Flexible master surgery scheduling: combining optimization and simulation in a rolling horizon approach

(Article begins on next page)

Flexible Master Surgery Scheduling: Combining optimization and simulation in a rolling horizon approach

Mariana Oliveira[1*]
mariana.m.oliveira@tecnico.ulisboa.pt; ORCID iD: 0000-0001-8739-7650
Filippo Visintin[2]
filippo.visintin@unifi.it; ORCID iD: 0000-0003-3262-6844
Daniel Santos[1]
daniel.rebelo.santos@tecnico.ulisboa.pt; ORCID iD: 0000-0003-0611-2460
Inês Marques[1]
ines.marques.p@tecnico.ulisboa.pt; ORCID iD: 0000-0002-1663-3713

[1]Centre for Management Studies, Instituto Superior Técnico, University of Lisbon
Av. Rovisco Pais 1, 1049-001 Lisbon, Portugal
[2]Information-Based Industrial Services Laboratory, Department of Industrial Engineering, University of Florence, PO Box 50134, Viale Morgagni 40/44, Florence, Italy

Abstract

Operating room managers are facing increasingly complex challenges, namely in complying with waiting time targets before surgery. This paper proposes a framework that combines optimization and simulation to generate dynamic master surgery schedules for a long planning horizon, in which the schedules are optimized by an integer programming model and the demand levels are modelled using the simulation model. The developed approach allows the resulting operating room plan to balance waiting lists as it assigns more time to the specialties with higher demand in terms of time needed to perform all the surgeries in the corresponding waiting lists. The analysis of the results obtained for the proposed flexible rolling horizon approach were proven robust, and were compared to static and flexible long-term approaches, the former not allowing flexibility and the latter using a deterministic update of the demand. Considering throughput, tardiness and waiting time, the flexible rolling horizon approach showed the best results, while the static one had the worst results.

Keywords: Health Services, Operating Rooms, Master Surgery Schedule, Strategic and Tactical Decisions, Optimization, Simulation

*Corresponding author.

Acknowledgements

The authors are grateful for the cooperation with a Portuguese hospital. They allowed us deep insights into the operating room scheduling processes and provided the data to conduct this study. This research is supported by the Portuguese National Science Foundation (Fundação para a Ciência e a Tecnologia, FCT) under project PTDC/EGEOGE/30442/2017, Lisboa-01.0145-Feder-30442, and a PhD scholarship, application reference 2020.09648.BD. The authors also acknowledge the anonymous reviewers for their comments, which helped us improve the quality of the manuscript.

Flexible Master Surgery Scheduling: Combining optimization and simulation in a rolling horizon approach

Abstract

Operating room managers are facing increasingly complex challenges, namely in complying with waiting time targets before surgery. This paper proposes a framework that combines optimization and simulation to generate dynamic master surgery schedules for a long planning horizon, in which the schedules are optimized by an integer programming model and the demand levels are modelled using the simulation model. The developed approach allows the resulting operating room plan to balance waiting lists as it assigns more time to the specialties with higher demand in terms of time needed to perform all the surgeries in the corresponding waiting lists. The analysis of the results obtained for the proposed flexible rolling horizon approach were proven robust, and were compared to static and flexible long-term approaches, the former not allowing flexibility and the latter using a deterministic update of the demand. Considering throughput, tardiness and waiting time, the flexible rolling horizon approach showed the best results, while the static one had the worst results.

Keywords: Health Services, Operating Rooms, Master Surgery Schedule, Strategic and Tactical Decisions, Optimization, Simulation

1 Introduction

The ageing of the worldwide population, the increasing demand for health services and the development of new and expensive technologies are creating increasingly complex challenges in the healthcare sector. Within a hospital, operating rooms are the center of costs and revenues, representing almost half of the economical transactions of the organization (Gao et al., 2013). Nonetheless, operating room planners have a huge challenge to efficiently schedule the operating rooms if they are not given adequate tools. Their work involves many variability and uncertainty factors and have substantial direct implications in the health status of patients waiting for surgery. Ideally, waiting time for surgery should not be longer than a predetermined maximum waiting time, which is an important healthcare access indicator. Thus, these waiting time limits should be considered to achieve a higher service level. To comply with waiting time targets, operating room time should be optimally shared among specialties to allow following surgical demand pattern and increase operating room utilization rates.

Operating room planning and scheduling decisions can be structured in three sequential levels (see, e.g. Zhu et al., 2019): first, in the long term, case mix problems decide on operating room time share among the surgical specialties (strategic level); second, in a medium term, master surgery scheduling assigns specific operating room slots to specialties (tactical level); and, finally, surgery scheduling problems select patients from a waiting list and obtain a short term schedule for these patients (operational level). This work integrates strategic and tactical decisions, by deciding on how much operating room time (strategic) and when (tactical) to assign to each surgical specialty. To address this problem, a systematic method is proposed to design dynamic master surgery schedules (MSSs) where the operating room time share among specialties is not given as input but is adjusted depending on the evolution of the demand pattern and the availability of staff. Considering dynamic MSSs, instead of the commonly used static ones, represents a trade-off between two opposite criteria: flexibility and stability. The former aims to allow changes in the weekly and monthly MSSs to better chase surgical demand whereas the latter aims to satisfy the staff (e.g. surgeons).

Accordingly, this paper introduces a novel framework to generate dynamic weekly MSSs for a long planning horizon. This new planning method follows fluctuations in the waiting lists of each surgical specialty by assigning more operating room time to the specialties with the most expected demand. The long planning period allows the surgeons to better manage their agendas and cope with changes in the MSS from week to week, which provides them satisfaction and improves the willingness to accept the resulting schedules. Moreover, in the literature, demand patterns are usually defined by the number of patients in the waiting list. This work, however, considers both the number of patients in the waiting list and the expected surgical duration for these patients. Thus, demand levels are defined in terms of time needed to perform all the surgeries registered in the waiting list. Finally, a simulation model is implemented to assess the impact of the proposed framework on the waiting list performance (i.e. waiting time and tardiness) over the long run.

Thus, this work presents original features in: (i) considering demand-chasing objectives in the strategic and tactical decision levels; (ii) studying waiting list dynamics and flexibility in MSSs; (iii) proposing an iterative combined optimization-simulation approach, in which the MSS is optimized by a mixed-integer linear programming model and the uncertain demand levels are modelled through simulation; and (iv) integrating optimization and simulation approaches to assess the impact of the proposed planning models in waiting list indicators. This approach is used to answer the three research questions addressed in this paper:

1. Is it possible to better chase surgical demand by optimizing and allowing for some flexibility (e.g. weekly and monthly changes) in the MSS?

2. Is it possible to better chase demand by combining optimization and simulation approaches when generating MSSs?

3. Is it possible to make the planning process overall more equitable (namely by giving higher priority to those patients whose due date has already expired or is going to expire)?

The remainder of this paper is structured as follows. Section 2 summarizes the state-of-the-art on the tactical decision level for operating room planning and scheduling, namely regarding simulation approaches. The problem under study and the proposed sequence of optimization and simulation approaches are detailed in Section 3. Section 4 discusses the results of the application of the proposed approach to the case study data, and provides answers to the research questions. Moreover, Section 5 discusses the managerial implications of using the proposed tool in a practical context. Finally, Section 6 concludes the paper.

2 Literature review

Most papers in operating room planning and scheduling focus on operational decisions, i.e. surgery scheduling (Zhu et al., 2019). These papers typically receive an MSS as input, meaning that patients are scheduled to slots assigned to the corresponding specialty (see, e.g. Schneider et al., 2020; M'Hallah and Visintin, 2019; Marques and Captivo, 2017; Addis et al., 2016). Thus, there is no flexibility here to manage the waiting list among specialties according to the surgical demand dynamic. This is the so-called block scheduling strategy. Other papers integrate surgery scheduling with patient prioritization to improve access to surgical services (Oliveira et al., 2020). To increase flexibility in waiting list management, some papers integrate tactical and operational decisions. An open scheduling strategy may be used where operating room time is shared according to the criteria considered for surgery scheduling (Molina-Pariente et al., 2018; Marques et al., 2012). Sometimes, this integration is also performed under a block scheduling strategy, where a cyclic MSS is designed together with a decision on the individual patients to be treated in each slot (Moosavi and Ebrahimnejad, 2020; Spratt and Kozan, 2016; Agnetis et al., 2014). The latter two papers consider a minimum and maximum number of slots to be assigned to each specialty but no information is provided on how these values are fixed. Although to a lesser extent in the literature, a combination of block and open scheduling strategies (modified block scheduling) intends to take advantage of a stable MSS but still allowing for some flexibility to accommodate changes in the demand (Kamran et al., 2019). However, there is still a lack of studies in the literature that focus on flexibility (Visintin et al., 2016).

In this paper, we focus on the integration of strategic and tactical decisions by deciding on how much operating room time and when to assign to each specialty. Marques et al. (2019) also focus on integrating these two decision levels. The multi-objective mixed integer programming model aims to level workload at downstream units, avoid sharing operating room time among different specialties, allocate operating room time according to the number of surgeons available and to recent utilization of operating room time. The latter is done by minimizing deviations of the weekly operating room time assigned to each specialty from the median value of the weekly time used by the specialty in the previous trimester. This procedure aims to stimulate productivity more than taking into account evolution of the surgical demand for each specialty. Penn et al. (2017) create cyclic MSS considering surgeons' availability and preferences, limited equipment availability and smoothing bed usage. Subject to these goals, the decisions on how much operating room time to assign to each specialty are only implicitly considered. Studies solely on tactical decisions often build a cyclic MSS considering a fixed operating room time to each specialty or group of surgeons stated in the strategic level of decision (case mix planning) (see, e.g. Beliën and Demeulemeester, 2007). These problems focus mainly on designing a fixed MSS which minimizes workload variability in downstream resources while considering uncertainty in the length of stay of the patients in these units. In all these situations, surgical demand is not fully considered for capacity allocating decisions and no attempts are done to chase this demand.

Integrated papers focusing on the tactical level tend more to include operational decisions such as the number of patients of a certain type to be scheduled in each slot. Surgery types are often sets of surgeries that are similar in economic and resource usage perspective. This is the case in, e.g., Bovim et al. (2020); Anjomshoa et al. (2018); Kumar et al. (2018); Dellaert and Jeunet (2017); Visintin et al. (2016); Cappanera et al. (2014); Banditori et al. (2013); van Oostrum et al. (2008). Once again, decisions on how much operating room time to assign to each specialty are not usually considered or do not take into account issues such as equity or waiting list length. Thus, surgical demand dynamic is not handled. Among these studies, it is important to highlight Visintin et al. (2016). The authors use simulation to investigate the effects associated with flexibility management of surgical teams, operating rooms and surgical units. However, this is based on scenarios obtained by modifying variables, parameters and constraints in a mixed integer programming model. Moreover, in most of these studies the surgical demand is not explicitly modelled and the assignment of operating room time to specialties aims at maximizing the number of surgeries planned or patient throughput (see, e.g. Banditori et al., 2013) often bounded by the availability of resources (i.e. surgical teams) or to balance workload at downstream units. In Dellaert and Jeunet (2017), a target number of patients for each type is to be achieved, although this number is set based on the average number of operated patients in past periods.

Although building a four-week cycle MSS which repeats along the year, Anjomshoa et al. (2018) starts from a base MSS and only allows a maximum number of differences between this base MSS and the proposed plan. Moreover, every six months slightly changes to the base plan are allowed. Similarly to our paper, Agnetis et al. (2012) also investigate the long-term trade-off between stability and flexibility for determining weekly MSSs. The authors claim that, in the latter, an MSS is dynamically adapted to the current state of the waiting lists. However, the assignment of slots to specialties is bounded by upper and lower limits which may arise from workload balancing goals and the number of available beds without considering the number of patients in the waiting list and the expected surgical duration required for these patients. Nonetheless, by integrating decisions on weekly MSSs and surgery scheduling (tactical and operational decision levels), the authors show that introducing a very limited degree of flexibility in the MSS can largely pay off in terms of resource efficiency and due date performance. This idea is further developed in our paper mostly to consider real-world features and explicit chasing demand strategies. First, we design weekly MSSs for a one-year planning horizon. The reason is that surgeons want to know and organize their agenda well in advance and only making this information available in the previous week is not acceptable by these operating room stakeholders. As a consequence, the dynamic of the waiting list is captured via simulation instead of being a result of the implementation of an optimal weekly surgical schedule. This supports the combination of optimization and simulation approaches followed in our paper. Finally, the assignment of slots to specialties follows the dynamic of the surgical demand in terms of the number of patients waiting and of the expected time required to perform the surgeries of these patients.

