UNIVERSITÀ DEGLI STUDI FIRENZE

DOTTORATO DI RICERCA IN
INGEGNERIA INDUSTRIALE

CICLO XXVII

COORDINATORE Prof. DE LUCIA

Sviluppo e implementazione di modelli e strumenti a supporto del processo di master surgical scheduling

Development and implementation of models and tools supporting the master surgical scheduling process

Settore Scientifico Disciplinare ING/IND 35

Dottorando
Dott. CARLO BANDITORI

____

Tutor
Ing. FILIPPO VISINTIN

____

Coordinatore
Prof. MAURIZIO DE LUCIA

____

Anni 2012/2014

Abstract

The operating theatre is one of the most critical functional area in a hospital. In fact, it drives most of the hospital admissions and it is responsible for most of its costs. Optimising the operating theatre operations, is therefore a primary concern for an increasing number of hospitals. In this regard, one of the most challenging problem that hospitals need to face is the planning and scheduling of the surgical activities. This thesis focuses on the master surgical scheduling (MSS) problem. Such a problem consists in the determination of (i) the specialty (or specialties) to assign to each operating room and session of each day of the planning cycle and (ii) the number and the typologies of surgeries that should be performed in each operating room session. A number of authors have proposed models to support such a process. However, most of them test the models, often on real data, but do not illustrate practical aspects of their implementations. This thesis concerns an action research study aiming at addressing this gap and thus at developing and implementing a MSS tool in a real context i.e. the Meyer children's hospital in Florence. As an action research, this study has a twofold objective: to solve a practically relevant problem and to contribute to the body of knowledge. In fact, first, it aims to implement a MSS tool at the Meyer hospital. Second, it proposes novel mixed integer programming models addressing the MSS problem and provides fresh insights about the its implementation process. These experience-driven evidences may be useful for researcher and practitioners to increase the chance to success in the transfer of a MSS model to their hospital settings.

Acknowledgments

This thesis is the result of my last three years of research within the IBIS Lab and the Meyer children’s hospital. All this would not have been possible without the contribution of all those people I have encountered during my PhD journey.

First of all, I would like to thank my supervisor Dr. Filippo Visintin for having constantly motivated and guided me during these years. What I know about research is mostly because of him. I am also very grateful to Dr. Paola Cappanera, for her kindness and for the valuable advices she gave me about my research. It has s been a pleasure working with both of them.

I want to thank the other (former or present) members of the IBIS Lab, Prof. Mario Tucci, Dr. Mario Rapaccini, Dr. Orlando Borgia and Dr. Filippo De Carlo for their precious advices and support.

I am grateful to Prof. Howard Harris for giving me the possibility to visit the University of South Australia. The two months spent in Adelaide have certainly made me grow both as a researcher and as a person. Thanks also to my new Australian friends, that made me feel at home even if I was on the other side of the world.

I would also like to express my sincere gratitude to all the people of the Meyer hospital, whose commitment in the project has been fundamental to achieve the results that are discussed in this dissertation. Thanks to Dr. Tommaso Langiano and Dr. Monica Frassineti, respectively former general and medical directors, who gave me the possibility to work with such an important organisation. Thanks to Mirco Gregorini, Simona Vergna and Dr. Maria Rosaria Agostino for their support. Finally thanks to Raffaella Lo Savio and the other UPAR operators for being so kind and helpful with me.

Many thanks are also for my PhD colleagues. Your support helped me to put things into perspective and to overcome the hardest times of this PhD journey. I will never forget all the moments spent together: the coffee breaks, the lunches, the night get-togethers.. So thanks to Mari, Elisa and Virginia, and above all, thanks to the colleagues with whom I have shared the most of these years, Isabella and Nelson. Thanks for being so good friends.

Thanks to my family, my parents and my brother for their constant presence and unconditioned support. Thank you for bearing me during all of these years.. It is my last thesis, I promise!

Last but not least, thanks Patrizia. You are the one who really knows what bearing me during “under-pressure” moments does mean. Thank you for your patience and for your invaluable support and advices. Thanks for being always by my side. Thanks.

Contents

Chapter 1 Introduction

The operating theatre (OT) is considered as the 'engine that drives the hospital' (Beliën et al., 2006). Its activities, in fact, greatly influence those of other departments and, consequently, the hospital performance as a whole (Cardoen et al., 2010). In addition, the OT is one of the most costly functional areas of a hospital (Denton et al. (2007), May et al. (2011)), and causes almost the 70% of all hospital admissions (Denton et al., 2007). Hospital managers are thus urged to maximise the patient throughput and the relevant revenues, rationalising the use of the hospital resources to contain costs. In this regard, there is unanimous consensus that the performance of the OT strongly depends on the way the surgical activities are planned (Litvak and Long (2000), Guinet and Chaabane (2003)). In the literature the surgical scheduling problem is typically seen as a three stages cascade process (Beliën and Demeulemeester, 2007): (i) the *case-mix planning*, i.e. the determination (usually on a yearly basis) of the total amount of operating room (OR) time to assign to each surgical specialty, (ii) the *master surgical scheduling* (MSS), i.e. the determination of the specialty (or specialties) to assign to each OR on each day of the planning horizon (e.g. two weeks or one month) and, sometimes, the specification of the number and typology of surgeries to be performed each day, and finally (iii) the *selection and sequencing* of patients who have to undergo a surgery.

In general, solving a surgical scheduling problem is noticeably complex. It requires the consideration of: (i) many different types of cases, characterised by different priority levels and requiring different procedures; (ii) many different types of resources, such as ORs, OR personnel (e.g., surgeons, anaesthetists and nurses), surgical and electro-medical equipment, postsurgical resources (e.g., ICU, post-surgical units); (iii) the randomness associated with patients' arrival, surgeries' duration and patients' length of stay (LoS) (May et al., 2000); and (iv) the conflicting priorities and preferences of the scheduling process stakeholders (Glouberman and Mintzberg, 2001). The complexity of such a process coupled with its significant economic and social impact has thus stimulated, in recent years, intensive research activities (Cardoen et al. (2010), Guerriero and Guido (2011), May et al. (2011)). The literature, indeed, abounds of models supporting the scheduling of surgical activities but there is the lack of contributions illustrating the models' implementations (Cardoen et al., 2010). In fact most of the authors test their models on real data, but do not show the practical aspects resulting from the transfer of the model in a real setting. Looking at the literature it is possible to notice that the surgical scheduling is not the only field of the health care sector affected by the lack of implementation of models (Brailsford et al., 2009). Barriers to implementation do exist in many other areas due to two causes: the scarce involvement of the stakeholders in the projects, which leads to a scarce understanding of the addressed problem; the need to publish in high quality journals, which brings academics to formulate more and more complex models, that are not implementable.

This thesis reports the results of a project whose aim was to develop and implement a tool supporting the second phase of the surgical scheduling problem, i.e. the master surgical scheduling. The study has been inspired by a real context, the Meyer hospital in Florence, which is one of the most renowned children's hospital in Europe. In order to overcome the aforementioned barriers to implementation the project has been organised as an action research (Coughlan and Coghlan, 2002). The high involvement of the researchers in the context under study, which is a characteristic of the action research methodology, on the one hand, has led to a better understanding of the process and thus to a scheduling tool that produces satisfactory solutions from the stakeholders' point of view. On the other hand, it has enabled a strong focus on the implementation results, that has allowed the actual transfer of the tool. This thesis contributes to the body of knowledge in two ways:

- it provides fresh insights about the implementation process of a MSS tool. In fact, the reflection process typical of the action research approach has allowed to rationalise what has been experienced and to highlight what factors and conditions can facilitate or thwart the MSS tool implementation;

- it provides different novel mixed integer programming models, which the tool was based on, that has been used to answer to other research questions about the MSS process emerged during the action research project. The formulation of these models and the relevant studies have been addressed using a model based research approach (Bertrand and Fransoo, 2002).

1.1 Structure of the dissertation

Figure 1 represents the structure of the thesis.

