between 7 and 11 hours per OR-day. The last distinction is made on the probability that one (group of) procedure(s) is handled *inside* the scheduled hours. This probability p is chosen to be 0.6, 0.75 or 0.9. Not all OR-days need to have the same working hours, because there are also procedures that take more than a working day to fulfill and there is staff that works in different shift. During the day there is more staff available, but in the evening there are also teams that can complete some procedures.

The SA heuristic has parameters that should be given. The number of non-improving iterations, we used to come up with the results below, equals 100. The cool rate r equals 0.95 and the stopping temperature θ equals 10.

10.3 Open Scheduling Strategy (OSS)

To construct a schedule with the OSS, the procedures from the procedure list were inserted one by one, using the strategy explained in the methods. The outcomes we show are based on 1000 simulations of all procedures in a week. The overtime and unused capacity are averages of those simulations.

In Table 3 we see that the overtime and unused capacity increase, while p is also increased, in the case of 7 working hours per OR-day. This can be explained by the procedures with planned slack that do not fit in the regular working hours. When the capacity is too limited and the planned slack is (extremely) high, there are more procedures which are already scheduled in overtime. In the case of 11 hours per working day we observe the opposite: the overtime and unused capacity decrease when p is increased. The schedules 110-9-0.9 and 110-7-0.9 are almost the same with respect to unused capacity, but the overtime is drastically lower in the first case. The choice of the schedule is an exchange between more working hours and probably less overtime.

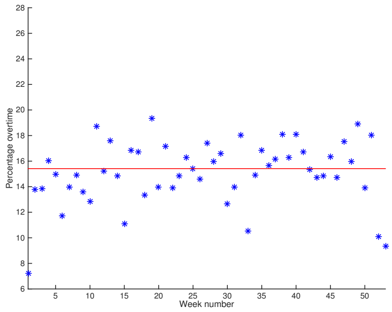
To see the dispersion of the overtime and unused capacity, we constructed scatter plots as can be seen in Figure 5 and 6. The red line is the average of the corresponding schedule. The plots of both percentages do not always move contrary. If the overtime decreases

# OR-days	Working hours per OR-day	p	% Overtime	% Unused capacity
110	7	0.6	15.4	25.6
110	7	0.75	15.8	27.9
110	7	0.9	22.7	38.5
110	8	0.6	11.0	31.5
110	8	0.75	9.8	31.6
110	8	0.9	12.0	35.8
110	9	0.6	9.6	38.2
110	9	0.75	7.3	37.3
110	9	0.9	7.9	39.0
110	11	0.6	7.6	48.3
110	11	0.75	5.3	47.4
110	11	0.9	3.2	46.5

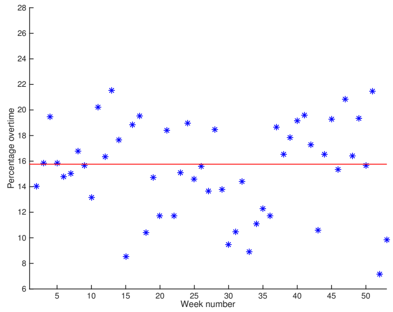
Table 3: Differences and outcomes in schedules OSS

from one week to another this does not say directly that the unused capacity should decrease, because of the total duration that can be decreased also in the following week. The scatterplots show a variety in overtime. It is possible that some types of procedures differ in duration in the trainingsset compared to test set. If those procedures occur multiple times a week, the schedule may be underestimating those durations. Another problem can be the limited OR-day capacity: procedures with a large planned slack have to be scheduled in overtime when the capacity is too limited. It can be seen that from the schedules with p is 0.6 the schedule with 11 hours per OR-day has (for most weeks) the smallest dispersion, while the one with only 7 hours per OR-day has larger dispersion, which is directly related with the limited capacity. It can also be seen that the deviation to a low percentage of overtime happens more often with more working hours together with more planned slack.

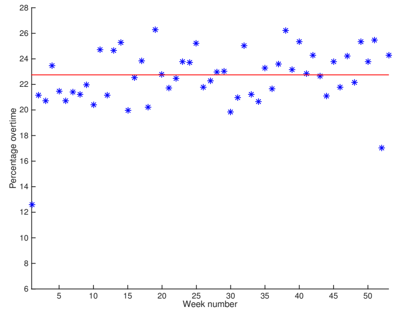
The most obvious part of Figure 6 are the outliers that can be seen in almost every schedule. The last weeks of the year have less procedures even as the weeks during the summer holiday. This results in more unused capacity. To exclude this observation we