In the master surgical scheduling literature, simulation has often been used to assess the performance of an optimization model solution. Banditori et al. (2013), for example, use simulation to assess the robustness of solutions of their mixed integer programming model and to dimension capacity slacks needed for the model to return robust solutions. Cappanera et al. (2014), instead, use simulation to compare alternative objective functions of their mixed integer programming model in terms of efficiency, balancing and robustness. Visintin et al. (2017)

advocate the use of discrete event simulation to adjust optimization model parameters prior to implementation. These studies address operational or tactical decision levels and use simulation either to (i) assess the short to medium-term effects of the implementation of deterministic optimization models when certain variables (e.g. surgical times and length of stay) are left to vary according with suitable probability distributions (typically, lognormal or empirical), or (ii) set model parameters to allow for robust solutions. In all these studies, the length of the simulation run corresponds to the time-horizon of the optimization model.

Zhang et al. (2009), instead, propose a Monte Carlo simulation, which considers a longer time horizon (100 weeks) and a random patient arrival process, to assess the quality of the solution returned by their optimization model. In their study, however, the optimization model (with a one-week planning horizon) is run at the beginning of the simulation and the solution is used as-is throughout the simulation run. Thus, the optimization model solution does not change depending on the demand. The aim of the paper is to assess how certain key performance indicators (e.g. throughput and operating room utilization) vary when using a "fixed" solution in a stochastic environment for a long time.

To the best of our knowledge, however, the operating room planning and scheduling literature lacks studies assessing how the implementation of a scheduling model may impact waiting list related performance - such as the cumulative patients tardiness - over the long run (Cappanera et al. (2019) performed such an assessment in a magnetic resonance imaging setting). This type of contributions are most needed, as the demand for elective procedures varies both in terms of volume and mix along the year (Visintin et al., 2017) and being capable of adjusting the MSS to take into consideration such a variation can lead to significant benefits.

Summing up, most operating room planning studies do not focus on flexibility issues in planning decisions, demand dynamics or assessment of scheduling model impact in the waiting list. Thus, this paper solves strategic and tactical problems by proposing a combination of optimization and simulation approach to study how allowing for some flexibility and following the dynamic demand in the optimization planning model impacts the results of the simulation model in terms of waiting time, tardiness and throughput.

3 Methods

This work deals with capacity allocation to adequately answer to surgical demand variation. Since operating room stakeholders prefer to know their agendas well in advance, the capacity allocation is made for a long planning horizon based on expected demand. Thus, this paper handles the assignment of operating room time among surgical specialties, i.e. it aims to build weekly MSSs for long planning horizons. The idea is to, on the one hand, regularly adjust capacity to expected demand and, on the other hand, provide the surgeons with information on the MSS changes as early as possible to increase receptiveness and implementation potential of the proposed approach.

When sharing operating room time among specialties, fairness is an important aspect to consider, however, it can be measured in many different ways. From the hospital point of view, the main interest should be to reduce waiting lists and, consequently, waiting times. Therefore, operating room times should be assigned to specialties based on their waiting lists. However, waiting lists are dynamic, with patients leaving and entering them every day. Thus, having a static MSS is certainly not the most efficient way to manage operating rooms. Many factors can be considered when measuring the time needed for each specialty, namely the length of waiting lists or the expected duration of surgeries being considered. The waiting list should also be updated based on the operating room production from the previous periods and considering the new entries.

In this section, an approach resulting from the combination of optimization and simulation is proposed to generate a MSS. The optimization model receives information about the waiting lists, expected surgery duration and availability of the staff and constructs an optimal MSS that serves as an input for the simulation model. This model simulates the patients that enter and exit the waiting list (new arrivals and scheduled patients, respectively).

Although the main rolling horizon approach we study in this paper works in the way described, we present more general optimization and simulation models that can also operate in different configurations. More precisely, the optimization model is presented as to provide a MSS for the whole year following deterministic updates of the waiting list (Section 3.2), while the simulation model can be used to simply test the outcome of the optimization model and assess the impact on the waiting list performance in the long run (Section 3.3). Section 3.1 defines the problem and introduces the notation.

3.1 Problem and definitions

This master surgery scheduling problem assigns each surgical specialty $s \in S$ to the available operating room time for each day $k \in K$ of the planning horizon. The operating room time is divided into slots that can be characterized by the month $m \in M$, week $w \in W$, working day $d \in D$, shift $b \in B$ and operating room $r \in R$. In each operating room r, the equipment is limited and it is only suitable for specialties $s \in S_r$. Each of the weekly available slots, $slots_w$, has a total duration of θ hours that should be assigned to the surgeries of each specialty, according to their expected duration dur_s. The assumption of equal duration for each slot is aligned with the literature and with practice in most hospitals. Each specialty should be scheduled, at least, mw_{sw} times in a month.

For a specialty to be assigned to a slot, staff must be also considered. Each surgeon $i \in I$ and anesthetist $a \in A$ has some constraints when to perform the surgeries. Only a surgeon with skills to perform surgeries of specialty s, $i \in I_s$, can be assigned to a slot of specialty s. Moreover, the daily availability of doctors a^{surgD}_{iwd} (takes value 1 if available and 0 if not available) and anesthetists a^{anestD}_{awd} (takes value 1 if available and 0 if not available) needs to be considered: in each shift, only a^{surg}_{swdb} surgeons of specialty s and a^{anest}_{wdb} anesthetists are available to perform surgeries. On the one hand, the law in most countries states a minimum number of surgeons δ^{surg} and anesthetists δ^{anest} to be assigned to each open slot. Conversely, each surgeon and anesthetist has a maximum number of weekly slots, ww^{surg}_i and ww^{anest}_a respectively, in which they can work on. For the staff, it is important to guarantee some stability in the schedule plans. Thus, a maximum number of different specialty assignments Δ^W_w is established between each week $w \in W_m$ of some month and the first week of the month. Moreover, a maximum number of changes Δ^M_m is also defined between each month.

Based on the MSS, patients are scheduled for surgery. In each slot assigned to specialty s, λ_s patients are expected to be scheduled. Therefore, the waiting list is dynamic. The number of patients of each specialty initially in the waiting list $inic_s$ must be adapted according to the expected production in the following period and the expected number of new entries in the waiting list ent_{sw}. Considering these three factors, it is possible to obtain the waiting list length update in the beginning of each week. Before and after surgery, the patient of specialty $s \in S_z$ goes through a series of up- and downstream units $z \in Z$, respectively. When a specialty is assigned to a slot, an expected number of beds e_{szk} are reserved for a maximum of n_{zs} days, before (for upstream units) or after (for downstream units) surgery. However, each unit has a limited available capacity c_{zk}.

Table 1 summarizes the notation used in the proposed optimization model. The optimization model has two main groups of decision variables and six groups of auxiliary variables. Variables t^-_{sw} and t^+_{sw} represent, respectively, the negative and positive deviations of the allocated time with respect to the dynamic target value for specialty s on week w (t_{sw}). Then, variable x_{swdbr} takes value 1 if specialty s is assigned to operating room r on week w, day d and shift b; and 0, otherwise. The auxiliary variables t_{sw} update the target time allocation for specialty s in week w, based on the value of p_{sw}, the auxiliary variable that defines the number of patients of specialty s on the waiting list in the beginning of week w. Moreover, f_{zk} update the expected number of patients in unit z on day k. The auxiliary variable y_{swdbr} takes value 0 if specialty s is assigned on week w to the same operating room, day d and shift b as the first week of the same month; and 1, otherwise. Similarly, j_{swdbr} takes value 0 if specialty s is assigned on week w to the same operating room r, day d and shift b as in the corresponding week on the first month of the planning horizon; and 1, otherwise.

Table 1: Indices, sets, subsets, parameters and variables for the mathematical model

Indices and Sets	
$s \in \mathrm{S}$	specialties
$m \in \mathrm{M}$	months
$w \in \mathrm{W}$	weeks
$d \in \mathrm{D}$	weekly working days
$k \in \mathrm{K}$	days in the planning horizon; the first day of the planning horizon is $k = 1$
$r \in \mathrm{R}$	operating rooms
$b \in \mathrm{B}$	shifts
$i \in \mathrm{I}$	surgeons
$a \in \mathrm{A}$	anesthesiologists
$z \in \mathrm{Z}$	up- and downstream units
Subsets	
W_m	weeks of month m; the first week of month m is w_{1m}
S_z	specialties that use unit z
S_r	specialties for which the operating room r has the necessary materials
I_s	surgeons of specialty s
$\mathrm{Z_u}$	upstream units
$\mathrm{Z_d}$	downstream units
Parameters	
ϕ_s	average waiting time of patients in the initial waiting list of specialty s
$slots_w$	total number of time slots (combination of day d, shift b and operating room r) in week w
θ	duration of each slot (in hours)
a^{surg}_{swdb}	number of surgeons of specialty s available on week w, day d and shift b
a^{surgD}_{iwd}	1, if surgeon i is available on at least one shift on week w and day d; 0, otherwise
a^{anest}_{wdb}	number of anesthesiologists available on week w, day d and shift b
a^{anestD}_{awd}	1, if anesthesiologist a is available on at least one shift on week w and day d; 0, otherwise
$inic_s$	number of patients of specialty s in the waiting list in the first day of the planning horizon
ent_{sw}	number of patients of specialty s entering the waiting list on week w
ww^{surg}_i	maximum weekly workload for surgeon i
ww^{anest}_a	maximum weekly workload for anesthesiologist a
mw_{sm}	minimum workload for specialty s on month m
Δ^M_m	monthly stability for month m
Δ^W_w	weekly stability for week w
e_{szk}	expected number of patients of specialty s is in unit z on day k before (for upstream units) or after (for downstream units) the surgery
dur_s	average duration of a surgery of specialty s (in hours)
λ_s	maximum number of patients operated per slot by specialty s
n_{zs}	maximum number of days that a patient of specialty s stays in unit z
c_{zk}	available capacity of unit z on day k
δ^{surg}	minimum requested number of surgeons available per slot assigned
δ^{anest}	minimum requested number of anesthesiologists available per slot assigned
Decision variables	
t^-_{sw}, t^+_{sw}	negative and positive deviations of the allocated time to the target value for specialty s on week w, respectively (compared to the target utilization value), respectively
Auxiliary variables	
x_{swdbr}	1, if specialty s is assigned to operating room r on week w, day d and shift b; 0, otherwise
t_{sw}	target time allocation for specialty s in week w
y_{swdbr}	1, if specialty s is not assigned on week w to the same operating room r, day d and shift b as the first week of the same month; 0, otherwise
j_{swdbr}	1, if specialty s is not assigned on week w to the same operating room r, day d and shift b as in the corresponding week on the first month of the planning horizon; 0, otherwise
f_{zk}	expected number of patients in unit z on day k
p_{sw}	number of patients of specialty s on the waiting list in the beginning of week w

3.2 Optimization model

The optimization model is as follows. Objective function (1) minimizes the deviations of the assigned operating room time to the target value, weighted by the initial average waiting time of patients in the respective waiting lists. Based on a block scheduling approach, constraints (2) force a maximum of one specialty to be assigned to each available slot. Moreover, when defining the MSS, a maximum number of slots is available in each week, as stated by constraints (3). Constraints (4) force specialties to only be assigned to operating rooms which have the necessary equipment or material. Constraints (5) guarantee that a slot is only assigned to a specialty if there is a minimum number of surgeons available. Furthermore, constraints (6) define that a surgeon can only be assigned to a slot if this individual surgeon is available during that slot and avoid assigning consecutive slots to the same surgeon. Constraints (7) restrict the weekly number of slots assigned to a surgeon. Similar constraints are defined for anesthetists in constraints (8)-(10). Constraints (11) establish a minimum number of slots to be assigned monthly to each specialty. Based on the MSS of the previous week, the expected number of patients to be scheduled in each slot, and the expected new entries in the previous week, constraints (12) update the length of the waiting lists of each specialty in the beginning of each week. Moreover, constraints (13) define the initial waiting lists. Constraints (14) specify the target value for the assignment of operating room time to specialties, based on their waiting list and expected duration of surgeries. Constraints (15) characterize the weekly positive and negative deviation of the operating room time target value for each specialty. Constraints (16) and (18) define the weekly and monthly stability measures (number of changes). Observe that they are written in their non-linear form, however they were linearized for the computational experiments as we show below. Constraints (17) and (19) limit the number of weekly and monthly changes on the MSS, respectively. Constraints (20) and (21) define the expected number of patients in each day and each up and downstream unit, respectively, based on the MSS of the previous and subsequent periods. Moreover, constraints (22) limit the beds used in each unit. Finally, constraints (23) and (24) set the domains for the decision and auxiliary variables.