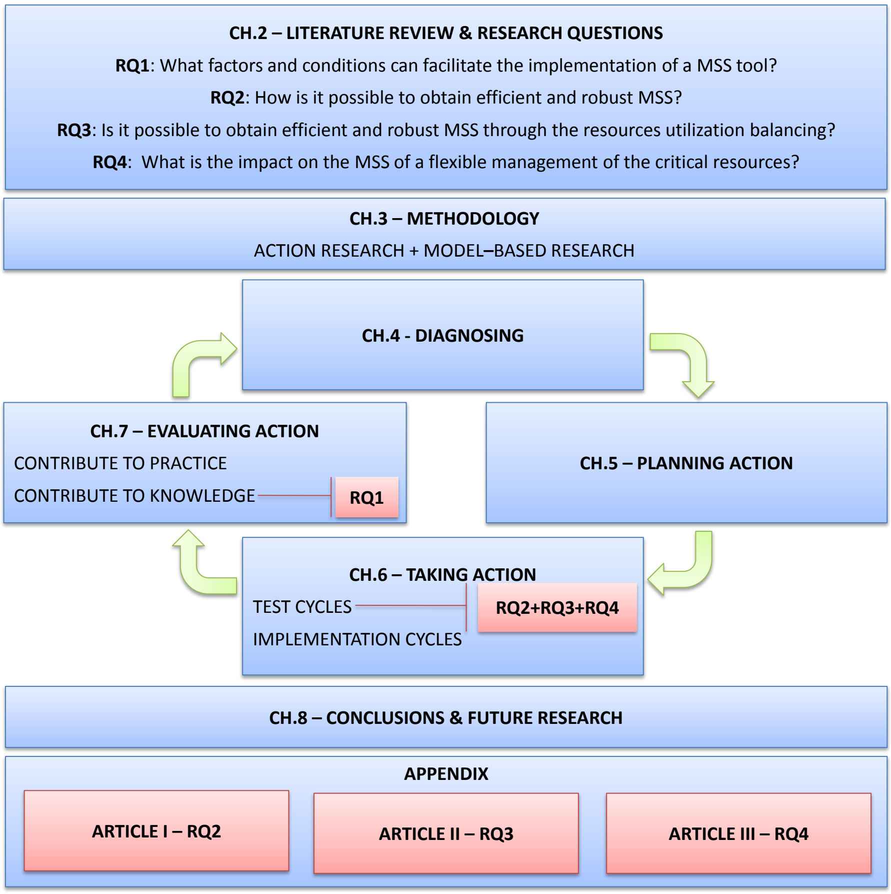


Figure 1 – Dissertation structure

In Chapter 2 the literature about the implementation of models in the health care sector and about the MSS problem is reviewed. In this chapter the research gaps are identified in the light of the existent literature and the research questions addressed in this dissertation are formalised.

Chapter 3 gives details about the methodology, i.e. a combination of action research and model-based research, adopted to answer to the research questions.

Chapters from 4 to 7 are organised as the cyclical phases of the whole action research project.

Chapter 4 concerns the context and the purpose of the project, i.e. what is the rationale for research and practice of the action research study, and the diagnosing phase, in which the problem to solve is identified after having gathered the relevant information.

Chapter 5 is about action planning, i.e. what actions are needed to solve what have been diagnosed.

Chapter 6 presents the actions undertaken. Single actions are conducted following the cyclical action research approach. The first cycles performed at this step aimed to develop the MSS model. The others pertain the implementation of the MSS tool. The studies conducted in three of the former cycles have answered to some research questions that emerged during the project and have been the object of three published research articles. With respect to these actions, the chapter reports only the relevant main points. The integral versions of the articles, e.g. the models' mathematical formulations, the experimental campaigns descriptions, the numerical results, are included in the Appendix.

Chapter 7 reports the evaluation of the actions undertaken during the project. The chapter reports the main results for practice, i.e. qualitative and quantitative impact of the implementation of the tool to the Meyer hospital, and for research, i.e. what are the lessons learned from the action research project, what insights about a MSS implementation process emerged.

Chapter 8 concludes the dissertation and outlines possible directions for future research.

Chapter 2 Literature review

As pointed out in the introduction, this thesis concerns an action research project aiming at developing and implementing a MSS tool in a real context.

For this reason the literature review is organised in two parts. The first concerns the contributions dealing with the implementation of quantitative models supporting decisions in the health care sector, with particular focus to the MSS field. Here the papers presenting models supporting the MSS process are identified and examined from the implementation perspective.

The second part is relevant to the MSS model development. The MSS models proposed in the papers identified in the first paragraph are analysed to understand to what extent they might be suitable for the hospital setting under study, and consequently to identify the literature gaps to be addressed. The development of the model has been a gradual process and required the formulation and the test of different preliminary model versions. In fact, since action research is "a series of unfolding and unpredictable events" (Coughlan and Coghlan, 2002), during the project new problems emerged, entailing different revisions of the model. In correspondence with each revision the literature review was further deepened, giving rise to other research questions.

Each of the following paragraphs of this chapter corresponds to a specific literature review round. In each paragraph it is highlighted:

- what specific problem emerged;
- what is the relevant literature;
- what are the literature gaps;
- what are the research questions.

2.1 Implementation of models for decision support in health care

In the last years, a number of authors have highlighted how the literature lacks of evidences of models implementation in the health care context. In his study, Wilson (1981) pointed out that only 16 out of 200 (8%) applications of computer simulation to health care problems achieved implementation. Referring to the 16 cases, he observed how (i) urgency of the decision making, (ii) timing of the project, (iii) availability of the relevant data and (iv) involvement of the organisation in the project are factors that positively contribute the implementation to be successful. Despite of the increasing of the contributions in this field in the last years, Brailsford et al. (2009) discovered that the implementation rate has not improved. In their review, they classify the studies on modelling in health care according to several dimensions, among which the level of implementation as well. In this regard they distinguish between (i) suggested (theoretically proposed by the authors), (ii) conceptualised (discussed with a client organisation), (iii) implemented (actually used in practice) and found that only the 5.3% of the examined studies belongs to the third category. Similar considerations are pointed out also by Eldabi (2009), who identifies three kinds of implementation barriers: (i) conflicting interest of stakeholders, (ii) lack of relevant tools and (iii) mismatching expectations. He affirms that these barriers are due to the “wicked” nature of the health care problems (Rittel and Webber, 1973) and that “tame approaches” aiming at finding a solution rather than a resolution are not suitable for this typologies of problem. He argues that the most of the contributions in the literature are characterised by elements such as (i) prescription of single solutions, (ii) back-office calculations, (iii) lack of transparency and (iv) lack of interactions with and between stakeholders, thereby making the process a modelling exercise. The lack of implementation is also claimed in the works by Harper and Pitt (2004), Proudlove et al. (2007), Brailsford and Vissers (2011), Mahdavi et al. (2013) and Virtue et al. (2013). Referring to the cited literature, it seems that the major causes of the detachment between models and real world problems in the health care field are:

- the scarce involvement of the stakeholders in the projects, thereby leading to a scarce understanding of the addressed problem;
- the needing to publish in high quality journals, that brings academics to formulate more and more complex models, requiring significant time to be formulated (usually not compatible with health care timescale) but that finally are not easy-to-use.

Both of these factors lead to models that are not suitable to address the real world problems, thus causing a low implementation rate.

All of these considerations are valid also for the models supporting the OT planning and scheduling. With respect to this particular topic, in their recent review, Cardoen et al. (2010) observe how the literature on surgical planning and scheduling lacks contributions in which authors show the results of the models' implementation. The most of the authors test the models, often on real data, but do not illustrate practical aspects of their implementations. Indeed, they do not exclude that the published models have not been implemented at a later stage, but they claim the fact that authors hardly provide results about the implementation process. They encourage authors to share their relevant experiences, because knowledge about the possible causes of a failure or the reason that lead to success, may be of a great value to the research community.