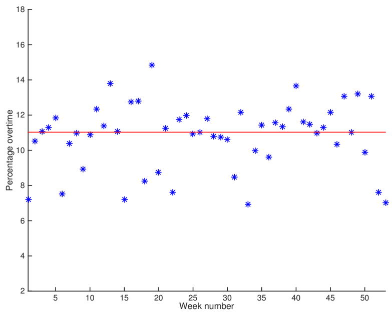
(a) 7 hour with 0.6 probability



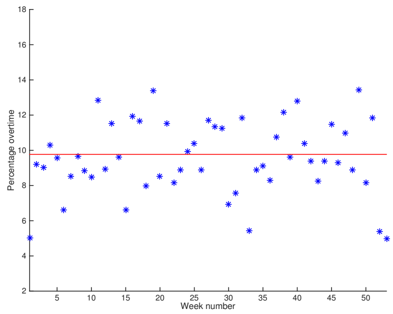
(b) 7 hour with 0.75 probability



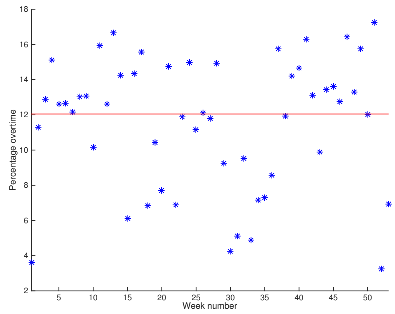
(c) 7 hour with 0.9 probability



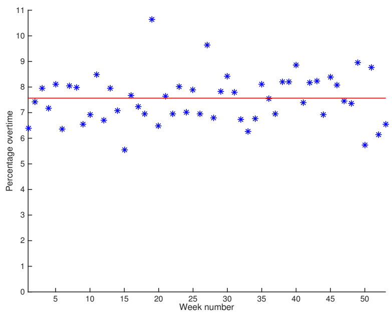
(d) 8 hour with 0.6 probability



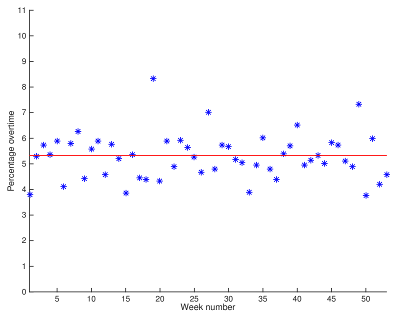
(e) 8 hour with 0.75 probability



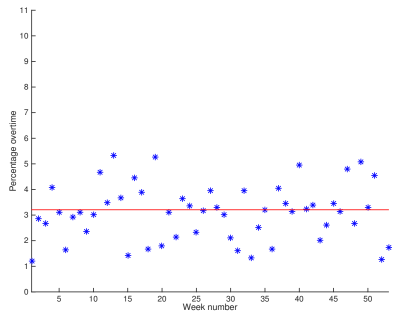
(f) 8 hour with 0.9 probability



(g) 11 hour with 0.6 probability



(h) 11 hour with 0.75 probability

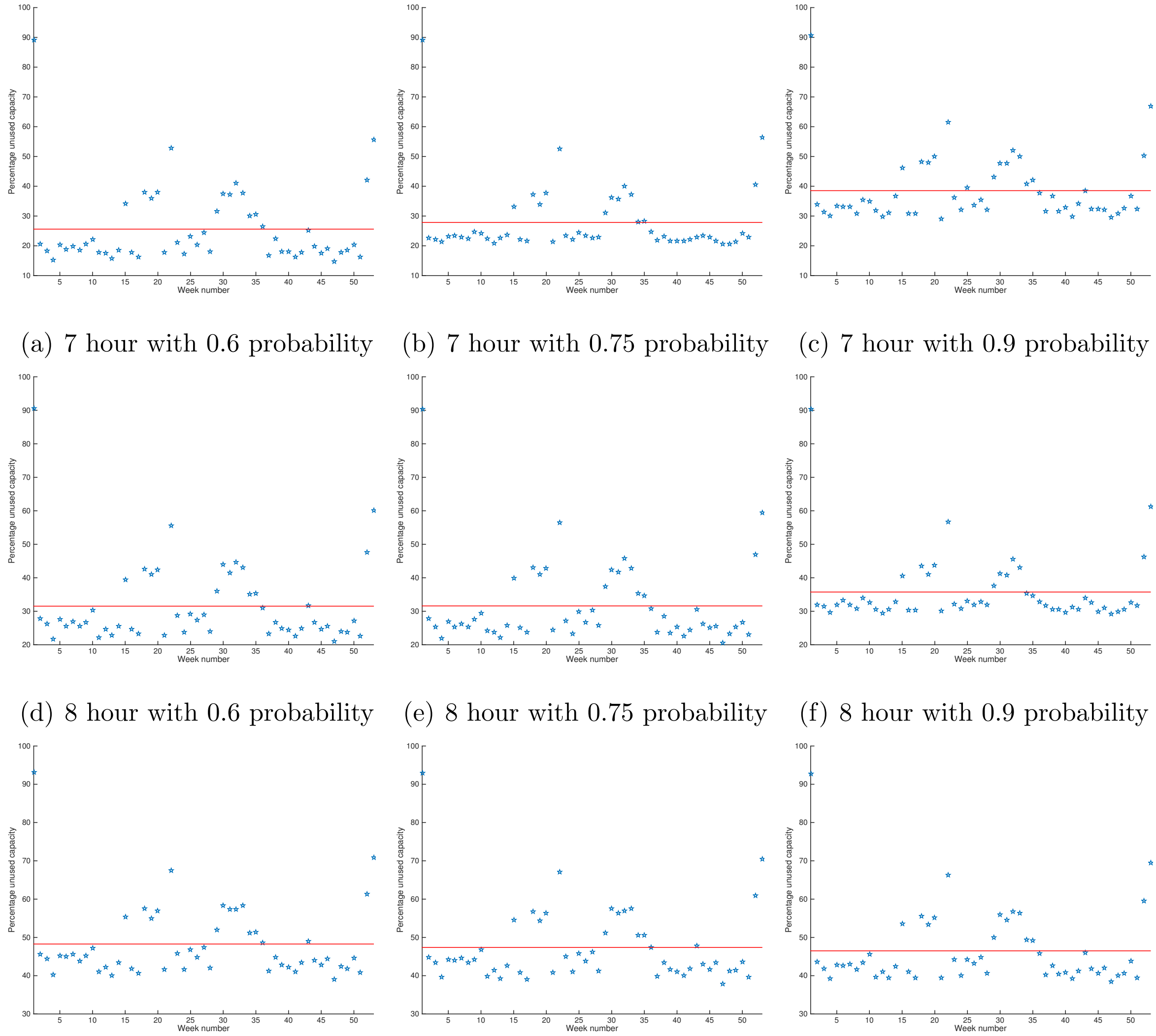


(i) 11 hour with 0.9 probability

Figure 5: Percentage overtime OSS schedules

could have compared with the scheduled procedures instead of comparing them with the proposed working hours, but we want to see the overall capacity of the ORs. This will be discussed later on.

Another important observation is the increase in unused capacity for both an increase in working hours as the increase in planned slack. Only in the case of 11 working hours per OR-day, we have a decrease in unused capacity, when increasing the planned slack. As mentioned earlier, this relates to the limited capacity. In the case of 11 hours, the procedures are mostly scheduled within the time windows (even most of the procedures with a duration longer than a regular working day). This results in less scheduled overtime together with less chance of procedures that occur outside the scheduled duration, which leads to a proportionally better used OR capacity.

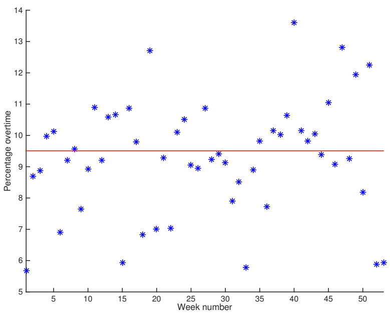


(a) 7 hour with 0.6 probability (b) 7 hour with 0.75 probability (c) 7 hour with 0.9 probability

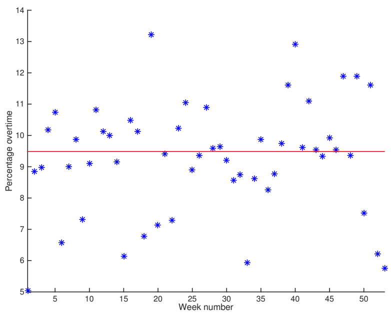
(d) 8 hour with 0.6 probability (e) 8 hour with 0.75 probability (f) 8 hour with 0.9 probability

(g) 11 hour with 0.6 probability (h) 11 hour with 0.75 probability (i) 11 hour with 0.9 probability

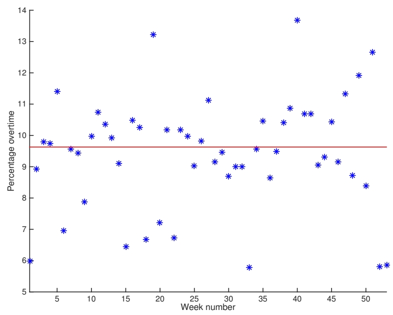
Figure 6: Percentage unused capacity OSS schedules



