iBMG Working Paper W2010.02

More levelled bed occupancy and controlled waiting lists using Master surgical schedules

Lanah Evers

Jeroen M. van Oostrum

Albert P.M. Wagelmans

Institute of Health Policy & Management

ERASMUS UNIVERSITEIT ROTTERDAM

INSTITUTE OF
HEALTH POLICY & MANAGEMENT

More levelled bed occupancy and controlled waiting lists using Master surgical schedules

Date of publication
October 2010

Authors
Lanah Evers
Jeroen M. van Oostrum
Albert P.M. Wagelmans

Contact
Erasmus University Rotterdam
Institute of Health Policy & Management
Tel. +31 10 408 8555
research@bmg.eur.nl
www.bmg.eur.nl

ERASMUS UNIVERSITEIT ROTTERDAM
INSTITUTE OF
HEALTH POLICY & MANAGEMENT

Levelled bed occupancy and controlled waiting lists using Master surgical schedules

Lanah Evers[1,3,4]
Jeroen M. van Oostrum[2]
Albert P.M. Wagelmans[1]

[1] Erasmus School of Economics, Erasmus University Rotterdam, the Netherlands
[2] Institute of Health Policy & Management, Erasmus University Rotterdam, the Netherlands
[3] TNO - Defense, Security and Safety
[4] Netherlands Defence Academy

Summary

Scheduling surgical patients is one of the complex organizational tasks hospitals face daily. Master surgical scheduling is one way to optimize utilization of scarce resources and to create a more predictable outflow from the operating room towards subsequent hospital departments.

The paper addresses two aims. First, we investigate the effect of the length of the planning horizon and other planning parameters in a master surgical scheduling approach on patients´ waiting time, schedule stability and hospital efficiency. Second, the master surgical scheduling approach is compared with a standard operating room planning approach on levelled bed occupancy.

The assignment of patients to a master surgical schedule is carefully described. Using real case data from a regional hospital in the Netherlands a simulation study is performed. The approach is applicable to any other hospital.

Results show that only the planning horizon has substantial influence on outcome parameters waiting time, schedule stability and hospital efficiency. We found that increasing the planning horizon increases patients' waiting time on the one hand, but also increases schedule stability and hospital efficiency on the other hand. Regarding our second aim, we found that using an MSS substantially decreases variability in bed occupancy levels.

Keywords: master surgical scheduling, hospital planning and scheduling, health care efficiency, operating rooms

1. Introduction

Scheduling surgical patients is one of the complex organizational tasks hospitals face daily. Surgical departments, nurses and the hospital administration all have their own goals and ideas on what is ‘optimal’. Physicians, for example, aim to maximize production and profits; while hospital administrations try to create stable and efficient schedules and patient flows through the hospital. From a quality of care perspective minimum waiting times are among the primary objectives. Traditional surgical scheduling approaches often solely focus on the operating room planning without accounting for other hospital resources. Hence, variability in demands throughout the hospital is created.

Master surgical scheduling (MSS) is one way to optimize utilization of scarce resources and to create a more predictable outflow from the operating room towards subsequent hospital departments (van Oostrum, Bredenhoff et al. 2010). In an MSS approach, patients are assigned to a recurrently executed surgical schedule, containing time slots for types of frequently performed elective surgical cases (van Oostrum, Bredenhoff et al. 2010, van Oostrum, Van Houdenhoven et al. 2008). After assigning patients to an MSS, a time slot may or not may be used due to variable demand. In case a slot remains unused, this is perceived as inefficient. However, using the time slot for another surgical case type creates variability that an MSS tries to avoid. The maximum time that a hospital plans ahead, the so-called planning horizon, is expected to affect the impact of unused time slots.

The aim of this study is twofold. First, we aim to determine the effect of the length of the planning horizon and other planning parameters in an MSS approach on patients´ waiting time, schedule stability and hospital efficiency. Second, we compare the MSS approach with a standard operating room planning approach on levelled bed occupancy. We perform this study with real case data from Beatrix Hospital, a regional hospital in The Netherlands. The structure of the paper is the following. In Section 2 we give background information on surgical case planning, master surgical scheduling and its counterparts in industry. In Section 3 we introduce a formal problem definition. Section 4 presents the methods that we apply in order to quantify the effects just mentioned. In Section 5 we present the case study and Section 6 contains concluding remarks.

2. Background

From an operations management perspective, an MSS in hospitals can be compared to a *Master Production Schedule* (MPS) in industry. An MPS determines production moments and production amounts. Factories that use such a schedule aim to stabilize their production flows. One form of using an MPS is to re-use a master schedule several times one after the other. Resulting from the MPS, a *material requirements plan* can be constructed that gives a description of the resource requirements for the production as scheduled in the MPS. The *material requirements plan* in a hospital incorporates requirements for surgery materials and post operative needs such as recovery and ward capacity.

Late changes in schedules are known to induce sometimes major changes in resource requirements at other stages of a production line. This leads to nervousness in the operational planning at various subsequent stages, a phenomenon that is well known in hospitals. For example, last minute changes in surgical schedules can cause severe nervousness at wards and other hospital departments. Hence, an MSS approach will only be successful when at the execution stage no need exists to adjust schedules.

Before we continue by introducing possible ways to avoid nervousness in an MSS approach, we review some relevant scientific literature on surgical scheduling and master surgical scheduling in particular. Thereafter, we introduce some approaches to deal with nervousness that are applied in industry. We finish by addressing how these approaches would work out in an MSS approach.

2.1. Surgical case scheduling

Operating room management and more specific operating room planning has been a popular topic among researchers. This has resulted in a wide set of approaches to this complex theme in hospital management (see for instance www.franklindexter.net). We refer for a complete and detailed discussion on various surgical scheduling approaches to (Cardoen, Demeulemeester et al. 2009). This paper classifies several characteristics of operating room planning and scheduling and gives a complete overview of research work on this topic.

One approach that has attracted attention is master surgical scheduling. Various researchers have defined this approach in slightly different ways, but all approaches have in common that operating room management is structured based on a tactical plan that is repetitively executed (van Oostrum, Van

Houdenhoven et al. 2008, Beliën, Demeulemeester 2007, Blake, Donald 2002, Vissers, Adan et al. 2005).