Referring to the studies addressed in the MSS field, the most relevant contributions published in peer-reviewed journals have been examined. These papers are classified according to the dimensions proposed by Brailsford et al. (2009) so to highlight the implementation rate of the MSS models to the real world.

Table 1 MSS models' implementation report

Suggested	Conceptualised	Implemented
Vissers et al. (2005)	Santibanez et al. (2007)	Blake and Donald (2002)
Said et al. (2006)	Testi et al. (2007)	
Tànfani and Testi (2010)	van Oostrum et al. (2008)	
	Zhang et al. (2008)	
	Beliën et al. (2009)	

As shown in Table 1, the most of the models are conceptualised, but only one of the examined papers reports the results of implementation. It cannot be assumed that the conceptualised models have not been implemented at all, but there is no evidence of the transfer of the solution to the studied context.

It is worth pointing out that discussions about the implementation of a MSS tool are by no means novel in the literature. van Oostrum et al. (2010) discuss the pros and the cons of the adoption of a MSS approach for the OR planning and scheduling and its suitability with different hospital organisation structures. However the authors do not provide insights on what factors may lead to success or not in implementing a MSS tool. This thesis aims to fill this literature gap, offering additional experience-driven fresh insights that may be useful for researcher and practitioners to increase the chance to success in the transfer of a MSS model to their hospital setting. One of the objectives of the thesis is thus to respond to the Cardoen et al. (2010) review's call for research, addressing the following research question (RQ1):

"What factors and conditions can facilitate the implementation of a MSS tool?"

2.2 Development of models supporting the MSS process

In the next paragraphs the models supporting the MSS process proposed in the literature are examined. Each paragraph represents a specific literature review round, each corresponding to a different stage of the action research project. From a round to another, the literature review has concerned a higher number of papers, including those that were not published yet at the previous rounds. Each literature review compares one model and the relevant study proposed in this thesis with the existing ones from different perspectives. The analysis are presented in tabular form, allowing to highlight the differences between the literature and the study that was addressed at that round. Since the models and the studies have been object of the published articles included in the Appendix, each table comprises a different article included in this thesis.

2.2.1 *The MSS models characteristics*

At this stage of the project, the problem was to create and test a MSS model reflecting the setting under study and able to produce schedules that are efficient, i.e. characterised by a high number of scheduled surgeries, and robust, i.e. immune with respect to the variability of surgical times (ST) and lengths of stay (LoS). For this reason the previously identified papers proposing MSS mathematical models are further analysed according to different dimensions, mainly concerning the characteristics of the addressed problem and on how such problems are solved. Specifically, building on the taxonomy/dimensions proposed by Cardoen et al. (2010), the papers are analysed according to the following dimensions: (i) *patient characteristics*, i.e., the typology of the patients scheduled (elective vs. non-elective, inpatient vs. outpatient); (ii) *performance criteria*, i.e., the optimised utility function (throughput, resource utilisation and so on); (iii) the *decision delineation*, which identifies the entity (specialty, patient, etc.) to which/whom the decision applies and the type of decision to support (e.g., the assignment of a specialty to a day vs. the assignment of a specific patient to a time slot); (iv) *research methodology*, which refers to the type of analysis (e.g., heuristic vs. exact optimisation) and to the solution techniques adopted (e.g., mathematical programming vs. simulation); (v) *type of constraints*, particularly the hard constraints that are considered (e.g., resource availability, demand, release/due date); (vi) *uncertainty*, which indicates if and how data randomness is managed; (vii) *applicability of the research*, which explains how the models have been tested (i.e., with real data, with realistic data, or not tested); (viii) a *planning horizon* indicating the time horizon on which the models have been applied. Dimensions (i), (v) and (vii) are taken as-is from Cardoen et al. (2010), while dimensions (ii), (iii), (iv) and (vi) have been adapted through the addition of more details in order to better position the study with respect to the literature. Finally, dimension (viii) has been introduced ex-novo. The review is organised and presented into tabular form (see Table 2), where rows represent the aforementioned dimensions, and each column represents one paper. Hence, each cell provides a brief description of a particular paper from a specific perspective. The first row of Table 2 represents the study, i.e. the model and the relevant test phase, conducted at this stage of the project. Such a row helps to highlight what are the differences existing between this thesis and the rest of the MSS literature.

Table 2 MSS literature review – Source: Banditori et al. (2013)

Article	Patient characteristics	Performance criteria	Decision delineation		Research methodology		Type of constraints		Uncertainty	Applicability	Planning horizon
			Scheduled 'object'	Decision details	Type of analysis	Solution technique	Resource	Others			
This thesis – Article I - Banditori et al. (2013)	Elective inpatients	- Throughput maximisation - Appropriate waiting lists consumption - Proper bed allocation	Specialties + procedure typologies + cases due dates	Date, time slot, OR	- Single criterion exact optimisation - Scenario analysis	- Mixed integer programming - Discrete event simulation	Units, surgical staff, equipment, regular OR time	- Procedures' due dates - Procedures mix	Deterministic (optimisation), stochastic (robustness test) ST and LoS	Tested on real and realistic data	1 month
Blake and Donald (2002)	Elective, not specified	Minimisation, for each specialty, of the OR time undersupply with respect to fixed targets	Specialties	Date, OR	Single criterion heuristic optimisation	- Mixed integer programming - Constructive heuristic	Surgical staff, equipment, regular OR time	Max and min n° of OR blocks per week to specialties	Deterministic	Tested on real data	1 week
Said et al. (2006)	Elective, not specified	Minimisation, for each specialty (or surgeon), of the gap between OR time demand and supply	Specialties/ surgeons + procedure typologies	Date, time, OR	Single criterion exact optimisation	Mixed integer programming	Surgical staff, regular OR time	Max and min n° of OR blocks per week to specialties	Deterministic	Randomly generated surgery duration and specialty/surgeon demand	1 week
Santibanez et al. (2007)	Elective, not specified	- Minimisation of the deviation among scheduled and target throughput - Minimisation of bed utilisation	Specialties + procedure typologies	Date, hospital, OR	- Single criterion exact optimisation - Scenario analysis	Mixed integer programming	Units, ICUs, surgical staff, equipment, regular OR time	- Throughput target - Schedule cyclicity	Deterministic	Tested on real data	1 month
Testi et al. (2007)	Elective inpatients	- Minimisation of the gap between specialty demand and supply - Fulfilment of the surgeons' preferences - OR overtime, resource utilisation, n° of shifted cases	Specialties, surgeons, patients	Date, time, OR	- Single criterion exact optimisation - Scenario analysis	- Mixed integer programming - Discrete event simulation	Surgical staff, regular OR time, OR overtime	Max and min n° of OR blocks per week to specialties	Deterministic (optimisation), stochastic (scenario analysis) ST and arrivals	Tested on real data	1 week

Table 3 MSS literature review – Source: Banditori et al. (2013)

