Due time driven surgery scheduling

Michael Samudra · Erik Demeulemeester · Brecht Cardoen · Nancy Vansteenkiste · Frank E. Rademakers

Abstract In many hospitals there are patients who receive surgery later than what is medically indicated. In one of Europe's largest hospitals, the University Hospital Leuven, this is the case for approximately every third patient. Serving patients late cannot always be avoided as a highly utilized OR department will sometimes suffer capacity shortage, occasionally leading to unavoidable delays in patient care. Nevertheless, serving patients late is a problem as it exposes them to an increased health risk and hence should be avoided whenever possible.

In order to improve the current situation, the delay in patient scheduling had to be quantified and the responsible mechanism, the scheduling process, had to be better understood. Drawing from this understanding, we implemented and tested realistic patient scheduling methods in a discrete event simulation model.

We found that it is important to model non-elective arrivals and to include elective rescheduling decisions made on surgery day itself. Rescheduling ensures that OR related performance measures, such as overtime, will only loosely depend on the chosen patient scheduling method.

We also found that capacity considerations should guide actions performed before the surgery day such as patient scheduling and patient replanning. This is the case as those scheduling strategies that ensure that OR capacity is efficiently used will also result in a high number of patients served within their medically indicated time limit. An efficient use of OR capacity can be achieved, for instance, by serving patients first come, first served. As applying first come, first served might not always be possible in a real setting, we found it is important to allow for patient replanning.

1 Introduction

It is a problem if patients wait longer for surgery than what is deemed to be optimal by their surgeons. In those cases, patients are said to have been served after the due time (DT) [57], which can pose a health risk. In one of Europe's largest hospitals, the University Hospital Leuven in Belgium, 34.6% of patients are served after their target DT. This is normal as a highly utilized OR department will sometimes suffer capacity shortage, occasionally leading to unavoidable delays in patient care.

Nevertheless, serving patients late should be prevented if possible, primarily from a medical standpoint, but also from a societal hidden cost perspective as patients in a worsened health condition are likely to require larger amounts of resources.

In order to improve the current situation, the lateness of patients had to be quantified and the primarily responsible mechanism, which is the patient scheduling process, had to be better understood. Drawing from this understanding, we implemented and tested realistic patient scheduling policies using a discrete event simulation model (DES). The results of the tests should

Michael Samudra, Erik Demeulemeester
Faculty of Economics and Business, KU Leuven
E-mail: samudra@kuleuven.be

Brecht Cardoen
Vlerick Business School

Nancy Vansteenkiste, Frank E. Rademakers
University Hospital Leuven

help surgeons and nurses to better understand the consequences of their patient scheduling related decisions.

In this text we describe some of the primary aspects and properties of the hospital's inpatient population, introduce the way patients are scheduled in reality and describe some of the major mechanisms that take place in the operating room (OR) department. We will therefore describe patient arrival patterns, the relationship between estimated and realized surgery durations, the applied rescheduling mechanisms on the day of surgery and the non-elective to OR allocation schema. Finally, we will introduce some of the manually applicable scheduling methods and show how they perform in the resulting simulation environment.

## 2 Problem description and literature review

The amount of time a patient can wait for surgery varies largely from case to case. It depends on many factors such as the general health condition of the patient, the speed at which the underlying disease is progressing, the endured pain level and the detrimental lifestyle effects.

One way to ensure that patients receive surgery within an acceptable time limit is to enforce waiting time targets, such as DT's. DT's can be set up by the authority of a larger geographic region such as a government (e.g., Australia and Canada [2, 4]) or can be defined by a lower level authority such as a hospital.

DT's were set up at the University Hospital Leuven by the surgeons of the hospital themselves and were determined on the basis of medical reasons. The DT is therefore a concept that has always existed and has been explicitly used, but has only been formalized recently. Formalizing it allows the hospital to use it as a benchmark criterion. Fig. 1 shows that a large part of the patient population is served before their DT and around one third of them is served after their DT. The figure is based on data covering the entire years 2012-2013 including all 13 disciplines (Table 1) that are served in the hospital's 22 inpatient ORs.

The DT is assigned to patients by the respective physician in charge. It is divided into 8 categories (Table 2) where categories 4 to 8 are used to classify electives and categories 1 to 3 are used to classify non-electives. The DT of elective patients is defined in weeks whereas the DT of non-electives, as they have to be served the latest within 24 hours after their admittance, is defined in hours.

As even the least urgent non-elective patients have to be served within 24 hours, there is no room for scheduling them in advance. Non-electives are therefore not planned and they are only included into the
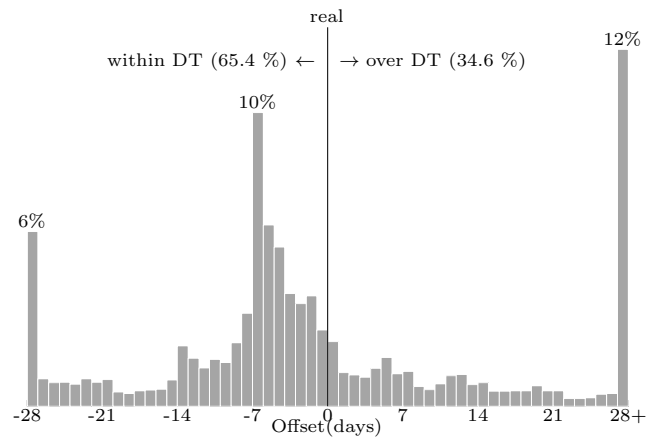
Fig. 1 The horizontal axis shows the number of days patients were served before or after their respective DT. The open-ended histogram shows that, for example, 10% of the patients are served exactly 7 days before their DT. The histogram does not cover those electives that have not been assigned a DT and thus do not need to be served within a time limit. Including them and assuming they are always served within their DT, the total percentage of patients served within DT changes to 76%.

Table 1 There are 13 disciplines served in the inpatient department.

GYN	Gynecology and obstetrics
Tx	Abdominal transplant surgery
ABD	Abdominal surgery
CAH	Cardiac surgery
NCH	Neurosurgery
ONC	General medical oncological
RHK	Plastic, reconstructive and cosmetic surgery
THO	Thoracic surgery
TRH	Traumatology
URO	Urology
VAT	Vascular surgery
MKA	Oral and maxillofacial surgery
NKO	Head and neck surgery

simulation model to test their impact on the execution of the elective schedule.

As Table 2 shows, the DT is defined as a time interval suggesting that it is best for a patient to get surgery only after a certain reference period. It might seem unreasonable to let patients wait unnecessarily, but it can be the case that they or their surgeons need time to prepare for the surgery. From a scheduling perspective, the end time of the interval is the determining factor.

The DT score of a discipline is calculated based on the weights associated to each DT category. The weights for DT categories 4 to 7 are 1, 1/2, 1/4 and 1/8 respectively. A weight of 0 is associated to DT category 8. The DT score of a discipline is the average DT weight assigned to their patients.

Table 2 The 8 DT categories used at the University Hospital Leuven.

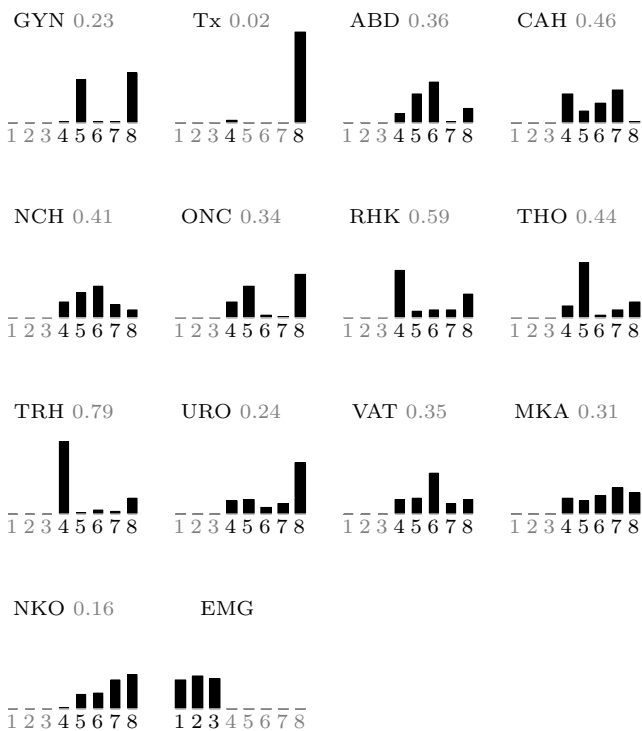|	Category	Target time
Non-elective	1	Instantly
	2	Up to 6 hours
	3	Today
Elective	4	1 week
	5	1 - 2 weeks
	6	2 - 4 weeks
	7	4 - 8 weeks
	8	no target time

Fig. 2 The distribution of DT categories varies largely across disciplines. The number in gray denotes the DT score. 'EMG' stands for non-elective.

Fig. 2 shows that both the DT score and the distribution of the DT categories is different for each discipline. For example, MKA covers an even spectrum of DT categories whereas, not surprisingly, for TRH the vast majority of patients is associated to DT 4 since wounds and injuries often need quick care. Correspondingly, TRH also has a high DT score.

The primary goal of our work is to increase the amount of patients served within their DT, thus served within the target time set by their surgeons. This goal can be achieved in three ways.

Firstly, it can be achieved by increasing capacity on the supply side by opening new ORs and hiring the additionally required personnel. Increasing existing OR capacities requires additional financial and spatial resources which in our setting are not readily available.

Secondly, it can be achieved by allowing more flexibility and using an open scheduling strategy, i.e., there is no Master Surgery Schedule (MSS) [5]. Allowing more flexibility in the schedule has advantages [56] and allows to better deal with occasional peaks in demand of single disciplines. Open scheduling is for the University Hospital Leuven, as for many other hospitals, not an option as it is important for the hospital to maintain a periodic and repetitive schedule. This allows surgeons to block certain weekdays for surgery while keeping other fixed days free for consultation, scientific work and teaching.

Thirdly, as considered in this research, it can be achieved by improving patient scheduling practices. We tested methods that are manually usable and thus do not involve a computer. This is done as surgeons (and nurses) at the University Hospital Leuven, and in Flanders (Belgium) in general [7], typically create patient schedules manually. Moreover, surgeons schedule their patients individually and therefore generally will not coordinate their schedules amongst each other.

In the literature, various other terms are used to identify typical OR related scheduling problems. Magerlein and Martin [32] distinguish between advance and allocation scheduling. Advance scheduling is the process of fixing a surgery date for a patient (patient to date assignment problem), whereas allocation scheduling determines the OR and the sequence (and starting time) of the surgeries.

Within advance scheduling, another distinction can be made between dynamic and static scheduling. In surgery scheduling, dynamic refers to a setting where a patient is given a surgery date at consultation time, whereas in the static case, surgeries are put on a waiting list. A subset of patients from this list are then scheduled at once at a certain point in time, e.g., at the end of each week. Therefore, in the static case patients to be scheduled are (assumed to be) known, whereas in the dynamic case, future arriving patients are (assumed to be) unknown.

These two problems are handled differently in the literature from a method perspective. For the static problem, the hospital can use an algorithm, typically an MP or improvement heuristic, that provides a schedule, i.e., the algorithm substitutes the scheduler. For the dynamic case, the hospital is usually using policies that are typically derived from a Markov decision process (MDP) or a DES which the scheduler (e.g., assistant of surgeon) should consider in daily practice. Dynamic scheduling can be used in settings where waiting lists are rare and waits are relatively short such as in our setting and in Belgium in general [49, 59]).

Another common distinction is made between block and open scheduling. In block scheduling, blocks (i.e., a combination of an OR and a day) are preallocated to a discipline or to a surgeon group. Surgeons are only allowed to book cases into blocks preallocated to them or their group. In open scheduling, surgeons are not limited by a block schedule and are therefore less restricted in choosing an OR for surgery. At the University Hospital Leuven block scheduling is used.

There are also various other aspects that one can use to distinguish between different surgery scheduling settings. For instance, Cardoen et al. [6] use descriptive fields, such as used performance measures and applied research methodology to categorize the literature. Alternatively, in the literature review of Guerriero and Guido [22], a selected number of articles are categorized according to the three commonly used hierarchical decision levels: strategic, tactical and operational.

In this paper we primarily focus on the patient to date assignment problem [8] which is heavily researched in the literature (we identified 64 journal articles written between 1999 and 2014 only). We filtered these articles for those that model arrivals stochastically (as we think this is important in dynamic scheduling). This left us with the following set of articles: [1, 9, 10, 11, 15, 20, 27, 28, 31, 35, 40, 42, 47, 52].

This list still contains static problems. Excluding these leaves us with a surprisingly small set of articles: [9, 10, 11, 20]. The fact that more methods are available for static surgery scheduling does not mean that dynamic scheduling is not popular in the health care literature as much has been published in the domain of appointment scheduling.

Unfortunately, as in appointment scheduling many strong assumptions are made, their results are not guaranteed to hold for our and other inpatient surgery settings in general. For example, in appointment scheduling it is mostly assumed that appointments have the same estimated durations (i.e., patients are scheduled to appointment slots of equal size). This assumption holds for some outpatient settings (e.g., imaging), but is generally not true for outpatient surgery and even less true for inpatient surgery. There are methods from appointment scheduling where this assumption is relaxed to some degree. For example, Gocgun and Ghate [20] use an MDP where surgeries of various fixed sizes can be allocated to a block (e.g., 4x1h surgeries and 2x2h surgeries). This allows to use the method more flexibly compared to methods with only one slot size.

In order to be able to capture the complexity inherent to surgery scheduling, we chose not to use an analytical approach (e.g., MDP) and instead use DES. DES is also used by Dexter et al. [9, 10] in a similar setting. They show that in order to maximize the utilization of the OR surgery blocks should be allocated on the basis of expected total hours of elective cases. Electives should then be scheduled into the first slot available within four weeks, otherwise they should be scheduled into overtime. There are also differences between their and our setting. At the University Hospital Leuven, while a highly utilized OR department is desired, it is not the primary target. A highly utilized OR can mean a large amount of overtime. Too much overtime is at this case hospital a problem as it means that nurses will have to work longer and might therefore not be allowed to work the next day (i.e., an overutilized OR department on one day can result in an underutilized OR department the next day). Consequently, unlike in Dexter et al. [9], we focus on objectives that primarily consider the DT.

Unlike for dynamic surgery scheduling, various solution approaches are provided in the literature for the static case. For example, Adan et al. [1] describe routines that can be followed step by step by schedulers. These routines are provided for three different problems: (1) determining the amount of optimal OR capacity to reserve for certain patient types, (2) assignment of patients to those previously reserved OR capacities, and (3) guidance of the decision making process on the day of the surgery.

The setting described in Adan et al. [1] is different from ours in several aspects. For instance, in their setting, electives are scheduled into regular OR time, whereas non-electives are scheduled into regular and slack time. In our setting, non-electives are not being scheduled since they either need to be served immediately (DT category 1), in which case scheduling is obsolete, or they can wait (DT categories 2 and 3), in which case they generally will, as in Azari-Rad et al. [3], be served in an OR of the corresponding discipline after the last elective of the day has been served.

Slack time is also used by Hans et al. [23] where the portfolio effect is used to minimize the risk of overtime. In their setting, the risk of overtime depends on the variance of the surgery durations. Grouping surgeries in a way that the total variance of the surgery durations assigned to ORs is decreased will result in a lower risk of overtime.

An example of a patient to date and a patient to OR assignment problem in an open scheduling setting is introduced by Fei et al. [16]. In such a setting it is generally also necessary to determine the patient sequence [17, 18]. A similar problem is solved by Lamiri et al. where patients are assigned to dates [27, 29] and also to ORs [26, 28].

A system where similarly to ours, also patient priorities play a role is introduced by Min and Yih [35, 37]. In their model, the priority of patients will translate into a cost value that is associated to not being scheduled in the next future period. If a patient is scheduled, an overtime cost may occur. The authors also introduce techniques that solve the patient to OR assignment problem including stochasticity in surgery durations, length of stay and non-elective arrivals [36].