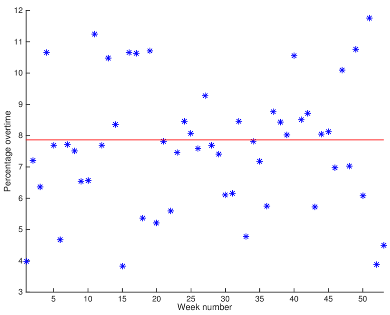
(a) 100 OR-days $p = 0.6$



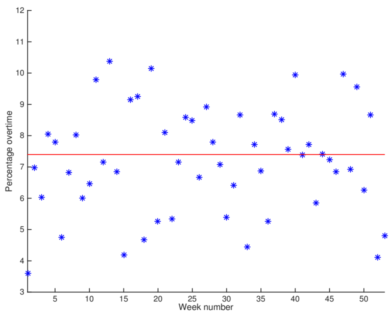
(b) 105 OR-days $p = 0.6$



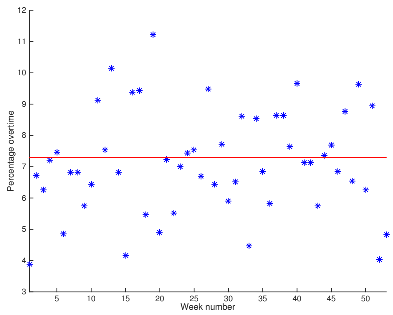
(c) 110 OR-days $p = 0.6$



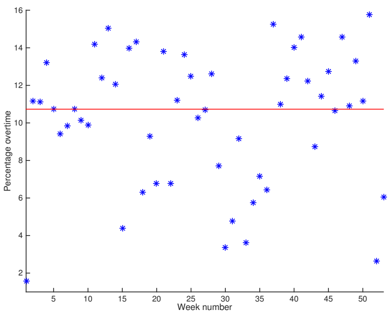
(d) 100 OR-days $p = 0.75$



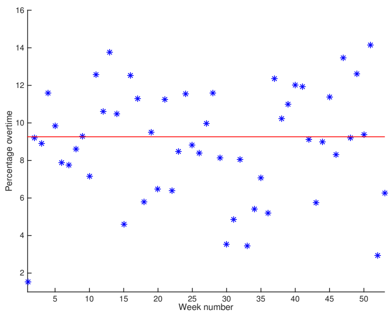
(e) 105 OR-days $p = 0.75$



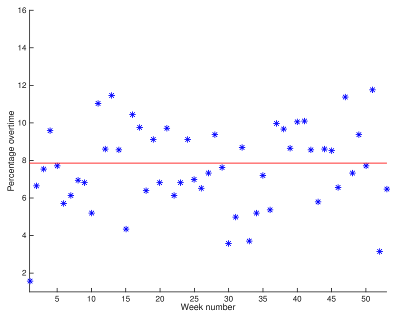
(f) 110 OR-days $p = 0.75$



(g) 100 OR-days $p = 0.9$



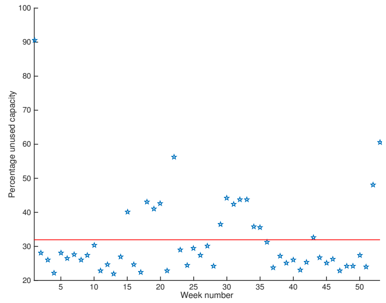
(h) 105 OR-days $p = 0.9$



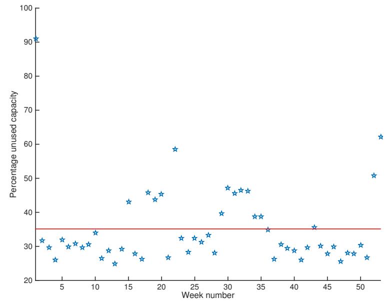
(i) 110 OR-days $p = 0.9$

Figure 7: Overtime of schedules with different # of OR-days, p, all 9 working hours

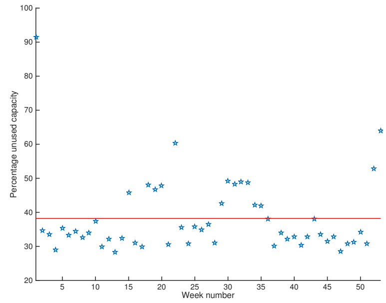
The difference in OR-days can be seen in Figure 7 and 8. When there is more planned slack, the difference becomes clear. With more planned slack there is more scheduled overtime if the capacity is too limited, which is the case in the schedules of 100 OR-days. When the planned slack is equally sized, the capacity restriction is apparently of less influence. Because of the objective of the least planned slack, the procedures are grouped as much as possible. Therefore the schedules of 110 OR-days will not be fully exploited. This can also be seen if we look at the unused capacity, which is much larger for the schedule 110-9-0.6 compared to 100-9-0.6.



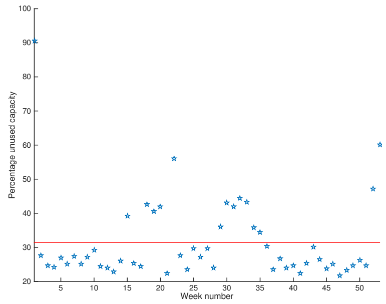
(a) 100 OR-days $p = 0.6$



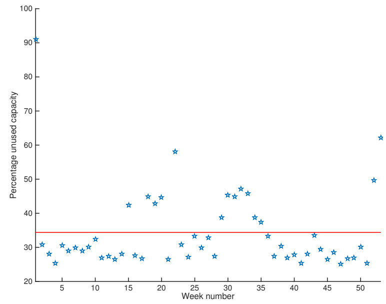
(b) 105 OR-days $p = 0.6$



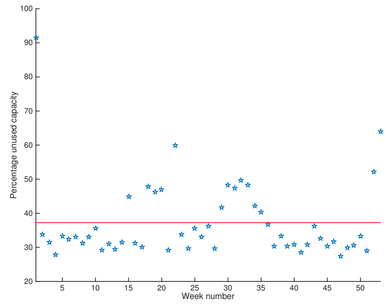
(c) 110 OR-days $p = 0.6$



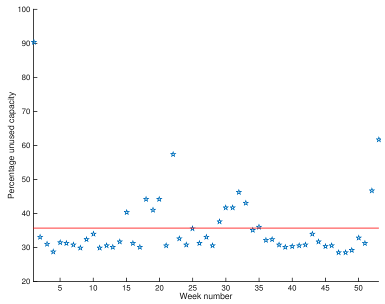
(d) 100 OR-days $p = 0.75$



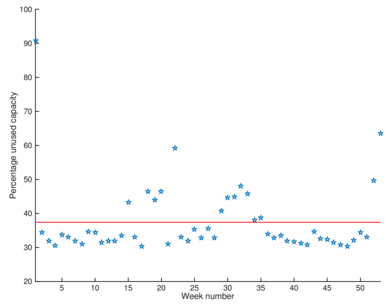
(e) 105 OR-days $p = 0.75$



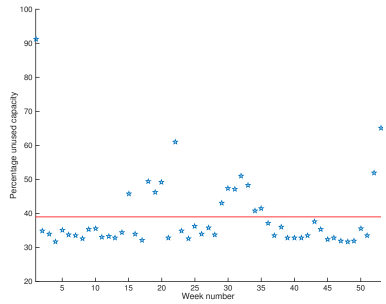
(f) 110 OR-days $p = 0.75$



(g) 100 OR-days $p = 0.9$



(h) 105 OR-days $p = 0.9$



(i) 110 OR-days $p = 0.9$

Figure 8: Unused capacity of schedules with different # of OR-days, p, all 9 working hours