$$\min \quad \sum_{s\in\mathrm{S}}\sum_{w\in\mathrm{W}} \phi_s \left(t^-_{sw} + t^+_{sw}\right) \tag{1}$$

$$\text{s.t.:} \quad \sum_{s\in\mathrm{S}} x_{swdbr} \le 1 \quad \forall w\in\mathrm{W}, d\in\mathrm{D}, b\in\mathrm{B}, r\in\mathrm{R} \tag{2}$$

$$\sum_{s\in\mathrm{S}}\sum_{d\in\mathrm{D}}\sum_{b\in\mathrm{B}}\sum_{r\in\mathrm{R}} x_{swdbr} \le slots_w \quad \forall w\in\mathrm{W} \tag{3}$$

$$\sum_{s\in\mathrm{S}\setminus\mathrm{S_r}} x_{swdbr} = 0 \quad \forall w\in\mathrm{W}, d\in\mathrm{D}, b\in\mathrm{B}, r\in\mathrm{R} \tag{4}$$

$$\delta^{surg}\sum_{r\in\mathrm{R}} x_{swdbr} \le a^{surg}_{swdb} \quad \forall s\in\mathrm{S}, w\in\mathrm{W}, d\in\mathrm{D}, b\in\mathrm{B} \tag{5}$$

$$\delta^{surg}\sum_{b\in\mathrm{B}}\sum_{r\in\mathrm{R}} x_{swdbr} \le \sum_{i\in\mathrm{I}_s} a^{surgD}_{iwd} \quad \forall s\in\mathrm{S}, w\in\mathrm{W}, d\in\mathrm{D} \tag{6}$$

$$\delta^{surg}\sum_{d\in\mathrm{D}}\sum_{b\in\mathrm{B}}\sum_{r\in\mathrm{R}} x_{swdbr} \le \sum_{i\in\mathrm{I}_s} ww^{surg}_i \quad \forall s\in\mathrm{S}, w\in\mathrm{W} \tag{7}$$

$$\delta^{anest}\sum_{s\in\mathrm{S}}\sum_{r\in\mathrm{R}} x_{swdbr} \le a^{anest}_{wdb} \quad \forall w\in\mathrm{W}, d\in\mathrm{D}, b\in\mathrm{B} \tag{8}$$

$$\delta^{anest}\sum_{s\in\mathrm{S}}\sum_{b\in\mathrm{B}}\sum_{r\in\mathrm{R}} x_{swdbr} \le \sum_{a\in\mathrm{A}} a^{anestD}_{awd} \quad \forall w\in\mathrm{W}, d\in\mathrm{D} \tag{9}$$

$$\delta^{anest}\sum_{s\in\mathrm{S}}\sum_{d\in\mathrm{D}}\sum_{b\in\mathrm{B}}\sum_{r\in\mathrm{R}} x_{swdbr} \le \sum_{a\in\mathrm{A}} ww^{anest}_a \quad \forall w\in\mathrm{W} \tag{10}$$

$$\sum_{w\in\mathrm{W}_m}\sum_{d\in\mathrm{D}}\sum_{b\in\mathrm{B}}\sum_{r\in\mathrm{R}} x_{swdbr} \ge mw_{sm} \quad \forall s\in\mathrm{S}, m\in\mathrm{M} \tag{11}$$

$$p_{sw} = p_{s,w-1} + ent_{s,w-1} - \sum_{d\in\mathrm{D}}\sum_{b\in\mathrm{B}}\sum_{r\in\mathrm{R}} \lambda_s x_{s,w-1,d,b,r} \quad \forall s\in\mathrm{S}, w\in\mathrm{W}\setminus\{1\} \tag{12}$$

$$p_{s1} = inic_s \quad \forall s\in\mathrm{S} \tag{13}$$

$$t_{sw} = p_{sw} dur_s \quad \forall s\in\mathrm{S}, w\in\mathrm{W} \tag{14}$$

$$\theta\sum_{d\in\mathrm{D}}\sum_{b\in\mathrm{B}}\sum_{r\in\mathrm{R}} x_{swdbr} + t^-_{sw} - t^+_{sw} = t_{sw} \quad \forall s\in\mathrm{S}, w\in\mathrm{W} \tag{15}$$

$$|x_{swdbr} - x_{sw_{1m}dbr}| = y_{swdbr} \quad \forall s\in\mathrm{S}, w\in\mathrm{W}_m\setminus\{w_{1m}\}, m\in\mathrm{M}, d\in\mathrm{D}, b\in\mathrm{B}, r\in\mathrm{R} \tag{16}$$

$$\sum_{s\in\mathrm{S}}\sum_{d\in\mathrm{D}}\sum_{b\in\mathrm{B}}\sum_{r\in\mathrm{R}} y_{swdbr} \le \Delta_w \quad \forall w\in\mathrm{W} \tag{17}$$

$$|x_{swdbr} - x_{sldbr}| = j_{swdbr} \quad \forall s\in\mathrm{S}, w\in\mathrm{W}_m, m\in\mathrm{M}\setminus\{1\}, l = w - \sum_{g<m}|\mathrm{W}_g|, d\in\mathrm{D}, b\in\mathrm{B}, r\in\mathrm{R} \tag{18}$$

$$\sum_{s\in\mathrm{S}}\sum_{w\in\mathrm{W}_m}\sum_{d\in\mathrm{D}}\sum_{b\in\mathrm{B}}\sum_{r\in\mathrm{R}} j_{swdbr} \le \Delta_m \quad \forall m\in\mathrm{M} \tag{19}$$

$$f_{zk} = \sum_{s\in\mathrm{S}_z}\sum_{b\in\mathrm{B}}\sum_{r\in\mathrm{R}}\sum_{l=0}^{\mathrm{n}_{zs}-1} e_{szl} x_{s,w,d+l,b,r} \quad \forall z\in\mathrm{Z_u}, k\in\mathrm{K}: k\to(w,d), w\in\mathrm{W}, d\in\mathrm{D} \tag{20}$$

$$f_{zk} = \sum_{s\in\mathrm{S}_z}\sum_{b\in\mathrm{B}}\sum_{r\in\mathrm{R}}\sum_{l=0}^{\mathrm{n}_{zs}-1} e_{szl} x_{s,w,d-l,b,r} \quad \forall z\in\mathrm{Z_d}, k\in\mathrm{K}: k\to(w,d), w\in\mathrm{W}, d\in\mathrm{D} \tag{21}$$

$$f_{zk} \le c_{zk} \quad \forall z\in\mathrm{Z}, k\in\mathrm{K} \tag{22}$$

$$t^-_{sw}, t^+_{sw}, f_{zk} \ge 0 \quad \forall s\in\mathrm{S}, w\in\mathrm{W}, z\in\mathrm{Z}, k\in\mathrm{K} \tag{23}$$

$$x_{swdbr}, y_{swdbr}, j_{swdbr} \in \{0,1\} \quad \forall s\in\mathrm{S}, w\in\mathrm{W}, d\in\mathrm{D}, b\in\mathrm{B}, r\in\mathrm{R} \tag{24}$$

In order to linearize constraints (16) and (18), it is sufficient to replace them with the following:

$$x_{swdbr} - x_{sw_{1m}dbr} \leq y_{swdbr} \quad \forall s \in \mathrm{S}, w \in \mathrm{W}_m \setminus \{w_{1m}\}, m \in \mathrm{M}, d \in \mathrm{D}, b \in \mathrm{B}, r \in \mathrm{R} \tag{25}$$

$$x_{sw_{1m}dbr} - x_{swdbr} \leq y_{swdbr} \quad \forall s \in \mathrm{S}, w \in \mathrm{W}_m \setminus \{w_{1m}\}, m \in \mathrm{M}, d \in \mathrm{D}, b \in \mathrm{B}, r \in \mathrm{R} \tag{26}$$

$$x_{swdbr} - x_{sldbr} \leq j_{swdbr} \quad \forall s \in \mathrm{S}, w \in \mathrm{W}_m, m \in \mathrm{M} \setminus \{1\}, l = w - \sum_{g<m} |\mathrm{W}_g|, d \in \mathrm{D}, b \in \mathrm{B}, r \in \mathrm{R} \tag{27}$$

$$x_{sldbr} - x_{swdbr} \leq j_{swdbr} \quad \forall s \in \mathrm{S}, w \in \mathrm{W}_m, m \in \mathrm{M} \setminus \{1\}, l = w - \sum_{g<m} |\mathrm{W}_g|, d \in \mathrm{D}, b \in \mathrm{B}, r \in \mathrm{R} \tag{28}$$

In fact, if $y_{swdbr} = 0$ (resp. $j_{swdbr} = 0$), then x_{swdbr} and $x_{sw_{1m}dbr}$ (resp. x_{swdbr} and x_{sldbr}) must both have the same value, either 0 or 1, which is equivalent to what is enforced with the non-linear constraints (16) (resp. (18)). Moreover, if x_{swdbr} and $x_{sw_{1m}dbr}$ (resp. x_{swdbr} and x_{sldbr}) have different values, then $y_{swdbr} = 1$ (resp. $j_{swdbr} = 1$). The only additional case not feasible for the non-linear constraints but that the linearized constraints allow is when both x_{swdbr} and $x_{sw_{1m}dbr}$ (resp. x_{swdbr} and x_{sldbr}) have the same value but $y_{swdbr} = 1$ (resp. $j_{swdbr} = 1$). This increases the solution space to include solutions where variables y_{swdbr} (resp. j_{swdbr}) do not assume their correct value, however it has no influence on the correctness of the limitation on the number of changes imposed by constraints (17) (resp. (19)) nor on the optimal value.

3.3 Simulation environment

Hereafter we describe the architecture of the simulation environment we used in this study. Such an environment integrates Rockwell Arena and R via VBA and it is made of 3 sub-models, as shown in Figure 1. More detailed information regarding general specifications and model verification and validation are given subsequently in Sections 3.3.1, 3.3.2 and 3.3.3, respectively. Furthermore, and as a complement to Figure 1, an overview of input data, performed actions, output and technology used in the optimization and simulation models and sub-models are presented in Appendices A and B.

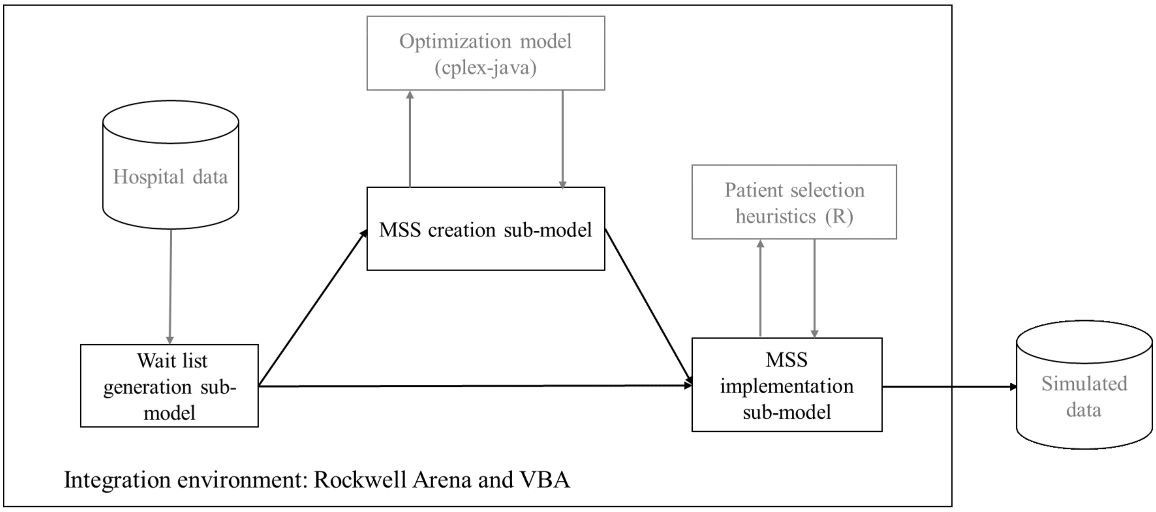

Figure 1: Simulation environment overview.