We focus on the surgery to date assignment step. We do not sequence and do not determine the start time of surgeries. Those two factors are not important in our setting as elective patients are available the whole day and surgeons usually "own" an OR for the entire day. Therefore a surgery starts directly once the preceding surgery is finished.

Additionally, we heavily focus on aspects that are important to get realistic results. This is true with regards to both the developed model and the tested scheduling methods. The model is realistic as we included all the aspects that we found to have a major effect on the results. This includes modeling aspects that relate to patient attributes (e.g., arrival, duration), to the structure of the setting (e.g., block assignment schema) and the processes (e.g., rescheduling, non-elective allocation schema). Also the tested scheduling methods are realistic as they reflect considerations or processes that are also important in reality.

Components of the model were created on the basis of hard data. For aspects that were not covered by the data, we relied on the insights of our contacts. They consist of a mix of people from the hospital that together have all the necessary insights. This includes, among others, the head surgeon, the head nurse, the responsible of the bed allocations and people from capacity management and the data gathering group.

Our contribution to the existing literature is therefore twofold. Firstly, we add to the sparsely addressed literature of dynamic advance surgery scheduling with stochastic arrivals. Secondly, we created a comprehensive model of the real hospital setting. We therefore modeled aspects of the real setting that are often not considered in the literature, but which we think are important to include.

3 Method

We will refer to different scheduling policies or methods as scheduling factors. The combination of those factors creates scenarios, which we then test in a DES model.

3.1 Model

We found that only a DES model is able to realistically capture all the aspects of the University Hospital Leuven's scheduling setting that we deemed to be important. The DES model incorporates all the aspects of the surgery setting of the University Hospital Leuven that we found to be vital. We included aspects that relate to the way surgeries are scheduled and replanned before the surgery date and are rescheduled on the surgery day itself. We also replicated the functions of the OR department. This includes, for instance, an implementation of the non-elective to OR allocation schema.

We created the DES model in MATLAB and Simulink. The combination of MATLAB with Simulink yields an environment where generic simulation building blocks can be used and, as part of the simulation model, regular MATLAB code can be executed. Simulation blocks are used to model, for instance, the surgery process in the OR, while MATLAB code is used to create models of more complex decision making processes, such as the process of patient scheduling and patient rescheduling.

We analyzed and imitated the real mechanisms encountered at the hospital. We made a minimal amount of modeling assumptions and used real data as the basis of all submodels. In cases where the data did not reveal enough about a process, we were helped by our contacts at the hospital who provided us with the missing knowledge.

The attributes of patients generated in the model realistically reflect the attributes of the inpatient population of the hospital. Patient attributes are: surgeon ID, arrival rate for each weekday, estimated and realized surgery duration and DT category. Discipline related attributes are: surgery start time bias (for the first surgeries of the day) and turnaround time.

The statistics for patient attributes are measured for each discipline separately. In the model, all patient attributes are generated on the basis of empirical distributions. Exceptions are the non-elective inter-arrival times and the realized and the estimated surgery durations. Non-elective inter-arrival times, for modeling purposes, are assumed to follow the exponential distribution for a given period. A period depends on the weekday and the daytime (daytime: 6 am to 10 pm, nighttime: 10 pm to 6 am). The relation between the estimated and the realized surgery durations is modeled using a statistic that is based on copulas [55].

3.2 Factors

The DT is one of the major scheduling related concepts at the University Hospital Leuven. We therefore tested how different scheduling factors impact it.

In our model, we imitate the reality of the hospital where patients are scheduled to a final surgery date during their consultation session. The surgeon or the administrative people with the input from the surgeon, find a suitable date and OR without a scheduling algorithm. Only those dates are considered on which the surgeon is assigned an OR. In the simulation model, as in reality, we ensure that, firstly, only patients associated to the same surgeon can be assigned to a particular OR and, secondly, a surgeon can only be assigned to one OR a day.

At the University Hospital Leuven an OR can be entirely filled up but is preferably not overbooked. However, there will be disciplines that occasionally overbook for a few hours. This is particularly true for CAH, NCH, THO, TRH, URO and NKO. These disciplines are in the model allowed to overbook, CAH by 2 hours and the remaining five by 1 hour. All other disciplines cannot overbook, i.e., the sum of the expected surgery durations assigned to their ORs can sum up to its total capacity (9 hours) but cannot exceed it.

Booking rules can vary from hospital to hospital. At some hospitals, ORs may never be fully booked or, conversely, can be overbooked. For example, at the Erasmus Medical Center in the Netherlands ORs are not fully booked and slack time is included. This ensures that the probability of overtime stays below a certain level [23].

In our setting, a surgery schedule is not necessarily fixed as surgeries can be replanned before the day of their surgery. Surgery replanning to earlier surgery dates can, for instance, be used to improve the usage of ORs. In reality, this is applied to 5.2% of the total patient population. Other reasons why surgeries are brought forward in the schedule are the worsening health condition of the patient or hospital related logistic reasons. We will focus on the capacity related advantages and investigate whether patient replanning, by utilizing unclaimed free short-term OR capacity, can improve OR usage. As the hospital generally tries to avoid excessive replanning, we also investigate whether the unused OR capacity can be filled up with new arrivals.

We grouped the different aspects of the patient scheduling process into three factors (Table 3). The first factor tests the use of the first come, first served (FCFS) strategy which assigns patients to the earliest possible surgery date regardless of their actual DT. The second factor tests the use of pushing lower urgency patients into the future leaving capacity free for higher urgency DT categories. The third factor tests the use of filling up unclaimed short-term free capacity. This is tested in two ways, firstly, by using patients arriving one day in advance, i.e., patients are allowed to be scheduled to exactly one day after their arrival, and secondly, by replanning patients from future dates to earlier dates. Replanning is done before any new elective arrival is registered for the current day.

Factor 1 (F1) is used to investigate whether it is beneficial to allow patients, up to certain DT categories, to be served FCFS. As table 3 shows, the factor can take 6 values: (1) none of the patients are served on a FCFS basis, (2) only DT category 4, (3) DT categories 4 and 5, (4) DT categories 4 to 6, (5) DT categories 4 to 7 or (6) DT categories 4 to 8 (all) patients are served FCFS. The factor allows patients of the included DT categories to be served as quickly as possible. Indirectly it also means that the patients of all DT categories served FCFS are treated equally. For example, if FCFS applies to patients up to DT category 6, then, from a scheduling perspective, DT category 5 and 6 are regarded to be equally urgent as DT category 4.

Similarly as done by Vijayakumar et al. [60] and Niu et al. [38] patients served FCFS will be assigned to the first date that has a suitable open OR available. In case that such an OR is not available, a new OR is opened. In our model, surgeries can only be allocated to ORs that are assigned to the corresponding discipline and to the corresponding surgeon. A surgery can be allocated to a new empty OR if the OR is assigned to the respective discipline in the MSS. The newly opened OR will be assigned to the surgery's surgeon and only accept those additional future surgeries that belong to the same surgeon.

Factor 2 (F2) is used to postpone less urgent surgeries, thereby creating short-term buffer capacity that can be used by more urgent patients. There are three strategies: schedule patients into the early, center or late part of their DT interval. With the early strategy, patients are assigned into the closer end of their DT interval. This is similar to FCFS with the restriction that patients can only be served after a certain reference period. With the center strategy, patients are scheduled as close to the middle of their DT interval as possible, i.e., the temporal distance between the selected date and (DT end + DT start) / 2 is minimized. With the late strategy, patients are scheduled into the end of their DT interval, that is, patients are served as late as possible while still within their DT. If there is no such date available, then a date after the patient's DT is chosen.

It is interesting to explicitly incorporate the DT into a scheduling strategy as serving patients closer to their due date is a concept that can intuitively feel advantageous to surgeons. This approach is also tested by Rizk and Arnaout [44].

Factor 3 (F3) is used to quantify the benefits of filling up unclaimed free short-term capacity. This is capacity that in the morning of the preceding day is still shown to be unclaimed and is therefore regarded to be in danger of being wasted. For example, if Wednesday morning the OR plan for Thursday shows 5 hours of unclaimed capacity, then we say that 5 hours of OR capacity is in danger of being wasted. We will refer to this type of capacity as next day free capacity. Next day free capacity can be occupied by patients from two different sources: firstly, new arrivals and, secondly, replanned patients. Replanning works similarly to a waiting list, where the replanning policy determines which patients to pick first from the list [33].

Factor 3, similarly to factor 1, applies to certain DT categories. It is applied to each discipline separately. In case of replanning, it is used in combination with the best-fit strategy. This means that from the list of eligible patients assigned to future dates (waiting list), those patients are replanned that make best use of the available free capacity. We use a replanning routine that is likely to be most often used in reality. We first replan the patient with the longest estimated surgery that still fits the next day free capacity. The second patient chosen will need to fit the remaining capacity. We continue this process until the left over free capacity does not allow to accommodate any further patient.

Next to best-fit, we also implemented a patient selection strategy that is based on an accumulating priority queue (APQ). In the APQ, patients accumulate priority as a linear function of their time in the queue and their priority [25, 50], i.e., their waiting time and their DT. The weight v_i associated to each patient is therefore:

$$v_i = \frac{(s_i - a_i)}{dt_i} \tag{1}$$

where a_i is the arrival day, s_i is the surgery day and dt_i is the due time in days of patient $i \in I$.

4 Hospital setting, modeling assumptions and validation

This section contains a description of the hospital setting. We describe patient attributes, the MSS, the non-elective allocation schema and the way patients are rescheduled. In order to validate the simulation model,

Table 3 As the table shows, each factor can take several values. The combination of the three factors forms a scheduling scenario. For example, a scheduling scenario is to serve patients up to DT category 5 on a FCFS basis, schedule the rest (DT categories 6 to 8) to the center of their respective DT interval and fill up next day capacity with the APQ (Eq. 1)

	Factor	Values
F1	FCFS	None, $\leq$ DT 4/5/6/7/8
F2	Into DT interval	Early, center, late
F3	Next day	None
	Fresh arrivals	$\leq$ DT 4/5/6/7/8
	Replanning patients	$\leq$ DT 4/5/6/7/8, APQ

we compare OR related performance measures to actual measurement data from the University Hospital Leuven.
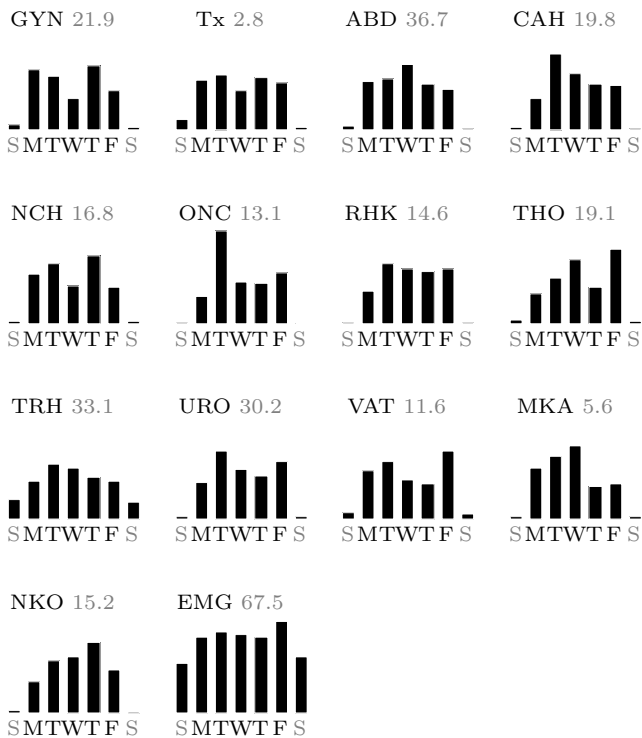
4.1 Patient arrivals

The arrival time of elective patients is the time point when their surgeon determines the need for surgery. This generally happens on weekdays anywhere during daytime. The arrival time of non-elective patients represents the time point when they are physically registered at the hospital. This can happen at any day and at any hour.

We model elective patient arrivals on the basis of a statistic that is based on rates, e.g., 5 CAH patients request surgery on a Monday (Fig. 3). For non-electives, a statistic is used that is based on inter-arrival times. This defines an exact time instance, e.g., a non-elective CAH patient arrives Monday at 2.21 pm.

Table 4 shows that the mean number of electives registered weekly for surgery is 240.35 with a standard deviation of 32.53 (column $\cup$ in the table). The average number of arrivals for an elective discipline is 18.49 with a standard deviation of 5.6 (column μ in the table).

As Table 4 shows, patient arrival numbers are highly variable. This is true for week to week (e.g., first week to second week of the year), day to day (e.g., Monday to Tuesday) and weekday to weekday (e.g., Tuesday to Tuesday) based comparisons. It is especially surprising that the weekday to weekday variation of patient arrivals is high. This might be counterintuitive as, given that surgeons have consultation times on a fairly regular basis (e.g., every Monday), one could assume that patient arrival numbers for the same weekday are more stable.

Interestingly, the week to week arrival variability differs strongly between disciplines. For MKA (and Tx)

GYN 21.9   Tx 2.8   ABD 36.7   CAH 19.8

NCH 16.8   ONC 13.1   RHK 14.6   THO 19.1

TRH 33.1   URO 30.2   VAT 11.6   MKA 5.6

NKO 15.2   EMG 67.5

Fig. 3 The average arrival rate of the 13 elective disciplines and non-electives. The number in gray represents the average weekly number of arrivals. The height of the column represents the % of patients that arrived on that day from the weekly volume.

it is very high in relation to the mean. This explains the large CV. Consequently one might wonder whether it is possible to provide timely service to MKA patients. Fortunately, MKA patients are generally not urgent (DT score of 0.31, Table 2). This allows to spread out arrivals from weeks with high loads to weeks with lower loads. The same could not be done by TRH as most of their patients must be served within 1 week. Fortunately, TRH has one of the most stable patient inflows and will therefore less frequently encounter weeks with very high loads.

One could assume that in reality disciplines with a patient mix that contains higher urgency patients or a larger arrival variability would generally provide less timely service to their patients when compared to the rest of the disciplines. Interestingly, we did not find any indications in the data that would support this theory.

It is worth to note that, in case a discipline covers a large population of DT 4 patients, not only the weekly, but also the daily arrival variability is important. Consequently, in the model, both discipline dependent weekly and also daily arrival variability needs to match reality. We ensure this by generating patients in two steps. In the first step, we determine for each discipline the total number of weekly arrivals. In the second step, the number of arrivals for each weekday is determined (Monday to Friday). This is done by selecting a realization of a week from a pool. The weeks in the pool were pre-generated using the empirical distributions observed in reality.

As a consequence of this two step procedure, for all disciplines, the difference between the model and the reality of arrival means and standard deviations are minimal (Table 4). A difference is present only if the union of all patients is considered. This difference can be explained by holidays. In reality, holidays in a week result in lower arrival numbers for all disciplines, i.e., weekly arrival numbers for disciplines correlate. In the model, holidays affect each discipline independently, therefore weekly arrival numbers for disciplines do not correlate. In regards to the results, this discrepancy will not matter because patients of different disciplines are scheduled into their own OR capacity, i.e., while the individual arrival variability of each discipline is important, this is less the case for the combined one.

In Table 4 it is interesting to observe that arrival means are generally not equal to their variance. This is the case for most disciplines, for the averages across disciplines (denoted in the table by μ) and for combined elective arrivals (denoted in the table by $\cup$). Furthermore, this is true for weekly arrival numbers, daily arrival numbers and weekday specific arrival numbers (Fig. 3). The fact that the arrival means and variances are not equal means that the arrivals, contrary to what is sometimes assumed in the literature, do not follow the Poisson distribution. Interestingly, this is even true for non-elective weekly arrival numbers. The two elective categories that seem to be exceptions are Tx and NKO.

A factor that has an influence on arrival variance are holidays. As Table 4 shows, the number of arrivals (first row) on holidays is lower than on normal days, but is by far not zero, i.e., patients are also on holidays scheduled for surgery. Important to note is the fact that excluding holidays will decrease the arrival variability only to a limited extent (e.g., the average standard deviation drops from 35.0 to 31.5). That is, excluding holidays will not yield arrival patterns that follow the Poisson distribution.