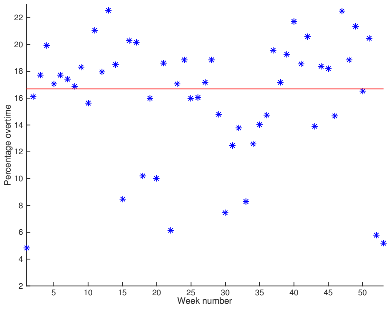
10.4 Block Scheduling Strategy (BSS)

The construction of an appropriate block schedule is an iterative process. In this research we assume a certain block schedule. The blocks are chosen from 2014 schedules where the total planned slack in the chosen week is the maximum of all weeks in the corresponding year. This block schedule has the most planned slack scheduled in 2014 and is constructed with data from the training set. We chose 2014, because the duration of the procedures in 2015 are unknown before those procedures have to be scheduled. Another possibility is to insert a block schedule in the model, whereafter the performance of that schedule can be calculated. The inserted block schedule should also be based on other insights. This possibility is useful for the Erasmus MC so they can find the probable performance of a block schedule. In this way block times can be proposed, whereafter the procedures from the procedure list are scheduled. This is done for each week with the same blocks for the specialisms, whereafter the measures could be calculated. The performance can be used as a motivation for rejecting or accepting the proposed block schedule.

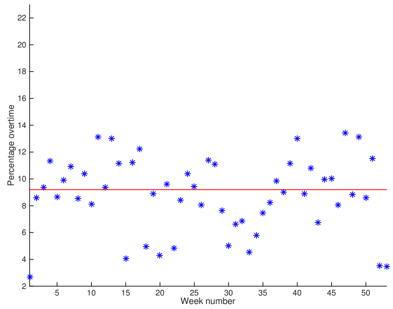
The blocks with the maximum planned slack of 2014 are used. Those results are shown in Table 4. We chose to use $p = 0.75$ because this performed well in the OSS and the execution time can be decreased when we do not use all different p values that are used in the OSS. It can be seen that the overtime decreases and the unused capacity increases when increasing the working hours, which is as can be expected.

# OR-days	Working hours per OR-day	p	% Scheduled overtime	% Unused capacity
100	7	0.75	22.6	28.0
105	7	0.75	19.3	28.1
110	7	0.75	16.6	28.8
100	9	0.75	9.3	32.8
105	9	0.75	9.3	35.9
110	9	0.75	9.2	38.7
100	11	0.75	6.7	43.1
105	11	0.75	6.6	45.7
110	11	0.75	6.5	48.1

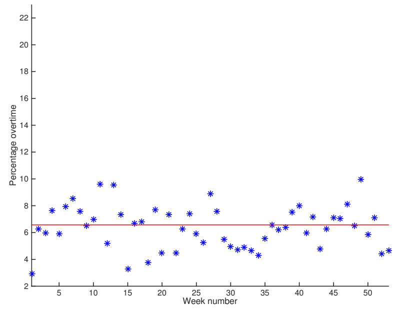
Table 4: Outcomes BSS



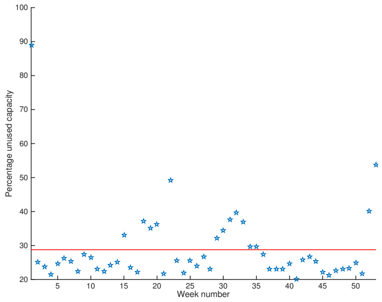
(a) Overtime: 7 hour, $p = 0.75$



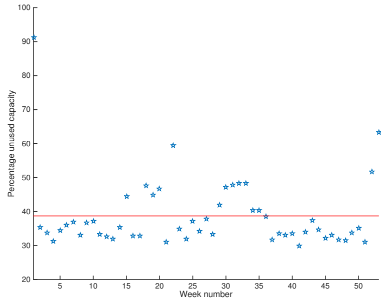
(b) Overtime: 9 hour, $p = 0.75$



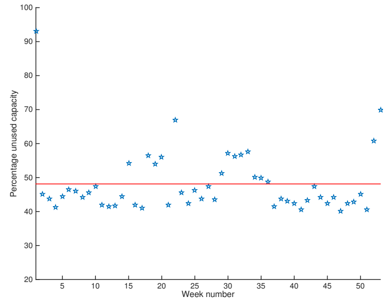
(c) Overtime: 11 hour, $p = 0.75$



(d) Unused capacity: 7 hour, $p = 0.75$



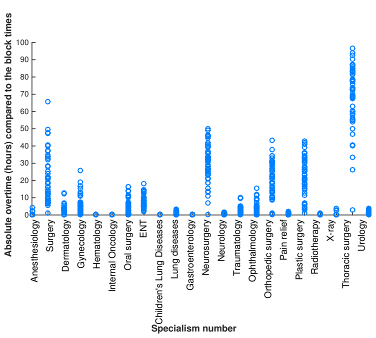
(e) Unused capacity: 9 hour, $p = 0.75$



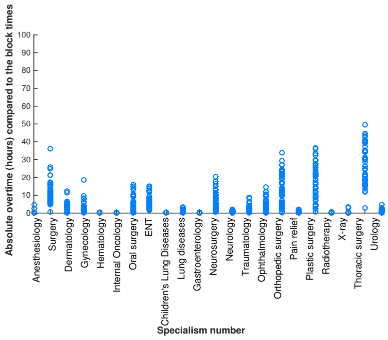
(f) Unused capacity: 11 hour, $p = 0.75$

Figure 9: Percentage overtime (above), unused capacity (under) BSS schedules with 110 OR-days

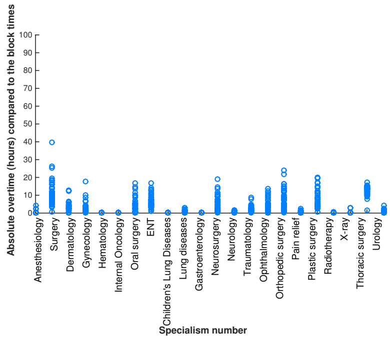
In Figure 10 we can see the absolute overtime per schedule per specialism. All different weeks are plotted per specialism. This way we can see which specialism causes the overtime of the schedule. When another schedule is proposed, we could take this information into account. We can see that the specialism Thoracic surgery has much overtime in schedule 110-9-0.75. An updated block schedule should include more block time for this specialism. The overtime of this specialism is due to the missing operation codes and the (on average) large duration of the procedures. Probably this overtime should be declined when the procedures can be divided in different types.