In this paper we use the definition as has been introduced by (van Oostrum, Bredenhoff et al. 2010, van Oostrum, Van Houdenhoven et al. 2008, Van Houdenhoven, van Oostrum et al. 2008). This approach consists of seven steps (van Oostrum, Bredenhoff et al. 2010):

1. Defining the scope of the MSS
2. Data gathering
3. Capacity planning
4. Defining a set of recurrent standard case types
5. Constructing the MSS
6. Executing the MSS
7. Updating the MSS

A master schedule is complete when the first five steps have been successfully done. Hence, the scope is clear (all shared resources involved in the planning process are known), all data is available, and there exist sound capacity plans such that on an aggregate level demand and supply are leveled. Moreover, standard surgical case types have been constructed for use at an MSS (van Oostrum, Parlevliet et al. 2008). Note that these case types have been constructed such that the variability in resource usage within a single case type is minimal and such that the total number of patients that would not fit into a standard case type is minimized.

After the first five steps a hospital knows its standard case types and the frequency that such a case type is performed per period. Furthermore, it knows the amount of capacity to be planned for elective patients for whom no standard case type will be available, so-called *elective slack capacity*. For example, this might be done by constructing one or more '*dummy surgery*' case types. At Step 5, a hospital has constructed its MSS such that operating room utilization is maximized and occupancy levels at subsequent hospital departments are leveled. Operating room capacity is reserved for emergency surgery by allocating slack at each operating room (Wullink, Van Houdenhoven et al. 2007). Hence, an MSS consists of standard surgical case types, elective slack capacity, and *emergency slack capacit*y.

At the operational level (Step 6), patients are assigned to standard case types of the MSS. We provide some definitions regarding the time elements in this planning process. A *planning cycle* is the period after which the MSS is repetitively executed. It contains a certain number of days. Assigning patients is restricted to the planning cycles that are available. For instance, a hospital might put a minimum time for elective patients between the moment of planning and the moment of execution of a surgical case, which is referred to as the *frozen horizon*. We assume that the frozen horizon contains an integer number of planning cycles. In addition, a hospital might put a restriction on the maximum time a patient is scheduled ahead, the *planning horizon*. Without loss of generality we assume that a hospital does the assignment of patients once per planning cycle. In case that not all patients can be planned, they are put on a waiting list. Patients who arrive between two planning moments are added to the waiting list. We assume that hospitals treat patients on the waiting list on a first come, first serve basis. As we will explain later, the length of the planning horizon is the main determinant for successful scheduling at this step.

Patients are assigned to the first available standard case type (slot) to which they belong. When there is no such slot available within the planning horizon (for any reason), a hospital tries to assign the patient to available elective slack capacity. If this still is not possible, but if it would be possible with a small amount of planned overtime, a hospital might do so. When, after all, a patient cannot be assigned, he/she remains on the waiting list to a next period. This assignment approach is summarized in Figure 1.

2.2. Avoiding nervousness in master production planning

Various researchers have studied the application of MPS in practice. In this subsection we discuss some relevant papers and focus on dealing with nervousness when using master production scheduling approach in industry.

Shirdharan, Berry et al. (2007, 1987) discuss that one way to resolve nervousness in production planning is to freeze a part of the planning horizon. Nervousness is expressed by schedule instability that can be measured as the amount of changes in the schedule between two consecutive cycles. Within a frozen horizon, the actual scheduling cannot be changed any longer for any reason. This also holds when eventually new information comes available, for instance about future demands. Hence, a frozen horizon gives the manufacturer more certainty about the production amounts in the near future. The researchers use computer simulation to investigate the effect of adopting freezing, and also the effect of various lengths of frozen horizons and the length of the planning horizon in general. The effects are measured by means of production and inventory cost and deterioration in customer service. The authors show that freezing up to 50 percent of the planning horizon has marginal effect on the measures just mentioned.

Both articles (Sridharan, Berry et al. 2007, Sridharan, Berry et al. 1987) consider a fixed dynamic future demand. ‘Fixed’ means that uncertainty about the demand levels was not taken into account and demand is `dynamic’ when the demand levels vary among the future periods. In reality however, demand forecasts of future periods might include forecast errors as is discussed by Lin and Krajewski (2007). They investigate the effects of demand uncertainty on several factors, including the stability of the MPS. Uncertainty in future demand results in additional costs. When customer satisfaction plays an important role, a reasonable amount of safety stock is required in order to compensate demands which exceed the expected demand level.

An additional way of improving performance is by considering rescheduling. In the literature different ways of rescheduling in production planning are mentioned (Yang, Jacobs 2007, Tang, Grubbström 2002, Hill, Berry et al. 2003). Within the context of an MPS rescheduling correlates to adding or deleting orders. It involves minimizing costs resulting from schedule changes related to lower-level items. The higher the number of levels the MRP contains, the more complicated the rescheduling optimization will get.

2.3. Stable master surgical scheduling

The previously discussed articles use the concept of schedule stability as measure for the variability in final schedules between subsequent planning cycles. We will adopt schedule stability as a measure to the MSS approach. Within this context schedule stability defines how much of the predetermined slots for surgical case types are actually used accordingly.

From the presented research it is clear that the length of a planning horizon and possibly freezing a part of the planning horizon has impact on the schedule stability. Freezing gives hospital management information about the operating room and personnel schedules in the coming period without possible changes. On the other hand, the frozen horizon results in an increase in the patients’ waiting time since available time slots within the frozen horizon are not available for new patients. Given this, we assume that hospitals take a minimal frozen horizon, often of one week.

Dealing with demand uncertainty is also an issue that has been incorporated in the MSS approach. Instead of using safety stock, hospitals either reserve slack capacity or leave capacity unassigned to a late moment to be able to adjust to variation in demand. Within an MSS approach, capacity is available to patients for whom no dedicated slots are available. A hospital may choose whether or not different medical specialties share this capacity in order to gain from economies of scale. The amount of capacity left over will be typical the result from the trade off between efficiency and the required ability to account for demand variability in the construction of an MSS.

Finally, rescheduling might be applied within an MSS approach as well as it is within MPS approaches. Patients might be moved to another slot if this is beneficial to a hospital. For example, patients that are assigned to slack capacity might be rescheduled to an empty surgical case type that comes available after the planning horizon has moved a cycle ahead.

3. Problem description

The two aims of this study are the following. First, we determine the effect of the length of a planning horizon in an MSS approach on patients´ waiting time, schedule stability and hospital efficiency. These three main performance measures are defined at the end of this section. Additionally, we investigate effects of sharing elective slack capacity by multiple medical specialties and the effects of rescheduling. Second, we compare the MSS approach with a standard operating room scheduling approach on levelled bed occupancy. First, we will introduce some notation and definitions and we will use these definitions in order to describe the heuristic that assigns patients to future time slots.