4.2 Non-electives

Every week around 70 non-elective patients, using around 160 hours of OR time (Fig. 6), get surgery at the hospital. This means that, if scheduled into regular OR time, they would occupy 3 to 4 ORs a day. This is a large number which explains their fundamental impact

Table 4 As the table shows, the arrival model will produce results that are close to reality (Δ values are small). The table also shows that arrivals generally do not follow the Poisson distribution (mean and variance are not equal). Weekly means are calculated on the basis of the 104 weeks of the years 2012 and 2013. Daily means are calculated on the basis of 520 days (5*104) whereas the values of weekdays are calculated on the basis of the corresponding 104 weekdays (e.g., all Tuesdays in 2012 and 2013). In the table, as an example, only Tuesday is shown. The mean value of all elective disciplines is denoted by 'μ', whereas the value considering electives in general is denoted by '$\cup$'.

			Non-Elective	GYN	Tx	ABD	CAH	NCH	ONC	RHK	THO	TRH	URO	VAT	MKA	NKO	μ	$\cup$
Weekly arrivals	real	μ	**67.6**	**21.9**	**2.8**	**36.7**	**19.8**	**16.8**	**13.1**	**14.6**	**19.1**	**33.1**	**30.2**	**11.6**	**5.6**	**15.2**	**18.5**	**240**
		var	118	38.8	2.9	84.7	50.1	36.7	15.9	39.1	39.4	51.8	46.1	20.8	9.9	18.5	35.0	1058
		σ	10.9	6.2	1.7	9.2	7.1	6.1	4.0	6.3	6.3	7.2	6.8	4.6	3.1	4.3	5.6	32.5
		CV	.16	.28	.61	.25	.36	.36	.30	.43	.33	.22	.22	.39	.56	.28	.35	.14
	model	μ	**67.4**	**22.1**	**2.7**	**36.6**	**19.6**	**16.8**	**13.4**	**14.9**	**19.2**	**32.8**	**30.4**	**11.6**	**5.7**	**15.0**	**18.5**	**241**
		var	75.1	40.5	2.7	85.4	45.4	38.4	16.5	41.6	39.7	51.9	44.8	21.8	9.8	18.0	35.1	481
		σ	8.7	6.4	1.6	9.2	6.7	6.2	4.1	6.5	6.3	7.2	6.7	4.7	3.1	4.2	5.6	21.9
		CV	.13	.29	.60	.25	.34	.37	.30	.43	.33	.22	.22	.40	.55	.28	.35	.09
	Δ	μ	**.20**	**-.16**	**.07**	**.04**	**.14**	**-.06**	**-.27**	**-.29**	**-.06**	**.23**	**-.21**	**-.07**	**-.11**	**.18**	**-.04**	**-.58**
		var	43.1	-1.7	.20	-.68	4.6	-1.7	-.59	-2.5	-.24	-.08	1.3	-1.1	.06	.47	-.14	577
		σ	2.2	-.13	.06	-.04	.34	-.14	-.07	-.20	-.02	-.01	.09	-.11	.01	.05	-.01	10.6
		CV	.03	.00	.01	.00	.01	-.01	.00	.00	.00	.00	.00	-.01	.01	.00	.00	.04
Daily arrivals	real	μ	**10.7**	**4.3**	**.53**	**7.2**	**3.9**	**3.3**	**2.6**	**2.9**	**3.8**	**5.7**	**5.9**	**2.2**	**1.1**	**3.0**	**3.6**	**46.5**
		var	11.6	6.1	.59	10.8	6.7	4.6	4.3	4.3	6.1	7.9	8.3	2.9	1.4	3.4	5.2	122
		σ	3.4	2.5	.77	3.3	2.6	2.2	2.1	2.1	2.5	2.8	2.9	1.7	1.2	1.8	2.2	11.0
		CV	.32	.58	1.5	.46	.66	.65	.79	.71	.66	.49	.49	.76	1.1	.61	.72	.24
	model	μ	**10.5**	**4.3**	**.51**	**7.2**	**3.9**	**3.3**	**2.7**	**3.0**	**3.8**	**5.7**	**6.0**	**2.2**	**1.1**	**3.0**	**3.6**	**46.7**
		var	11.0	6.6	.56	11.6	7.5	5.3	4.5	5.1	6.8	8.0	8.3	3.3	1.5	3.4	5.6	97.3
		σ	3.3	2.6	.75	3.4	2.7	2.3	2.1	2.3	2.6	2.8	2.9	1.8	1.2	1.9	2.3	9.9
		CV	.31	.59	1.5	.47	.70	.69	.80	.76	.69	.50	.48	.81	1.1	.62	.74	.21
	Δ	μ	**.16**	**-.03**	**.02**	**.01**	**.02**	**-.02**	**-.06**	**-.06**	**-.01**	**.03**	**-.05**	**-.02**	**-.02**	**.03**	**-.01**	**-.15**
		var	.65	-.46	.03	-.80	-.85	-.66	-.27	-.87	-.63	-.09	-.02	-.39	-.12	-.01	-.40	24.7
		σ	.10	-.09	.02	-.12	-.16	-.15	-.06	-.20	-.12	-.02	.00	-.11	-.05	.00	-.08	1.2
		CV	.00	-.02	-.01	-.02	-.04	-.04	-.01	-.05	-.03	-.01	.00	-.04	-.02	-.01	-.02	.03
Weekday arrivals (TUE)	real	μ	**10.8**	**4.6**	**.60**	**7.3**	**5.9**	**4.0**	**4.8**	**3.5**	**3.4**	**7.0**	**7.9**	**2.6**	**1.4**	**3.1**	**4.3**	**56.1**
		var	12.5	5.8	.63	9.1	8.8	4.6	5.4	4.2	3.6	8.4	7.9	2.9	1.7	2.8	5.1	106
		σ	3.5	2.4	.79	3.0	3.0	2.1	2.3	2.0	1.9	2.9	2.8	1.7	1.3	1.7	2.2	10.3
		CV	.33	.52	1.3	.41	.50	.54	.48	.59	.56	.41	.35	.66	.96	.54	.61	.18
	model	μ	**10.6**	**4.7**	**.62**	**7.2**	**6.0**	**4.1**	**4.9**	**3.5**	**3.3**	**7.0**	**7.7**	**2.5**	**1.4**	**2.9**	**4.3**	**55.9**
		var	10.2	6.2	.69	10.4	11.4	5.4	5.3	4.3	3.9	8.2	8.5	3.3	1.8	2.4	5.5	76.7
		σ	3.2	2.5	.83	3.2	3.4	2.3	2.3	2.1	2.0	2.9	2.9	1.8	1.3	1.6	2.2	8.8
		CV	.30	.53	1.4	.45	.56	.56	.47	.59	.61	.41	.38	.71	.96	.54	.62	.16
	Δ	μ	**.19**	**-.10**	**-.02**	**.09**	**-.06**	**-.15**	**-.08**	**-.03**	**.14**	**-.01**	**.22**	**.03**	**-.01**	**.18**	**.01**	**.20**
		var	2.3	-.42	-.06	-1.3	-2.6	-.76	.08	-.12	-.32	.22	-.55	-.42	-.05	.39	-.45	29.1
		σ	.34	-.09	-.03	-.21	-.41	-.17	.02	-.03	-.08	.04	-.10	-.12	-.02	.12	-.08	1.5
		CV	.03	-.01	-.01	-.03	-.06	-.02	.01	.00	-.05	.01	-.02	-.05	-.01	.01	-.02	.03

on the hospital's OR department. In order to realistically model this impact, we analyzed both their arrival patterns and the discipline dependent way they are allocated to ORs (Fig. 4).

Non-electives arrive with the highest rate during daytime on weekdays. We call those time intervals high impact periods as this is also the time when non-electives have the largest impact on the elective schedule.

In the DES model, we explicitly model high impact periods, i.e., non-elective inter-arrival times will depend on the day of the week (Fig. 3) and the time period of the day. There are two time periods, (1) daytime is between 6 am and 10 pm and (2) nighttime is between 10 pm and 6 am. Arrival ratios will be around 3.4 times higher during daytime than nighttime.

Another important component of the model is the discipline dependent non-elective OR allocation
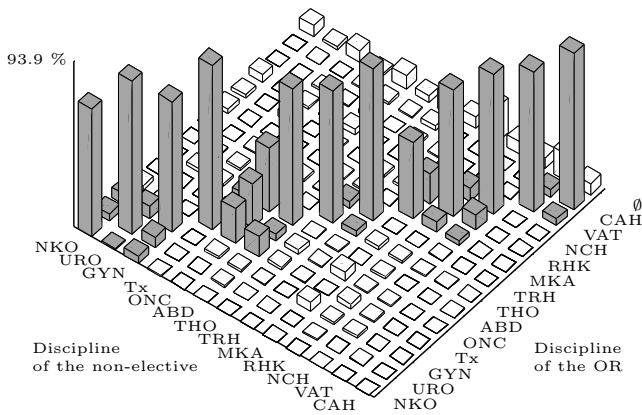
Fig. 4 During high impact periods, non-electives are generally assigned to an OR that serves electives of the corresponding discipline (diagonal columns). Occasionally, non-electives can be served in ORs that are not assigned to any discipline (marked with ∅). Disciplines in the figure are grouped on basis of their cluster (in gray). A cluster contains 4-6 ORs that form a physical unit. The figure shows that non-electives can sometimes enter ORs that serve a discipline the same cluster.

schema. As shown in Fig. 4, non-electives of all DT categories are during high impact periods generally assigned to an OR of the corresponding discipline. For example, an open wound patient brought to the hospital is generally assigned to an OR that is occupied by electives from TRH. Non-elective ONC patients are the only exception to this rule as they are frequently served in ORs allocated to ABD or Tx.

In the DES model, non-electives can during high impact periods generally only enter an OR that serves patients of their discipline. An exception is made for MKA, ONC and Tx as there will be weekdays on which they have no OR assigned to them. In those cases MKA patients are assigned to empty ORs. Tx can always occupy OR number 7 even if the OR was originally closed on the day, whereas ONC patients can enter ORs of ABD and Tx. Those exceptions are based on our findings in the hospital data and thus imitate the real practice.

In the model, we also distinguish between DT category 1 and DT categories 2 and 3. DT category 1 patients have to be served immediately and are assigned to the next possible suitable open OR that serves their discipline, that is, at the next break-in-moment [30]. Contrarily, DT category 2 and 3 patients will be left to the end of the elective schedule. This is also often happening in reality and serves the interest of the surgeons as it allows them to finish all their electives before any non-elective is started.
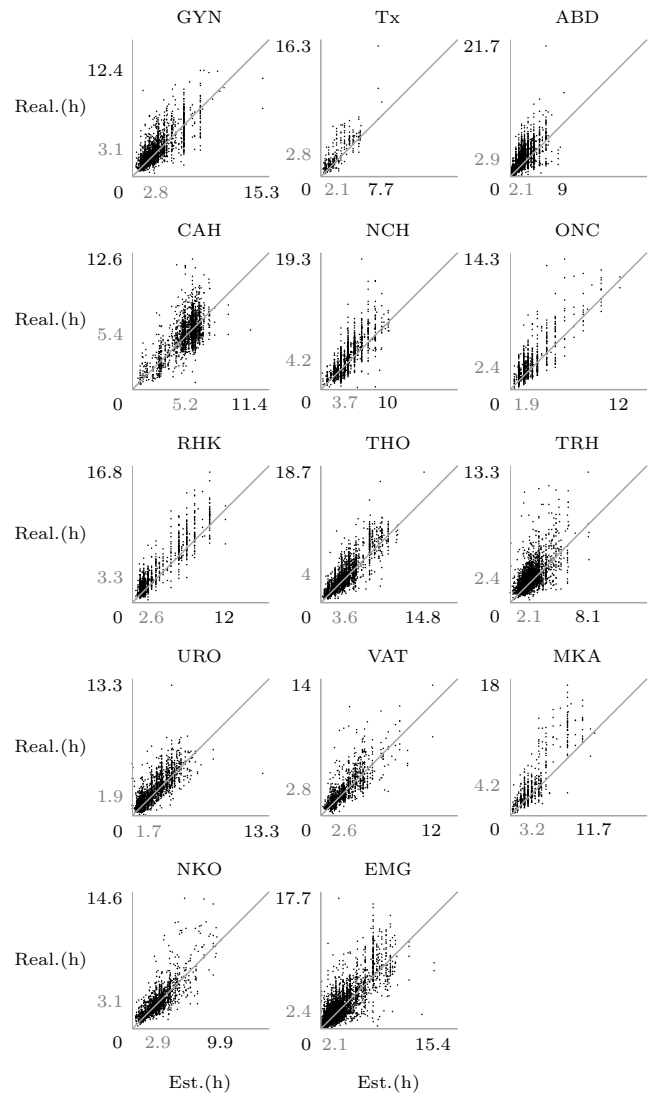
Fig. 5 The first axis represents surgery duration estimates, whereas the second axis represents their realization. Points on the same vertical line are either surgeries of the same type or they are discrete values that were selected manually by surgeons (e.g., 1 hour and 30 min). The numbers in gray represent the mean values.

4.3 Surgery duration

The surgery duration of a patient is defined as the time that elapsed between the moment the patient is rolled into the OR and the time when the patient leaves the OR (Fig. 5). It does not include cleaning time. If the patient is in the OR, it includes the setup time. Generally, if the setup time is specific to the patient, then it is included into the surgery duration.

The estimated surgery duration (Table 5), suggested to the surgeon, is based on the mean of the realized surgery durations of previous similar OR sessions. The surgeon can then accept or overrule this value.

Table 5 Surgery durations can often not be accurately estimated (high standard deviation of the error) and generally, as in Travis et al. [54], are underestimated. This is a problem as large estimation errors lead to OR overtime, case cancellations and generally decreased efficiency of OR resources [58].

| | | Non-Elective | Elective | | | | | | | | | | | | | μ | $\cup$ |
			GYN	Tx	ABD	CAH	NCH	ONC	RHK	THO	TRH	URO	VAT	MKA	NKO		
Realized	μ	**2.4**	**3.1**	**2.8**	**2.9**	**5.4**	**4.3**	**2.4**	**3.3**	**4.0**	**2.5**	**1.9**	**2.8**	**4.2**	**3.1**	**3.3**	**3.2**
	σ	*2.0*	*1.8*	*2.0*	*1.6*	*1.7*	*2.4*	*1.9*	*2.8*	*2.5*	*1.3*	*1.4*	*1.8*	*3.4*	*1.8*	*2.0*	*2.1*
	CV	.86	.56	.70	.54	.32	.55	.78	.86	.62	.53	.75	.63	.80	.57	.63	.67
Est.	μ	**2.1**	**2.8**	**2.1**	**2.1**	**5.2**	**3.8**	**1.9**	**2.6**	**3.6**	**2.1**	**1.7**	**2.6**	**3.2**	**2.9**	**2.8**	**2.7**
	std	1.8	1.6	1.4	1.1	1.3	1.7	1.5	2.4	2.1	.89	1.2	1.5	2.1	1.5	1.6	1.8
	CV	.84	.57	.68	.52	.25	.45	.77	.93	.59	.42	.71	.59	.67	.49	.59	.65
Δ	μ	**.29**	**.30**	**.70**	**.78**	**.16**	**.50**	**.48**	**.69**	**.46**	**.33**	**.18**	**.20**	**1.0**	**.15**	**.46**	**.42**
	std	1.2	1.2	1.2	1.2	1.3	1.4	1.0	1.1	1.3	1.0	.77	1.1	1.9	1.1	1.2	1.2
	CV	4.0	4.0	1.7	1.5	8.0	2.9	2.1	1.6	2.9	3.1	4.2	5.7	1.9	7.3	3.6	2.8

Each discipline performs different types of surgeries. Each of those surgery types is assigned a unique identifier that generally contains an ICD-9 code and a local component. ICD-9 codes on themselves can be too restrictive to accurately describe a procedure and thus need this additional local component. Surgeries with the same identifier will represent similar procedures and will consequently have a similar length.