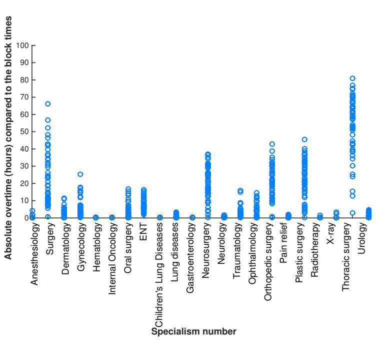
(a) Unused capacity: 100-7-0.75



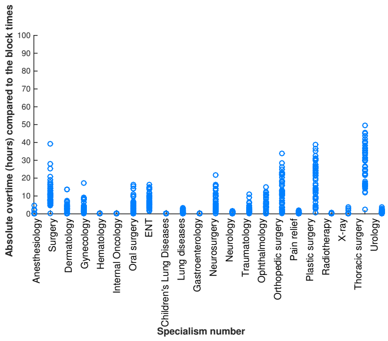
(b) Unused capacity: 100-9-0.75



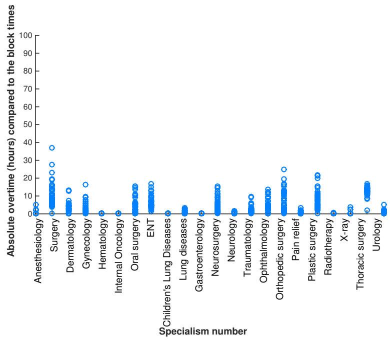
(c) Unused capacity: 100-11-0.75



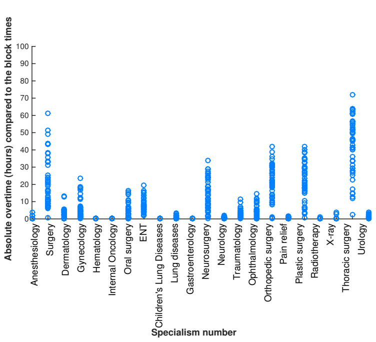
(d) Unused capacity: 105-7-0.75



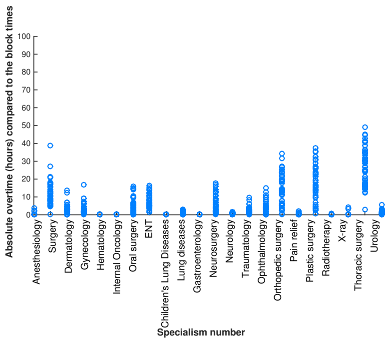
(e) Unused capacity: 105-9-0.75



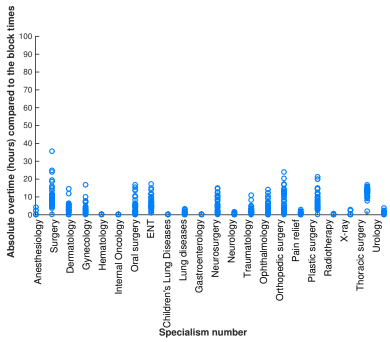
(f) Unused capacity: 105-11-0.75



(g) Unused capacity: 110-7-0.75



(h) Unused capacity: 110-9-0.75



(i) Unused capacity: 110-11-0.75

Figure 10: Absolute overtime of each specialism in the scheduled block times plotted for each week

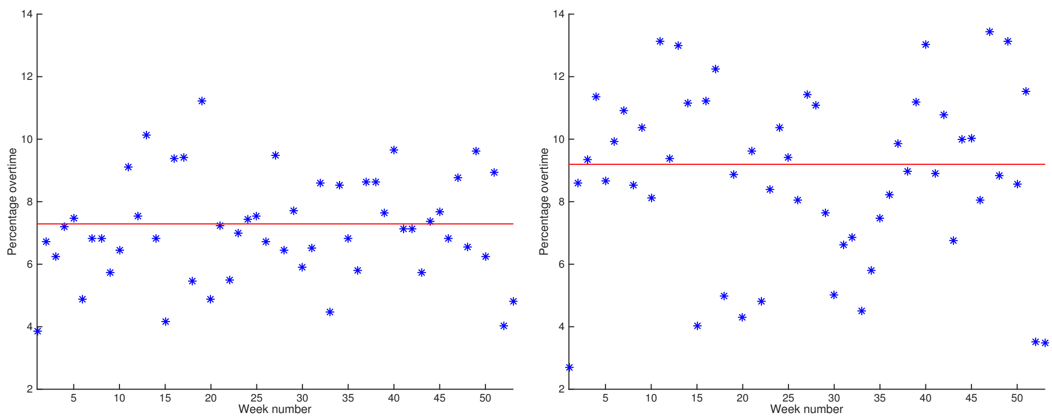
We can use other block times in the block schedules and calculate the performance of those schedules. The proposed block schedules should not only be optimized in terms of overtime and unused capacity of the ORs, but for example also in terms of needed staff and ICU beds. The block times that are now constructed with the use of data from 2014 can be used as a starting point, from where we make little changes. This has to be done together with the Erasmus MC.

10.5 Outcomes OSS and BSS compared

In case of the OSS we construct a schedule that changes every week. Dependent of the number and types of procedures that occur in a week a schedule is computed. As can be seen from the previous results from Table 3 and Table 4, the performance of the OSS is better compared to the BSS. Not only the overtime and unused capacity are smaller compared to BSS, but the dispersion is also less. The standard deviation of the percentage overtime can be seen in Table 5. Graphically, the difference can be seen in Figure 11. The OSS schedule is adapted for each week, what results in less outliers.

Schedule	OSS	BSS
110-7-0.75	3.9	4.6
110-9-0.75	1.7	2.9
110-11-0.75	0.9	1.6

Table 5: Standard deviation of percentage overtime



(a) OSS: 110-9-0.75

(b) BSS: 110-9-0.75

Figure 11: Two schedules compared on percentage overtime

The computation times are compared also. We measured the time of the first-fit heuristic and the SA heuristic for each week that is scheduled. The total computation time in seconds for scheduling the whole year is given in Table 6. It can be seen that the number of non-improving iterations causes much more computation time, but also the cool rate r is important for the computation time. When the cool rate is close to 1, there are much more steps needed, before the temperature is 'cooled down' below the stopping temperature. With more non-improving iterations and a higher cooling rate, more steps are taken to come to the final outcomes, which results in a higher computation time, but probably also a better objective.