Article	Patient characteristics	Performance criteria	Decision delineation		Research methodology		Type of constraints		Uncertainty	Applicability	Planning horizon
			Scheduled 'object'	Decision details	Type of analysis	Solution technique	Resource	Others			
van Oostrum et al. (2008)	Elective, not specified	- Minimisation of the required ORs - Bed occupancy levelling	Procedure typologies	Date, OR	- Multi-criteria exact optimisation - Multi-criteria heuristic optimisation	- Mixed integer programming - Column generation	Units, ICUs, OR overtime	Throughput target	Deterministic LoS, stochastic ST	Tested on real data	1–2 weeks, 1 month
Zhang et al. (2008)	- Elective, inpatients and outpatients - Non-elective, emergency cases	- Minimisation of the patients' LoS - Minimisation of OR time undersupply to specialties	Specialties	Date, time, OR	- Single criterion exact optimisation - Scenario analysis	- Mixed integer programming - Discrete event simulation	Surgical staff, equipment, regular OR time	Specialty demand (elective, non-elective)	Deterministic (optimisation), stochastic (scenario analysis) ST and arrivals	Tested on real data	1 week
Adan et al. (2009)	Elective inpatients	Minimisation of the deviation between realised and target resource utilisation	Procedure typologies	Date	- Multi-criteria exact optimisation - Scenario analysis	Mixed integer programming	Units, ICUs, nursing staff, regular OR time	- Throughput target - Additional restrictions	Deterministic ST Deterministic IC nursing load Stochastic LoS	Tested on real data	1 month
Beliën et al. (2009)	Elective inpatients	- Bed occupancy levelling - Schedule cyclicity - Minimisation of OR sharing among different specialties	Surgeon	Date, time, OR	- Multi-criteria exact optimisation - Multi-criteria heuristic optimisation	- Goal programming - Simulated annealing	Regular OR time	Surgeon demand	Deterministic (multinomial distribution for the n of patients per OR block and patient LoS)	Tested on real data	1–2 weeks
Tànfani and Testi (2010)	Elective inpatients	Minimisation of patients' waiting time	Patients	Date, OR	Single criterion heuristic optimisation	Constructive heuristic	Units, ICUs, surgical staff, regular OR time, OR overtime	Additional restrictions	Deterministic	Tested on realistic data	1 week

Looking at Table 2 and Table 3, it can be noted that the proposed model exhibits decision variables that are similar to those used in Santibanez et al. (2007). However, the two models differ in several aspects. The most important is that the proposed model takes into account the cases' due dates and, consequently, allows—to a certain extent—the exertion of control over the hospital's waiting list. Another important feature is that it actually schedules *procedure typologies* instead of cases. Such a characteristic is shared by half of the reviewed papers. However, none of these deals explicitly with cases' due dates. While due dates are, indeed, considered in Tànfani and Testi (2010), their model assigns OR time slots to actual patients (instead of to procedure typologies) and assumes a planning horizon of one week. As such, their model is unsuitable for monthly planning.

Finally, another important contribution of this part of the thesis is that it addresses ST and LoS uncertainty. In fact one of the aim of the study was to create a model able to produce solutions robust against their variability. Several other authors have incorporated LoS or ST uncertainty into their models (see Cardoen et al., 2010, p. 928). For example, van Oostrum et al. (2008) proposed an optimisation model where a constraint is inserted to keep the probability of realising an OR overtime from exceeding a defined threshold. Specifically, they exploited portfolio optimisation theory (Hans et al., 2008) to reduce the time required to complete a surgical session. In addition, they mitigated the effects of LoS variability through the proper balancing of bed usage. Other authors (e.g., Testi et al. (2007), VanBerkel and Blake (2007), Zhang et al. (2008)) have instead utilised simulation to evaluate, ex-post, the robustness of schedules produced by optimisation models. However none of them have considered simultaneously the variability of both ST and LoS, thereby leading to the following research question (RQ2):

"How is it possible to obtain efficient and robust MSS?"

2.2.2 *Balancing objective functions in the MSS field*

The solutions offered by the first model, despite being efficient and robust, exhibited a scarce balancing of the daily utilisations of the ORs and the post-surgical bed units (hereinafter beds). This fact was considered unacceptable by the hospital, making necessary to include the resources utilisation balancing as criterion in the objective function of the model. Moreover, balanced solutions should be more robust. In fact, in general, if the daily utilisation profiles of ORs and beds are nicely balanced there should be some idle resources to absorb the unexpected peaks caused by ST and LoS variability (Beliën et al., 2009). Hence, utilisation balancing would have contributed also to increase the robustness of the solution, but the extent to which this would have happened was unknown.

Consequently a new literature review round was addressed, comprising only those studies in the MSS fields dealing with resource balancing issues . The papers were analysed according to the following seven dimensions: (i) *balancing criteria,* i.e. the criterion adopted to balance resource utilisation; (ii) *balanced resources*, i.e. the resources whose utilisation is balanced; (iii) *solution technique*, i.e. the typologies of model/s adopted to solve the problem addressed; (iv) *type of analysis*, i.e. the approach followed to solve the problem; (v) *uncertainty*, that indicates if the parameters used in the model/s are deterministic or stochastic and, in this latter case, if the effect of randomness is assessed ex-post via simulation; (vi) *types of distributions*, i.e. empirical, theoretical or both, used to model the stochasticity of ST and/or LoS; (vii) *investigated setting,* i.e. the number and the type (real and/or realistic) of hospital settings where the proposed models are tested, and the number of dimensions (experimental factors) used to differentiate the settings from each other. Dimensions (iii) and (iv) are taken as is from the review scheme given by Cardoen et al. (2010). Dimensions (v) and (vii) has been adapted by adding some details. Dimensions (i), (ii) and (vi) have been developed ex-novo. The review is organised in tabular form and presented in Table 4. Each column of the table represents one dimension, while each row represents a paper. As in the previous paragraph, in order to emphasise the differences between this study and the related literature, the first row represents the specific study of the thesis. Moreover the table comprises also an article (Banditori et al., 2014) in which very preliminary results of the present study were reported.

As it can be observed in Table 4, the minimisation of the maximum utilisation is the most common balancing criterion. It can involve one or more resources, thus respectively entailing the minimisation of a single maximum value of daily utilisation or the sum of the maximum daily utilisation values.

Referring to the balanced resources, beds (belonging to a single or multiple units/hospitals) are considered in all of the examined papers. In these papers, the major aim of the balancing is to reduce the bed utilisation variability thus to prevent schedule disruptions and patient cancellations. In addition, some authors also consider other resources (e.g. IC beds, IC nurses). In the literature, OR balancing is only addressed by Adan et al. (2009) and Banditori et al. (2014). Most of the examined papers deal with the LoS randomness. Instead, ST randomness is only considered by van Oostrum et al. (2008) and Banditori et al. (2014).

Table 4 Balancing objective functions in the MSS field – Source: Cappanera et al. (2014)

Article	Balancing criteria	Rresources	Uncertainty	Distributions	Investigated settings
This thesis – Article II - Cappanera et al. (2014)	Minimisation of the maximum daily utilisation Minimisation of the difference between the maximum and the minimum daily utilisations Minimisation of the sum of the quadratic deviations from a threshold	ORs Beds of a single unit	Stochastic ST (ex-post) Stochastic LoS (ex-post)	Empirical Theoretical	1 real setting 26 realistic settings 3 experimental factors (Beds/ORs ratio, OR utilisation rate and case MIX)
Santibáñez et al. (2007)	Minimisation of the sum of the maximum daily utilisations	Beds of different hospitals	Deterministic ST Deterministic LoS	None	1 real setting
van Oostrum et al. (2008)	Minimisation of the maximum daily utilisation	Beds of different units	Stochastic ST Deterministic LoS	Empirical	1 real setting 35 realistic settings 3 experimental factors (Planning horizon, N° of ORs. N° of bed types)
Adan et al. (2009)	Minimisation of the deviation from a target utilisation	ORs Medium care beds IC beds IC nurses	Deterministic ST Deterministic IC nursing load Stochastic LoS	Empirical Theoretical	1 real setting
Beliën et al. (2009)	Minimisation of the weighted sum of the quadratic mean and variance of the utilisations	Beds of different units	Stochastic Los	Empirical	1 real setting 1 realistic setting 1 experimental factor (Planning horizon)
Chow et al. (2011)	Minimisation of the sum of the maximum daily utilisations	Beds of different units	Deterministic ST Stochastic LoS (ex-post)	Empirical	1 real setting
Carter and Ketabi (2012)	Minimisation of the sum of the maximum daily utilisations	Beds of different units	Deterministic ST Stochastic LoS	Theoretical	1 real setting
Banditori et al. (2014)	Minimisation of the maximum daily utilisation Minimisation of the difference between the maximum and the minimum daily utilisations	ORs Beds of a single unit	Stochastic ST (ex-post) Stochastic LoS (ex-post)	Empirical	1 real setting 4 realistic setting 1 experimental factor (OR utilisation rate)

In sum, within the MSS literature, only the study of Banditori et al. (2014) explores how to obtain MSSs that are robust against *both* ST and LoS variability, by balancing *both* beds and ORs. However, such a study presents a number of shortcomings. First and foremost, it compares two balancing criteria, namely the minimization of the maximum value and the minimization of the difference between the maximum and the minimum values, but does not explain *why* in certain conditions and for certain performances one criterion performs better than the others. Second, the study's computational campaign includes only a limited number of very similar hospital settings and it is based on empirical distributions only. These facts clearly hamper the external validity of the study findings that are, indeed, very context-specific. Third, the study of Banditori et al. (2014) does not consider a fairly well known balancing criterion, i.e. the minimization of the sum of the squared positive deviations of the values from a fixed threshold (Sen et al., 1996). The study addressed in this part of the project aims to fill this gap addressing the following research question (RQ3):

"Is it possible to obtain efficient and robust MSS through the resources utilisation balancing?"