3.1 Notation and definitions

Let P denote the set of patients. Each patient requires one specific surgical case, which is in an MSS approach linked to a so called standard surgical case. The set of standard surgical cases is denoted by I. Let indicator z_{pi} be 1 if patient $p \in P$ requires standard surgical case $i \in I$, 0 otherwise. The planned duration of a standard surgical case $i \in I$ is represented by d_i. S denotes the set of surgical departments. The set I^s represents all surgical cases $i \in I$ that are performed by surgical department $s \in S$. It holds that $\bigcup_{s \in S} I^s = I$. Furthermore, let J denote the set of operating rooms and let T be the set containing all days within one planning cycle of the MSS. The planning cycle has a predetermined length $|T|$. We denote operating room $j \in J$ on day $t \in T$ of the planning cycle by OR-day(j,t). Each OR-day(j,t) has a certain amount of time capacity cap_{jt} for surgeries to be performed. We assume that emergency slack is allocated to all operating rooms (Wullink, Van Houdenhoven et al. 2007). We refer to the amount of time reserved for emergency slack by e_{jt} for OR-day(j,t).

We assume that an MSS is given in which time slots are reserved for types of frequently performed elective surgical cases within every OR-day of the planning cycle. We introduce the parameters $a_{ijt} \in IN$ that denote the number of surgical case types i that are included in OR-day(j,t) of the MSS. The parameters a_{ijt} are a result of an automated scheduling algorithm (van Oostrum, Van Houdenhoven et al. 2008). Hospitals may choose to restrict the number of different surgical departments on a single OR-day (j,t). Surgical cases $i \in I$ are incorporated in an MSS on their rounded down expected frequency per cycle. Hence, some may not be prescribed to any OR-day (i.e. $\sum_{j\in J}\sum_{t\in T} a_{ijt} = 0$). Based on the expected durations of these surgery types, as well as the expected duration of left-over fractions of recurring surgical cases, the MSS contains dummy time slots which may be filled by any type of elective surgery in the planning process. Any remaining capacity is added to the dummy time slots, which is hence of size $c_{jt} = cap_{jt} - e_{jt} - \sum_{i\in I} a_{ijt} d_i$. Consequently, there are finally three types of time slots: the ones for emergency surgery, the ones for the standard surgical cases and the time slots containing elective slack (dummy) capacity.

The notation is consistent with definitions introduced by Van Oostrum, Van Houdenhoven et al. (2008). For a method to obtain standard surgical cases we refer to Van Oostrum, Parlevliet *et al.* (2008), and for methods to obtain a particular MSS we refer to Van Oostrum, Van Houdenhoven et al. (2008).

Regarding the time horizon and planning process, let K be the set containing all future planning cycles. A hospital schedules elective patients a certain number of planning cycles ahead, restricted by the planning horizon $h \in IN$. In practice, hospitals schedule elective patients a minimum time in advance, modeled by the frozen horizon. We define $f \in IN$ as the number of frozen cycles (by definition $f < h$). Without loss of generality we assume that once per cycle, at the beginning of a cycle $k \in K$, patients are scheduled within the time window $[k+f, k+h-1]$. Figure 2 illustrates these settings for a MSS with a cycle length of one week, a planning horizon of four weeks and a frozen horizon of one week. Patients that are accumulated at the beginning of week 1 may be scheduled within the OR-days of weeks 2,3 and 4. The process of scheduling is repeated every cycle, such that the horizon rolls forward every cycle. Hence, we have modeled a so-called rolling horizon. The waiting list at the beginning of cycle k is defined as $W_k \subseteq P$. The actual planning is defined by $x_{pjtk} \in \{0,1\}$, where $x_{pjtk} = 1$ if patient $p \in P$ is scheduled at OR-day (j,t) in cycle k and 0 elsewhere.

3.2 Assignment heuristic

At the operational level patients are to be assigned to slots of an MSS. We have modeled this assignment of patients to MSS slots by the heuristic presented below. Note that this heuristic is a detailed application the process depicted in Figure 1.

For all cycles $k \in K$ do

- Accumulate patients that have arrived up to and including cycle $k-1$ who have not been scheduled yet, on the waiting list W_k
- For all patients $p \in W_k$ do
    - o If an empty slot for patient p is available in the planning horizon (i.e. if there is a $k' \in \{k+f,...,k+h-1\}$ and a $j \in J$ and a $t \in T$, such that $a_{ijt} - \sum_{p \in P} z_{pi} x_{pjtk'} \geq 1$) then
        - ▪ Assign patient p to the first available slot at (j,t) in cycle k'. That is, set $x_{pjtk'} = 1$.
        - ▪ Else If unused slack capacity is available in the planning horizon for surgical case i (i.e. if there is a $k' \in \{k+f,...,k+h-1\}$ and a $j \in J$ and a $t \in T$, such that $c_{jt} - \sum_{p \in P} d_i z_{pi} x_{pjtk'} \geq d_i$) then
            - • Assign patient p to the first available slack capacity at (j,t) in cycle k'. That is, set $x_{pjtk'} = 1$.
            - • Else If slack capacity plus some proportion δ of the emergency slack capacity e_{jt} is sufficient to schedule surgery type i (i.e. if there is a $k' \in \{k+f,...,k+h-1\}$

and a $j \in J$ and a $t \in T$, such that $c_{jt} + \delta \cdot e_{jt} - \sum_{p \in P} d_i z_{pi} x_{pjtk'} \geq d_i$) then

- o Assign patient p to the first available slack capacity at (j,t) in cycle k'. That is, set $x_{pjtk'} = 1$.
- o Else place patient p on waiting list W_{k+1} for cycle k+1

In order to further improve the effectiveness of the MSS, rescheduling some of the patients after their initial assignment, might be considered. One way to apply rescheduling would be to shift patients from elective slack capacity within a certain predetermined replanning horizon of length r, with $r \leq h - f$ to a newly available time slot explicitly meant for the surgery type that the patient requires, in case such a slot is still available when time has 'rolled on' to the next cycle. When multiple patients of a certain surgery type are scheduled in elective slack capacity within the replanning horizon, selecting the one that is scheduled the furthest ahead in time gives the least increase in the overall average patients' waiting time. Therefore, the replanning horizon is checked from back to front. Applying this rescheduling approach would then extend the above heuristic as follows:

- For all recurrent surgical cases (i.e. all cases $i \in I$ that have $\sum_{j \in J} \sum_{t \in T} a_{ijt} > 0$
  - o Do while there is a patient of type i scheduled within elective slack capacity of the replanning horizon AND there is a time slot available specifically for type i in the last cycle of the planning horizon (i.e. there is a $k' \in \{k+h-2, k+h-3, ..., k+h-r-1\}$ and a $j \in J$ and a $t \in T$, such that $\sum_{p \in P} z_{ip} x_{pjtk'} > a_{ijt}$ AND there is a $j \in J$ and a $t \in T$, such that $\sum_{p \in P} z_{ip} x_{pjt(k+h-1)} < a_{ijt}$
    - ▪ Shift patient from slack capacity to the available recurrent time slot of the last cycle of the MSS. That is, for the patient p that was scheduled in elective slack set $x_{pjtk'} = 0$ and set $x_{pjt(k+h-1)} = 1$.