As Table 6 shows, the log-logistic distribution provides, starting from the tested parametric distributions, the best fit on surgery types. In the literature sometimes the log-normal distribution is used as it provides a better fit than the normal distribution [51]. Also in our setting, the log-normal distribution clearly dominates the normal distribution. Interestingly, however, the log-logistic distribution outperforms both of them. Firstly, the log-logistic distribution fits all of the surgery types, whereas the log-normal distribution fits 97.7% of the surgery types. Additionally, based on the AIC criterion, the log-logistic distribution provides, amongst the tested distributions, the best fit in 31.8 % of the cases. For the log-normal distribution this is true for 2.7 % of the cases whereas the normal distribution never provides the best fit.

Despite the fact that surgery types seem easy to work with, there is a factor that prevents their use. We will often not have a sufficiently large enough sample size to reliably estimate the parameters of a distribution. The problem would remain if we would analyze more than two years of data. A larger total sample size would likely include new unseen surgery types which might again have a low sample count.

Because of the aforementioned factors, we model surgeries on a higher level, namely on the level of the discipline (Fig. 5). This avoids the problem of low counts, but unfortunately it also introduces a new prob-

Table 6 The fit of a distribution on a surgery type is evaluated by the Kolmogorov-Smirnov test. For each surgery type, a ranking is created using the Akaike information criterion (AIC). Only those surgery types were included into the analysis that are performed at least 10 times during the years 2012-2013. This covers 78% of all surgeries.

Distribution	Fits	Best fit (AIC)
log-logistic	100	31.8%
logistic	98.5	3.5%
log-normal	97.7	2.7%
gamma	96.7	5%
birnbaumsaunders	96.3	8%
inverse gaussian	96.2	26.6%
nakagami	94.5	5.3%
weibull	92.8	9.7%
rician	92.2	1.3%
normal	91	0%
extreme value	75.2	4.3%
rayleigh	41.1	1.8%

lem, namely multimodality. This is the case when disciplines cover several surgery types, which typically have a different mean duration. Unimodal parametric distributions (such as described in Table 6) do not work well on multimodal data. Methods that do work are based on, e.g., a kernel density estimator (KDE) or a Gaussian mixture model (GMM).

As Fig. 5 shows, the estimation error (which is the difference between the realized and the estimated durations) tends to increase with the length of the surgery. This aspect, amongst others, is captured by a bivariate distribution. In Table 7 we compare the goodness of fit of a few bivariate models on the data. The first distribution in the table is a purely parametric model, the bivariate GMM. The remaining models are based on the theory of copulas.

Both GMMs and copula based models have their benefits and drawbacks. A GMM assumes that all the

data points are generated from a mixture of a finite number of Gaussian distributions. In reality, this assumption might not be true for surgery durations.

Copulas are not constrained to distributions with Gaussian mixes. Copulas provide a way to describe joint distributions by separating the estimation of the marginal distributions of the random variable from the dependencies between them. Unfortunately, copulas such as the Gaussian- or (Student) t-copula come with their own set of restrictions as they can perform less well on multimodal data [53].

In order to model realized and estimated surgery durations at the discipline level, a model is needed that can handle multimodality and is flexible with regards to the assumptions made on the underlying distribution. Such a model was developed by Tewari et al. [53] and is a combination of GMMs and copulas using a class of functions called Gaussian Mixture Copula (GMC) functions. In Table 7 we compare such a GMC based model with a bivariate GMM, a Gaussian copula and a Student-t copula.

From Table 7 we see that the bivariate GMM performs well with regards to some disciplines. For the disciplines where the bivariate fit is bad, the marginal fit on estimated durations is bad as well. The bad fit is likely a consequence of the fact that estimates can have a discrete component.

On the contrary, as Table 7 shows, copula models can provide a good fit on the marginals but do not perform well with regards to the bivariate fit. This shows that the method fails to correctly capture the connection between the realized and the estimated durations. More specifically, both the Gaussian and the t-copulas seem to fail because of the multimodal aspect of the data.

A method that provides a good fit on both the bivariate distribution and on the marginals is the GMC model. As generally with copula methods, it is also in this case of critical importance to choose suitable marginal distributions. For example, choosing a lognormal marginal distribution for the realized durations clearly yields a bad fit (Table 7). Two marginals that work well are the univariate GMM and the KDE. The KDE we found to work well is the fixed bandwidth method described by Shimazaki and Shinomoto [48].

In the simulation model, we use the described GMC model with a univariate GMM for realized and a KDE method for estimated marginals. Alternatively, a purely empirical distribution could be used. This, however, would provide us with a less general model as the duration generator for some of the disciplines with lower sample counts (e.g., Tx) would produce reoccurring duration values. Methods that do not seem to work are

based on the bivariate GMM and all copula methods where other than GMM or KDE marginals are chosen.

4.4 Capacity allocation

In the literature, the OR planning process is commonly divided into three levels: strategic, tactical and operational [5]. At the strategic level, a certain amount of OR capacity is allocated to each discipline. This relates to the patient case mix as the hospital decides for each discipline on the number of future patients it wants to serve. At the tactical level, the MSS is created, this is a 1 or 2 week cyclic plan where to each weekday and OR a specific discipline or surgeon is assigned. At the operational level, surgeons assign patients to their own OR sessions. There are hospitals where the ORs are planned differently, generally however a schema similar to the one described is followed.

There are many criteria that can guide the creation of the MSS. Typically, the arrival caseload average and its variability are factors that are considered [24]. Additional factors can relate to tradition, i.e., if a discipline generally received a lot of capacity, they might also get more capacity in the future.

In the simulation model the MSS is predetermined and therefore static. This also means that the capacities assigned to each discipline are fixed for each week. The fixed weekly capacities averaged out over all weeks equals the average capacities of the University Hospital Leuven's final MSS. For example, if in reality on average 5.5 ORs a week are used by NKO then in the simulation model an MSS with 2 cycles is used where NKO in one week is assigned 5 and in the other week is assigned 6 ORs.

There will be slack capacity in the system as the supply of OR time is larger than the demand. The slack capacity shown in Table 8 is based on duration estimates and therefore relates to the planning phase. The table also shows that if non-elective capacity arriving during high impact periods is deducted from the available OR capacity then the total slack capacity is reduced to 10%. Interestingly, this is also the value that is suggested to work best by M'Hallah and Al-Roomi [34].

Table 8 shows that we allocate to some disciplines a different amount of capacity in the model as is in reality. For instance, the weekly capacities assigned to CAH and VAT were both reduced by 2 ORs (18h). This is done to get a more up to date system as also in reality, from the second half of 2013 on, their assigned capacity decreased.

On a side note, it is also from a capacity perspective important to create a realistic model of the non-elective

Table 7 The numbers in the table represent the p-values of the goodness of fit tests of various bivariate models and their marginals. As the Kolmogorov-Smirnov test is only applicable to continuous distributions, we use a χ^2 test. For the joint, bivariate distribution, a two-sample two-dimensional χ^2 test is used whereas for the marginals (realized and estimated) a two-sample one-dimensional χ^2 test is used. The bins in the two-dimensional case are based on a 10*10 grid of bins, whereas in the one-dimensional case on 10 bins. Bins with count lower than 5 are merged with neighboring bins. Distributions where all p-values are in bold (> 0.05) provide a good fit with the data and can therefore be used in the model.

Method		Non-Elective	GYN	Tx	ABD	CAH	NCH	ONC	RHK	THO	TRH	URO	VAT	MKA	NKO
GMM-Biv	joint	<.001	<.001	**.73**	.02	**.91**	**.19**	.02	<.001	**.6**	**.9**	**.79**	**.92**	**.81**	**.06**
	real.	**.26**	**.29**	**.53**	**.92**	**.86**	**.55**	**.22**	.02	**.19**	**.89**	**.13**	**.91**	**.53**	**.75**
	est.	<.001	<.001	**.09**	<.001	**.5**	<.001	<.001	<.001	.03	**.07**	**.18**	**.31**	**.11**	**.83**
Gauss.	joint	<.001	.05	**.11**	.02	<.001	<.001	<.001	<.001	<.001	0	<.001	<.001	<.001	**.37**
	GMM-real.	**.47**	**.15**	**.99**	**.22**	**.51**	**.65**	**.94**	**.47**	**.2**	**.08**	**.51**	**.46**	**.73**	**.07**
	KDE-est.	**.64**	**.76**	**.19**	**.52**	**.83**	**.72**	**.52**	**.86**	**.95**	**.72**	**.72**	**.3**	**.54**	**.3**
Stud.-t	joint	<.001	**.08**	**.06**	**.16**	<.001	<.001	0	<.001	<.001	**.4**	<.001	**.43**	<.001	**.51**
	GMM-real.	**.53**	**.53**	**.86**	**.51**	**.12**	**.26**	**.95**	**.53**	**.26**	**.47**	**.48**	**.48**	**.79**	**.26**
	KDE-est.	**.83**	**.58**	**.23**	**.88**	**.78**	**.93**	**.43**	**.56**	**.98**	**.58**	**.43**	**.34**	**.48**	**.46**
GMCM	joint	<.001	**.2**	**.1**	<.001	<.001	.04	.03	<.001	<.001	<.001	<.001	**.13**	<.001	**4**
	logN-real.	<.001	**.06**	**.12**	<.001	<.001	**.15**	<.001	<.001	<.001	<.001	<.001	**.14**	<.001	<.001
	KDE-est.	**.64**	**.74**	**.62**	**.88**	**.97**	**.88**	**.94**	**.87**	**.82**	**.75**	**.23**	**.38**	**.68**	**.84**
GMCM	joint	<.001	.05	**.87**	**.13**	**.14**	<.001	.01	**.31**	**.98**	**.29**	**.65**	**.38**	**.86**	**.32**
	GMM-real.	**.99**	**.29**	**.68**	**.95**	**.33**	**.87**	**.63**	**.46**	**.75**	**.38**	**.99**	**.49**	**.8**	**.71**
	GMM-est.	<.001	0	.01	<.001	<.001	**.35**	**.11**	<.001	.04	.05	**.64**	**.45**	**.52**	**.27**
GMCM	joint	**.37**	**.95**	**.9**	**.12**	**.56**	**.33**	**.91**	**.23**	**.93**	**.19**	**.43**	**.11**	**.25**	**.55**
	KDE-real.	**.95**	**.99**	**.71**	**.31**	**.88**	**.41**	**1**	**.58**	**.22**	**.41**	**.92**	**.67**	**.56**	**.79**
	KDE-est.	**.51**	**.84**	**.76**	**.24**	**.66**	**.93**	**.86**	**.45**	**.84**	**.56**	**.94**	**.62**	**.88**	**.88**
GMCM	joint	**.4**	**.75**	**.75**	**.95**	**.81**	**.41**	**.31**	**.92**	**.48**	**.69**	**.11**	**.6**	**.73**	**.3**
	GMM-real.	**.59**	**.79**	**.91**	**.36**	**.28**	**.86**	**.68**	**.8**	**.8**	**.72**	**.58**	**.85**	**.9**	**.91**
	KDE-est.	**.98**	**.56**	**.35**	**.61**	**.8**	**.61**	**.93**	**.58**	**.56**	**.78**	**.09**	**.58**	**.8**	**.85**

Table 8 The table shows that the total capacity available at all ORs are on average 824 hours a week. Out of this capacity, 20% is slack capacity. Slack capacity is based on the available OR capacity and the estimated caseload (i.e., caseload based on estimated durations). Slack capacity is part of the MSS and is different from the capacity that is used to protect against overtime at some hospitals (usually inserted at the end of the daily schedule). The slack capacity that remains after reducing the available OR capacity with the expected non-elective caseload is denoted by a '*'. The difference in open capacity in reality and in the model for Tx is only theoretic as Tx electives will in reality on average only be scheduled into around 2 ORs a week (18 hours), the rest (87% of slack) is generally used to accommodate urgent transplant patients. The difference for VAT and CAH reflects recent changes in the capacity allocation schema.

		GYN	Tx	ABD	CAH	NCH	ONC	RHK	THO	TRH	URO	VAT	MKA	NKO	μ	$\cup$
Reality	open cap.	67.8	46.3	97.5	128	68.0	32.1	55.1	73.8	74.3	57.0	57.1	20.9	46.0	63.4	824
	slack %	8%	87%	21%	19%	8%	22%	31%	7%	6%	9%	47%	14%	3%	22%	20%
	slack % *	6%	58%	3%	11%	-4%	17%	26%	-4%	-5%	4%	39%	10%	0%	14%	10%
Model	open cap.	68.0	18.0	97.5	110	68.1	32.1	55.3	74.0	74.5	57.2	39.3	21.0	46.1	58.5	761
	slack %	9%	67%	21%	6%	8%	22%	32%	7%	6%	9%	23%	15%	4%	18%	14%

Electives				Non-electives	
46.8	6%	NKO	2%		3.2
57.7	8%	URO	3%		4.1
68.9	9%	GYN	2%		2.5
173.5 h in Cluster A					9.7 h
7.9	1%	Tx	13%		21.4
31.4	4%	ONC	1%		2.4
106.4	14%	ABD	25%		39.6
145.6 h in Cluster B					63.3 h
77.6	10%	THO	13%		20.3
81.3	11%	TRH	12%		18.6
159.0 h in Cluster C					38.9 h
23.9	3%	MKA	1%		2.2
48.1	6%	RHK	3%		4.7
71.5	9%	NCH	12%		19.3
143.5 h in Cluster D					26.3 h
32.7	4%	VAT	4%		6.8
106.6	14%	CAH	10%		16.2
139.3 h in Cluster E					23.1 h
760.9 h	100%		100%		161.3 h

Fig. 6 The values are based on realized durations and show how much average weekly capacity is used by elective and non-elective disciplines. For planning purposes less capacity is booked since duration estimates are generally underestimated (Table 8).

OR assignment schema. Fig. 6 shows that the elective load on different clusters is different. A balanced load on clusters is only observable if non-electives are included. This is fair to do as non-electives usually enter ORs that are assigned to the elective discipline itself or to the disciplines cluster (Fig. 4), i.e., they either way contribute to the clusters' load. This shows that also from a modeling perspective it is important to allocate non-electives realistically as, e.g., a random assignment would yield a false load on clusters and disciplines. This would lead to a false view on rescheduling and on OR and patient related performance outputs.

4.5 Rescheduling

There are two major reasons why regular OR time is not always enough to serve all planned electives. Firstly, it frequently happens that surgeries take longer than estimated (Fig. 5). Secondly, a non-elective arrival, generally DT category 1, can demand immediate access to an OR and thus postpone the execution of the OR's elective schedule (Fig. 4). In those cases it can become necessary to reschedule elective patients in order to avoid excessive overtime.

We distinguish between two basic types of rescheduling actions: surgery reassignment and surgery cancellation. A surgery is reassigned if, on the day of the surgery, it is moved from the originally planned OR to another OR. The surgery is, however, still performed on the originally planned date. On the contrary, a cancelled surgery will be performed on a later date and is assigned to the surgeon's next OR session. This is done even if the next session is already fully booked.

We make a clear distinction between rescheduling and replanning. Rescheduling is done on the day of the surgery itself and is used to avoid excessive overtime. It is not part of the advance patient scheduling process, but a component of the simulation model. We therefore do not test different rescheduling policies and only model the current practice found at the hospital. In contrast, replanning a surgery is done before the actual surgery date and is therefore part of the patient scheduling process.

In our setting, we focus on rescheduling as an action to control some aspect of OR overtime. In some hospitals, a limit on the OR overtime is enforced. For instance, in a Spanish setting described by Pulido et al. [41] this limit is 2 hours of overtime. For work beyond that limit, surgeons and nurses are not paid, giving them an incentive to rather reschedule a surgery than to go over the set time limit. Other hospitals may control the risk of going into overtime as they ensure that an OR goes into overtime only in a certain percent of the cases. This is done, among others, so that nurses will only occasionally have to work longer hours. Other, mostly profit oriented, hospitals may trade off the cost of paying for overtime staffing and the profit gained from performing a surgery in overtime. At the University Hospital Leuven, depending on the hour of the day, a limit is set on the number of ORs that are allowed to stay open in overtime, i.e., 8 ORs out of the 22 ORs may be running overtime beyond 6 pm, 4 beyond 7 pm and only 2 beyond 8 pm. Those 2 ORs remain open the entire night and serve incoming non-elective patients.