2.2.3 *Resource management policies in the MSS field*

At this stage the problem was to quantify the impact of different critical resources management policies on the MSS efficiency. The MSS literature is here analysed according to the resources that the models consider and to the way the models manage these resources, i.e. the degree of flexibility with which the resources are managed. Also this review round is organised in tabular form. Each column of Table 5 represents a resource, while each row represents a model. In each cell, it is specified if and how the resource is modelled. When a resource is not explicitly considered in the model, the cell contains "NEC." In order to emphasise similarities and differences between this part of the thesis and the related literature, a row representing the study is added.

When a study proposes both flexible and rigid approaches to manage a resource, either the alternatives are reported in the table. Table 5 reveals that most of the authors considered three main critical resources in their models: *surgical teams, ORs* and *units' beds*. Therefore, the remainder of this paragraph will focus on these resources.

Table 5 Resources management policies in the MSS field – Source: Visintin et al. (2014)

Article	Surgical teams	ORs	Surgical units' beds	Other resources
This thesis – Article II - Visintin et al. (2014)	Number of sessions per surgical specialty bounded on a daily and on a weekly basis Session assignment performed: - Once and then considered as fixed (low flex) - Every time the MSS is produced, but only limited changes are allowed with respect to a predefined assignment (high flex)	Fully interchangeable ORs Two sessions per day/OR Sessions: - Dedicated (low flex) - Mixed (high flex)	Three types of surgical units (one day surgery unit and two regular units). - All units are dedicated to specific patient types, no mismatch allowed (low flex) - Regular units are pooled (high flex)	NEC
Blake et al. (2002)	Number of sessions per surgical specialty bounded on a daily and on a weekly basis Session assignment performed once and then kept constant in the following period	Partially interchangeable ORs One session per day/OR Mixed sessions	NEC	Medical equipment
Vissers et al. (2005)	NEC	Fully interchangeable ORs One session per day/OR Mixed sessions	Two types of surgical units (ICU and regular unit) Dedicated units, no mismatch allowed	ICU nursing staff
Santibáñez et al. (2007)	Number of sessions per surgical specialty bounded on a daily and on a monthly basis Session assignment performed once and then considered as fixed	Partially interchangeable ORs One or two sessions per day/OR Mixed sessions	Two types of surgical units (SCU and regular unit) Dedicated units, no mismatch allowed	NEC
van Oostrum et al. (2008)	NEC	Fully interchangeable ORs One session per day/OR Mixed sessions	Two types of surgical units (ICU and regular unit) Dedicated units, no mismatch allowed	NEC
Beliën et al. (2009)	Number of sessions per surgical specialty bounded on a weekly basis Session assignment performed once and then considered as fixed	Fully interchangeable ORs One or more sessions per day/OR Mixed sessions	Several types of surgical units Dedicated units, no mismatch allowed	NEC
Tànfani and Testi (2010)	Number of sessions per surgical specialty bounded on a weekly basis Session assignment performed every time MSS is produced	Fully interchangeable ORs One or two sessions per day/OR Mixed sessions	Two types of surgical units (ICU and regular unit) Dedicated units, no mismatch allowed	NEC
Agnetis et al. (2012)	Number of sessions per surgical specialty bounded on a daily and on a weekly basis Session assignment performed: - Once and then considered as fixed (low flex) - Every time the MSS is produced, but only limited changes are allowed with respect to a predefined assignment (medium flex) - Every time the MSS is produced without limiting the changes allowed with re-spect to a predefined assignment (high flex)	Partially interchangeable ORs One or two sessions per day/OR Dedicated sessions	NEC	NEC

Surgical teams, i.e. the teams of surgeons belonging to the same specialty that actually carry out surgeries are considered explicitly in all but two models (i.e. the model of Vissers et al. (2005) and van Oostrum et al. (2008)). In the remaining works, the availability of surgical teams is modelled by limiting the number of sessions that each surgical specialty can perform on a weekly basis and/or daily basis. Based on these constraints, almost all models assign sessions to specialties, thereby identifying when a surgery team will potentially operate in the planning horizon (session assignment). In addition, some models (Santibáñez et al., 2007, van Oostrum et al., 2008, Banditori et al., 2013) also determine the type and/or the number of surgeries that surgical teams will execute in each session (surgery types assignment). In (Agnetis et al., 2012), instead, one of the proposed models assumes that the session assignment has already been done and, consequently, supports the surgery types assignment only. Most studies suggest that the session assignment should be carried out once and should not be changed frequently (Guerriero and Guido, 2011). The underlying assumption of these studies is that it is not technically feasible to change the session assignment on a monthly (or more frequent) basis because it would make it very complex for surgeons to coordinate their activities inside and outside the OT (van Oostrum et al., 2010). Nonetheless, Agnetis et al. (2012) demonstrate that small and frequent changes in the session assignment can yield substantial benefits and that these benefits are higher than those associated with large yet less frequent changes. Therefore, the authors argue that a limited amount of flexibility in managing surgical teams can produce benefits that are higher than the organisational cost of implementing this solution. For that reason, this latter case was included in the study and compared to the case where the session assignment is considered as already having been performed.

Contrary to surgical teams, ORs are considered as critical in all the reviewed models. However, different authors model these resources in different ways. A first distinction is between interchangeable and partially interchangeable ORs. The former can host every type of surgery; the latter, instead, can host only a limited subset of surgeries and/or specialties. A second distinction pertains to how OR time is divided into sessions. Some authors consider one session per OR per day van Oostrum et al. (2008), some consider two (Santibáñez et al., 2007) or more (Beliën et al., 2009) sessions per OR per day, while others allow both daily sessions and shorter sessions (Agnetis et al., 2012). A third distinction concerns the types of surgery that can be performed in the same OR session. For example, Agnetis et al. (2012) distinguishes two macro-types of surgeries: general surgeries and day surgeries. The former includes all the procedures leading to a LoS of at least two days (one night), and the latter includes those procedures associated with a LoS of just one day. Based on this distinction, Agnetis et al. (2012)'s model allows only dedicated sessions, meaning that within the same session it is not possible to execute both day-surgeries and general surgeries. Instead, he most of the other models allow mixed sessions where these types of surgeries can coexist. While the interchangeability of an OR depends on the structural characteristics (e.g. the presence of certain equipment) of the OR itself, hospital managers have more degrees of freedom in deciding how to subdivide the OR time. Nonetheless, this decision is influenced by the actual number of surgical teams available for each specialty. For example, all-day-long sessions cannot be planned for those specialties relying on less than two surgical teams per day (except in extraordinary cases, one team cannot operate for the entire day). The decision to organise dedicated or mixed sessions, instead, is generally free. The literature suggests that surgeons usually prefer dedicated sessions; surgeons, in fact, can reduce surgery time because of the repetitive nature of their work (Hans et al., 2008). On the other hand, a mixed session makes the scheduling process less constrained and as such, it potentially allows scheduling a greater number of surgeries. In this study, both options are explored.