The first sub-model simulates the process by which the hospital put patients needing an elective procedure in the hospital waiting list (waiting list generation process). Based on provided hospital data, every day, for each specialty, this sub-model generates a number of *form entities* which is randomly sampled from the interval [N-0.3N, N+0.3N] where N is the number of arrivals for that day and specialty in some year b. *Forms entities* represents the form filled-in by surgeons when they prescribe elective procedures. These forms report attributes of both the patient (name, age, address, type, etc..) and the surgery s/he needs to undergo (specialty, ICD9CM diagnosis and procedure code, priority class). Upon creation, each *form entity* is assigned with a specialty. The other attributes are assigned to the form by randomly sampling the attributes of patients of the same specialty truly arrived in year b. Such a pragmatic approach to the waiting lists generation, allows obtaining realistic demand instances by taking into consideration the seasonal variation - both in terms of volume and mix – typically affecting the demand for elective surgeries (Visintin et al., 2017). After being assigned with their attributes, *form entities* are placed into a queue which represents the hospital waiting list - hereafter *waiting list queue*. At the beginning of the simulation, the model creates the *form entities* corresponding to the waiting list at the last day of year $b-1$ (*initial waiting list*), assigns them their attributes, and places them in the *waiting list queue*.

The second sub-model simulates the MSS creation process. In this sub-model, an auxiliary *trigger entity* is created every T days (which should be less than |K| days in the rolling horizon approach and equal to |K| days when using the simulation model to test long-term optimization solutions) and triggers the optimization model. The optimization model is executed in shell, meaning that while the optimization model runs, the simulation model is frozen, and the simulation clock does not advance. The simulation model creates an input file with all the data needed to instantiate the optimization model and runs it. This data includes: (i) the current waiting list; (ii) the number of beds available in each ward; and (iii) a simple forecast of the number of patients that will join the waiting list in the following T days. A seasonal naïve method is used to forecast this number, i.e. it is assumed to be equal to the number of patients that entered the waiting list, in the same time period, in the previous year. Once the optimization model finds a solution the simulation model reads the solution (i.e. the MSS) and saves it in an internal variable. Eventually, the *trigger entity* is disposed of.

The third sub-model simulates the MSS implementation process. In this sub-model, a *week entity* is generated every week and it is subsequently split into a number of *slot entities* equal to the number of slots in the MSS for that week. Each *slot entity*, according to the MSS, is assigned with a date, an operating room, and a specialty. Each *slot entity* waits until the simulated time matches its date attribute, and then it triggers the patient selection heuristic. The heuristic is coded in R and its pseudo-code is detailed in Algorithm 1. Based on the output of the heuristic, the model browses the *forms entities* in the *waiting list queue*, select the patients to insert into each slot and removes the *form entity* from the queue. The *slot entity* is then split into a number of *patient entities* equal to the number of selected patients. These *patient entities* seize the operating room and bed they are assigned to and release these resources after a time which is sampled from the probability distributions of surgical time (ST) and length of stay (LOS), respectively. Eventually, *patient entities* are disposed of. For both ST and LOS we fitted a triangular distribution, both obtained by asking a surgeon to estimate, for each surgery type (identified by its ICD9-CM code) the minimum, modal and maximum value of the these variables. This was done since the detail of the real historical data with respect to surgery duration is not enough to fit empirical or lognormal distributions (as done in Cappanera et al. (2014)). As suggested by Law (2006), when empirical data is missing, using triangular distributions allows: (i) accounting for the variability of the variables; (ii) considering their skewness (surgical time and length of stay are positively skewed - see e.g. Cappanera et al. (2014)); and (iii) avoiding extremely large (or small) values that are very unlikely to occur in reality (which is a problem arising when using lognormal distributions - Cappanera et al. (2014)).

Algorithm 1 Heuristic used in the MSS implementation sub-model.

Select all patients whose arrival date is earlier than the current date and whose specialty corresponds to the specialty (s) assigned to the slot.
Order the selected patients according to their due date in descending order (patients with approaching or expired due date first).
Calculate, for each selected patient p, the cumulative surgical time ($cumSumST(p)$).
Extract the set of patients for which the cumulative surgical time is less than or equal to the slot duration multiplied by a target utilisation factor (in our case 80%).
if The number of patients in the selected set is greater than or equal to the maximum number of surgeries allowed in a slot (λ_s) **then**
    Select the first λ_s patients.
else
    Calculate the slot residual time ($resTime$) as the difference between the slot duration ($slotTime = 0,8 \times \theta$) and the cumulative surgical time ($cumSumST$) of the last selected patient p ($resTime = slotTime - cumSumST(p)$).
    Order the remaining patients according to their surgical time first and due date second in descending order (patients with longer surgical time and approaching/expired due date first).
    while The number of selected patients is less than λ_s **do**
        Select the first patient for which surgical time is still smaller than the residual time (i.e. the patient with the largest surgical time that still fits).
        Update the residual time.
    end while
end if

The scheduling heuristic proposed is both equitable and long-term oriented. In fact, it firstly selects the patients with approaching or expired due date. When there is no more space to accommodate these patients, it tries to fill in the slot with the longest possible surgeries thereby avoid leaving long-lasting surgeries in the waiting list and compromising the hospital efficiency in the long run. As demonstrated in Banditori et al. (2013), fixing a target utilisation factor smaller than 100% when scheduling surgical slots, allows obtaining robust solutions, i.e. solutions that once implemented give rise to limited or null overtime.

3.3.1 General specifications of the model

The simulation model has 3 entity types, 11 resources and 5 queues. Entities are subdivided into form entities, patient entities, week entities and auxiliary entities. The latter are used to trigger the optimization model and the patient selection heuristic, as well as for the collection of statistics. The resources in the model represent the wards where patients can stay before surgery (pre-surgery ward), after surgery (surgery1, surgery2, surgery3, ICU, pediatrics and outpatient wards), and the operating rooms (OR1, OR2, OR3 and OR4). The queue objects include the mentioned waiting list queue, where form entities are placed on hold before being scheduled, and other queues associated with the operating rooms, the pre-surgery ward, the post-surgery wards and the ICU.

In addition to the internal variables of Arena, the model has 33 user-defined variables (5 of which are multi-dimensional) and 58 entity attributes, used to manage the entity flow and for measuring system performance. The model logic is coded using Arena templates (98 blocks in total) even if most of it was coded using Arena objects in the embedded VBA editor.

3.3.2 Model verification

Model verification is the process by which the modeler checks whether the outputs of each sub-model and the behavior of the model as a whole conforms with their expectation (Manuj et al., 2009). Such a process requires debugging of any errors in programming logic and code.

In our case, the verification process required debugging first the optimization model and the patient selection heuristic. For this purpose, we used the standard debug features of Java and R. Afterwards, we verified whether the simulation output was consistent with the output of both the optimization model and the patient selection heuristic. This involved systematically comparing the output files of the latter with the output files of the simulation model using a purposely created script code written in R. We verified, for example, that in the simulation output the slots were assigned to specialties which were consistent with the MSS returned by the optimization model, and also that the number of patients processed in each slot was consistent with the results of the heuristic. Finally, by using the standard debug features of Arena and VBA, we double-checked the model logic, the flow of entities and the resource status throughout the simulation run. To improve the quality of the verification process, the model logic was checked by two people other than the one who coded the model.

3.3.3 Model validation

Model validation is the process by which the modeler determines whether a simulation model is an accurate representation of the system under study (Law, 2006). Validation is always desirable, but unfortunately, it is not always possible (Fishman and Kiviat, 1968; Visintin et al., 2014). In our study, we were unable to validate the model. Comparing real hospital patient flow with the simulation output would have required, on the one hand, to know the MSS implemented for each week of the year and, on the other hand, the criteria followed by the hospital staff for patient selection and sequencing. With respect to the former, although the MSS of the hospital is supposed to be stable during the year, it frequently undergoes untracked changes. In addition, patients are assigned to surgical sessions and sequenced by hand without following any reproducible algorithm. Nonetheless, as suggested by Manuj et al. (2009), we thoroughly discussed the simulation results with the hospital management which judged them as reasonable and the overall model credible. This led us to conclude that the model had satisfactory face validity (Banks, 1988).

4 Results

In this section we present and discuss the results of applying our methodology to real data from a Portuguese hospital. Historical data on elective inpatients of seven specialties in the year 2017 was used to plan for the year of 2018. A description of the instance data is given in Section 4.1. The main approach proposed is the flexible rolling horizon approach, which integrates optimization and simulation. However, for a more thorough computational experiment and to define a ground for comparison, we start in Section 4.2 by analyzing the solution obtained by a flexible long-term approach, where the optimization model is used to plan the whole year at once, and compare it to a static long-term case, which mimics what is currently done in reality. This comparison allows us to answer the first research question. In Section 4.3 we analyze the solution obtained by the flexible rolling horizon approach and compare it to the flexible long-term approach, thus answering the second research question and reinforcing the answer of the first research question. Finally, in Section 4.4 we present some statistical tests to support our conclusions, including an analysis of the robustness of our methodology. Jointly, these three subsections provide a complete answer to the last research question.

The optimization model was implemented in Java using the API from CPLEX version 12.8.0. The simulation model was implemented using Rockwell Arena and R integrated via VBA. For each approach, 30 simulation runs were performed. We determined this number by following the approach suggested in Rossetti (2015). More precisely, we calculate the error associated with the point estimate of the average tardiness as follows. Let $avgT_r$ be the average patient tardiness in replication r, $M(avgT)$ the mean value of $avgT_r$ across n replications (i.e., the point estimate) and $s(avgT)$ the corresponding standard deviation. We calculate the half-width of the average tardiness as $h = t_{\alpha/2,n-1} \frac{s(avgT)}{\sqrt{n}}$, where we chose $\alpha = 0.05$ and, as mentioned, $n = 30$, with an error of $100 \times \frac{h}{M(avgT)}$. In the worst scenario, we obtain $h = 0.25$ with an error of 0.12%, meaning that we can be 95% confident of estimating the true average tardiness within more or less 0.25 days. We also decided to adopt a zero-length warm-up time. This decision is justified for two main reasons. Firstly, we have assumed that at the beginning of the simulation the resources in the system were all empty. This is reasonable, as we have two resources, operating rooms and beds. At the beginning of each simulated day, operating rooms are empty while beds are occupied by patients hospitalized in the previous days. However, since our simulation starts on 1 January, after the Christmas holiday when no elective procedures are performed, assuming that beds are empty (or nearly empty) is definitely reasonable. Secondly, the simulation run is very long (one year). As a result, the steady-state performance of the system is not significantly affected by initial bed occupancy (and, in any case, the hypothesized negligible occupancy makes sense as we explained).

All tests were ran in a computer with an Intel Core i7-4930 3.40 GHz processor and with 32GB of RAM. The static long-term solution is optimal and was obtained after less than one second of computation. The flexible long-term solution has an optimality gap of 0.5% after less than five minutes of computational time, however it still achieves a lower objective function value than the static case. With respect to the flexible rolling horizon approach, optimal solutions are achieved in every week in short computational times of at most thirty seconds.

4.1 Instance description

In this study we consider elective inpatients of 7 surgical specialties, to be scheduled during one year, in 4 operating rooms. Both morning and afternoon shifts, with a 360-minute duration each (288 minutes after applying the 80% utilization buffer), can be planned according to the availability of 41 surgeons, 10 anesthesiologists and beds in pre-wards, the intensive care unit and wards. Tables 2 and 3 show, respectively, statistics concerning the initial waiting list in 1 January 2018 and the expected arrivals throughout the year of 2018 (namely mean and standard deviation across 52 weeks estimated, as we mentioned, from historical data of the year 2017).

Table 2: Initial waiting list statistics reported by the case study hospital (1 January 2018).

	Waiting Time		Tardiness		
Specialties	Average	Total	Average	Total	Length
General Surgery	281.7	248751.0	228.2	201540.9	883
Ophthalmology	28.0	56.0	8.5	17.0	2
ORL	292.3	104335.0	236.0	842370.0	357
Orthopedics	183.6	55801.0	132.2	40179.3	304
Pediatric Surgery	229.0	229.0	170.0	170.0	1
Plastic Surgery	148.8	14282.0	92.9	8914.2	96
Urology	340.4	81026.0	288.5	68669.9	238
Total	268.2	504480.0	214.6	403728.3	1881

Table 3: Expected patient arrivals statistics.