At the hospital, cancellations are less often carried out than surgery reassignments. Understandably, the hospital's head anesthesiologist is more reluctant to cancel a surgery than to reassign it to another OR. This is the case as it is frustrating for patients to be cancelled. Patients being reassigned to another OR is not a problem. As a matter of fact, they might not even notice it.

At the hospital, between one and two patients are cancelled on a daily basis and more than six are OR reassigned. For an elective patient this means a probability of 3.4% to be cancelled and 13.1% to be reas-

signed. The hospital targets a cancellation rate of 2% in the future.

In the literature the term cancellation can also cover actions where surgeries are cleared from the schedule already prior to the surgery day [14]. Using this definition, Dexter et al. [12] show that inpatients are more often cancelled than outpatients. Moreover, Epstein and Dexter [13] conclude that cancellations do not need to be necessarily interpreted as a system failure as scheduling cancelled patients does not increase the variability in the OR workload. Therefore, some amount of cancellation is not only unavoidable but can be also desirable.
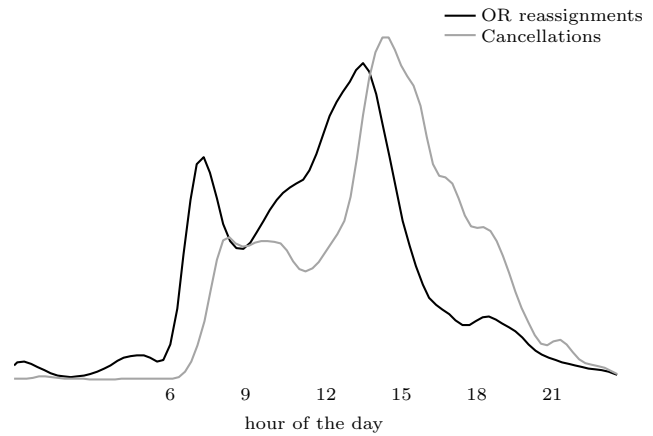
Most OR reassignments and cancellations are performed during the early morning and, especially, during the early afternoon hours (Fig. 7). Similarly to our findings, also Epstein and Dexter [14] show a histogram with 2 peaks for cancellations on the day of surgery, one peak in the early morning and one in the afternoon. In their setting, however a larger peak is present in the morning (around 7 and 8 am) than in the afternoon.

In the simulation model, we imitate the behavior of the hospital's head anesthesiologist. The head anesthesiologist makes rescheduling decisions continuously throughout the day. Therefore, in the model, each full hour from 8 am to 9 pm we identify those ORs that are expected to run into overtime. From the identified ORs it is then checked whether surgeries can be moved to other ORs. An OR can only accept a surgery if, including the new surgery, the OR is still estimated to close before OR closing time (4.45 pm). In case a surgery cannot be reassigned to another OR it may get cancelled.

Whether it is believed that an OR goes into overtime depends on the OR closing time estimate and on the reliability of this estimate. The estimate is more reliable if it is later in the day as consequently a larger number of surgeries will already have been performed.

In the simulation model, the degree to which it is believed that an OR will go into overtime depends on two functions. Their product will determine whether a patient can only be OR reassigned or may also be cancelled. The first function increases linearly with the time of the day, while the second function represents the OR closing time estimate. In other words, the later it is in the day, the more it is believed that the OR closing time estimates are correct and consequently, in case the OR closing time estimates indicate overtime, the more likely it becomes that a rescheduling action is carried out.

Importantly, surgeries cannot freely be moved between two ORs. In the simulation model (as in reality) a strict hierarchy is followed. Firstly, it is preferred to reassign a patient to an OR that serves the same disci-

Fig. 7 Rescheduling decisions are made throughout the day. The PDFs have two peaks, one in the morning and one in the afternoon. The larger peaks in the afternoon are at 2 pm and 3 pm for OR reassignment and cancellation respectively.

pline. Alternatively, a surgery can be reassigned to an OR of another discipline as long as it is within the same cluster. Less favorable but possible is to move surgeries across clusters A, B, C and D (Fig. 6). CAH can only be reassigned to its own ORs and NCH surgeries will not leave the cluster to which they are assigned.

4.6 Model assumptions and validation

Our results are usable in the real setting as we ensured that the model is credible and valid. Model credibility is concerned with "developing in potential users the confidence they require in order to use a model and in the information derived from that model" [45]. We created the model based on the data of the University Hospital Leuven and through numerous meetings with the management (head surgeon, head nurse, etc.) confirmed that we have the right understanding of both the data and the setting. Our model is consequently credible to the people at the hospital.

Model validation is the "substantiation that a computerized model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model" [46]. We validated our model by comparing the simulation results with real hospital data. We think that there are 3 aspects that are of key importance and thus have to be validated: (1) a realistic capacity to discipline allocation schema (Table 8), (2) a realistic arrival caseload process (Table 9) and, in order to ensure that the hospital processes are modeled accurately, (3) the validation of key hospital related performance measures (Fig. 8). As we already dealt with the first point, we will only focus on points 2 and 3.

The arrival caseload is the amount of OR hours that have to be scheduled for surgery in the current or future weeks, which is different from the actual surgery caseload assigned to that week. Whereas the surgery caseload depends largely on the fixed MSS and is thus fairly stable and predictable, the arrival caseload is more variable as it depends both on stochastic patient arrival numbers and on their stochastic surgery duration lengths.

Table 9 shows that the model is realistic as both averages and standard deviations reflect reality (small Δ). This is true for the realized arrival caseload, for the estimated caseload and for the error between them. The error is important as it contributes to the uncertainty that differentiates a planned from a realized schedule.

In order to ensure that the hospital processes are modeled accurately, we validate some of the key hospital related performance measures (Fig. 8). The results confirm that the model is valid. There are four measures where there is a statistical difference. In reality (1) lower utilization, (2) less overtime, (3) more undertime and (4) more daily open ORs are experienced.

In case of measures (2) and (3), the difference is statistically speaking significant but small enough to not be of practical relevance. The reasons for (1) and (4) can be explained by the fact that in reality less than 9 hours of capacity might be allocated to an OR-day. However, as it is difficult to identify those OR-days, we will assume they are always assigned the full 9 hours. The measured utilization in reality will thus be lower than in the model. Similarly, as in the model and in reality the same amount of total capacity is used, shorter opening hours entail that the OR is open on more days.

4.7 Simplifications

We tried to simplify the arrival model, without success. For instance, modeling patient arrivals with a Poisson distribution causes the std. of the average caseload per disciplines to decrease from 20.2 hours (measured in reality) to 15.7 hours. For some disciplines it will be 8 hours lower than in reality, which is almost an entire OR. This means that using a Poisson arrival process may lead to a system that is much more stable than it is in reality. This can lead to misleading results. This is especially true with regards to DT 4 patients which, as they are required to be served within a week, are more sensitive to short-term capacity shortages.

Other simplifications involve the surgery duration model, where we tried to model durations in the univariate space and fitted a parametric distribution on
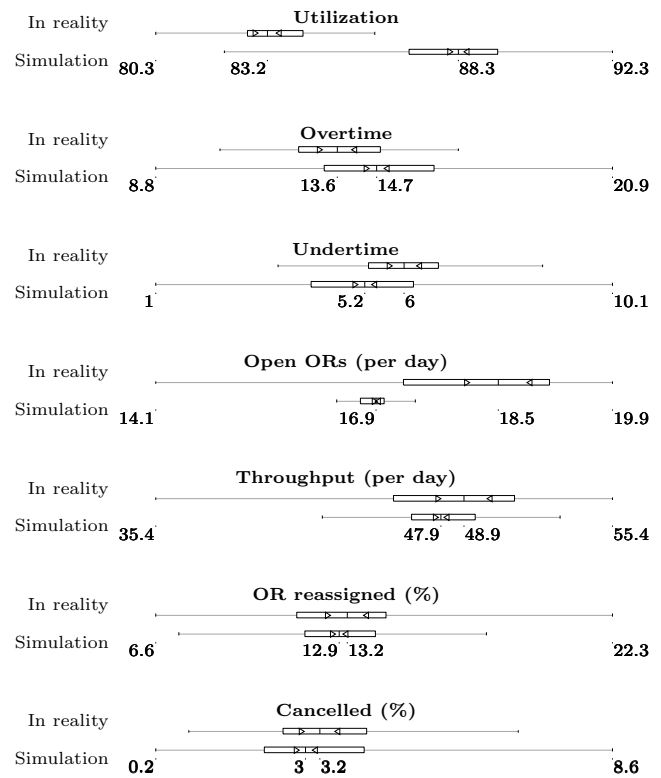
Fig. 8 The central mark represents the median and the edges of the boxes the 25th and the 75th percentiles. If the intervals, represented by triangular markers, do not overlap, then the two medians are significantly different at the 5% significance level.

both realized and estimated durations independently. The chosen parametric distribution is, for each discipline, the one with the best AIC value. The tested distributions include, amongst others, the ones described in Table 6. Interestingly, this would lead to good results with regards to the average estimation error of the weekly caseload. The problem however is the standard deviation, which increases to 14.6 hours from the 5.5 hours measured in reality.

As realized and estimated surgery durations cannot be modeled independently, we also tested whether a univariate fit on the mismatch between the two might bring the desired results, i.e., fit on realized durations and on the estimation error itself. Estimated durations are then the sum of the two. On the positive side, this method gives generally smaller errors than if any of the other previously mentioned simplifications are used. On the negative side, there will be extreme cases as, for instance, the estimated weekly arrival caseload for RHK would show a standard deviation of 42.2 hours instead of the 23.2 hours measured in reality. This is especially a problem for RHK as a large part of their patient population belongs to DT 4. We consequently think that it is

Table 9 Caseload of weekly arrivals. The modeled arrival caseload matches the real one.

			Non-Elective	GYN	Tx	ABD	CAH	NCH	ONC	RHK	THO	TRH	URO	VAT	MKA	NKO	μ	∪
Realized (h)	real	μ	**160**	**68.6**	**7.8**	**106**	**106**	**71.2**	**31.3**	**48.0**	**77.3**	**80.9**	**57.5**	**32.6**	**23.8**	**46.6**	**58.3**	**758**
		σ	31.5	20.7	6.0	28.3	39.1	27.8	11.8	23.2	28.5	19.3	15.2	14.3	14.6	14.1	20.2	138
		CV	.20	.30	.76	.27	.37	.39	.38	.48	.37	.24	.26	.44	.61	.30	.40	.18
	model	μ	**159**	**69.7**	**7.6**	**106**	**105**	**71.1**	**31.8**	**49.0**	**78.2**	**80.5**	**57.6**	**32.7**	**23.9**	**46.5**	**58.5**	**760**
		σ	26.4	21.5	5.5	28.3	36.6	28.2	11.6	23.6	27.4	19.3	15.1	14.2	15.1	15.1	20.1	80.3
		CV	.17	.31	.73	.27	.35	.40	.36	.48	.35	.24	.26	.43	.63	.33	.40	.11
	Δ	μ	**.97**	**-1.0**	**.28**	**-.09**	**.79**	**.10**	**-.47**	**-1.1**	**-.89**	**.36**	**-.06**	**-.13**	**-.14**	**.14**	**-.17**	**-2.2**
		σ	5.2	-.80	.44	.02	2.5	-.39	.25	-.36	1.0	.03	.02	.13	-.46	-1.0	.11	57.8
		CV	.03	-.01	.03	.00	.02	-.01	.01	.00	.02	.00	.00	.01	-.02	-.02	.00	.08
Est. (h)	real	μ	**141**	**62.1**	**5.9**	**77.3**	**103**	**62.8**	**25.1**	**37.9**	**68.5**	**70.1**	**52.0**	**30.3**	**18.0**	**44.4**	**50.6**	**657**
		σ	26.3	18.7	4.1	21.4	37.1	23.9	9.9	18.8	25.3	16.2	13.8	13.2	10.8	13.0	17.4	113
		CV	.19	.30	.70	.28	.36	.38	.40	.50	.37	.23	.27	.44	.60	.29	.39	.17
		μ	**138**	**62.5**	**5.7**	**77.2**	**102**	**62.4**	**25.1**	**38.4**	**68.8**	**69.3**	**51.8**	**30.1**	**18.0**	**44.1**	**50.4**	**656**
		σ	23.2	19.5	4.1	20.4	35.4	24.3	9.0	18.9	23.8	16.0	13.2	12.9	10.9	13.7	17.1	69.3
		CV	.17	.31	.72	.26	.35	.39	.36	.49	.35	.23	.26	.43	.60	.31	.39	.11
	Δ	μ	**2.1**	**-.39**	**.20**	**.04**	**.90**	**.40**	**-.06**	**-.59**	**-.28**	**.74**	**.21**	**.13**	**-.09**	**.34**	**.12**	**1.5**
		σ	3.1	-.84	.05	.96	1.7	-.33	.88	-.15	1.5	.19	.58	.31	-.07	-.78	.31	44.2
		CV	.02	-.01	-.02	.01	.01	-.01	.04	.00	.02	.00	.01	.01	.00	-.02	.00	.07
Est. error (h)	real	μ	**19.7**	**6.5**	**1.9**	**28.7**	**3.1**	**8.4**	**6.3**	**10.1**	**8.8**	**10.9**	**5.5**	**2.3**	**5.8**	**2.3**	**7.7**	**101**
		σ	11.5	6.2	2.5	10.1	5.8	6.6	4.2	6.1	6.3	6.7	4.0	4.0	5.0	4.4	5.5	32.7
		CV	.58	.96	1.3	.35	1.9	.78	.67	.60	.71	.62	.73	1.7	.86	2.0	1.0	.33
	model	μ	**20.9**	**7.2**	**1.9**	**28.8**	**3.2**	**8.7**	**6.7**	**10.6**	**9.4**	**11.2**	**5.8**	**2.6**	**5.9**	**2.5**	**8.0**	**104**
		σ	9.3	5.6	2.4	10.3	5.5	6.9	4.5	6.2	6.7	6.5	4.1	4.7	5.6	4.6	5.7	22.4
		CV	.45	.79	1.3	.36	1.7	.79	.67	.59	.71	.58	.72	1.8	.95	1.9	.99	.22
	Δ	μ	**-1.2**	**-.64**	**.08**	**-.13**	**-.11**	**-.30**	**-.41**	**-.48**	**-.61**	**-.38**	**-.27**	**-.27**	**-.04**	**-.20**	**-.29**	**-3.7**
		σ	2.2	.60	.16	-.26	.32	-.32	-.26	-.16	-.42	.25	-.14	-.69	-.59	-.15	-.13	10.2
		CV	.14	.17	.03	-.01	.17	-.01	.00	.01	.00	.04	.01	-.09	-.09	.10	.03	.11

not possible to include any of the previously mentioned simplifications without substantially changing the setting.

5 Results

Results are easiest analyzed if the effect of each factor representing a scheduling method can be isolated from the other 2 factors (i.e., averaged out). This corresponds to an analysis of the main individual effects. It is important to note that such an analysis can yield misleading results in the presence of interaction effects between any two factors and the particular performance measure. Unfortunately, we noticed that with most of these performance measures there are significant two-way (and three-way) interaction effects present at a 5% significance level. As a consequence, we analyzed factors in combination. This can be done with interaction plots (e.g., Fig. 9). In order to represent all results in a consistent way, we use interaction plots for all results

(also for performance measure where no significant interaction effect was present). As the interaction plots show, a full factorial design with 216 scenarios is used: 6 (factor 1) * 3 (factor 2) * 12 (factor 3).