Finally, post-surgical beds units, i.e. the facilities where patients are cared for following surgical procedures, are considered in six out of eight models. These units are usually classified based on the intensity of care required by the hospitalised patients: e.g. intensive care units, day-surgery units, regular units. Moreover, these units are characterised by a given capacity that is expressed in terms of the number of beds. Certain hospitals (e.g. the Meyer hospital) allocate patients to the regular units based on the specialty. Such a practice makes it easier and faster for surgeons to control and visit their hospitalised patients. Different models assume different numbers of units and unit types. All the reviewed models constrain each type of patient to be hospitalised into a specific unit. In general, the literature (Vincent et al., 1998) suggests that it is risky to accommodate patients requiring thorough care in units characterised by reduced nursing staff or that are physically located far away from the intensive care unit. Thus, units should be pooled only if they are characterised by similar care settings, which is the flexible practice explored in this study. Banditori et al. (2013)'s model, instead, violates this recommendation and allows bed mismatches whenever they allow increasing the OT throughput.

According to Table 1, it can be argued that flexible practices are considered in several studies. However, no study proposes an analysis that investigates how different flexible practices can interact, thereby leading to the following research question (RQ4):

"What is the impact on the master surgical schedule of a flexible management of the critical resources?"

2.2.4 *Multi-criteria approaches to the MSS problem*

During the project it emerged how the MSS problem is fundamentally characterised by different objectives. Hence, it was decided to address it through a multi-criteria approach. Looking at Tables 2-5 it can be noticed how some authors have proposed multi-criteria approaches to the problem (van Oostrum et al., 2008, Adan et al., 2009, Beliën et al., 2009), however none of their models resulted to be suitable to describe the characteristics of the Meyer hospital MSS problem. Consequently, in this thesis, based on the information gathered during the project, a novel goal programming model for the MSS is proposed.

Chapter 3 Methodology

The methodology used in the project is a combination of qualitative and quantitative techniques. The main project is organised as an action research project and, as such, aims to contribute both to knowledge (developing novel models to support the MSS problem and developing understanding about MSS models implementation) and practice (implementing a MSS tool at the Meyer hospital). However, while action research is used to guide the whole project and its cycles are the means through which developing new understanding about the MSS tool implementation, the models were created following a model based research approach.

The next paragraphs give an overview on action research and model based research, with particular emphasis on the application of these methodologies in the operations management field. This section ends showing *why and how* these two methodologies have been combined to deal with the problem addressed in this thesis.

3.1 Action research

Kurt Lewin is considered by the scientific community as the father of action research. In fact, the term "action research" can be found for the first time in the works he conducted together with his associates in the 1940s in the social sciences field. Despite action research methodology has been applied mostly in this field, in the last years it has been successfully applied also to address operations management issues (Westbrook, 1993, Karlsson and Åhlström, 1996, Bennett and Lee, 2000, Hales and Chakravorty, 2006, LaGanga, 2011, Carvalho et al., 2014). This can be justified by the fact that operations management, as the social sciences, often involves people and organisation thus making this approach suitable to address the relevant problems (Westbrook, 1995).

As survey based and case based, action research is an empirical research methodology. However these methodologies mainly differ in the way the researcher is involved in the organisation object of the study. In the surveys and in the case studies the researcher is a detached observer and does not influence the processes of the context under study. Instead an action researcher is directly involved in the context and engages in the research together with the people of the client organisation. One definition of action research is given by (Shani and Pasmore, 1985, p. 439):

> "Action research may be defined as an emergent inquiry process in which applied behavioural science knowledge is integrated with existing organizational knowledge and applied to solve real organizational problems. It is simultaneously concerned with bringing about change in organizations, in developing self-help competencies in organizational members and adding to scientific knowledge. Finally, it is an evolving process that is undertaken in a spirit of collaboration and co-inquiry."

According to Coughlan and Coghlan (2002), the major characteristics of this methodology are:

- research in action, rather than research about action;
- participative;
- concurrent with action;
- a sequence of events and a an approach to problem solving.

Hence, an action research study has a twofold aim: to solve a practically relevant problem and to contribute to the body of knowledge. The researcher is not a mere observer of the system as in the traditional positivist science: he engages with the client organisation and acts like a facilitator of the change in the organisation. Action research is "a series of unfolding and unpredictable events". These events unfold following a cyclical path, comprising: (i) diagnosing the problem, (ii) planning the required actions, (iii) taking the planned action, (iv) evaluating the outcomes of the performed actions. During the unfolding of these cycles the researcher contributes to practice helping the organisation to solve a problem and contribute to knowledge standing back form the action, reflecting on it and on its outcomes. This reflection process is the core of the action research, since its results, that are the lessons learned from the action research project, represent the contribute of the researcher to the body of knowledge.

The characteristics of action research has been also analysed by Gummesson (2000). Specifically he outlines the following ten major characteristics:

1. action researchers take action: the researcher is not a mere observer of the system but he engages with the client organisation and acts like a facilitator of the change in the organisation;
2. action research always involves two goals: an action research study has a twofold aim, to solve a practically relevant problem and to contribute to the body of knowledge;
3. action research is interactive: the researcher and the people of the organisation under study work together and react to the contingent events happening during the project
4. action research aims at developing holistic understanding;
5. action research is fundamentally about change: in the sense that it is applicable to manage the change in an organisation;
6. action research requires on understanding of the ethical framework, since ethical principles must be considered because actions may impact on the people of the organisation
7. action research can include all types of data gathering methods: both qualitative, e.g. interviews, and quantitative, e.g. surveys, are allowed;
8. action research requires a breadth of pre-understanding;
9. action research should be conducted in real time, as action research is an unfolding series of events; however retrospective action research is also acceptable;
10. the action research paradigm requires its own criteria.

As mentioned, action research works as a cyclical process. It comprises one pre-step and four basic main steps. The pre-step concerns the context and the purpose of the study. At this phase the researcher, based on his knowledge about the specific context and about business organisations, is called to answer to the following questions:

1. what is the rationale for action?
2. what is the rationale for research?

Specifically, the action researcher must respectively:

1. understand what is the need and the desirability of the project in the organisation, what are the forces driving the necessity to change and thus establish collaborative relationships with those who have or need to have the ownership of the project; hence, an action research team, composed by the researchers and member of the organisation, has to be arranged;
2. understand why this action research is worth studying, thus why action research is the suitable methodology for the problem under study and what is the expected contribution to the body of knowledge.

The main steps, instead, represent the core of the action research project. As action research is participative, it is important that during all these steps researchers and members of the client organisation adopt a collaborative approach, working together within the action research team. The members of the organisation are the ones who know the context under study best and thus know what will work or not. Their involvement is crucial because they will be ones that will implement the developed solution and know how to manage resistance best. The four steps are:

1. diagnosing: that concern in identifying what are the issues, on the basis of which the actions will be planned and taken. At these stage "hard" data, e.g. statistics about resources utilisation, and "soft" data, e.g. people perceptions, are gathered through different methods, e.g. direct observation, interviews and discussions;
2. planning action: at this step action is planned in the light of the results of the pre-step and the diagnosis;
3. taking action: after the planning phase, the action is implemented in the context of study;
4. evaluate action: outcomes of the previous implementing phase are here analysed; the results of this assessment feed the following diagnosing step leading to the following cycle.

During the unfolding of the action research cycles as described, the researcher has to stand out from the action and to reflect on what is going on. This meta learning process is a continuous inquiring process, that the researchers perform to create knowledge about what they are experiencing. This process is fundamental to create actionable knowledge and to achieve the objective of contributing to the body of literature.

In order to explain how an action research project unfolds, Coghlan and Brannick (2005) use the image of a clock. An action research project is composed of different cycles. These cycles have a different time span, and are performed concurrently.

> "The hour hand, which takes twelve hours to complete its cycle, may represent the project as a whole which may take several years to complete its cycle. The minute hand, which takes an hour to complete its cycle, may represent phases or particular sections of the project. The second hand, which completes its cycle in a minute, may represent specific actions within the project, such as a specific meeting or interview. As in the clock, where the revolutions of the three hands are concurrent and where the revolutions of the second hand enable the revolutions of the minute hand and the revolutions of the second and minute hands enable the completion of the hour hand, the short-term action research cycles contribute to the medium term cycles which contribute to the longer-term cycle."