3.3 Performance Measures

We consider the following performance measures: patients' waiting time, schedule stability, and hospital efficiency. Below each of the measures is defined.

Patients' waiting time

The waiting time of a patient is defined as the time (expressed in weeks) between the submission of a request for surgery and the actual time of the execution of surgery. We are interested in the distribution of the patients' waiting time, and more specifically in the average waiting time and the proportion of patients waiting for 8 weeks or less. The latter is taken into account since in practice, governmental regulations enforce that certain proportions of patients do not have to wait longer than some cut-off value or values.

Schedule stability

We measure schedule stability in two ways. The first measure $Afill$ indicates to what extent the final OR schedule fits to the MSS. Hence, we measure the proportion of slots that is ultimately filled with appropriate patients. For cycle $k \in K$ we have $Afill = \frac{\sum_{i \in I} \sum_{j \in J} \sum_{t \in T} \min\left(a_{ijt}, \sum_{p \in P} x_{pjtk} z_{pi}\right)}{\sum_{i \in I} \sum_{j \in J} \sum_{t \in T} a_{ijt}} 100\%$.

Here, the numerator represents the number of patients scheduled in recurring elective time slots and the denominator represents the total amount of recurring time slots contained in the MSS. Recall that time slots for standard surgical cases may remain empty when the number of patients requiring a specific surgical case is less than expected. In that case $\sum_{p \in P} x_{pjtk} z_{pi} < a_{ijt}$ holds.

Note that increasing the planning horizon will ultimately result in an asymptotic upper bound of 100 percent appropriate use of standard surgical cases since cases are scheduled at the MSS based on their rounded-down frequency.

As a second measure for schedule stability we define $AinA$ as the proportion of patients that are scheduled in an appropriate standard surgical case, contrary to being assigned to elective slack capacity. For cycle $k \in K$ we have

$$AinA = \frac{\sum_{i\in I}\sum_{j\in J}\sum_{t\in T}\min\left(a_{ijt}, \sum_{p\in P} x_{pjtk} z_{pi}\right)}{\sum_{p\in P}\sum_{j\in J}\sum_{t\in T} x_{pjtk}} \cdot 100\% .$$

Here, again the numerator represents the number of patients scheduled in recurring elective time slots, while now the denominator represents the total amount of elective patients scheduled in week k. Recall that patients will be scheduled in elective slack when the number of patients of type i arriving in week k exceeds the amount of recurrent slots for that type. In that case $\sum_{p\in P} x_{pjtk} z_{pi} > a_{ijt}$ holds. Since surgical cases are scheduled at the MSS based upon their rounded-down frequency, for each type $i \in I$ more patients arrive on average than the number of recurrent slots that are available in the MSS for that type. Therefore, as we increase the planning horizon, this measure will not asymptotically converge to 100 percent.

Hospital efficiency

We distinguish three measures of hospital efficiency. First the frequency and duration of planned overtime at operating rooms, second the fluctuation in bed occupancy at wards subsequent to surgery.

Hospitals tend to apply fuzzy approaches to capacity restrictions at operating room departments. In the defined assignment heuristics we accounted for this. We therefore define as measures for overtime the frequency of violating the available capacity and the average amount of time per surgery that is planned in overtime (which we will denote by FPO and DPO respectively).

Fluctuation in bed occupancy causes extra costs since ward staff is either under utilized or extra staffing is required. The latter imposes often direct costs for hiring flexible capacity shortly in advance. We assume that a bed can only be used for one patient during a day. Bed occupancy is measured on a daily basis. Given these data our performance measure is the fluctuation represented by the standard deviation of the daily fluctuation of bed occupancy (which is denoted by $SDBO$).

4. Solution Approach

In this section we will address the approach that we use to investigate the effect of scheduling decisions as well as the effect of the MSS compared to a standard operating room scheduling approach.

4.1 Experimental design

Aim 1: effect of the planning parameters

In order to determine the effect of the planning parameters, which is the first aim of this study, we introduce different scenarios. A scenario is defined by the following planning parameters: the planning horizon (denoted by h), whether or not rescheduling is applied (as described in section 3.2), and whether or not the elective slack capacity may be shared among the different hospital departments (denoted by ss). If rescheduling is applied, the replanning horizon (denoted by r) needs to be defined.

We define a so-called basic scenario which we will use as a benchmark against alternative scenarios. As for the settings of this basic scenario we choose a planning horizon of five cycles (e.g. 5 weeks), we disregard the possibility of rescheduling and we apply the distinction between the elective slack capacity restrictively assigned to one specialty. In the alternative scenarios, we vary only one of these input parameters from the basic scenario. This allows us to report the effect of each input parameter of the scheduling approach separately. Table 1 presents all alternative scenarios.

In all scenarios, we will simulate the flow of incoming patients by a Poisson arrival distribution with arrival rate λ, where λ represents the average number of patients arriving per cycle.

Aim 2: comparing bed occupancy of MSS and standard planning

As for the second aim of this study, we compare the MSS approach with a standard operating room planning approach. The standard operating room planning approach that we will use as a benchmark to test the MSS approach against, is the First-Come, First-Serve (FCFS) approach (Dexter, Macario et al. 1999). Using a FCFS approach, patients are scheduled in the first available OR-day according to the order in which they arrive.

We expect the variation in the number of daily arriving patients to have substantial influence on the bed occupancy levels, not only for the FCFS approach, but also to some extent for the MSS approach. Therefore, we will measure the standard deviation of the bed occupancy level resulting from both the FCFS approach and the MSS approach (with the settings of the basic scenario), by assuming different arrival distributions of patients. Besides a Poisson distribution, we test four additional arrival distributions. We will use a constant arrival of λ patients per cycle and three different gamma distributions as to achieve increasing variances of the arrival rate. The gamma distributions have an expected value of λ and have variances of 0.5λ, 2λ and 4λ respectively.