In order to be aligned with the two week MSS cycle used at the hospital, all performance measures are batched on a two week period basis, i.e., one batch covers two weeks. Mean values shown in the results will therefore be the mean values of individual two week batch means. Similarly, also the shown standard deviations will relate to the variability between those batches.

5.1 OR and patient related performance measures

We distinguish between OR related performance measures (e.g., utilization, overtime and undertime and patient related performance measures (e.g., patient waiting time and the ratio of patients that are served within their respective DT).

Results for each performance measures are visualized using a 3-way interaction plot. Each interaction plot contains four dimensions, three correspond to the three scheduling factors (Table 3) and one to the respective performance value (y axis). Each point represents one scenario, thus a combination of the three factors.

Table 10 shows that many of the scheduling factors have a significant main effect on OR related performance measures (p-values are smaller than 0.05). Nevertheless, as the standard deviation between different scenarios is very small, those effects do not bear any practical relevance. A similar observation can be made in Fig. 9 for overtime. The figure shows that overtime values change between 14.6% and 15.1%. Small differences like these are for the hospital of little practical importance.

The reason why overtime does not depend on the chosen scheduling strategy can be explained as follows. As demand closely matches supply and as surgeons can fill up their ORs fully, ORs will irrespective of the patient scheduling strategy be fully booked and therefore highly utilized (around 87.9-88.1%). The fact that the ORs are highly utilized is bound to lead to a substantial amount of average overtime. The exact amount is however independent of the patient scheduling strategy (Fig. 9) but determined by factors such as the estimation error of surgery durations (Table 5). This also means that overtime might not be avoided without sacrificing the efficient use of OR time. And vice versa, OR time might not be efficiently used without overtime.

Because scheduling factors practically speaking do not affect OR related performance measures, they are excluded from any further analysis. This allows us to concentrate on patient related performance measures only. We will focus on three in particular: the percentage of patients served within their DT, the average patient waiting time and the weighted DT cost (Fig. 10).

5.2 Percentage of patients served within DT

The percentage of patients who are served within the medically advised time limits set by their surgeons, the DT, is an important indicator for the hospital. Whether patients are served within DT depends on their arrival and surgery date. For example, if it is on a Monday determined that a DT category 4 patient needs surgery then the latest date which is within the DT is the Monday the following week. Later days, regardless of the exact number, are considered to be after the DT. As DT category 8 patients are not given any hard deadline, they are excluded from the calculations.
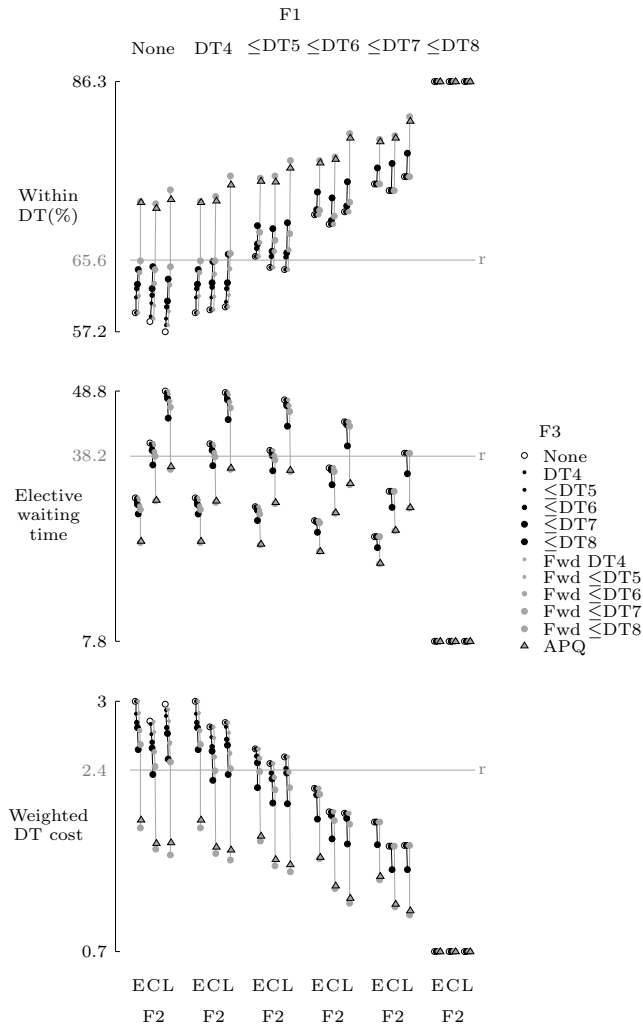
Currently, at the hospital, around 65% of patients are served within their DT. Further decomposing this

Table 10 Real values are compared, in specific, to the FCFS strategy and, in general, to all scenarios. The standard deviation of the scenario means (e.g., 0.05 for utilization) has a different meaning than the mean of the scenario standard deviation (e.g., 1.65 for utilization). The former shows how much the value of a performance measure differs between different scenarios, whereas the latter shows how much variability is generally present between the batches within scenarios.

		utilization %	overtime %	undertime %	open slot daily mean	throughput daily mean	reassigned %	cancelled %
real	μ	83.4	13.6	6.0	18.0	47.9	13.1	3.4
	σ	1.3	1.6	1.2	1.5	4.2	2.8	1.1
	CV	.02	.11	.19	.08	.09	.21	.33
FCFS	μ	88.0	14.8	5.2	16.9	48.0	12.9	3.2
	σ	1.8	2.1	1.5	.17	2.0	1.8	1.4
	CV	.02	.14	.29	.01	.04	.14	.45
scen. avg.	μ	88.0	14.8	5.3	16.9	48.0	12.9	3.2
	σ	.05	.07	.10	.01	.03	.17	.08
	CV	.00	.00	.02	.00	.00	.01	.02
scen. std.	μ	1.65	1.97	1.47	0.17	1.83	1.67	1.44
p-value main effect	F1	.96	<.001	<.001	<.001	1.0	<.001	.02
	F2	.01	.06	.00	.01	.40	<.001	<.001
	F3	.00	.54	.03	.71	.19	<.001	.21
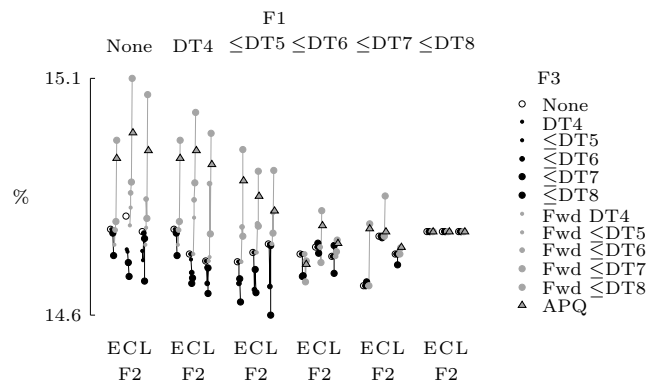
Fig. 9 The amount of overtime is, from a practical perspective, independent of the chosen patient scheduling strategy as the minimum and maximum values are very close. Each point in the figure represents a scenario, that is, a combination of three factors. The exact realization of each factor 1-3 is defined by: the label on the top (F1), the label on the bottom (F2) and the marker (F3). The full markers represent scenarios where fresh arrivals (black), existing patients (gray) are used for replanning. For example, the most left triangle represents a scenario where: no DT category is served FCFS (F1), surgeries are served in the early part of their DT interval (F2) and replanning uses the APQ method (F3). The second axis shows the respective performance value, which in this case is overtime.

Fig. 10 Patient related performance measures. The horizontal line shows the real values measured at the hospital ('r').

Table 11 All factors have a significant main effect on all three performance measures. Additionally, the scenario means shows a large CV value, which means that they also practically speaking have a large influence on the results. The fact that the average standard deviation within scenarios (indicated by 'Scenario std.') is high for all three measures, shows that there can be large differences between different two week periods.

		DT offset (%)	Elective waiting time	Weighted DT cost
real	μ	65.6	38.2	2.4
	σ	4.4	4.0	.43
	CV	.07	.11	.18
FCFS	μ	86.3	7.8	.71
	σ	5.3	1.1	.32
	CV	.06	.14	.46
scen. avg.	μ	71.5	30.8	1.9
	σ	8.7	12.4	.71
	CV	.12	.40	.37
scen. std.	μ	5.97	1.86	0.55
p-value main effect	F1	<.001	<.001	<.001
	F2	<.001	<.001	<.001
	F3	<.001	<.001	<.001

anymore, scheduling patients in the very last moment into this capacity avoids that the replanned patients occupy future OR capacity that might be needed for other patients.

Fig. 10 also shows that filling up next day free capacity by replanning and thus bringing patients forward in the schedule is considerably more effective than using fresh arrivals. This is the case as an OR can only serve surgeries from one particular surgeon. This restricts the number of usable new arrivals to a limited set. More patients are generally available for replanning.

In contrast, factor 2 shows only a minimal effect. This means that scheduling patients into the early, center or late part of their DT does not make a substantial difference.

5.3 Patient waiting time

Patient waiting time is one of the classical performance metrics used in the literature. The waiting time of a patient equals the number of days between the date the decision for surgery was made and the date the surgery was performed. The decision for surgery is made when the surgeon and the patient meet for consultation and a form is filled out with the details of the surgery. This type of waiting time which is usually measured in days is also called indirect waiting time.

Elective patients at the University Hospital Leuven wait 38.2 days on average (Fig. 10). Further decomposition by DT category shows that for DT category 4, 5,

result by DT category shows that 81.2% of DT category 4 patients are served within their DT, making it the most efficiently served DT category at the hospital. For patients with DT categories 5, 6 and 7, the respective values are 52.6%, 59.5% and 67.2%.

From Fig. 10 we see that as more DT categories are served FCFS (Factor 1), the percentage of patients served within their DT is increasing. This means that serving patients from a specific DT category FCFS is beneficial as, on the one hand, it naturally helps patients from the specific DT category and, on the other hand, seems to have at most a limited detrimental effect on patients from other DT categories.

Similarly, also scheduling patients 'next day' (Factor 3) is beneficial. This implies that it is crucial to save capacity that might remain unused. OR capacity that remains unused is wasted and cannot be recovered

Table 12 Waiting time by DT category.

		DT4	DT5	DT6	DT7	DT8
	μ	8.1	21.6	40.1	51.7	75.1
real	σ	1.8	3.8	9.6	9.3	15.1
	CV	.22	.18	.24	.18	.20
	μ	6.5	8.4	7.0	10.1	7.9
FCFS	σ	1.3	1.9	1.3	2.4	1.7
	CV	.19	.23	.19	.23	.22
	μ	10.7	16.4	16.3	33.5	68.6
scen. avg.	σ	2.3	4.8	7.5	16.1	33.5
	CV	.21	.29	.46	.48	.49
scen. std.	μ	2	3.76	1.87	2.5	1.29
p-value	F1	<.001	<.001	<.001	<.001	<.001
main effect	F2	<.001	<.001	<.001	<.001	<.001
	F3	<.001	<.001	<.001	<.001	<.001

6, 7 and 8 it is 8 days, 21.6 days, 40.1 days, 51.7 days and 75.1 days respectively (Fig. 11).

Improvements with regards to waiting time can be achieved in an intuitive and straightforward way. Firstly, by scheduling more DT categories on a FCFS basis. Secondly, by serving patients in the earlier part of their DT. Thirdly, by allowing patients to be served next day. Replanning patients is also with regards to the waiting time more effective than using fresh arrivals. As shown by Fig. 11, gains apply to all DT categories.

In Fig. 10 it is interesting to note that improvements generally remain severely limited as long as FCFS (or next day) is only applied to patients up to DT category 7. Real benefits are only realized once DT category 8 patients are included. This shows that the way DT category 8 patients are scheduled is important, but it also highlights some of the drawbacks of using patient waiting time as a performance measure as it is heavily determined by low urgency patients.

5.4 Weighted DT cost

The degree to which the DT is obeyed can be measured in several ways. A straightforward method is to simply determine the ratio of patients who are served within DT. However, this does not provide any information on the extent to which patients are late once they are over their DT. This information is provided by the weighted DT cost which implicitly considers the tail of the distribution.

The weighted DT cost is based on the idea that patients that passed their DT should be served quickly. The more urgent the patients initial DT category, the fewer days they should be allowed to be served after their DT. Therefore, the cost function is proportional to the number of days a patient is served after the DT
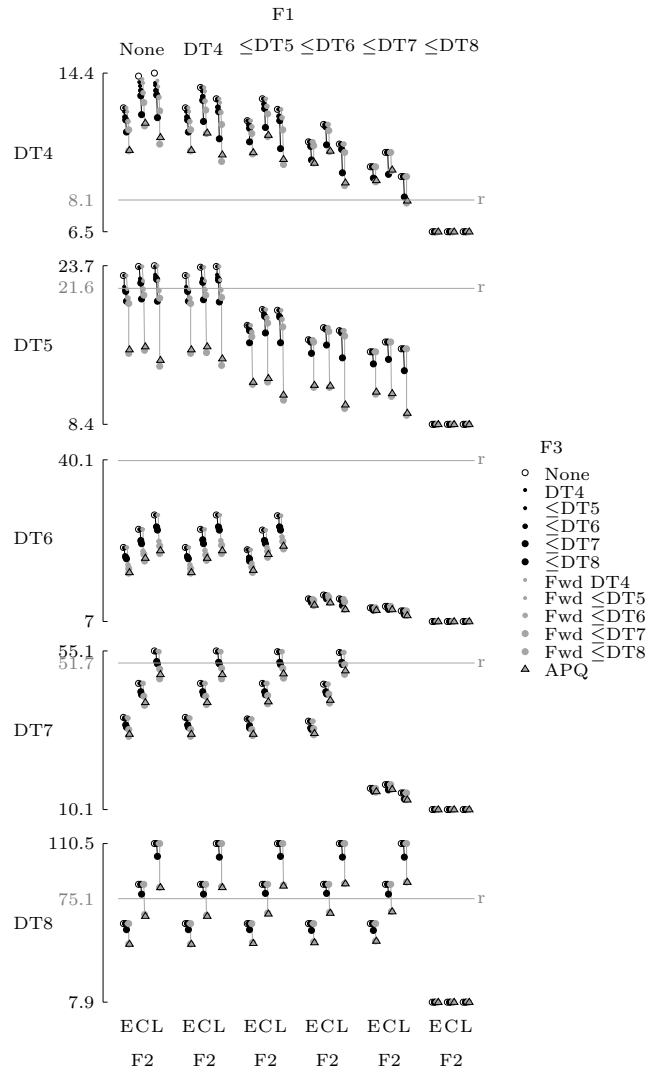
Fig. 11 The waiting time of each elective DT category.

but inversely proportional to the patients initial DT in days. It is defined as:

$$v_i = \begin{cases} \frac{7}{dt_i}(s_i - (a_i + dt_i)) & s_i - a_i > dt_i \\ 0 & otherwise \end{cases} \quad (2)$$

$$V = \frac{\sum_{i \in I} v_i}{|I|} \quad (3)$$

This cost is zero for patients that are served within their DT. This reflects the idea that from a patient outcome perspective the time when a patient is served does not matter as long as it is before the DT. Moreover, patients that are served after the DT will, at different points in time, eventually be exposed to a similar health risk. For example, a patient with a DT of 1 week who is served 1 week late is associated with the same cost/risk

as a patient with a DT of 4 weeks who is served 4 weeks late. The penalties for each day served late for DT categories 4 to 7 are 1, 1/2, 1/4 and 1/8 respectively. No penalty is linked to DT category 8 as they are not given a time limit and their health conditions should generally not worsen over time. A similar idea is used in Riise and Burke [43], who describe a Norwegian setting where violations of due dates are regarded to be one measure of a hospital's efficiency.

In reality, surgeries can be performed late for other than scheduling related reasons. We assume that this is the case for patients who wait for a longer time than 5 times their DT. Those patients are consequently excluded from the cost formula.

From Fig. 10 we see that FCFS also performs well with regards to the weighted DT cost. As more DT categories are scheduled FCFS, the average weighted DT cost decreases. This suggests that the benefit of scheduling patients of less urgent DT categories FCFS compensates for the resulting possible delays of more urgent DT category patients. For example, the benefit of providing DT category 5 patients quick access to the OR compensates for the occasionally caused delays of DT category 4 patients.