	Patient arrivals	
Specialties	Mean	Standard deviation
General Surgery	18.2	5.4
Ophthalmology	2.0	1.3
ORL	3.5	2.0
Orthopedics	12.0	4.4
Pediatric Surgery	1.2	0.4
Plastic Surgery	3.7	2.0
Urology	3.9	2.2

As shown in Table 2, the waiting list on 1 January 2018, i.e. the initial waiting list, had a total of 1,881 patients waiting for surgery, with an average waiting time of 268.2 days and an average tardiness of 214.6 days. Regarding waiting time, urology presents the higher average, while general surgery presents the higher total waiting time. Concerning tardiness values, urology also presents the higher average values, but instead ORL presents the higher total. Ophthalmology is the specialty with lower values in both average and total values of waiting time and tardiness. Finally, general surgery has the longest waiting list, while pediatric surgery has the shorter one. As for Table 3, we can see that general surgery and orthopedics present the two highest mean arrivals per week (18.2 and 12.0 patients, respectively), whereas pediatric surgery and ophthalmology have the lowest mean arrivals per week (1.2 and 2.0 patients, respectively).

4.2 Flexible long-term solution analysis

In this section we compare the solution obtained by a flexible long-term approach with the solution obtained using a static case setting, designed to replicate what is currently done in reality. The static approach results from fixing every stability parameter Δ_m^M and Δ_w^M to 0, thus, essentially, not permitting any changes. Conversely, the flexible approach is equivalent to setting the parameters to $+\infty$, that is, allowing any changes to be made. The generality of the optimization model allows for other settings to be easily tested, however we do not do so in this discussion.

Clearly, from the perspective of the optimization model, the optimal solution of the flexible approach is not worse than the static case, since it is less constrained (changes are allowed). In fact, although not proven to be optimal, the solution obtained by the flexible approach has a lower objective function value than the optimal solution of the static case. We establish exactly how better the latter is by using the simulation model to do the comparison on key performance indicators (KPIs). All values represent an average across the 30 simulation runs. We start with total waiting time, total tardiness (in days) and total throughput, shown in Figure 2 (results for the flexible rolling horizon approach are also included but we discuss them in the next subsection). This figure shows the values of the corresponding KPIs at the end of each week of the planning period (of one year), and are accumulated values for all patients, both scheduled and unscheduled, at the end of the corresponding week, with week 0 representing the initial waiting list. We present more detailed comparison results in Table 4, with the corresponding mean (m) and standard deviation (sd) values. In this case, to simplify, we only show the average of the 30 simulation runs at end of the planning period. Included are the total tardiness, total waiting time and throughput, also present in Figure 2, as well as the total tardiness divided into scheduled and unscheduled patients and the average tardiness per patient.

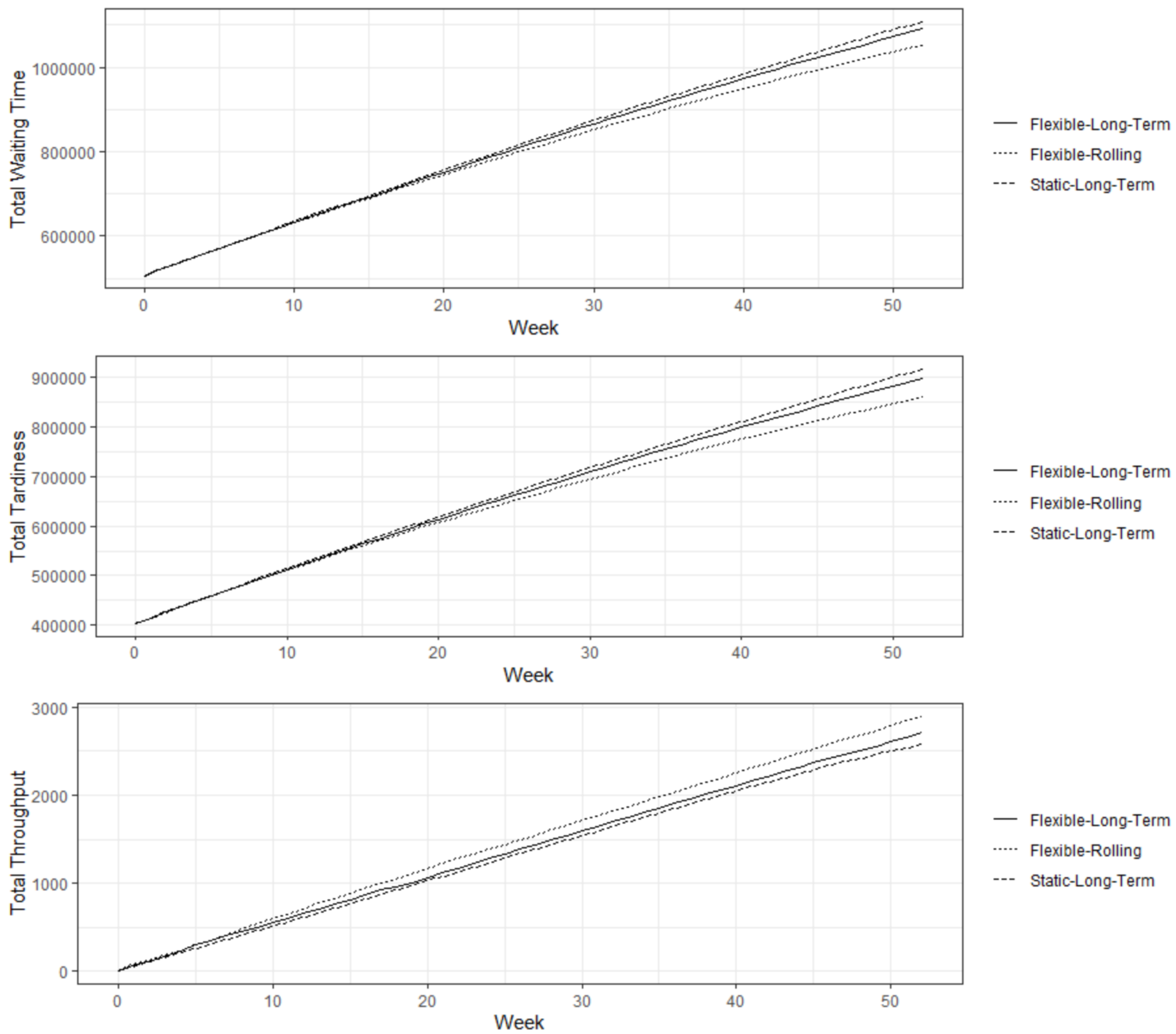

Figure 2: Comparison of the three approaches on average total waiting time, average total tardiness and average throughput.

Table 4: Average KPI values for the comparison of the long-term static and flexible approaches.

	Total tardiness		Total Waiting Time		Throughput	
Model	m	sd	m	sd	m	sd
Flexible long-term	897235.7	3150.3	1090995.2	3703.1	2707.2	7.2
Static long-term	916601.4	3373.9	1108760.1	4094.1	2574.0	6.9

The results of Figure 2 show that both total tardiness and total waiting time have a tendency to increase throughout the year. Nevertheless, the solution of the flexible approach presents lower (thus, better) values than the one of the static approach in almost every week. Note that this was not necessarily expected since the optimization model does not directly optimize for total tardiness and total waiting time, and it indicates that patients are scheduled sooner on average in the solution of the flexible approach. At the end of the planning period, where the difference is the largest, a total tardiness of approximately 897235.7 days is observed for the flexible case (see Table 4), which is a reduction of 2.11% from the total tardiness of 916691.4 days observed for the static case. As for the total waiting time, the final values are of approximately 1090995.2 and 1108760.1 days, respectively, corresponding to a decrease of 1.60% from the static to the flexible approach. The results of Table 4 confirm that the average throughput is higher in the solution of the flexible case with an increase of an average of 133 additional patients scheduled throughout the year. Finally, we observe that the solution of the flexible approach is not proven optimal, thus there is the potential for further improvements.

To conclude this section, we recall that the flexible approach has as a (potential) disadvantage the reduced stability provided by a fixed MSS. However, clear and quantifiable benefits can be achieved with respect to important waiting list management KPIs. We showed with these results the expected gains in the trade-off between stability and flexibility, even if only using a deterministic update of the waiting list as a consequence of not using the simulation model in an integrated way. We discuss the trade-offs between stability and flexibility in more detail in Section 5.

4.3 Flexible rolling horizon solution analysis

We now compare the solution of the flexible rolling horizon approach to the solution of the flexible long-term approach. A comparison to the static case is not done since it was clearly outperformed by the latter. The practical difference between both approaches is that in the flexible long-term case the waiting list evolution, as well as the average waiting time, are updated in a deterministic way, that is, patients scheduled for surgery are assumed to have undergone surgery when updating the waiting list. Conversely, in the flexible rolling horizon case, the uncertainty with respect to the number of surgeries that are actually performed (within the ones planned) is taken into account when updating the waiting list via the simulation model. As the waiting lists are updated, the respective average waiting time values are also updated throughout time.

We compare both solutions based on the total waiting time, tardiness and throughput at the end of each week of the planning period, as depicted in Figure 2 and in Table 5, with a similar structured as Table 4.

Table 5: Average KPI values for the comparison of the flexible long-term and rolling horizon approaches.

	Total tardiness		Total Waiting Time		Throughput	
Model	m	sd	m	sd	m	sd
Rolling horizon	859399.1	3413.5	1052868.3	3824.1	2889.5	10.9
Flexible long-term	897235.7	3150.3	1090995.2	3703.1	2707.2	7.2

The results of Figure 2 show that the solution of the rolling horizon approach leads, in the long run, to a better performance. The explanation for this difference is the following. Recall that the optimization model adjusts the MSS to the demand and corresponding average waiting time of each specialty in the current waiting list. By using a deterministic update of the waiting list, the patients scheduled for surgery are all assumed to have their surgery performed, thus, in the following week the MSS will adjust to completely new patients. However, by using an update based on the simulation model, some of the patients of the previous week may not have had their surgery performed so the adjustment of the MSS may be different. Therefore, in the long-run, the perceived demand of the deterministic adjustment starts to become too different from the actual demand. Note that this was expected and is the main advantage of the rolling horizon approach.

As for the results of Table 5, they allow us to quantify the improvements of the solution of the rolling horizon approach at the end of the planning period. Observe that the average throughput increased by 182 patients, however, the real benefit is seen in the reduction of 4.22% in the total tardiness and of 3.50% in total waiting time. An important final observation is that the optimization model is always able to obtain an optimal solution in each week in the flexible rolling horizon approach, thus, the difference to the solution of the flexible long term approach (which is not proven optimal) may narrow. Nevertheless, this may be seen as a further advantage of the rolling horizon approach, since it is more computationally tractable.

4.4 Statistical tests

In this section we present a statistical analysis of our results to support our conclusions. First, we discuss statistical tests to assess the differences with respect to total tardiness, total waiting time and throughput among the three approaches. For this purpose, we performed, for all three KPIs, a one-way ANOVA using the approach as a factor (or independent variable).

In order to perform the ANOVAs, we first checked the assumptions of normality of the residuals and homogeneity of the variances. With respect to the former, we carried out the Shapiro-Wilk normality test and failed to reject the null hypothesis of the residuals being normally distributed for both KPIs (W = 0.989 and p-value = 0.688 for total tardiness; W = 0.990 and p-value = 0.753 for total waiting time, W = 0.976 and p-value = 0.099 for throughput). As for the latter, we performed the Levene's test and failed to reject the null hypothesis of the variances being equal (F = 0.005 and p-value = 0.996 for total tardiness; F = 0.220 and p-value = 0.803 for total waiting time; F = 2.331 and p-value = 0.103 for throughput).

The ANOVAs show that all three KPIs are significantly different across the three approaches (F = 2311 and p-value = 0.000 for total tardiness; F = 1628 and p-value = 0.000 for total waiting time; F = 10333 and p-value = 0.000 for throughput). Moreover, Tukey's post-hoc tests show that all three KPIs for the flexible rolling horizon approach are significantly smaller than the ones associated with the flexible long-term approach and with the static long-term approach, and also that all three KPIs are significantly smaller for the flexible long-term approach than for the static long-term approach (p-value = 0.000 in every case).

Another important aspect to analyze is the robustness with respect to overtime of our solutions. For this purpose, we present some results in Table 6 with respect to the planned slot length and overtime taken from the sample of all slots across the 30 simulations runs, namely average, standard error and a 95% confidence interval.

Table 6: Planned slot length and overtime values for the sample of all slots across the 30 simulation runs.