4.2 Simulation

The complexity of this scheduling problem, including the use of slack capacity justifies the use of simulation. To investigate the effect of a certain scenario, in each iteration the following steps are executed:

1. A realisation of the number of arrivals in one cycle is randomly drawn from the predetermined arrival distribution. Subsequently patients are randomly drawn from an empirical data set that includes all relevant patient data.
2. Patients on the waiting list, combined with the new arrivals, are scheduled if possible, within the appropriate time slots of the current planning horizon as described in the assignment heuristic in section 3.2. If allowed, rescheduling may be applied.

Upon completion of a simulation run, the performance measures are calculated.

A number of iterations need to be performed, before the system will arrive at a steady state. To determine the length of this so-called warm-up phase of one simulation run, we have recorded the number of patients in the system after each iteration. By the number of patients in system we mean the number of patients scheduled within the future planning horizon, together with the patients on the waiting list. Naturally, the longer the planning horizon is chosen, the longer the warm-up phase of the simulation process takes. Next, a certain number of cycles succeeding the warm-up phase are considered to extract the performance measures resulting from the given scheduling approach from. Since the performance measures of successive cycles are not independent, we repeat this simulation process n times in order to construct confidence intervals around the performance measures. A $(1-\alpha)\%$ confidence interval around measure X is constructed by

$$\left[\overline{X} - z_{\alpha/2}\frac{s}{\sqrt{n}}, \overline{X} + z_{\alpha/2}\frac{s}{\sqrt{n}}\right],$$

where $\overline{X}$ is the average and s is the standard deviation of performance measure X over the n independent simulations and $z_{\alpha/2}$ is the z-statistic of the standard normal distribution (for large values of n). This implies that $(1-\alpha)\%$ of the confidence intervals constructed in this manner, actually contain the real parameter that we are trying to estimate. Increasing the number of repetitions will decrease the size of the interval. In case a specific size of the interval would be requested, the required number of runs n could be obtained using the formula just presented, were s would be taken as the standard deviation of a preliminary chosen number of runs.

5. Case study

In this section we present a case study performed for the Beatrix Hospital, the Netherlands, to answer our two research objectives as formulated in Section 3. Section 5.1 describes the data of this case study. Section 5.2 presents the results the case study. Section 5.3 analyses the results.

5.1 Data

The data that is used in this research originates from the urology and general surgery department of the Beatrix Hospital, the Netherlands ($s=1$ denotes general surgery and $s=2$ denotes urology). The dataset consists of all patients for both departments in one year, including 1862 unique patients and 200 different surgical cases. Of each case we have the surgery duration and the length of stay at the ward. The data implies that on average 16.3 beds are occupied.

The Beatrix Hospital determines norm utilization β_s for elective surgery that both departments should obtain ($\beta_1 = 0.75$ and $\beta_2 = 0.81$). The remainder of the capacity is allocated as emergency slack to OR-day schedules. Hence, $e_{jt} = (1-\beta_s)cap_{jt}$ gives the amount of emergency slack. The parameter for allowing fuzzy use of capacity constraints was set to $\delta = 0.5$. The cycle time chosen for this MSS is one week and the number of production weeks per year is 46.

Applying clustering to the original data (van Oostrum, Parlevliet et al. 2008) resulted coincidentally in thirteen different standard surgery case types for both departments. Table 2 shows the result after clustering the unique cases into standard surgical case types. Subsequently, the MSS is constructed by means of optimisation with regard to the OR utilisation and the stability of its resulting demand for hospital beds as described in (van Oostrum, Van Houdenhoven et al. 2008).

5.2 Results

Aim 1: effect of the planning horizon

Table 3 shows the effects of all aforementioned scenarios on patients' waiting time, schedule stability and hospital efficiency. All scenarios produce significantly different performance measures, based on non-overlapping 95% confidence intervals. To provide some insight in the length of the intervals, Table 4 contains the confidence intervals of the performance measures resulting from the basic scenario.

Table 3 shows that an increase of the planning horizon leads to longer patients' waiting time. The increments of the waiting are decreasing. The proportion of patients that experience a waiting time of 8 weeks or less decreases with longer planning horizons. The schedule stability is high and only marginal increases by planning horizons of 7 weeks or longer. Most of the slots (over 95% for planning horizons of four weeks and longer) are filled. The proportion of cases that is assigned to the right slots is consequently also high. Finally, regarding hospital efficiency, increasing the planning horizon decreases the average number of surgical cases that is scheduled in planned overtime, while its average duration increases. $SDBO$ shows that the variation in the bed occupancy levels decreases as we increase the planning horizon.

Besides the effects of the planning horizon, Table 3 also allows us to address the effects of replanning and sharing elective slack capacity between the two surgical departments. Increasing the replanning horizon in the replanning process slightly increases waiting time and, at the same time, slightly increases schedule stability and hospital efficiency. Sharing elective slack capacity between both departments is of marginal influence on some performance measures.

We found some remarkable results when analysing the distribution of waiting times. To characterize the distribution of the patients' waiting time, Figure 3 shows the relative frequencies for patients that have to wait a certain number of weeks, when a planning horizon of 50 weeks is used. Maybe less realistic, but we show the result of this specific setting to enlarge the effect of the MSS on the waiting time distribution. A large proportion of patients either wait relatively short or relatively long for surgery, resulting in a waiting time distribution with an unusual shape.

To examine this issue further, we look at the waiting time distribution of individual surgery types separately. In doing so, we distinguish two groups of surgery types i which each have their own typical shape: 1) the recurrent surgery types that are explicitly scheduled in the MSS (which have

$\sum_{j \in J} \sum_{t \in T} a_{ijt} > 0$) and 2) the surgery types for which no specific time slot is contained within the MSS (and thus have $\sum_{j \in J} \sum_{t \in T} a_{ijt} = 0$). The waiting time distribution for the first type shows increasing amounts of long waiting time near the end of the planning horizon and some decreasing, relatively small frequencies for the short waiting times. The waiting time distributions for the second type on the other hand, show that longer waiting times have strictly decreasing probabilities. The explanation for this is that for patients of the fist type, the whole planning horizon is checked for available time slots of their own type. When no such time slot is available, the patient is assigned to elective slack capacity. Since the amount of available slots specifically meant for the first type in the MSS is by construction lower than the average number of such patients arriving per cycle, all these time slots within the planning horizon will gradually become occupied as time rolls on. At that point, the only time slots available specifically for these patients are the new ones of the MSS that are added at the end of the planning horizon as time rolls on. In case the number of available time slots is not sufficient for the arrivals within one cycle, some of these patients will be scheduled within the fist available elective slack capacity. Because the amount of available slack capacity exceeds the expected required slack capacity, these cases are assigned near the beginning of the planning horizon. Patients of the second type are assigned to the first available elective slack capacity right away, hence they are assigned near the beginning of the planning horizon. Combining these two possibilities, results in the pattern of Figure 3.