Schedule	Cool rate r	# non-improving iterations	OSS		BSS	
			First-Fit	SA	First-Fit	SA
110-9-0.75	0.8	10	28	71	27	45
110-9-0.75	0.8	100	28	749	27	468
110-9-0.75	0.8	1000	28	7634	27	4717
110-9-0.75	0.9	10	28	148	28	92
110-9-0.75	0.9	100	28	1528	27	947
110-9-0.75	0.95	10	28	303	27	190
110-9-0.75	0.95	100	28	3048	26	1884

Table 6: Computation times in seconds of the different schedules for the initial and improvement heuristics

The absolute objectives can be seen in Table 7. These objectives are the average absolute scheduled overtimes in minutes for one week. It can be seen that the SA heuristic improves much more with more non-improving iterations and a cool rate closer to one. The objectives in the first-fit heuristic of the BSS differ because of the blocks that are dependent of the outcomes of the OSS. In case of 1000 non-improving iterations with a cool rate of 0.8 it can be seen that the objective is increased slightly in case of BSS. This can be due to the dependency of the OSS schedule. Apparently, the blocks that are used in the case of 1000 iterations are worse compared to the blocks of 100 iterations. This is a clear example of the interdependency of the block times and the final schedule.

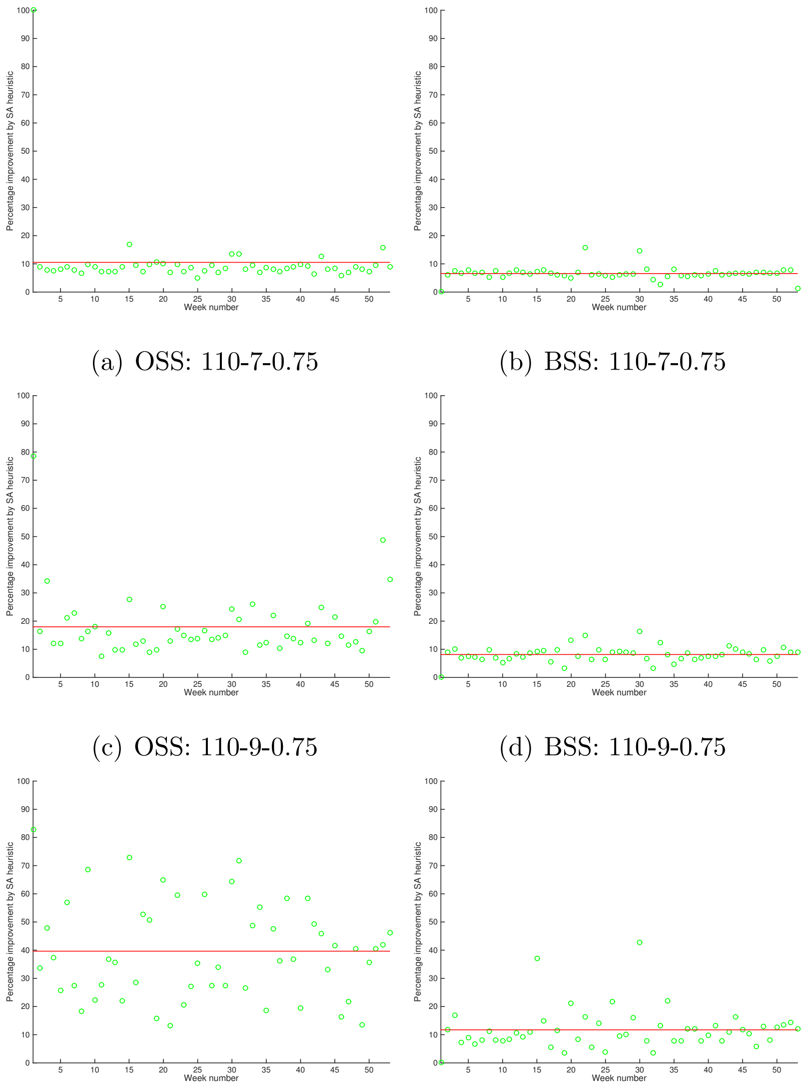
Schedule	Cool rate r	# non-improving iterations	OSS		BSS	
			First-Fit	SA	First-Fit	SA
110-9-0.75	0.8	10	3249	3181	4807	4725
110-9-0.75	0.8	100	3249	2793	4810	4483
110-9-0.75	0.8	1000	3249	2647	4993	4581
110-9-0.75	0.9	10	3249	3072	4857	4708
110-9-0.75	0.9	100	3249	2722	4827	4470
110-9-0.75	0.95	10	3249	2948	4828	4581
110-9-0.75	0.95	100	3249	2686	4751	4383

Table 7: Absolute objectives in minutes of the different schedules after the initial and improvement heuristic (average over the different weeks)

10.6 Performance SA heuristic

The performance of the SA heuristic is measured in improvement of objective compared to the objective after the first-fit heuristic. The initial objective is calculated after the first-fit heuristic and the best objective is calculated after the SA procedure took place for scheduling the procedures in the given time blocks. In Figure 12 the improvement of the SA heuristic is shown with respect to the initial solution. The SA heuristic leads to more improvement in the OSS schedules. This is probably because the OSS schedules do not have the preassigned blocks and therefore more possibilities to improve compared the the BSS schedules.

We chose to use a cooling rate of 0.95 with 100 non-improving iterations. The parameter θ from Section 8.3.2 is chosen to be 10. The problem consists of 100 up to 110 OR-days where mostly 22 specialisms have to operate. If for example each specialism has 10 procedures in a week, there are 220 procedures that have to be scheduled. Not all procedures could be handled in any OR-day, but if they can, there are at most 100^{220} possibilities. Just a fraction of the possibilities is examined with the SA heuristic. Possibly when adapting the SA heuristic with different parameters we can improve the schedules still a bit more.



(a) OSS: 110-7-0.75 (b) BSS: 110-7-0.75

(c) OSS: 110-9-0.75 (d) BSS: 110-9-0.75

(e) OSS: 110-11-0.75 (f) BSS: 110-11-0.75

Figure 12: Improvement of SA heuristic in objective

10.7 Measures from current situation 2015

In 45.0% of the working hours in 2015, the ORs are not used for elective cases. They were empty or an emergency procedure took place. There are reserved ORs for emergency cases, so we assume that those where handled in different ORs. Besides 14.9% of the total elective duration is handled in overtime. It is desirable that the overtime percentage and the unused capacity are both as small as possible under the restriction that there is room for the emergency cases. This is also discussed in the next section. The capacity of the ORs in in the current situation $28 \cdot 7.5 \cdot 5 = 1050$ hours. When using all ORs in the new situation, this is equal to 22 ORs with working hours of 9.5 hours.

11 Discussion

In this section we will discuss some restrictions and possible extensions of this research. There are assumptions made to deal with this comprehensive problem. These assumptions should be justified, which is also done in the remainder of this section. After the assumptions are discussed, we mention some explanations for differences in results.