	Planned slot length		95% confidence interval	
	Average	Standard error	Lower bound	Upper bound
Rolling horizon	239.23	2.41	234.31	244.15
Long-term	236.00	1.82	232.27	239.73
Long-term (static)	232.99	2.04	228.82	237.16

	Overtime		95% confidence interval	
	Average	Standard error	Lower bound	Upper bound
Rolling horizon	1.11	0.28	0.53	1.69
Long-term	1.17	0.24	0.67	1.66
Long-term (static)	0.99	0.28	0.43	1.56

The results of Table 6 show that our approach of planning the slots to at most 80% of the actual length (360 minutes) results in a very robust setting. In fact, the average overtime in any of the approaches is close to 0, with the confidence intervals indicating that it is very unlikely that it could be higher than two minutes. In fact, the 0.95 sample quantiles in all three approaches are 0.00, whereas the 0.99 sample quantiles are 41.14, 42.81 and 35.96 minutes, respectively for the flexible rolling horizon, the flexible long-term and the static long-term approaches.

Finally, we performed another ANOVA (after the appropriate Shapiro-Wilk and Levene's tests). The results indicate F = 3.331 and p-value = 0.0404, thus it is inconclusive whether there are significant differences or not. Tukey's post-hoc tests reveal that the approaches that may differ significantly are the static long-term and the flexible long-term (p-value = 0.034). Nevertheless, despite statistically significant differences for a type I error of 0.05, we believe these are mostly due to the fact that the static long-term approach has a higher underutilization as a result of fewer scheduled patients, so these differences are not relevant in practice (less than a seven-minute difference in the 0.99 sample quantiles between the static long-term and the flexible long-term approaches). These results suggest that our comparisons are fair since the solutions produced are valid from the point of view of the hospital and are similarly robust.

5 Managerial Insights

In this section we discuss managerial implications of using the approaches proposed in this work in a practical context. For this purpose, we start by performing a comparison of our results to what is shown by an analysis of real data in Section 5.1. Afterwards, we provide a critical discussion of the trade-off between stability and flexibility in Section 5.2. Finally, in Section 5.3, we study the impact of a scenario of increased demand in the comparisons established between the models.

5.1 Comparison with real data

To assess the potential benefits of the use of our models in practice, we compared, via simulation, the performance that the case hospital could have achieved in 2018 using our approach to the performance indicated by the data provided by the hospital in the same year. The statistics concerning the initial and final waiting lists (i.e., for unscheduled patients) in the year 2018 reported by the hospital, and organized by specialty, are presented in Tables 2 (in Section 4.1) and 7, respectively.

Table 7: Final waiting list statistics reported by the case study hospital (31 December 2018).

	Waiting Time		Tardiness		
Specialty	Average	Total	Average	Total	Length
General Surgery	242.7	190258.8	190.5	149337.2	784
Ophthalmology	12.0	12.0	0.0	0.0	1
ORL	336.2	91453.5	280.9	76394.2	272
Orthopedics	122.8	17680.3	80.7	11616.5	144
Pediatric Surgery	0.0	0.0	0.0	0.0	0
Plastic Surgery	66.9	6758.8	21.6	2181.2	101
Urology	387.2	91763.5	336.1	79649.9	237
Total	258.6	397926.9	207.4	319178.9	1539

In 1 January 2018, the reported initial waiting list had 1,881 patients (of the 7 surgical specialties under study), with an average tardiness and waiting time of 214.6 and 268.2 days, respectively. In 31 December 2018, the final waiting list had 1,539 patients, with an average tardiness and waiting time of 207.4 and 258.6 days, respectively, which suggests a clear improvement during the year. This performance was obtained by allotting 15 slots per week to elective patients of the considered specialties. As we mentioned, to perform a fair comparison between our simulated results and the results reported by the hospital, we ran our simulation considering the same initial waiting list and the availability of a maximum of 15 slots for each week. Table 8 shows the values of the same KPIs of Tables 2 and 7 but with respect to the final simulated waiting lists. More precisely, for each KPI we report the mean value (m) and the standard deviation (sd) obtained across 30 simulation runs for the three proposed approaches.

Table 8: Final waiting list statistics of the three proposed approaches: rolling horizon (RH), flexible long-term (FLT) and static long-term (SLT).

	Waiting Time				Tardiness					
	Average		Total		Average		Total		Length	
Model	m	sd	m	sd	m	sd	m	sd	m	sd
RH	114.0	2.6	129137.5	4199.4	74.2	2.3	84002.3	3434.4	1132.3	13.2
FLT	153.4	1.4	202337.6	3307.7	112.5	1.2	148310.5	2617.5	1318.6	12.6
SLT	163.7	1.9	237301.3	4275.5	124.4	1.7	180355.6	3547.0	1449.5	11.0

The results of Table 8 show that all three approaches proposed in this work outperform the planning of the hospital for the year of 2018 in all KPIs. In particular, the final waiting list of the static long-term approach has an average of 90 patients less than the hospital, the flexible-long term has an average of 220 patients less than the hospital and, finally, the rolling horizon has an average of 407 less patients than the hospital. To reinforce the observed differences, we performed statistical tests to test the null hypothesis of the final simulated value of each KPI reported in Table 8 being smaller than the final one reported by the hospital in Table 7. Prior to these tests, we checked the normality hypothesis using the Shapiro-Wilk test. For all but one combination of KPI and approach, namely tardiness in the flexible rolling horizon approach, we failed to reject the null hypothesis of the KPI being normally distributed (with p-values consistently above 0.05). Therefore, we used t-tests in all cases except in the combination mentioned before, where instead we used a Wilcoxon signed rank test. In all tests we obtained a p-value below 0.01, indicating that the results produced by any of the three approaches could lead to a significant improvement of the performance.

5.2 Trade-off between stability and flexibility

This work studies the implementation of flexibility into the MSS. Flexibility can lead to an overall significant improvement of the system, as our results suggest, particularly those presented in the previous section, when compared to the standard way of a static MSS. Nevertheless, clear trade-offs exist, which we analyze in this section.

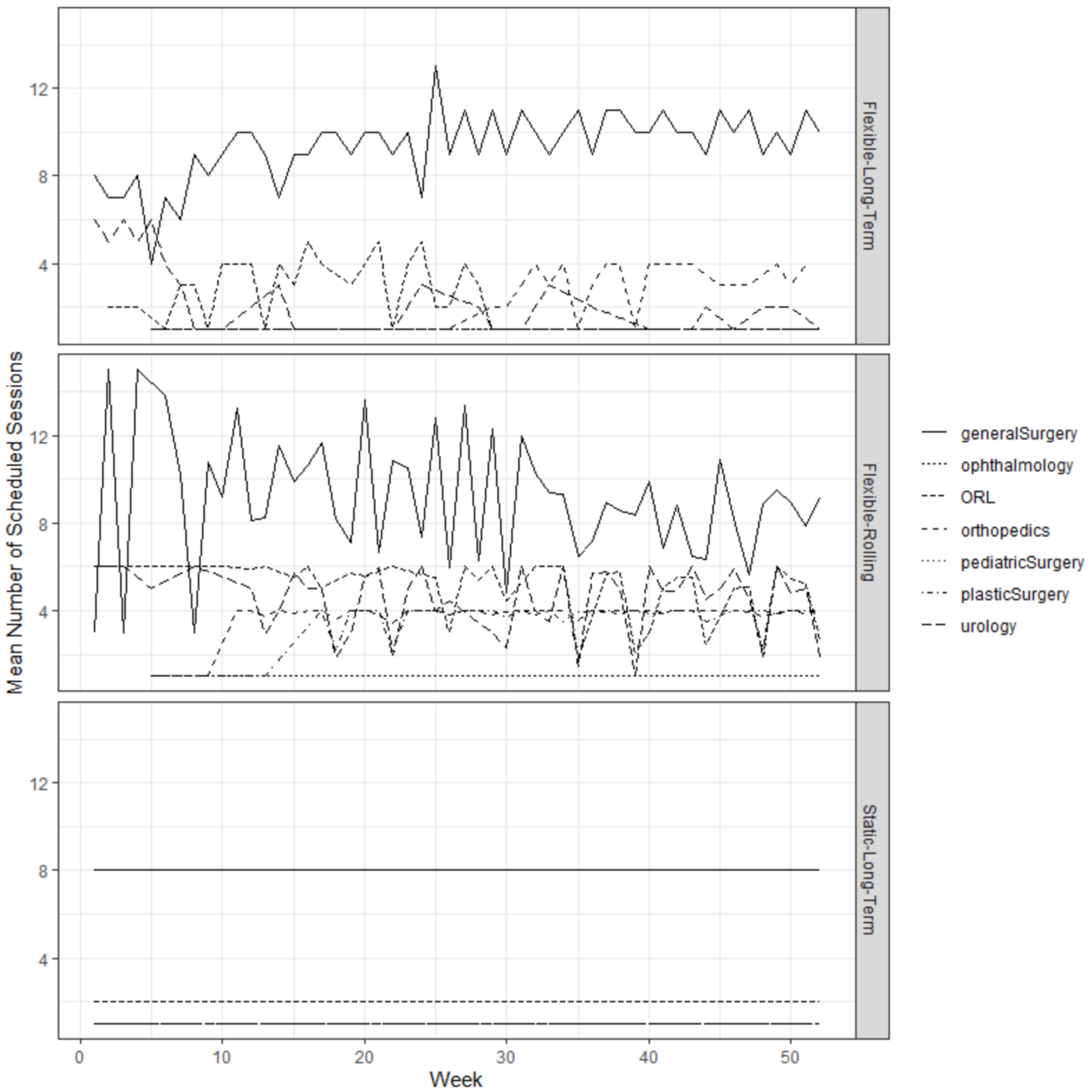

Figure 3: Mean number of scheduled slots per specialty for the three approaches.

To support this analysis, we present in Figure 3 the number of scheduled slots per specialty and per week (out of the 15 available) for all three approaches, where in the case of the flexible rolling horizon, the value is an average of the 30 simulations runs. This number allows us to quantify flexibility in the MSS to a certain extent. As expected, Figure 3 shows that, in the static long-term approach, the number of weekly slots reserved for each specialty remains constant. Conversely, the two flexible approaches constantly readjust the number of scheduled slots, with the flexible rolling horizon showing greater variability. For example, considering general surgery, one can observe that for the static approach 8 weekly slots are planned, with the flexible long-term approach a minimum of 4 and a maximum of 13 slots are scheduled, while with the rolling horizon one, the average number of planned slots ranges from 3 to 15. Additionally, and again looking at the example of general surgery, the flexible long-term approach has an average of 9.37 slots per week with a standard deviation of 1.55 and the rolling horizon approach has an average of 9.09 slots with a standard deviation of 2.89.

On the one hand, a static MSS allows for easier planning of resources at an operational level and, as a consequence, a more fixed routine for staff, leading to a greater satisfaction of hospital employees. On the other hand, as our results show, the throughput is substantially lower. More precisely, the throughput of the static long-term approach is, on average, 315.5 and 128.2 less than the flexible rolling horizon and the flexible long-term approaches, respectively, which translates to approximately 6 and 2.5 fewer patients per week. These results were obtained with the same patient selection and sequencing methods, thus, the additional throughput can be attributed exclusively to the flexibility in the MSS.