The benefit of FCFS is the largest if replanning is not allowed. Similarly, replanning is able to partly compensate for the case when applying FCFS is not entirely possible. Consequently, if FCFS is not applicable in reality, it is important to allow for replanning. This, as shown by Fig. 10, should include patients from DT category 8. This is interesting as DT category 8 does not contribute directly to the DT measure (surgeries of the category have a weight of 0). But their surgeries, if replanned, free up future capacity that can be used by surgeries from DT categories that do contribute to the DT cost measure.

One might wonder why FCFS outperforms replanning. FCFS is a regular planning procedure and has to obey the MSS only. Contrary, replanning is more restricted than a regular planning procedure as surgeries cannot be replanned to empty ORs, i.e., a surgery can only be replanned into an OR that has already been assigned to the corresponding surgeon. Replanning surgeries into empty ORs would require the hospital to provide full staffing for entire ORs from one day to another. This is something that we generally would like to avoid.

From Fig. 10 we see that the APQ does not outperform the best-fit strategy (the triangle and the large gray dot overlay). This is surprising as the APQ ensures that urgent patients in danger of running late are replanned first. It is thus tailored to perform well with regards to the weighted DT cost. Further analysis shows that the APQ and best best fit strategies are not performing statistically differently with regards to any of the three tested performance measures (Table 13). This implies that the benefits of replanning are not a result of cost reductions associated to individual patients saved from running late. Instead, it performs well because it saves capacity that otherwise would be wasted. Consequently, the replanning procedure does not need to consider the DT.

Interestingly, factor 2 will not make a difference except if replanning is not allowed. In this case, the early strategy will perform best. This is in line with the results of Gocgun and Puterman [21] where they show that scheduling chemotherapy patients into earlier slots outperforms the other tested policies.

From Table 14, showing the decomposition of the cost by DT category, we see that the highest cost is in the model associated to DT 4. This is to be expected. Interestingly, however, in reality, the highest cost is associated to DT 5. This is unexpected and shows that in reality DT category 4 is efficiently handled. It may also mean that DT category 4 patients might be in certain situations overly prioritized resulting in exaggerated delays of patients of DT categories 5 and 6. Improving the scheduling of those two categories might therefore lead to the largest benefits for the hospital.

Table 14 also shows that DT category 4, in absolute terms, is better handled in reality than in any simulated scenario including FCFS. It seems that some surgeons may always keep some slack capacity reserved for DT 4. The occasional capacity loss might then translate into decreased service levels provided to DT 5 and 6 patients.

5.5 Discipline specific insights

In Table 15 we highlight some of the discipline specific aspects of the results. As the table shows, the FCFS strategy always performs better than the average scenario (FCFS μ is always better than scen. μ). FCFS will also generally give better results than what is currently measured at the hospital. This does not necessarily mean that the FCFS strategy, if implemented, would perform necessarily best in reality as there could be discipline specific constraints that are not captured by the model but are important in practice. Nevertheless, it is an indication that it could be beneficial generally for disciplines to schedule patients to closer dates and as a rule not to leave any capacity unused.

Interestingly, some of our specific results do not necessarily apply to all disciplines. For example, there will be disciplines for which the replanning strategies (if they include DT category 8) will perform very similarly or even better than FCFS (e.g., GYN).

Table 13 One-way anova shows that the APQ and best-fit strategies do not perform significantly differently (all p-values are > 0.05). This is true for any combination of factors 1 and 2.

F1	None			DT4			≤DT5			≤DT6			≤DT7			≤DT8		
F2	E	C	L	E	C	L	E	C	L	E	C	L	E	C	L	E	C	L
Within DT(%)	.89	.33	.02	.89	.31	.04	.48	.18	.08	.61	.55	.26	.6	.54	.32	1	1	1
Waiting time	.11	.54	.04	.11	.26	.17	.32	.28	.25	.65	.73	.31	.58	.85	.66	1	1	1
Weighted DT cost	.16	.23	.01	.16	.15	.03	.35	.18	.1	.67	.56	.2	.44	.46	.25	1	1	1

Table 14 Weighted DT cost decomposed by DT category.

		DT4	DT5	DT6	DT7
real	μ	1.4	3.2	3.1	1.8
	σ	.81	.77	.69	.83
	CV	.60	.24	.22	.48
FCFS	μ	1.6	.60	.02	.00
	σ	.81	.61	.05	.01
	CV	.51	1.0	2.0	6.6
scen. avg.	μ	3.0	2.7	.45	.09
	σ	.89	1.3	.25	.07
	CV	.29	.47	.56	.81
scen. std.	μ	1.23	1.1	0.25	0.1
p-value main effect	F1	<.001	<.001	<.001	<.001
	F2	<.001	<.001	<.001	<.001
	F3	<.001	<.001	<.001	<.001

5.6 Discussion

In this section, we first discuss the reasons why the efficient use of OR capacity results in good patient related performance values. Secondly, we discuss points that relate to the practical applicability of the scheduling factors. We regard a scheduling method applicable if, firstly, surgeons are likely and willing to use them and, secondly, the scheduling methods allow enough time for patients to prepare for the surgery.

Algorithms that are most interesting to the hospital are the ones that consider the DT and are manually usable by the surgeons. This is the case as surgeons prefer to plan their patients themselves. This is unlikely to change in the near future as surrendering patient scheduling to a central authority would mean that surgeons would lose part of their independence and thus influence on their own work schedule. At the moment, there is no central hospital wide patient scheduling system in place.

Additionally, it would be difficult to convince all surgeons of the benefits of using a computer and optimization software to schedule their patients. This is one of the major reasons why formulating the patient scheduling problem as an optimization problem would, in our setting, be of less use.

The scheduling methods we tested are all manually usable. More involvement is required for the APQ driven replanning procedure. This is however not a problem as we showed that the APQ, in our setting, does not bring any benefits.

There are some aspects of the real setting that we did not model, but which we also do not deem important. They relate, on the one hand, to surgeon and patient preference related factors and, on the other hand, to downstream facilities. Surgeon preference is, for example, to have only one difficult surgery (e.g., hip replacement) on a day. Similarly, also the number of children can be restricted. This is done as patients before their surgery are not allowed to eat for a certain amount of time, which is more difficult to do for children. It is therefore best to serve one child first in the morning. Those factors are important to consider when scheduling patients, but excluding them in the simula-

Comparing results from the model with reality, we see that a discipline where the performance difference is large is ABD. As the results in Table 15 suggest, the discipline should theoretically be able to handle its patient load very well. This is the case as, its arrival caseload is stable (low CV value), the estimated durations are short and little variable, its surgery urgency mix is low (low DT score) and it seems to have enough slack capacity.

One of the reasons why ABD might perform worse than expected could relate to the fact that they accommodate a large amount of non-elective patients. At the hospital, they might therefore be more wary of fully utilizing their available capacity and instead leave more slack.

A discipline that seems to have difficulties accommodating it's surgeries is MKA. The discipline's major problem factor seems to be the highly unstable arrival caseload (CV is 0.6). This explains the low amount of patients that are served within the DT both in reality and in the model. They could perform better if strategies would be in place that allow them to flexibly control their weekly number of ORs. This might help them to be better equipped for weeks with high loads.

Table 15 How well a discipline performs depends on many factors, such as the arrival caseload variability and the available slack capacity. DT related performance measures for Tx cannot be interpreted as most of their patients are from DT category 8.

|		GYN	Tx	ABD	CAH	NCH	ONC	RHK	THO	TRH	URO	VAT	MKA	NKO	∪
DT score		.23	.02	.36	.46	.41	.34	.59	.44	.79	.24	.35	.31	.16	.41
Duration est.		2.8	2.1	2.1	5.2	3.8	1.9	2.6	3.6	2.1	1.7	2.6	3.2	2.9	2.7
CV arrival caseload		.30	.70	.28	.36	.38	.40	.50	.37	.23	.27	.44	.60	.29	.17
Slack %		8%	87%	21%	19%	8%	22%	31%	7%	6%	9%	47%	14%	3%	20%
DT offset(%)	real	52.6	-	60.8	69.3	55.3	71.7	93.8	51.0	86.3	47.1	87.4	29.9	51.3	65.6
	FCFS	55.0	-	99.2	72.9	86.6	84.5	89.0	91.0	86.6	95.7	92.1	65.4	95.2	86.3
	scen.	37.7	-	93.5	63.9	76.3	60.5	76.3	76.9	69.8	57.5	83.1	54.1	74.0	71.5
Waiting time	real	83.9	19.5	40.1	31.9	37.2	16.0	44.4	31.1	12.6	35.5	27.6	74.2	57.4	38.2
	FCFS	15.1	3.4	3.9	13.0	8.2	6.9	4.9	6.9	4.9	5.0	7.3	20.3	9.5	7.8
	scen.	50.9	62.5	17.8	25.4	20.4	39.6	24.6	21.9	17.4	50.5	23.6	42.4	44.1	30.8
Weighted DT	real	2.6	-	2.8	2.1	3.7	1.9	.20	3.3	.86	4.3	.80	5.4	3.2	2.4
	FCFS	2.1	-	.02	2.4	.57	.49	.46	.29	.35	.09	.37	2.9	.15	.71
	scen.	5.6	-	.26	3.2	1.5	3.2	1.4	1.0	1.3	3.6	1.1	3.7	1.8	1.9

tion model is unlikely to change the conclusion of our results in a major way.

At some hospitals, capacity problems at downstream facilities such as the ICU and the PACU cause OR blocking and therefore have a detrimental effect on OR usage. It could be interesting to include those aspects into a future version of our simulation model. At the University Hospital Leuven's inpatient OR department, partly due to recent changes, blockage at downstream facilities does not pose a problem.

There are restrictions to our model that we do regard as important but did not include into the model. They mostly relate to the fact that some of the tested scheduling factors might require that a surgery is performed on brief notice, i.e., within days. This might be a problem as the patient might need some days to mentally prepare for the surgery, sort out practicalities such as getting leave from work or needs to start/stop medication. Additionally, there are cases where it is advised to wait some time before the surgery, e.g., for some knee surgeries. It is therefore important to test whether some of the scheduling factors require a larger number of patients to be served on a brief notice than in reality. For the sake of simplicity, we restrict our analysis to patients that have one day to prepare, thus need to be served next day.

Interestingly, the percentage of patients served next day in the simulation model is never larger than in reality. This is the case as all scenarios in Fig. 12 lie under the line representing the real setting. This suggests that all scheduling factors are valid and applicable in reality.

In order to get a more refined picture, we investigated how often patients need to be ready for next day
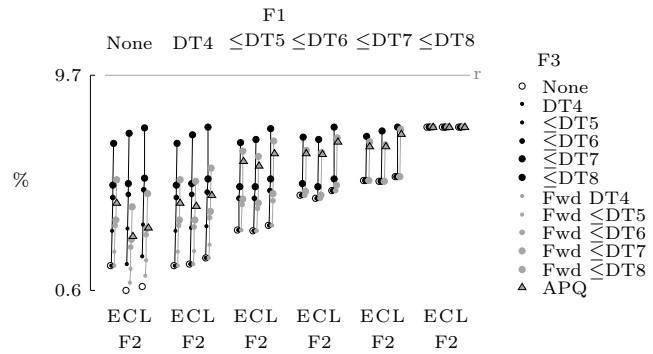

Fig. 12 Probability of a patient to be served next day. The figure shows that none of the scheduling scenarios requires an overly large proportion of patients to be available for next day surgery.

surgery from each DT category. This will be, in reality, frequently the case for DT category 4 patients. It is also generally true that DT category 4 patients are quicker and more efficiently served in reality than in the simulation model.

Our analysis suggest that FCFS and next day is consistently applicable to DT categories 4 and 5. This might not be the case for DT category 6 and 7 (our model suggests a larger number of patients served next day than reality). This suggests that patients with a less critical condition might, in reality, not always be able or willing to get surgery quickly. Consequently, factors FCFS and next day may only be applicable consistently to DT categories 4 and 5.

One of the general trends observable in the results is that the effectiveness of scheduling factors in utilizing OR capacity will determine how good it performs. This
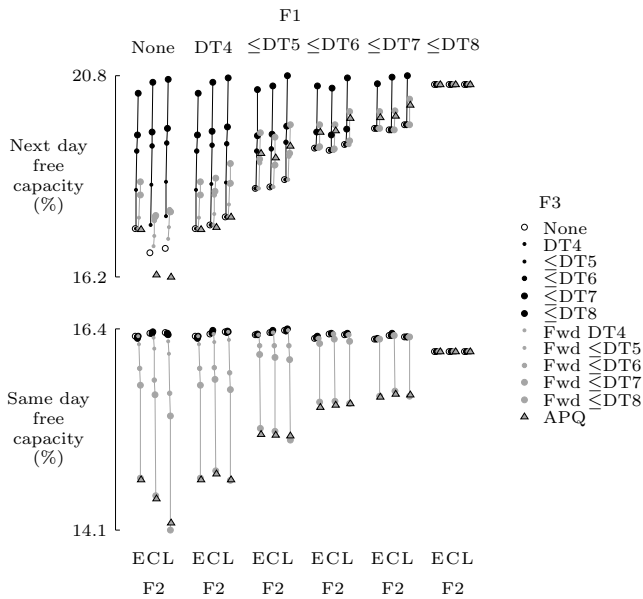
Fig. 13 The percentage of next day free capacity is observed each day morning and relates to the amount of unplanned capacity of the following day before any replanning was done. Same day free capacity relates to the amount of unplanned capacity of the same day. Same day free capacity, as patients cannot be planned for the same day, corresponds to the day's final amount of unplanned free capacity.

is shown as, firstly, FCFS which is a strategy that disregards the DT but ensures good use of OR capacity, performs very well. Secondly, replanning using the best-fit method outperforms the APQ method. This means that avoiding the waste of OR capacity is more important than saving individual patients from going over their DT. Thirdly, as shown by Fig. 13, scheduling methods that perform well also yield the highest amount of average next day free capacity. A high amount of average free next day capacity means that there is little blockage and thus OR capacity is used efficiently.

Identifying the efficient use of OR capacity as the dominant driver of patient related performances also explains some of the seemingly counterintuitive results we got. We noticed that including higher DT categories to be served FCFS does not hinder fast service of lower DT categories, e.g., scheduling DT category 5 FCFS does generally not result in less patients from DT category 4 served next day. In reality, there may be cases when DT category 5 patients do hinder quick service of DT category 4 patients. This, however, is counterbalanced by the fact that serving DT category 5 FCFS results in an improved use of OR capacity that in return also benefits most DT category 4 patients.

6 Future work

There are two main extensions planned to our work. Firstly, testing scheduling methods that use protection levels and, secondly, including a rejection mechanism into the model.

Protection levels split OR capacity into parts that can only be used by surgeries of certain DT categories. This may ensure that patients, within each DT category, will generally be served in a timely manner. For example, assuming that on average three DT category 4 patients arrive a day, capacity to serve them could be kept reserved on a daily basis.

Capacity can also be reserved in a dynamic way so that the amount of capacity protected for a certain DT category on dates in the near future is lower than for dates further in the future. This can compensate for cases when more than the expected number of patients arrive, as more capacity is protected for them on later dates. Similarly, when fewer than the expected number of patients arrive, capacity is less likely to be wasted as on dates in the near future less capacity is protected. Capacity can also be fully released on 'next days' so that it can be used by any DT category.

It is challenging to use protection levels in an inpatient setting as surgery durations are highly variable (Fig. 5). For instance, it is difficult to reserve the right amount of capacity for each DT category. If too little capacity is reserved, patients with long surgery durations will not fit. If too much capacity is reserved, some capacity will sometimes be wasted.