This simile will be useful in the last paragraph of this section to explain how the research project of this thesis is organised.

3.2 Model-based research

Also described in the operations management literature as analytical modelling (Meredith et al., 1989), quantitative model based research is a methodology "where models of causal relationships between control variables and performance variables are developed, analysed or tested" (Bertrand and Fransoo, 2002). The relationships between independent and dependent variables are here meant as causal and strongly quantitative: a change of value alpha in the independent variable provokes a variation of f(alpha) in an independent variable. Mathematical modelling and simulation are examples of techniques used in this research methodology.

With the purpose to classify the operations management model-based research literature, Bertrand and Fransoo (2002) distinguish between axiomatic and empirical research and between descriptive and normative research (Table 6).

Table 6 Model based research

	Descriptive	Normative
Empirical	ED	EN
Axiomatic	AD	AN

Source: Bertrand and Fransoo (2002)

Axiomatic research is mainly driven by the model itself. The model is usually an idealised version of the real-context and is formulated with some assumptions that are justified by the extent to which these assumptions do not affect the purposes of the study. The objective of the researcher is to find solutions within the defined model and make sure that these solutions provide insights into the structure of the problem as defined by model itself. It can be both descriptive, i.e. aiming at understanding and describing the causal relationships between the variables in the process under study, and normative, i.e. aiming at predicting the effect of new strategies and policies, improving the results available in the literature.

Instead empirical research is guided by empirical findings and measurements. Here the aim of the researcher is to guarantee that the model accurately replicates observations and actions in reality and thus to remove the assumptions owned by the axiomatic approaches. As the axiomatic research, empirical research can be both descriptive and normative.

In order to highlight the differences among these four approaches to the model based research, Bertrand and Fransoo (2002) use the model proposed by Mitroff et al. (1974). The model concerns the operational research typical approach in solving problems. Specifically four cyclical steps are identified (Figure 2):

1. conceptualisation, where the problem addresses is conceptualised and decisions about what variables have to be included in the model are taken;
2. modelling, concerning the model formulation according to the conceptualised problem;
3. model solving, concerning finding a solution to the formulated model;
4. implementation, which means the transfer of the model results to the studied real context.

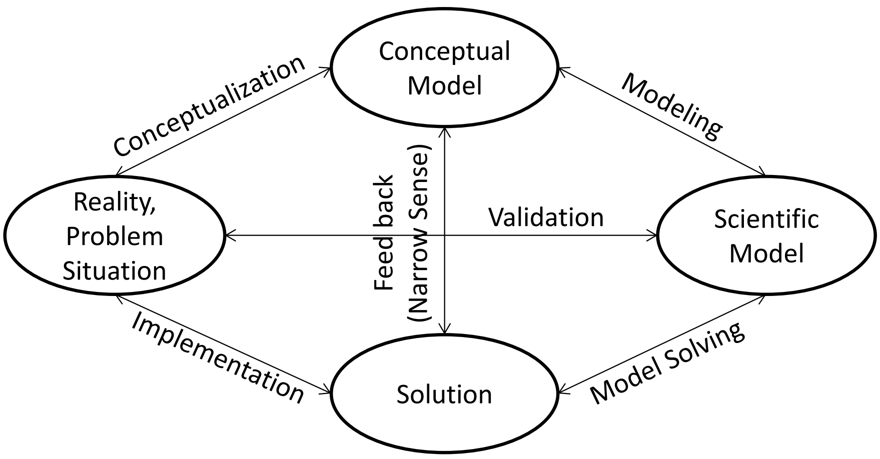


Figure 2 Mitroff model - Source: Mitroff et al. (1974)

As pointed out by Mitroff et al. (1974), a research cycle can begin and end at any step of the cycle and sometimes "shortcuts" in the cycle happen.

According to the Mitroff et al. (1974)'s model, the model based research typologies can be described as follows (Figure 3):

- axiomatic descriptive research (AD) is a cycle comprising only modelling, as the researcher focuses on developing model to better understand the model itself, based on a conceptual model often taken from the literature;
- axiomatic normative (AN) is a cycle comprising both model formulation and model solving; the researcher, in addition to conceptualize and formulate the model, focuses on finding solutions to the model and to feed them back to the model formulation;
- empirical descriptive (ED) is a cycle comprising conceptualization, modelling and validation
- empirical normative (EN) is the most complete cycle comprising all the steps of the cycles model.

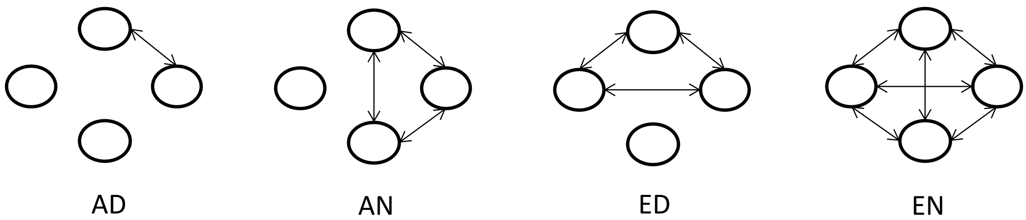


Figure 3 Model based research cycles – Based on Mitroff et al. (1974)

3.3 Why combining action and model-based research

This thesis is based on a project aiming at implementing a MSS tool in a real hospital setting. Two main issues arise from this statement:

1. formulating the MSS model on which the tool is be based on;
2. implementing the MSS tool in the real hospital setting under study.

Apart from the fact that both issues lead to rethink the MSS process (due to the introduction of the tool), the specific problems arising from these issues are very different in nature. The first issue requires to identify the MSS characteristics, to formulate a model that replicates, under certain assumptions, the actual process and to test the performance of the model. The second issue requires to deal with the impact that the introduced change might have on the organisation, to manage the resistance to change of the involved members of the organisation and to reflect on the outcomes. As such, it requires to adopt different methodologies to address the two issues.

In order to address the first issue, based on the setting and on the general characteristics of the MSS problem recognised in the literature, the specific MSS problem must be conceptualised and modelled. The model must then be tested to provide useful feed-back to fine tune the model formulation. Hence, the axiomatic model-based methodology seemed to be the most suitable approach to address the first issue. However, since the MSS conceptual model has been based on elements both from literature and the real context, the conceptualisation phase was indeed a relevant phase of the adopted approach. As a result the followed research methodology slightly differs from the classical axiomatic normative approach (Figure 4).

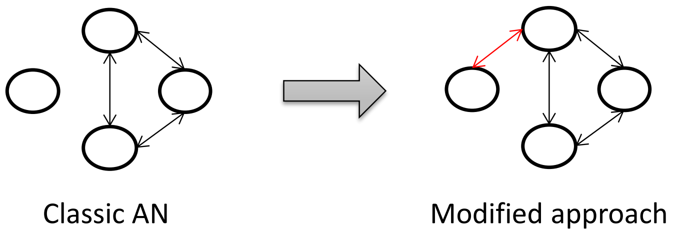


Figure 4 Modified AN model based research cycle – Based on Mitroff et al. (1974)

Instead, in order to address the second issue, the action research methodology was adopted. Since action research is characterised by a strong participation of the researcher in the organisation, it seemed to be the best approach to overcome the barriers to implementation. In addition, this high involvement would likely lead to a better understanding of the process and thus to a MSS model that produces satisfactory solutions from the stakeholders' point of view. Finally the reflection process typical of the action research approach will allow to better rationalise what has been experienced, in order to highlight what factors and conditions have facilitated or thwarted the MSS tool implementation process.