Aim 2

Table 5 shows the differences between the use of an MSS and a standard FCFS planning approach with respect to the variation in bed occupancy levels. For the MSS, the results are based on the basic scenario settings. Different arrival distributions are used as described in the previous section.

Table 5 shows that arrival distributions with higher variances result in higher variability in bed occupancy levels for both planning approaches. Also, Table 5 shows that using an MSS decreases the standard deviation of the bed occupancy level almost by a factor two. While the arrival rates and the planning approach result in different standard deviations of the bed occupancy levels, naturally all settings result in the same average bed occupancy, which is 16.3 beds occupied per week.

6. Conclusion and discussion

Regarding our first aim, we have been able to quantify the effects of the planning horizon, rescheduling and sharing elective slack among different departments on waiting time, schedule stability and hospital efficiency. We found that increasing the planning horizon increases patients' waiting time on the one hand, but also increases schedule stability and hospital efficiency on the other hand. Moreover, only marginal influences are found for rescheduling and sharing elective slack by departments.

Regarding our second aim, comparing MSS with a standard First Come, First Serve approach, we found that using an MSS results in a substantial decrease in variability in bed occupancy levels. This shows the great benefits of the use of an MSS for hospitals as it creates more predictable flows of patients from the operating room department to subsequent hospital departments.

This research shows that it is possible to work with an MSS, while maintaining short waiting lists. In addition, it shows the major benefits for variability reduction on wards which is among one of the main drivers for inefficiencies in hospitals. Our methodology for analysis of two departments of the Beatrix hospital can be applied to any other hospital. Based upon the results hospital management should decide upon the best planning horizon for their hospital, given the trade-off between waiting times, schedule stability, and hospital efficiency.

References

BELIЁN, J. and DEMEULEMEESTER, E., 2007. Building cyclic master surgery schedules with leveled resulting bed occupancy. *European Journal of Operational Research,* **176**(2), 1185-1204.

BLAKE, J.T. and DONALD, J., 2002. Mount Sinai hospital uses integer programming to allocate operating room time. *Interfaces,* **32**(2), 63-73.

CARDOEN, B., DEMEULEMEESTER, E. and BELIЁN, J., 2009. Operating room planning and scheduling: A literature review. *European Journal of Operational Research,* **201**(3), 921-932.

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.

HILL, J.A., BERRY, W.L. and SCHILLING, D.A., 2003. Revising the master production schedule in sequence dependent processes. *International Journal of Production Research,* **41**(9), 2021-2035.

LIN, N.P. and KRAJEWSKI, L., 2007. A model for master production scheduling in uncertain environments. *Decision Sciences,* **23**(4), 839-861.

SRIDHARAN, S.V., BERRY, W.L. and UDAYABHANU, V., 2007. Measuring master production schedule stability under rolling planning horizons. *Decision Sciences,* **19**(1), 147-166.

SRIDHARAN, V., BERRY, W.L. and UDAYABHANU, V., 1987. Freezing the master production schedule under rolling planning horizons. *Management Science,* **33**(9), 1137-1149.

TANG, O. and GRUBBSTRÖM, R.W., 2002. Planning and replanning the master production schedule under demand uncertainty. *International Journal of Production Economics,* **78**(3), 323-334.

VAN HOUDENHOVEN, M., VAN OOSTRUM, J.M., WULLINK, G., HANS, E., HURINK, J.L., BAKKER, J. and KAZEMIER, G., 2008. Fewer intensive care unit refusals and a higher capacity utilization by using a cyclic surgical case schedule. *Journal of critical care,* **23**(2), 222-226.

VAN OOSTRUM, J., BREDENHOFF, E. and HANS, E., 2010. Suitability and managerial implications of a Master Surgical Scheduling approach. *Annals of Operations Research,* **178**(1), 91-104.

VAN OOSTRUM, J.M., PARLEVLIET, T., WAGELMANS, A.P.M. and KAZEMIER, G., 2008. A method for clustering surgical cases to allow master surgical scheduling. *Conference proceeding ORAHS 2008*.

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.

VISSERS, J.M.H., ADAN, I. and BEKKERS, J.A., 2005. Patient mix optimization in tactical cardiothoracic surgery planning: a case study. *IMA Journal of Management Mathematics*.

WULLINK, G., VAN HOUDENHOVEN, M., HANS, E.W., VAN OOSTRUM, J.M., VAN DER LANS, M. and KAZEMIER, G., 2007. Closing emergency operating rooms improves efficiency. *Journal of medical systems,* **31**(6), 543-546.

YANG, K.K. and JACOBS, F.R., 2007. Replanning the master production schedule for a capacity-constrained job shop. *Decision Sciences,* **30**(3), 719-748.