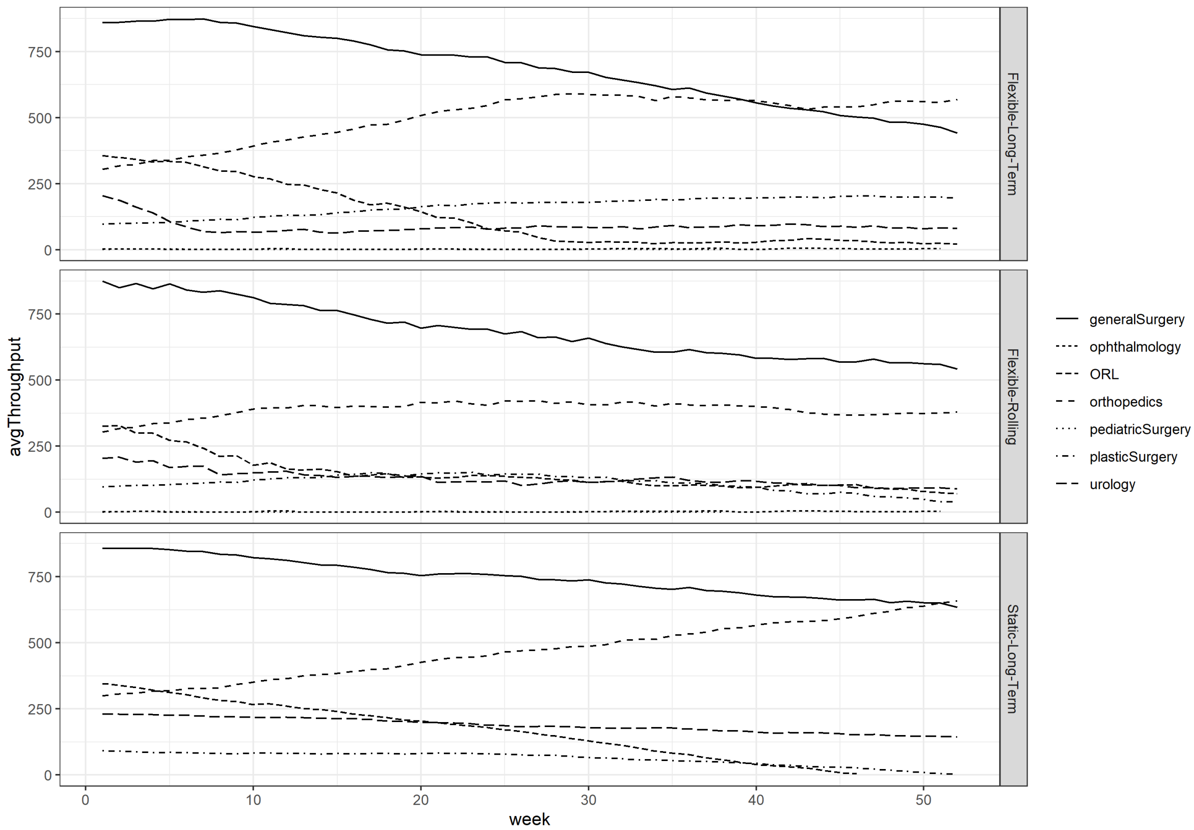

Figure 4: Evolution of waiting lists, by specialty, for the three approaches.

A better throughput originates shorter waiting lists in the end, however, the main advantage of the flexibility introduced by our approaches is that the capacity is allocated to the specialties with larger backlogs, particularly in the initial weeks of the planning year, thus contributing to a more equitable scheduling process. Figure 4 presents the evolution of the waiting list along the planning year for all three approaches. As can be seen, the waiting list in the first weeks has a substantial imbalance since some specialties have a much higher number of patients waiting for surgery, as is the case of general surgery. Whereas the static long-term approach takes considerably longer to decrease this backlog, the flexibility of the two other approaches allows the imbalance to be corrected much faster. Taking the case of orthopedics as another example, we can observe that, in the static long-term case, there is a prominent increase of the waiting list, however, with both flexible approaches, the long-term and the rolling horizon one, it is possible to stabilize this growth around weeks 28 and 10, respectively. The added throughput along with this backlog reduction property make it so the flexible approaches obtain much better indicators with respect to total tardiness and total waiting time.

In conclusion, stability in the MSS is an important feature which reduces the complexity of operational planning in the operating room. Sacrificing this for a flexible approach can only be considered if solid proof that it would be worth it can be provided. The flexible rolling horizon approach presents the most extreme case of flexibility, since the MSS is updated every week based on the waiting list at the beginning of that week. As a different option, we propose the flexible long-term approach, where changes are allowed but the MSS is planned based on the expected demand. The computational study conducted allows both models complete flexibility, although the optimization model is prepared to limit the number of changes on a weekly or monthly basis. We do not believe complete flexibility in the MSS to be feasible in a practical context, however, our results clearly show that allowing for some degree of flexibility should be considered and studied in a real setting.

5.3 Impact of increased demand

To further complement our computational study and comparisons, we tested how scenarios of increased demand, specifically of specialties with lower demand, could affect the results in terms of waiting time, tardiness and waiting list evolution. We selected the three specialties with fewer patients in the initial waiting list (see Table 2), namely ophthalmology, pediatric surgery and plastic surgery, and increased the expected number of patients in the arrival list by 10% and 20% distributed equally by these three specialties. In other words, to the already expected patient arrivals, we added new generated patients with characteristics sampled from the initial waiting list of the three mentioned specialties up to 10% and 20% more patients. The reason for choosing these specialities is that their expected demand along the year is low, thus, the flexible approaches, which chase the demand, are expected to have worse performance, particularly the long-term one.

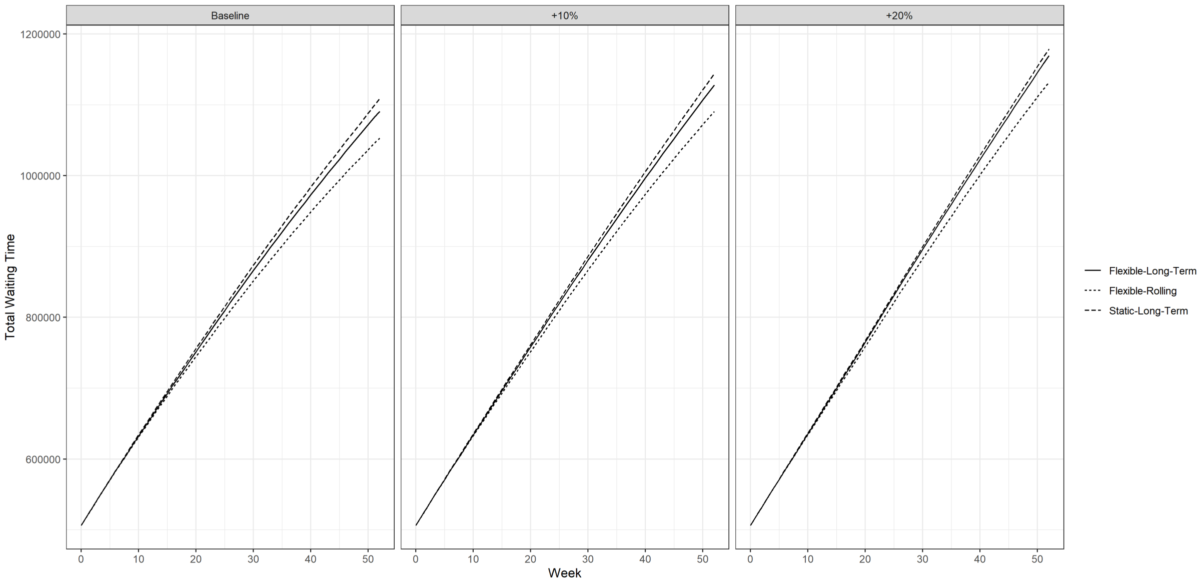

Figure 5: Comparison of the three approaches on average total waiting time in scenarios of increased demand.

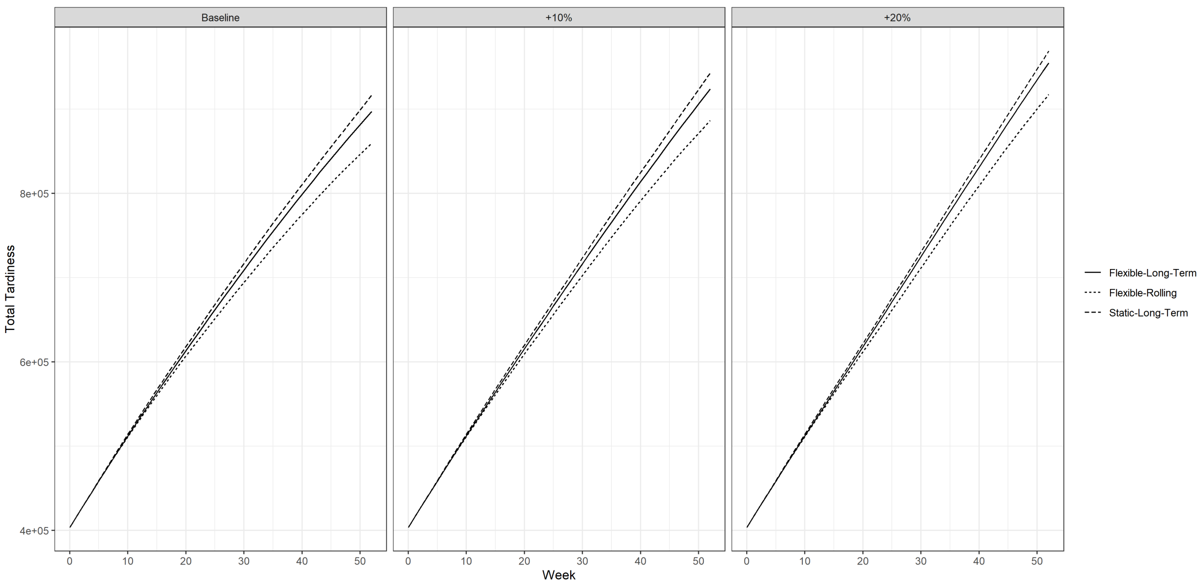

Figure 6: Comparison of the three approaches on average total tardiness in scenarios of increased demand.

Figures 5 and 6 show the results with respect to waiting time and tardiness, respectively, where the baseline case corresponds to the original arrival list (i.e., 0% increased demand). As expected, since the arrival list is the original one with additional patients, performance deteriorates in both KPIs in all three approaches compared to the baseline. In particular, the flexible long-term approach shows a closer (but still better) performance to the static long-term approach when demand increases.

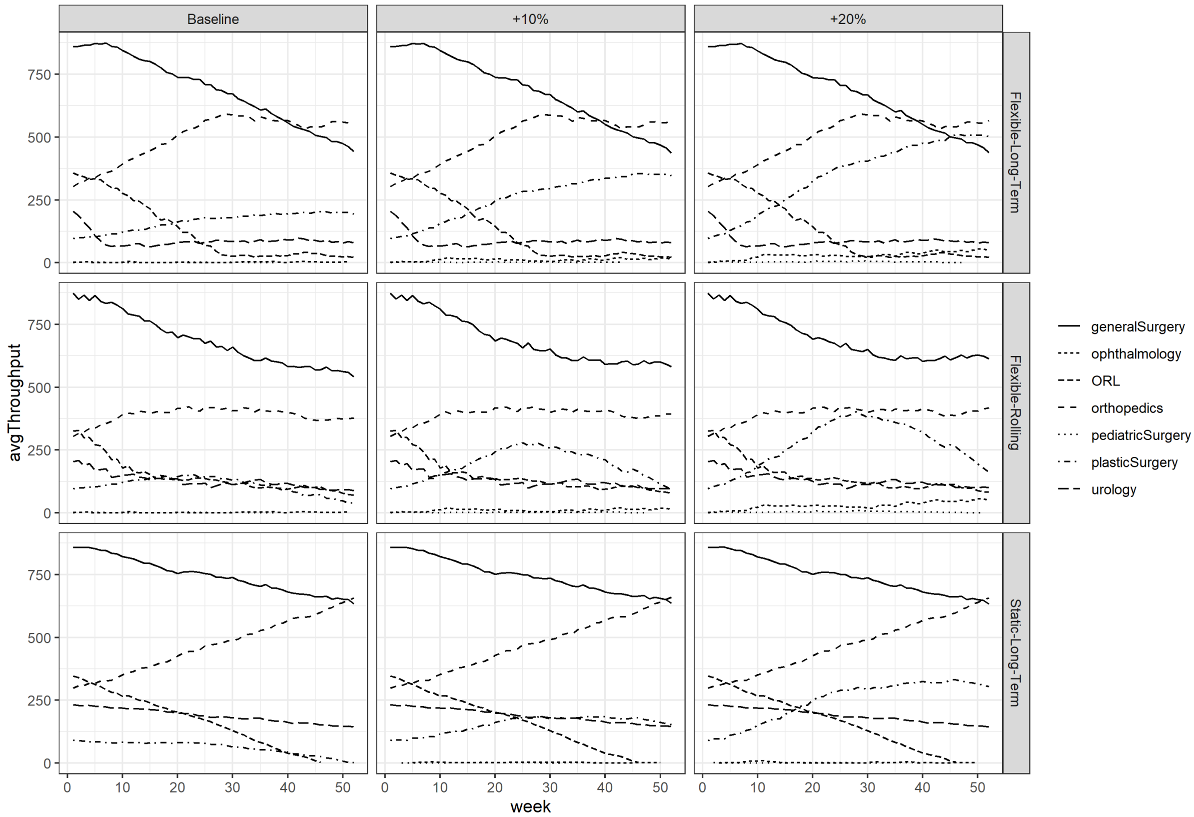

Figure 7: Evolution of waiting lists, by specialty, for the three approaches in scenarios of increased demand.