Protection levels introduce additional inefficiencies by dividing the capacity of complete ORs. This is a problem as surgeries will never perfectly fit them and unused capacity within the protection levels is wasted. This can have a strong detrimental effect on disciplines that generally have few ORs available (e.g., MKA) or have long surgeries (e.g., CAH). Wasting OR capacity should be avoided by all means as we showed that this will have a negative effect on patient related performance measures. Consequently, protection levels might be less suited to inpatient settings in general. They might be more effectively used in settings where patients need a regular amount of time such as appointments at the doctor, appointments for medical imaging [39], or the scheduling of surgeries in an outpatient setting (where surgeries are more standardized).

The patient rejection mechanism constitutes a further extension to our model. This would include patients into the model that originally intended to get surgery at the University Hospital Leuven but ended up getting surgery at another hospital. This new model would consequently use patient data that does not

necessarily constitute a feasible schedule, i.e., demand might be larger than supply. Including a patient rejection mechanism into the model can also help us to assess whether the benefits of using the FCFS strategy might be offset by suddenly having more patients requesting surgery.

In the literature, the patient rejection mechanism is generally modeled as a trade-off function where cost factors such as waiting time, OR overtime and OR opening costs are balanced against the profit gained from surgeries. In order to model the patient rejection mechanism of the University Hospital Leuven we need to overcome three challenges.

Firstly, we need to define an appropriate cost structure. This is challenging as besides the academic and medical relevance there are also other considerations determining the value of a surgery. Such values are related to the monetary return (defined by the reimbursement tariff), the fact that a specific expertise is present in a certain hospital and government related regulations.

Secondly, it is challenging to model the patient rejection process itself as it is only partly controlled by the hospital. Patients can legally not be rejected. As the University Hospital Leuven is for some surgeries regarded to be the most qualified hospital in the country, patients may be reluctant to go to another hospital even if asked to do so. On the contrary, there might be patients who registered at several hospitals and therefore can cancel at any arbitrary time.

Thirdly, it might be challenging to get a realistic understanding of the rejection process as we are missing the necessary data. Patients that are immediately rejected will not enter the hospitals data system. For example, some patients that went for consultation to a surgeon who has a long waiting list might have been convinced to register for surgery at another hospital. Since they never entered the hospital's data system, we have no information on their number and on their attributes.

A first approach to understand the rejection process at the hospital is to get an understanding of the value surgeons attach to each surgery type. A possible way to get this value is by performing a Delphi study amongst the surgeons.

7 Conclusion

It is a complex task to manage the ORs because, amongst others, there is a need to manage ORs efficiently while it is also important to provide an adequate level of responsiveness to non-electives [19]. The degree to which non-electives can impact the elective schedule varies from hospital to hospital. For the case of the University Hospital Leuven, we determined that the simulation model of the OR department needs to include these non-electives since they have a large impact on the OR department and the elective schedule. This is the case as non-electives instead of entering an arbitrary empty or lowly utilized OR will often be assigned to ORs that serve the corresponding discipline (Fig. 4). These ORs can already be fully planned with electives. This combined with the fact that some disciplines need to serve a large amount of non-electives (Fig. 6) means that their ORs can be heavily utilized. This can lead to overtime and elective rescheduling.

Elective rescheduling is a component that also needs to be included into the simulation model. It contributes to the fact that OR related performance measures, such as overtime, will not depend on the chosen patient scheduling strategy. The rescheduling model determines, firstly, how patients can be reassigned to ORs of different disciplines and, secondly, imitates the timing of decisions made in reality. We found that in reality most surgery reassignments and cancellations happen continuously throughout the entire day.

There are straightforward managerial implications of our results for the surgeons of the University Hospital Leuven and for schedulers of similar hospitals. A high number of patients are served within their DT if OR capacities are efficiently used. Therefore, FCFS, which is a strategy that makes good use of OR capacities, will perform well. FCFS might not always be applicable in reality as patients from less urgent urgency classes may not always be available for surgery on short notice. Therefore, it is important to allow for patient replanning. One might intuitively think that replanning high priority patients is advantageous, but we showed that it is better to replan those ones that best fill out free next day OR capacity. This is the case as the major benefit of replanning stems from saving OR capacity, i.e., valuable capacity from the originally planned date is exchanged for less valuable next day capacity that is in danger of being wasted.

We also deduced that capacity should be regarded as critical and its efficient use be kept as one of the main priorities both during patient scheduling and patient replanning procedures. This also entails that surgeons should in a timely manner release ORs that they do not use so that other surgeons can use them.

Our results also entail that evaluating the capacity assigned to surgeons should not only be based on the average waiting time but also on the used scheduling strategy. This is the case as low average waiting times do not necessarily mean that surgeons have more capacity available than needed, but it can also mean that they

use a scheduling strategy where surgeries are scheduled to close dates and therefore capacity is used efficiently. This behavior should not be punished. Similarly, Dexter et al. [9] argue that determining an appropriate amount of block time and selecting a method to schedule cases into a surgeon's blocks must be done simultaneously.

Results that are specific to the case hospital, show that DT category 4 patients are in reality overprioritized. Assigning a relatively higher priority to the other DT categories could lead to larger gains with regards to the weighted DT cost.

Our study has several limitations. This is a single center study. The hospital under consideration contains highly utilized ORs. Moreover, the hospital is situated in the Belgian context, which means that waiting times are relatively short [49, 59] and therefore using dynamic scheduling is justified. Moreover, the sequencing step is not of importance as surgeons usually own an OR for an entire day and therefore the exact surgery time does not matter, i.e., the next surgery starts when the previous one is finished.

It is also important to remember that many of our results are easy to interpret because scheduling decisions were shown to be independent of OR performance measures (e.g., overtime). This might not necessarily be true in any type of setting. For example, if surgeries cannot start before their planned start time, setting appropriate start times will likely have an effect on OR performances. In those cases a trade-off has to be found between OR and patient related performances.

In our study we included arrivals that were served by the hospital and we therefore have a feasible schedule. A higher arrival rate, however, could modify our conclusions. This is challenging to model, details were discussed in the future research section.

Neither did we investigate whether the conclusions hold in the case that non-electives are served in separate ORs. In those cases, on the one hand, less capacity would remain for electives but, on the other hand, it would bring more stability to the elective schedule. Whether this stability could offset the drawback of having less capacity is of interest to the hospital. We think the same conclusions would hold for this case, as OR related performance measures would likely remain independent of the used scheduling strategy.

There are two major extensions planned to the model. Firstly, the patient rejection process and, secondly, patient scheduling strategies that involve protection levels. We are currently looking into ways to get a realistic understanding of the rejection process. Patient rejection at the hospital is difficult to model as, on the one hand, it is only partly controlled by the hospital itself and, on the other hand, it is difficult to define

an appropriate cost structure. We are also investigating scheduling strategies that include protection levels, that is, scheduling strategies where OR capacity is reserved for a given DT category.

References

1. Adan I, Bekkers J, Dellaert N, Jeunet J, Vissers J (2011) Improving operational effectiveness of tactical master plans for emergency and elective patients under stochastic demand and capacitated resources. European Journal of Operational Research 213:290–308
2. AIHW (2013) Australian hospital statistics: National emergency access and elective surgery targets 2012. Tech. rep., Australian Institute of Health and Welfare
3. Azari-Rad S, Yontef AL, Aleman DM, Urbach DR (2013) Reducing elective general surgery cancellations at a Canadian hospital. Canadian Journal of Surgery 56:113–118
4. Barua B, Esmail N (2013) Waiting your turn: Wait times for health care in Canada. Tech. rep., Fraser Institute
5. Beliën J, Demeulemeester E (2007) Building cyclic master surgery schedules with leveled resulting bed occupancy. European Journal of Operational Research 176:1185–1204
6. Cardoen B, Demeulemeester E, Beliën J (2010) Operating room planning and scheduling: A literature review. European Journal of Operational Research 201:921–932
7. Cardoen B, Demeulemeester E, Van der Hoeven J (2010) On the use of planning models in the operating theatre: Results of a survey in Flanders. International Journal of Health Planning and Management 25:400–414
8. Demeulemeester E, Beliën J, Cardoen B, Samudra M (2013) Operating room planning and scheduling, Springer
9. Dexter F, Macario A, Traub RD, Hopwood M, Lubarsky DA (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. Anesth Analg 89:7–20

10. Dexter F, Macario A, O'Neill L (2000) Scheduling surgical cases into overflow block time: Computer simulation of the effects of scheduling strategies on operating room labor costs. Anesthesia and Analgesia 90:980–988

11. Dexter F, Macario A, Lubarsky DA (2001) The impact on revenue of increasing patient volume at surgical suites with relatively high operating room utilization. Anesthesia and Analgesia 92:1215–1221

12. Dexter F, Maxbauer T, Stout C, Archbold L, Epstein RH (2014) Relative influence on total cancelled operating room time from patients who are inpatients or outpatients preoperatively. Anesth Analg 118:1072–80

13. Epstein RH, Dexter F (2013) Rescheduling of previously cancelled surgical cases does not increase variability in operating room workload when cases are scheduled based on maximizing efficiency of use of operating room time. Anesth Analg 117:995–1002

14. Epstein RH, Dexter F (2015) Management implications for the perioperative surgical home related to inpatient case cancellations and add-on case scheduling on the day of surgery. Anesth Analg 121:206–18

15. Everett JE (2002) A decision support simulation model for the management of an elective surgery waiting system. Health Care Management Science 5:89–95

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

17. Fei H, Meskens N, Combes C, Chu C (2009) The endoscopy scheduling problem: A case study with two specialised operating rooms. International Journal of Production Economics 120:452–462

18. Fei H, Meskens N, Chu C (2010) A planning and scheduling problem for an operating theatre using an open scheduling strategy. Computers & Industrial Engineering 58:221–230

19. Ferrand YB, Magazine MJ, Rao US (2014) Managing operating room efficiency and responsiveness for emergency and elective surgeries—a literature survey. IIE Transactions on Healthcare Systems Engineering 4:49–64

20. Gocgun Y, Ghate A (2012) Lagrangian relaxation and constraint generation for allocation and advanced scheduling. Computers & Operations Research 39:2323–2336

21. Gocgun Y, Puterman ML (2014) Dynamic scheduling with due dates and time windows: an application to chemotherapy patient appointment booking. Health Care Management Science 17:60–76

22. Guerriero F, Guido R (2011) Operational research in the management of the operating theatre: A survey. Health Care Management Science 14:89–114

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

24. Hosseini N, Taaffe KM (2014) Allocating operating room block time using historical caseload variability. Health Care Management Science

25. Kleinrock L (1964) A delay dependent queue discipline. Naval Research Logistics Quarterly 11:329–341

26. Lamiri M, Dreo J, Xiaolan X (2007) Operating room planning with random surgery times. In: Proceedings of the 3th IEEE Conference on Automation Science and Engineering, pp 521–6

27. Lamiri M, Xie X, Dolgui A, Grimaud F (2008) A stochastic model for operating room planning with elective and emergency demand for surgery. European Journal of Operational Research 185:1026–1037

28. Lamiri M, Xie X, Zhang SG (2008) Column generation approach to operating theater planning with elective and emergency patients. IIE Transactions 40:838–852

29. Lamiri M, Grimaud F, Xie X (2009) Optimization methods for a stochastic surgery planning problem. International Journal of Production Economics 120:400–410

30. van der Lans M, Hans EW, Hurink JL, Wullink G, Van Houdenhoven M, Kazemier G (2006) Anticipating urgent surgery in operating room departments, University of Twente

31. Ma G, Demeulemeester E (2013) A multilevel integrative approach to hospital case mix and capacity planning. Computers & Operations Research 40:2198–2207

32. Magerlein JM, Martin JB (1978) Surgical demand scheduling: A review. Health Services Research 13:418–433

33. Marques I, Captivo ME, Vaz Pato M (2014) A bicriteria heuristic for an elective surgery scheduling problem. Health Care Management Science

34. MHallah R, Al-Roomi AH (2014) The planning and scheduling of operating rooms: A simulation approach. Computers & Industrial Engineering 78:235–248

35. Min D, Yih Y (2010) An elective surgery scheduling problem considering patient priority. Computers & Operations Research 37:1091–1099

36. Min D, Yih Y (2010) Scheduling elective surgery under uncertainty and downstream capacity con-

straints. European Journal of Operational Research 206:642–652

37. Min D, Yih Y (2014) Managing a patient waiting list with time-dependent priority and adverse events. RAIRO - Operations Research 48:53–74

38. Niu Q, Peng Q, Y ElMekkawy T (2013) Improvement in the operating room efficiency using tabu search in simulation. Business Process Management Journal 19:799–818

39. Patrick J, Puterman ML, Queyranne M (2008) Dynamic multipriority patient scheduling for a diagnostic resource. Operations Research 56:1507–1525

40. Persson M, Persson JA (2009) Health economic modeling to support surgery management at a swedish hospital. Omega-International Journal of Management Science 37:853–863

41. Pulido R, Aguirre AM, Ortega-Mier M, Garcia-Sanchez A, Mendez CA (2014) Managing daily surgery schedules in a teaching hospital: A mixed-integer optimization approach. BMC Health Services Research 14

42. Rachuba S, Werners B (2014) A robust approach for scheduling in hospitals using multiple objectives. Journal of the Operational Research Society 65:546–556

43. Riise A, Burke E (2011) Local search for the surgery admission planning problem. Journal of Heuristics 17:389–414

44. Rizk C, Arnaout JP (2012) ACO for the surgical cases assignment problem. Journal of Medical Systems 36:1891–1899

45. Sargent R (2005) Verification and validation of simulation models

46. Schlesinger S, Crosby R, Cagne R, Innis G, Lalwani C, Loch J, Sylvester R, Wright R, Kheir N, Bartos D (1979) Terminology for model credibility. SIMULATION 32:103–104

47. Sciomachen A, Tanfani E, Testi A (2005) Simulation models for optimal schedules of operating theatres. International Journal of Simulation 6:26–34

48. Shimazaki H, Shinomoto S (2010) Kernel bandwidth optimization in spike rate estimation. Journal of Computational Neuroscience 29:171–182

49. Siciliani LMB, Moran V (2013) Waiting Time Policies in the Health Sector. OECD Publishing

50. Stanford D, Taylor P, Ziedins I (2014) Waiting time distributions in the accumulating priority queue. Queueing Systems 77:297–330

51. Strum DP, May JH, Vargas LG (2000) Modeling the uncertainty of surgical procedure times - comparison of log-normal and normal models. Anesthesiology 92:1160–1167

52. Testi A, Tanfani E, Torre G (2007) A three-phase approach for operating theatre schedules. Health Care Management Science 10:163–72

53. Tewari A, Giering MJ, Raghunathan A (2011) Parametric characterization of multimodal distributions with non-gaussian modes. In: 2011 IEEE 11th International Conference on Data Mining Workshops (ICDMW),, pp 286–292

54. Travis E, Woodhouse S, Tan R, Patel S, Donovan J, Brogan K (2014) Operating theatre time, where does it all go? a prospective observational study. Bmj 349:g7182

55. Trivedi P, Zimmer D (2006) Copula modeling: An introduction for practitioners. Foundations and Trends® in Econometrics 1:1–111

56. Van Houdenhoven M, van Oostrum JM, Hans EW, Wullink G, Kazemier G (2007) Improving operating room efficiency by applying bin-packing and portfolio techniques to surgical case scheduling. Anesthesia and Analgesia 105:707–714

57. Vansteenkiste N, Lamote C, Vandersmissen J, Luysmans P, Monnens P, De Voldere G, Kips J, Rademakers FE (2012) Reallocation of operating room capacity using the due-time model. Medical Care 50:779–784

58. van Veen-Berkx E, Bitter J, Elkhuizen SG, Buhre WF, Kalkman CJ, Gooszen HG, Kazemier G (2014) The influence of anesthesia-controlled time on operating room scheduling in dutch university medical centres. Can J Anaesth 61:524–32

59. Viberg N, Molin R, Forsberg B (2011) Swedish waiting times for health care in an international perspective. Tech. rep., sALAR

60. Vijayakumar B, Parikh PJ, Scott R, Barnes A, Gallimore J (2013) A dual bin-packing approach to scheduling surgical cases at a publicly-funded hospital. European Journal of Operational Research 224:583–591