Finding a schedule that meets all wishes in such a complex problem is unlikely. The main question during this research is the following: "what working hours are needed in the new situation, in which the procedures are handled without too much overtime." Because there are much uncertainties for the new situation, we tried to come up with some possibilities and recommendations. That is why we chose to make use of two strategies from which the outcomes can be compared. The ORs are grouped in the new building of the Erasmus MC. To compare the results, we constructed the measures in the same way as in the current situation. The decisions for a schedule should be specified further, whereafter a choice can be made.

The problem of the Erasmus MC consists of 110 OR-days with 22 specialisms, which results in an extremely large amount of possibilities for a schedule. Besides, the non-identical ORs cause more dependencies in outcomes. It may be relevant for further research to investigate the influence of the number of non-improving iterations in combination with the cooling rate of the SA heuristic extensively. The choice of the two parameters had to do with the computation times together with the number of possible schedules we wanted to show.

The OSS has less restrictions as the BSS, while the assigned blocks are changed over time in case of OSS. This gives a better insight in the possible capacity of the new building of the Erasmus MC, because the BSS is an iterative process that should be optimized later on. Therefore we constructed a model where proposed blocks can be inserted, whereafter the possible performance can be calculated. The choice of a block schedule for a whole year is also chosen because we should propose a starting point, from where some decisions can be taken. This way we tried to gain knowledge, which leads to better decisions when the new building is taken in operation.

An other interesting question is the development in the number of procedures. Because of the unknown policy of the Erasmus MC, with respect to the surgeries that should be fulfilled, we did not take the growth (or decline) of procedures into account. There is a lot of discussion what procedures should be handled in an university hospital. This increases the uncertainty about the types of procedures that will be handled. The growth per specialism and operation codes for the coming year(s), can be inserted in the constructed model. This should be added in the construction of the procedure list.

A problem we observed, is the missing operation codes for the specialism 'Thoracic surgery'. All procedures of this specialism are generalized, because of the missing operation codes. Probably the schedules would be much more accurate when we had these data, especially because this specialism is responsible for a large part of the total duration. It can be seen in the outcomes that 'Thoracic surgery' is responsible for a large part of the overtime. For further research it would be valuable to find another point of distinction for this specialism. It is recommended to register uniformly over all the specialisms. That way the comparison (in all kinds of questions) is more reliable.

As mentioned earlier, it is valuable to investigate if changing some parameters of the SA heuristic results in even better outcomes. Some of the outcomes (like the decrease in overtime, while decreasing the number of OR-days available) do not make sense. This can be caused by the specific OR-days that are removed: if those removed OR-days are used minimally and the procedures can be handled somewhere else, this may result in less overtime. In Section 10 we also discussed that these outcomes can be caused by the changed input and that possibly not enough of the solution space is investigated. When the parameters of the SA heuristic are better adjusted, these unexpected results can possibly be solved. To find best methods or adjust the methods more closely, it is advisable to start with a much smaller problem including a smaller solution space.

The total duration can decrease when a procedure is added and scheduled in an OR where already much procedures of this specialism are handled. This is owing to the log-normal distribution. It can be favored that more procedures are scheduled in one OR even if it is already in overtime. The probability p of no overtime per group of procedures is earlier met in that case. This is not mentioned in the paper Hans et al. (2008), because they used a normal distribution for the durations. The problem that rises, is solved by using a slightly different objective. Instead of minimizing the planned slack in scheduling, we minimized the scheduled overtime.

The proposed lognormal distribution is tested with the Anderson-Darling goodness of fit test. The largest part, of the different types of procedures, did not reject a lognormal distribution. However, there were also types of procedures that rejected the null hypothesis. Finding the best distribution for surgical procedure durations is a different research. On the other hand, it is of high importance for the reliability of the outcomes for this research. Therefore, it can be relevant to investigate if several distributions can be used to model the durations. The sum distribution that is used in this research, will be much more difficult when using several distributions for different types of procedures. However, it is shown that surgical procedures fit a lognormal distribution in more cases compared to the normal distribution, which supports the outcomes of Strum et al. (2000).

The measures we computed, will possibly result in better outcomes in practice. In our simulation there are no canceled procedures or replaced procedures, which only will increase the usage of the ORs (less overtime and less unused capacity). In fact, there were definitely changes made during the day. We chose not to take these adjustments into account, because we favor underestimating the capacity slightly. This way we are sure we do not get capacity problems when overlooked issues pass by.

The measures of the current situation are constructed under a totally different situation. Nevertheless, under the assumption that no emergency procedures are handled in the 'elective ORs' we can use the measures as a comparison. What kind of improvement can be gained in the new situation? This is not a comparison in methodology, but can be used as a motivation for the staff.

12 Conclusion

This research is initialized by the Erasmus Medical Center to get insight in the capacity and restrictions of the upcoming ORs. Two strategies are used to find the possible working hours of the new ORs in which the surgical procedures can be handled without too much overtime. Obviously these working hours should be as tight as possible (when all assumptions are satisfied), because OR capacity is one of the main costs in hospitals. The insights we try to give may lead to possible choices of the Erasmus Medical Center. We favor using slightly less capacity than what is actually available. Finally, this will not result in huge capacity problems.

The main question we wanted to answer during this research is what working hours, for the ORs, are needed in the new situation to fulfill all surgical procedures without too much overtime? The answer depends on the definition of 'too much'. The Erasmus Medical Center can take the possibilities, of different schedule inputs, into account.

We used two strategies of scheduling, namely the open scheduling strategy(OSS) and the block scheduling strategy(BSS). OSS is a scheduling strategy where the specialisms of the surgical procedures can have different working hours over a period, while this is fixed in BSS. The durations of the surgical procedures are based on the training set, which consist of data from 2005-2014. For the OSS the blocks are constructed by scheduling the procedures with the minimal amount of overtime. This is done with the use a first-fit heuristic, whereafter this solution is improved with simulated annealing.

In the BSS the scheduling is slightly different. The blocks are chosen with the use of the OSS schedules. After the construction of the blocks, the procedures are scheduled with the use of a first-fit heuristic. Thereafter this solution is improved with simulated annealing, while minimizing the overtime. Each schedule is compared with the outcomes of a simulation of the procedures. The surgical procedures are simulated with the durations based on the test set, which consist of data from 2015. This results in the measures, overtime and unused capacity, for each created schedule.

The overtime and unused capacity in the OSS schedules are always lower than the proposed BSS schedules. This is just as expected, because the OSS has less restrictions. The BSS strategy should be optimized further to reach results that are more in line with the OSS outcomes. The outcomes within the same strategy for different input parameters change mostly as expected. The percentage overtime rises when the capacity is decreased, together with a decrease of unused capacity and vice versa. It happens that the capacity is decreased and the percentage overtime also declines. This is discussed in the previous section.