These results suggest that constructing a flexible MSS based on the expected demand may not be sufficient to handle increased unexpected demand better than the standard static approach. An important note, however, is that, to ensure a greater degree of flexibility, the flexible long-term approach does not set a weekly minimum number of slots for any specialty, whereas the static long-term inherently has at least one slot per week attributed to each specialty, even if none or few patients require surgery. In this situation of increased demand of the three lower demand specialities, the static long-term is able to more easily accommodate a part of this increased demand, however, in a situation where these slots are not needed due to lack of patients, valuable operating room time may be lost. On the other hand, the approach of the flexible rolling horizon of adjusting the MSS on a weekly basis allows handling both cases of when demand is lower or higher than expected, which is reflected in its stable performance. Additionally, the waiting list evolution of Figure 7 shows that the flexible approaches are still able to ensure a more equitable and balanced scheduling process.

Finally, we performed more ANOVAs that show that the differences in both waiting time and tardiness (and also throughput) across the three approaches are still statistically significant for both the 10% and the 20% increased demand cases (all p-values below 0.001).

6 Conclusions

This paper answers both strategic and tactical decisions of operating room planning and scheduling. It represents a strong scientific contribution as it uses the expected duration to characterize demand instead of just the waiting list length, considers a long planning horizon, introduces flexibility in the MSS, and combines optimization and simulation approaches to accurately assess and define demand variation.

The first research question “Is it possible to better chase surgical demand by optimizing and allowing for some flexibility?” is answered through a comparison of a static and a flexible long-term approach of the optimization model. Using the simulation environment to assess the optimization results, it is possible to conclude that allowing weekly and monthly changes in the MSS leads to better results in terms of waiting time and tardiness. The results also show that, although the static approach reduces complexity of operational planning as well as providing the benefit of increased stability for employees, it schedules fewer patients. On the other hand, the flexible approach, despite presenting more variability in the workload, are able to assign more surgical time to specialties with larger backlogs of patients.

The flexible long-term approach is also compared with a flexible rolling horizon approach, in which the optimization and simulation run iteratively one week at a time, to answer the second research question “Is it possible to better chase demand by combining optimization and simulation approaches when generating MSSs?”. In this case, with the rolling horizon approach it is possible to have even higher values of throughput and, particularly, lower values of waiting time and tardiness. Furthermore, when compared to real data from 2018 reported by the hospital, all three proposed approaches (static, flexible long-term and rolling horizon) outperform the real case in terms of waiting time, tardiness and throughput.

Considering the obtained results, it is also possible to answer the third research question “Is it possible to make the planning process overall more equitable?”. Since it was shown that both flexible approaches are able to reduce tardiness when compared to the static method, one can state that the process is more equitable. The obtained results also show that, as flexible approaches assign more time to specialties with larger backlogs, they are able to balance surgical waiting lists. This is an important result as it can constitute a powerful tool in complying with waiting time targets for surgical patients. In addition, if the demand is higher than expected, the performance of the three models tends to deteriorate and, particularly, the advantage of the flexible long-term approach over the static one reduces. However, both flexible approaches, and most importantly the flexible rolling horizon approach, are still shown to be able to provide more equitable results with respect to waiting list balance.

The lack of detail of the real data obtained from the hospital did not enable to fit any theoretical distribution regarding surgery duration and length-of-stay of the patients, which is a limitation of our study. Moreover, a more detailed trade-off analysis between stability and flexibility can be performed to assess the number of monthly and weekly changes that should be allowed to obtain maximum benefit from the proposed framework in terms of chasing demand and ultimately providing timely access to surgical care. In addition, the operating room planning and scheduling literature lacks sophisticated methods to capture preferences or availability of the surgical staff which may undermine the practical implementation of such frameworks. This research avenue is strongly recommended. Finally, the proposed framework can be extended and adapted to be applied to other contexts such as appointment scheduling.

References

Addis, B., Carello, G., Grosso, A., and Tànfani, E. (2016). Operating room scheduling and rescheduling: a rolling horizon approach. *Flexible Services and Manufacturing Journal*, 28:206 – 232.

Agnetis, A., Coppi, A., Corsini, M., Dellino, G., Meloni, C., and Pranzo, M. (2012). Long term evaluation of operating theater planning policies. *Operations Research for Health Care*, 1(4):95 – 104.

Agnetis, A., Coppi, A., Corsini, M., Dellino, G., Meloni, C., and Pranzo, M. (2014). A decomposition approach for the combined master surgical schedule and surgical case assignment problems. *Health Care Management Science*, 17(1):49 – 59.

Anjomshoa, H., Dumitrescu, I., Lustig, I., and Smith, O. J. (2018). An exact approach for tactical planning and patient selection for elective surgeries. *European Journal of Operational Research*, 268(2):728 – 739.

Banditori, C., Cappanera, P., and Visintin, F. (2013). A combined optimization–simulation approach to the master surgical scheduling problem. *IMA Journal of Management Mathematics*, 24(2):155–187.

Banks, J. (1988). *Handbook of Simulation: Principles, Methodology, Advances, Applications, and Practice*. John Wiley & Sons, Inc.

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

Bovim, T. R., Christiansen, M., Gullhav, A. N., Range, T. M., and Hellemo, L. (2020). Stochastic master surgery scheduling. *European Journal of Operational Research*, 285(2):695–711.

Cappanera, P., Visintin, F., and Banditori, C. (2014). Comparing resource balancing criteria in master surgical scheduling: A combined optimisation-simulation approach. *International Journal of Production Economics*, 158:179–196.

Cappanera, P., Visintin, F., Banditori, C., and Feo, D. (2019). Evaluating the long-term effects of appointment scheduling policies in a magnetic resonance imaging setting. *Flexible Services and Manufacturing Journal*, 31(1):212–254.

Dellaert, N. and Jeunet, J. (2017). A variable neighborhood search algorithm for the surgery tactical planning problem. *Computers & Operations Research*, 84:216–225.

Fishman, G. S. and Kiviat, P. J. (1968). The statistics of discrete-event simulation. *SIMULATION*, 10(4):185–195.

Gao, N.-n., Liu, Z.-x., and Li, Y.-f. (2013). Estimating the hidden costs of operating room with time-driven activity-based costing. In Qi, E., Shen, J., and Dou, R., editors, *The 19th International Conference on Industrial Engineering and Engineering Management*, pages 87–94, Berlin, Heidelberg. Springer Berlin Heidelberg.

Kamran, M. A., Karimi, B., Dellaert, N., and Demeulemeester, E. (2019). Adaptive operating rooms planning and scheduling: A rolling horizon approach. *Operations Research for Health Care*, 22:100200.

Kumar, A., Costa, A. M., Fackrell, M., and Taylor, P. G. (2018). A sequential stochastic mixed integer programming model for tactical master surgery scheduling. *European Journal of Operational Research*, 270(2):734 – 746.

Law, A. (2006). *Simulation Modeling and Analysis*. McGraw-Hill.

Manuj, I., Mentzer, J., and Bowers, M. (2009). Improving the rigor of discrete-event simulation in logistics and supply chain research. *International Journal of Physical Distribution & Logistics Management*, 39:172–201.

Marques, I. and Captivo, M. E. (2017). Different stakeholders' perspectives for a surgical case assignment problem: deterministic and robust approaches. *European Journal of Operational Research*, 261(1):260–278.

Marques, I., Captivo, M. E., and Barros, N. (2019). Optimizing the master surgery schedule in a private hospital. *Operations Research for Health Care*, 20:11 – 24.

Marques, I., Captivo, M. E., and Pato, M. V. (2012). An integer programming approach to elective surgery scheduling. *OR Spectrum*, 34(2):407–427.

M'Hallah, R. and Visintin, F. (2019). A stochastic model for scheduling elective surgeries in a cyclic master surgical schedule. *Computers & Industrial Engineering*, 129:156 – 168.

Molina-Pariente, J. M., Hans, E. H., and Framinan, J. M. (2018). A stochastic approach for solving the operating room scheduling problem. *Flexible Services and Manufacturing Journal*, 30:224 – 251.

Moosavi, A. and Ebrahimnejad, S. (2020). Robust operating room planning considering upstream and downstream units: A new two-stage heuristic algorithm. *Computers & Industrial Engineering*, 143:106387.

Oliveira, M., Bélanger, V., Marques, I., and Ruiz, A. (2020). Assessing the impact of patient prioritization on operating room schedules. *Operations Research for Health Care*, 24:100232.

Penn, M., Potts, C., and Harper, P. (2017). Multiple criteria mixed-integer programming for incorporating multiple factors into the development of master operating theatre timetables. *European Journal of Operational Research*, 262(1):194–206.

Rossetti, M. D. (2015). *Simulation modeling and Arena*. John Wiley & Sons.

Schneider, A., van Essen, J., Carlier, M., and Hans, E. W. (2020). Scheduling surgery groups considering multiple downstream resources. *European Journal of Operational Research*, 282(2):741 – 752.

Spratt, B. and Kozan, E. (2016). Waiting list management through master surgical schedules: A case study. *Operations Research for Health Care*, 10:49 – 64.

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

Visintin, F., Cappanera, P., and Banditori, C. (2016). Evaluating the impact of flexible practices on the master surgical scheduling process: an empirical analysis. *Flexible Services and Manufacturing Journal*, 28:182–205.

Visintin, F., Cappanera, P., Banditori, C., and Danese, P. (2017). Development and implementation of an operating room scheduling tool: an action research study. *Production Planning & Control*, 28(9):758–775.

Visintin, F., Porcelli, I., and Ghini, A. (2014). Applying discrete event simulation to the design of a service delivery system in the aerospace industry: A case study. *Journal of Intelligent Manufacturing*, 25:1–18.

Zhang, B., Murali, P., Dessouky, M., and Belson, D. (2009). A mixed integer programming approach for allocating operating room capacity. *Journal of the Operational Research Society*, 60:663–673.

Zhu, S., Fan, W., Yang, S., Pei, J., and Pardalos, P. (2019). Operating room planning and surgical case scheduling: a review of literature. *Journal of Combinatorial Optimization*, 37:757–805.

Appendices

Appendix A: Overview of the proposed optimization model.

Model	Input data	Performed action	Output	Technology
Optimization model	Value of the sets and parameters needed to run the optimization model (see Section 3.1)	Solve the optimization model described in Section 3.2	Optimization model solution (MSS)	cplex-Java

Appendix B: Overview of each component of the proposed simulation approach.

Models	Input data	Performed action	Output	Technology
Wait list generation sub model	Patients in the waiting list at the beginning of year b; Patient arrivals in year b-1; Surgical times and LOS for each surgery type	Creation of a demand stream based on the arrivals of year b-1	Waiting list queue: queue of *form entities* representing patients waiting for surgery and storing all the patients (name, age, address, type, etc..) and surgery attributes (specialty, ICD9CM diagnosis and procedure code, priority class, due date)	Rockwell Arena, VBA
MSS creation submodel	Waiting list queue data; Forecast of the number of patients that will join the waiting list in the following T days; Optimization model solution	Scan the waiting list queue and the available hospital resources (beds, OR, ICU,etc.) and create the input files with the set and parameters needed by the optimization model; Triggers the optimization model *in shell*; Reads the optimization model solution and saves the corresponding MSS in an Arena variable	Optimization model input data; Arena matrix variable storing the Optimization output (MSS)	Rockwell Arena, VBA
MSS implementation submodel	Arena matrix variable storing the Optimization model's output; Waiting list queue data; The output of the patient selection heuristic	Triggers the Patient selection heuristic *in shell*; Reads the solution of the heuristic; Picks the form entities in the Waiting list queue according to the heuristic solution, seizes the resources needed to process them (OR, Beds) for a time sampled from a suitable distribution, records all the time stamps relevant to the patient journey and eventually dismisses the patient	Input data for the patient selection heuristic: Patients waiting for surgery and their attributes, MSS and Scheduled slots duration. Output files recording, for each processed entity, its attributes and the start and end time of each process step it was involved in. For each resource, records its utilization statistics	Rockwell Arena, VBA
Patient selection heuristic	Patients waiting for surgery and their attributes; MSS; Scheduled slot duration	Assign patients in the waiting list to a suitable slot	List of patients to fill in each scheduled slot of the planning horizon	R