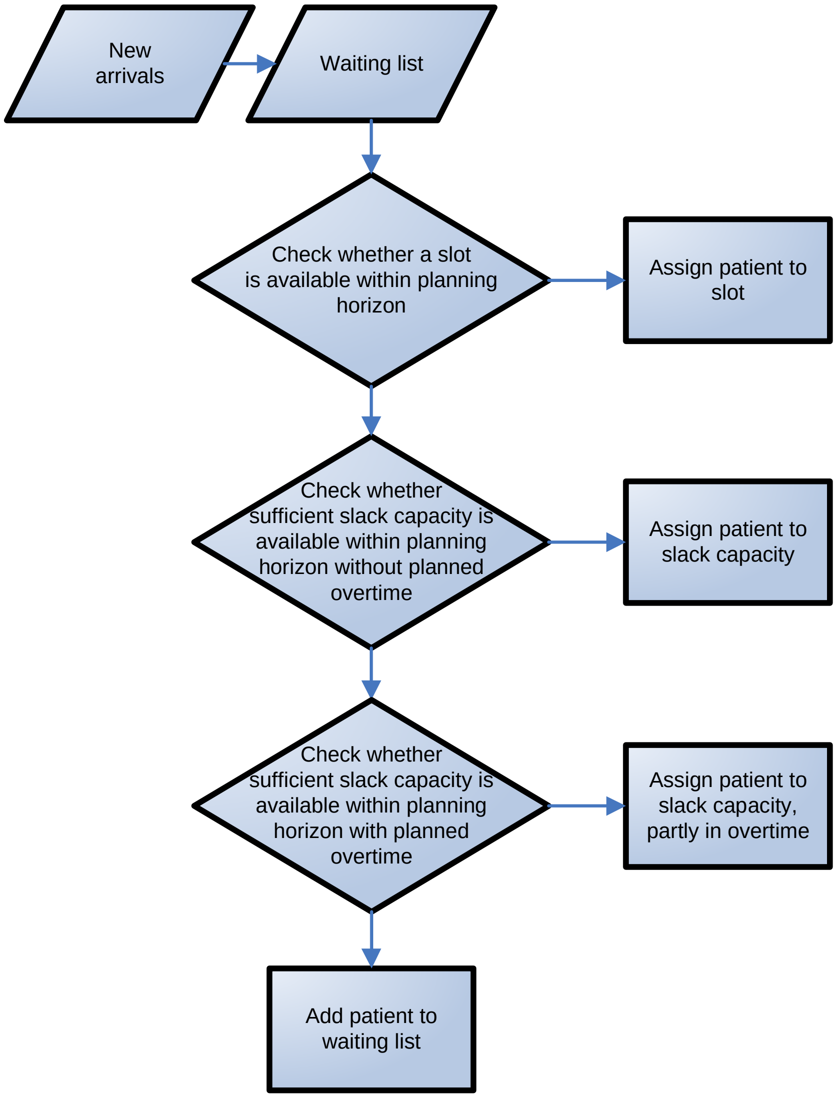

Figure 1: Heuristic for assigning patients to MSS

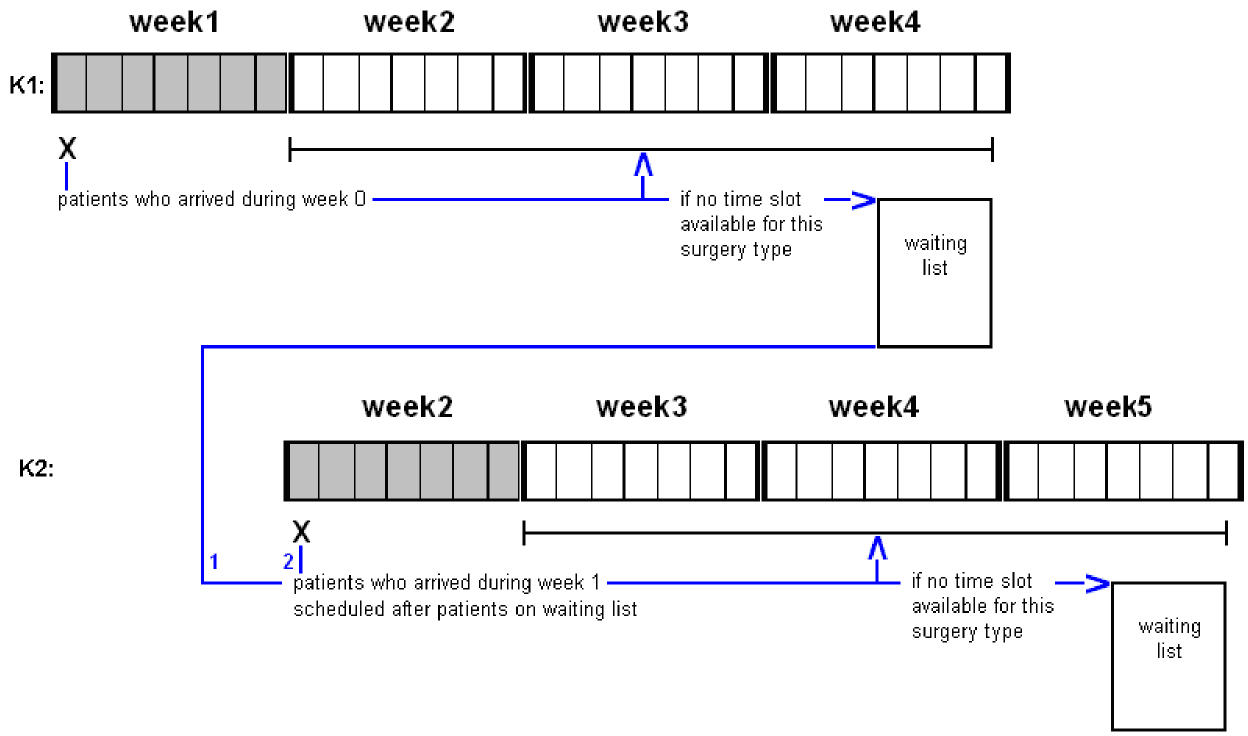

Figure 2: The rolling horizon planning process, at cycle *k*=1 and *k*=2. The cycles are of length one week (containing 7 days, *T*={1,2,…, 7}. A frozen horizon of *f*=1 (denoted by grey days) and a planning horizon of *h*=4 is applied.