The OSS schedules with 9 working hours per OR-day have slightly less total capacity as the current situation, because the number of OR-days also decreases, which results in less available hours. With the used methods we find a schedule with 7.3% overtime and 37.3% unused capacity (schedule 110-9-0.75). Under the assumption that there are no emergency cases scheduled in the 'elective ORs' from the current situation and there were no intentionally scheduled elective cases outside working hours, we increase OR usage in the new situation. This result is gained by a combination of the centralization of the ORs, together with the proposed methodology in this research. The disadvantage of this strategy is the varying working hours over the different weeks, in contrast to the current situation. Therefore we also tried the BSS.

The BSS is an iterative process where block times are proposed, whereafter the performance is measured, new block times are proposed and so on. We constructed block times based on information in the training set, which should be updated in combination with other insight from the Erasmus Medical Center. With the BSS schedule, including all 110 OR-days and 9 working hours per OR-day, we get better measures compared to the current situation. The measures are however worse than the OSS schedule. We find an overtime of 9.2% and an unused capacity of 38.7%. The assumptions from the previous paragraph, with respect to the current situation, are doubtful. However, these outcomes can be improved further when the block times are iteratively improved. This should be done in dialog with the Erasmus MC.

The Erasmus Medical Center should choose what measures are favored. The BSS is still desirable above the OSS, especially because of the goodwill of the staff that needs to be secured. Besides, it is advised to increase the working hours per OR-day. The 7.5 hours, that are used in the current situation, will lead to an undesirable amount of overtime. Around 9 working hours per OR-day should result in acceptable measures. Other knowledge like the policy of surgical procedures that are handled in the Erasmus MC, together with the growth in the number of procedures can be included in the model. That way we can come up with a model that fits the demands of the Erasmus MC even better.

As questioned in this research: the procedures will definitely fit in the new building without more time outside the working hours when keeping the total capacity equally sized. On the other hand, the working hours should be chosen on several insights. The capacity of available staff or ICU beds may be a bottle neck in the new situation. There is always room for new research.

Hopefully the steps, assumptions and decisions that are taken during this research are understandable and inspiring to read. Therefore a short fact that hopefully turns into practice: "A lot of my research time is spent daydreaming - telling an imaginary admiring audience of laymen how to understand some difficult scientific idea" (Leonard Susskind).

References

Anderson, T. W. and Darling, D. A. (1954). A test of goodness of fit. *Journal of the American statistical association*, 49(268):765–769.

Cardoen, B., Demeulemeester, E., and Beliën, J. (2010). Operating room planning and scheduling: A literature review. *European Journal of Operational Research*, 201(3):921–932.

Denton, B. T., Miller, A. J., Balasubramanian, H. J., and Huschka, T. R. (2010). Optimal allocation of surgery blocks to operating rooms under uncertainty. *Operations research*, 58(4-part-1):802–816.

Dexter, F., Macario, A., Traub, R. D., Hopwood, M., and Lubarsky, D. A. (1999). An operating room scheduling strategy to maximize the use of operating room block time:

computer simulation of patient scheduling and survey of patients' preferences for surgical waiting time. *Anesthesia & Analgesia*, 89(1):7–20.

Eglese, R. (1990). Simulated annealing: a tool for operational research. *European journal of operational research*, 46(3):271–281.

Fei, H., Chu, C., and Meskens, N. (2009). Solving a tactical operating room planning problem by a column-generation-based heuristic procedure with four criteria. *Annals of Operations Research*, 166(1):91–108.

Fenton, L. F. (1960). The sum of log-normal probability distributions in scatter transmission systems. *Communications Systems, IRE Transactions on*, 8(1):57–67.

Hans, E., Wullink, G., Van Houdenhoven, M., and Kazemier, G. (2008). Robust surgery loading. *European Journal of Operational Research*, 185(3):1038–1050.

Massey Jr, F. J. (1951). The kolmogorov-smirnov test for goodness of fit. *Journal of the American statistical Association*, 46(253):68–78.

Melouk, S., Damodaran, P., and Chang, P.-Y. (2004). Minimizing makespan for single machine batch processing with non-identical job sizes using simulated annealing. *International Journal of Production Economics*, 87(2):141–147.

Pham, D.-N. and Klinkert, A. (2008). Surgical case scheduling as a generalized job shop scheduling problem. *European Journal of Operational Research*, 185(3):1011–1025.

Stephens, M. A. (1976). Asymptotic results for goodness-of-fit statistics with unknown parameters. *The Annals of Statistics*, pages 357–369.

Strum, D. P., May, J. H., and Vargas, L. G. (2000). Modeling the uncertainty of surgical procedure times: comparison of log-normal and normal models. *Anesthesiology*, 92(4):1160–1167.

Van Oostrum, J. M., Van Houdenhoven, M., Hurink, J. L., Hans, E. W., Wullink, G., and Kazemier, G. (2008). A master surgical scheduling approach for cyclic scheduling in operating room departments. *OR spectrum*, 30(2):355–374.

A Specialism information

Specialism	Permitted ORs	Favored ORs
Anesthesiology	All, but 12	None
Surgery	All, but 12	1 & 2
Dermatology	All, but 12	None
Gynecology	All, but 12	None
Hematology	All, but 12	None
Internal Oncology	All, but 12	None
Oral surgery/ Maxilla facial surgery	All, but 12	None
Ear-Nose-Throat (ENT)	7, 9 & 10	7, 9 & 10
Pediatric surgery	All, but 12	None
Children Lung Diseases	All, but 12	None
Pediatric oncology	All, but 12	None
Lung diseases	All, but 12	None
Gastro logy	All, but 12	None
Neurosurgery	All, but 12	4
Neurology	All, but 12	None
Accident medicine	All, but 12	1
Ophthalmology	16	16
Orthopedic surgery	12	12
Pain relief	All, but 12	None
Plastic surgery	All, but 12	None
Psychiatry	All, but 12	None
Radiotherapy	5	5
Rheumatology	All, but 12	None
X-Ray	6	6
Thoracic surgery	18, 19, 20, 21 & 26	18, 19, 20, 21 & 26
Urology	7, 9, 10, 11 & 13	7, 9, 10, 11 & 13

Table 8: Specialism OR information

In total there are 22 ORs that are named with the numbers 1 until 21, except for the last OR. The last OR is named 26, as can be seen in Table 8 for Thoracic surgery.

B Emergency data

	training set (2005-2014)	test set(2015)
Total number of emergency operations	24,346	2401
Total duration of emergency procedures (minutes)	$3.3 \cdot 10^6$	$3.6 \cdot 10^5$
Average duration (minutes)	136	148
Standard deviation (minutes)	110	116
Maximum duration (minutes)	1676	1145
Minimum duration(minutes)	1	8
Number of registered specialisms	19	16
Number of registered operation codes	684	310

Table 9: Information about emergency cases