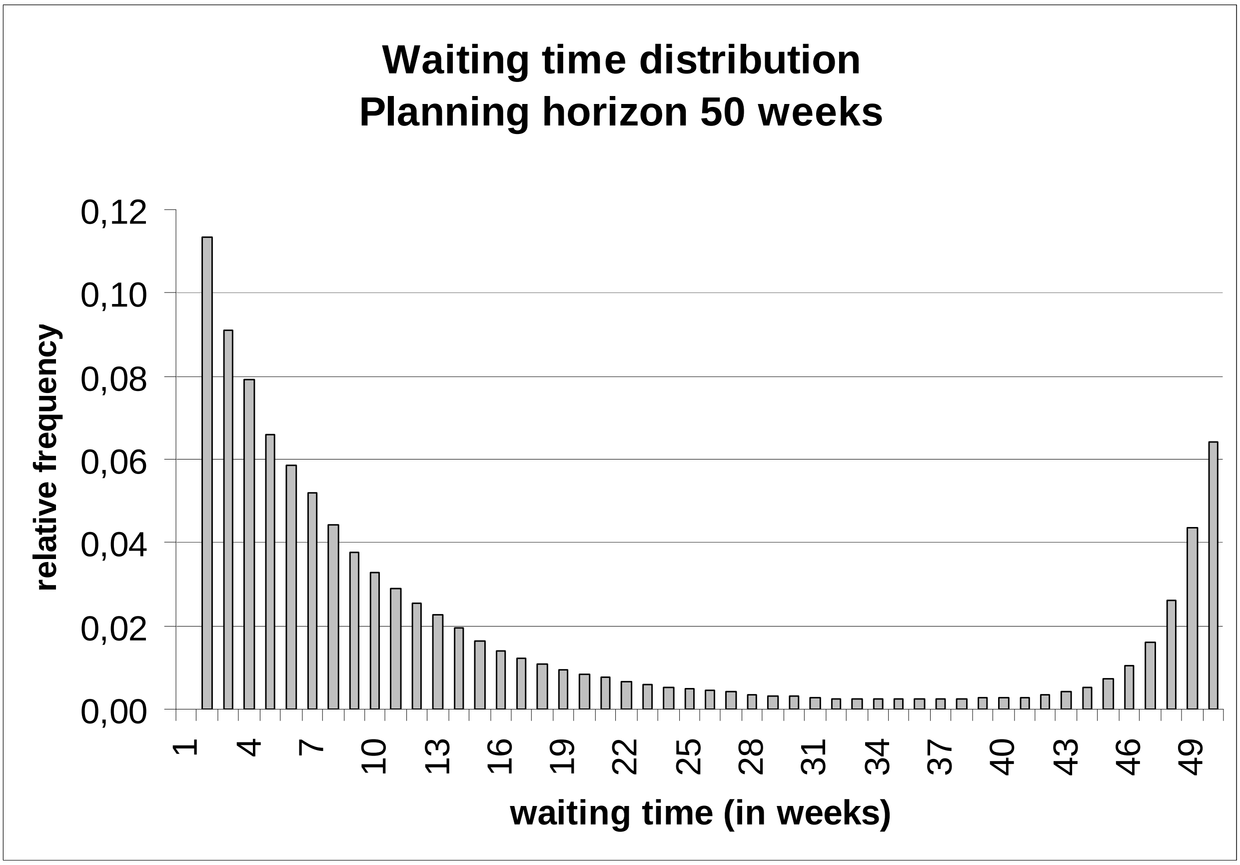

Figure 3: waiting time distribution for *h*=50.

Planning parameters	Basic scenario	Alternative options
Planning horizon (h)	5	{2,3,… ,14,15,25,50}\{5}
Replanning horizon (r)	0	{1,2,3,4}
Share elective slack capacity (ss)	No	Yes

Table 1: scenario definitions

	General surgery				Urology		
	Frequency per year	Frequency in MSS	Average duration (minutes)		Frequency per year	Frequency in MSS	Average duration (minutes)
Different standard surgical case types	463	10	60	Different standard surgical case types	137	2	54
	418	9	41		94	2	65
	184	4	70		93	2	37
	156	3	85		60	1	53
	59	1	196		14	0	169
	58	1	146		10	0	333
	50	1	83		6	0	121
	10	0	223		6	0	298
	10	0	217		4	0	131
	8	0	285		3	0	108
	7	0	65		3	0	186
	3	0	67		2	0	124
	2	0	423		2	0	421

Table 2: Standard surgical cases in Beatrix hospital with their annual frequency, their rounded down weekly frequency (contained in the MSS), and the expected durations.

Scheduling Approach			Performance Measures						
			Waiting Time		Schedule Stability		Hospital Efficiency		
h	*r*	*ss*	Average	< 8 weeks	Afill	AinA	FPO	DPO	SDBO
(weeks)	(weeks)	(Yes/No)	(weeks)	(%)	(%)	(%)	(surgeries)	(minutes)	(beds)
basic scenario									
5	0	No	3,5	100,0	96,9	89,7	1,4	88,0	3,3
alternative scenario's									
2	0	No	2,2	100,0	86,4	79,9	4,4	48,1	3,7
3	0	No	2,6	100,0	93,0	86,1	2,5	58,8	3,5
4	0	No	3,0	100,0	95,7	88,5	1,6	79,7	3,3
6	0	No	4,0	100,0	97,7	90,4	1,3	90,8	3,2
7	0	No	4,5	100,0	98,1	90,7	1,3	92,3	3,2
8	0	No	4,9	100,0	98,4	91,0	1,3	92,9	3,2
9	0	No	5,4	87,7	98,6	91,2	1,3	93,2	3,2
10	0	No	5,9	77,3	98,7	91,4	1,3	93,3	3,2
11	0	No	6,3	69,0	98,8	91,5	1,3	93,4	3,2
12	0	No	6,8	63,7	98,9	91,6	1,3	93,4	3,2
13	0	No	7,2	60,0	99,1	91,6	1,3	93,5	3,2
14	0	No	7,7	56,8	99,1	91,7	1,3	93,5	3,2
15	0	No	8,1	55,0	99,1	91,7	1,3	93,5	3,2
25	0	No	11,6	45,9	99,4	92,0	1,3	93,5	3,1
50	0	No	19,0	44,1	99,5	92,1	1,3	93,9	3,1
5	1	No	3,6	100,0	97,1	89,9	1,3	88,6	3,3
5	2	No	3,6	100,0	97,2	89,9	1,4	87,8	3,3
5	3	No	3,6	100,0	97,2	90,0	1,4	87,2	3,3
5	4	No	3,7	100,0	97,4	90,1	1,4	87,1	3,2
5	0	Yes	3,5	100,0	96,9	89,7	1,4	88,4	3,3

Table 3: Experimental results of the case study for all scenarios as defined in Section 4.1. For all scenarios, a Poisson arrival distribution with λ=40 is used. The results are based on *n*=100 independent simulation runs, each with a warm-up phase of 800 cycles, followed by a run length of 2000 cycles.

Patients' Waiting Time		Schedule Stability		Hospital Efficiency		
Average	< 8 weeks	Afill	AinA	FPO	DPO	SDBO
[3.51, 3.53]	[100.00, 100.00]	[96.90, 96.98]	[89.68, 98.76]	[1.40, 1.41]	[87.66, 88.33]	[3.25, 3.26]

Table 4: confidence intervals (rounded to two decimals) for the basic scenario

Arrival distribution	SDBO FCFS	SDBO MSS basic scenario
Constant	6.29	3.20
Gamma (40,20)	6.41	3.23
Poison (40)	6.55	3.25
Gamma (40,80)	6.78	3.32
Gamma (40,160)	7.14	3.46

Table 5: comparison of the standard deviation of the bed occupancy levels (SDBO) between a FCFS approach and an MSS approach (basic scenario), for different arrival distributions with increasing variability. Expected arrival rate is 40 patients per week for all distributions.

Erasmus University Rotterdam

INSTITUTE OF HEALTH, POLICY & MANAGEMENT

Visiting address
Burgemeester Oudlaan 50
3062 PA Rotterdam

Postal address
P.O. Box 1738
3000 DR Rotterdam
The Netherlands

Telephone	+31 10 408 85 55
Fax	+31 10 408 90 94
Internet	www.bmg.eur.nl
E-mail	research@bmg.eur.nl

ERASMUS UNIVERSITEIT ROTTERDAM
INSTITUTE OF
HEALTH POLICY & MANAGEMENT