As a final remark it is worth to point out that the use of two or more research methodologies to address different aspects of the same problem in the operations management field has been recently encouraged by Sodhi and Tang (2014). In their paper they state that the different stages of the research stream, i.e. (i) awareness, (ii) framing, (iii) modelling and (iv) validation, can be all addressed through a single research approach, but this would lead to two major problems: (i) "island of methodology" and (ii) disconnection from practice. They argue that these problems can be avoided adopting different research methodologies to address the different phases, given the fact that each research methodology is more suitable that the others to face a specific phase of the research stream. For this reason they finally encourage doctoral students to adopt different research methodologies when approaching new problems.

3.4 How action and model-based research are combined

Figure 5 reports a scheme of the research project on which this thesis is based on.

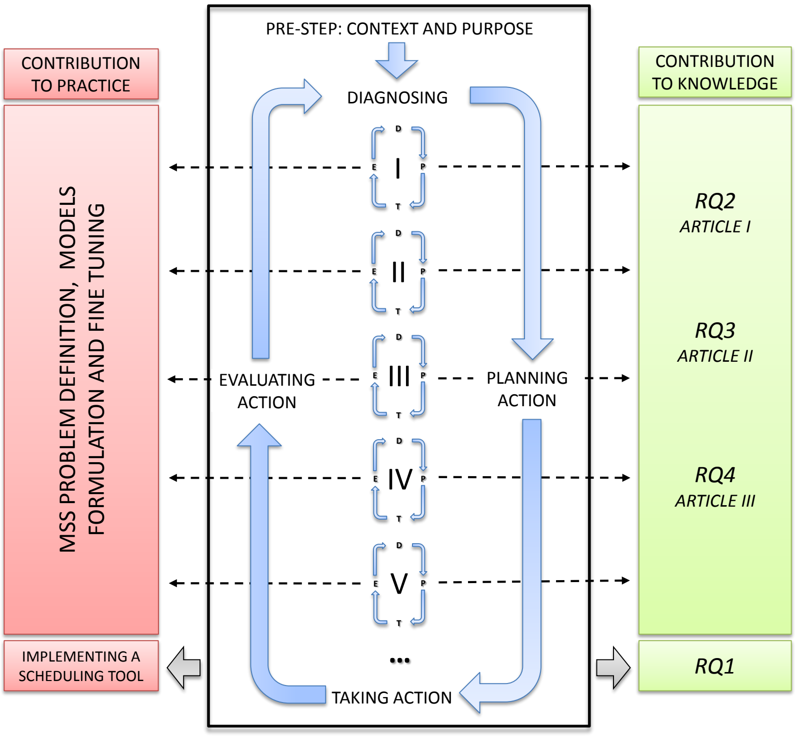


Figure 5 Action research project scheme

The whole project is organised as an action research project. In the central part of Figure 5 it is possible to observe the main cycle that represents the project (in the clock simile by Coghlan and Brannick (2005) the hour hand). Its outcomes, i.e. implementing a MSS tool and providing insights on such a process, in the lower part of the figure. As pointed out in the literature review chapter, these two outcomes are respectively practically and scientifically relevant. During the main cycle, other smaller cycles have been undertaken (in the clock simile by Coghlan and Brannick (2005) the minute hand). Each cycle represents a different MSS model formulation and test and their activities were undertaken following a axiomatic model based research approach. To some extent a parallel between the phases of the mod-el-based and the action research in the smaller cycles can be made. Conceptualization can be consider as the diagnosing phase, that uses the output of the previous evaluation phase to define the problem to be solved. The model formulation can be represented by the planning phase in which it is established how the problem would be solved. The model solving, both for test or implementation purposes, can be seen as the taking action phase, in which the effort is focused on solving the problem. Feed-back phase can be considered as the evaluation phase. As a results, someone can argue that these phases can be conducted regardless the action research approach and its cycles. This can be true. However, the need to integrate the model based approach in the action research cycles arises from two facts. First, often models fail to be implemented because their solutions are not satisfactory. This is often due to a scarce understanding of the problem by the researchers. The strong involvement of the organisation members typical of the action research projects can overcome this issue. The problem conceptualisation, the model formulation, the model solving and the feed-back, i.e. each cycle, were conducted within the action research team thus ensuring the solutions provided by the model to be both satisfactory and implementable. Second, the inquiring process typical of the action research approach undertaken during these cycles allowed to reflect on what characteristics of a MSS model make its solutions more implementable. This process has been fundamental to respond to the first research question of this thesis and thus to contribute to the body of knowledge.

As the main cycle, the smaller cycles have produced practically and scientifically relevant results as well. Details about these results will be given in the further sections.

The next chapters of this thesis are organised following the typical action research cycle that comprises four basic steps (Coghlan and Brannick, 2005):

1. diagnosis;
2. planning action;
3. taking action;
4. evaluating action.

Chapter 4 Diagnosing

4.1 Context and purpose

4.1.1 *What is the rationale for action?*

This thesis is based on a research project involving the IBIS lab research group and the Meyer hospital. The hospital top management was committed to optimise the operations of the OT in terms of resource utilisation and throughput. Improving these performance in fact leads to:

- increase the revenues – since hospitals in Italy are subjected to the DRG reimbursement system (Fattore and Torbica, 2006), the higher the patient throughput, the higher the total incomes;
- decrease the costs – principally extra-costs associated to personnel that must work overtime;
- increase the patients satisfaction – because a higher throughput brings to a reduction of the patients' waiting times.

Since the performance of the OT strongly depends upon how its activities are scheduled, the hospital management decided to focus on the surgical planning and scheduling process. At Meyer hospital such a process was not optimised and after a first analysis it seemed that there was a vast room for improvement.

The project started in 2011 with the aim to implement, and thus to transfer to the hospital, a MSS tool, through which the planner of the hospital would have been able to produce optimised surgical schedules. Since it was decided to organise the project as an action research project, an action research team was set up. The team was composed by the IBIS lab researchers and by the members of the organisation that are mostly involved in the surgical scheduling process, that are:

- the general and the medical director, who committed the study;
- the OT manager, who is responsible for the assignment of OR sessions to the surgical specialties;
- the beds manager, who is responsible for the allocation of patients to the post-surgical units;
- one member of the planning department, who is responsible for reporting scheduling issues of the planning department to the OT and the beds manager.

However, other members of the organisation, like surgeons or the planning department personnel, have been occasionally involved in the project.

4.1.2 *What is the rationale for research?*

As pointed out in the literature review chapter, the problem of models implementation in the health care context, but generally speaking, the problem of disconnection of the research from practice, is actually a significant issue that operations management researchers are called to face before starting a project. Action research is indeed a methodology that can overcome this issue. In fact, its aim is to provide outcomes both for practice and research. Such a twofold aim is pursued through a strong and continuous collaboration between researchers and practitioners that ensures the practical relevance of the study.

With specific regard to the surgical planning and scheduling models, Cardoen et al. (2010) point out how literature lacks of contributions in which authors show the results of the implementation of their models to the real setting under study. The continuous interaction of the researcher with the process stakeholders (that is a characteristic of the action research) allows to better understand the characteristics of the problem and thus to develop models that are easily implementable. The reflection process during the project allows find out more about what are the conditions that facilitate a surgical scheduling model implementation. Finally the models and the studies are novel with respect to the existent literature.

In summary this action research study contributes to knowledge in three ways:

- it gives fresh insights about what characteristics of a MSS model make it easy-to-implement;
- it provides experience-driven understanding about what conditions can facilitate a MSS tool implementation process;
- it presents new MSS models and shows the results of novel studies performed through these models,

thus responding to the research questions pointed out in the literature review chapter.

4.2 Diagnosing

The aim of this phase was to understand the surgical planning and scheduling process of the Meyer hospital. The relevant data have been gathered through different methods including direct observation, interviews and the analysis of the information systems.

The direct observation focused on both the activities of planning department and the activities performed in the OT. The interviews involved all the hospital members of the action research team, the planning department personnel and some members of the OT staff, i.e. surgeons, anaesthetists and nurses. The analysis of information systems concerned hospital data base (for those information that were digitalised) and documental analyses (for those information that were not digitalised and thus that are managed "on paper").

This data gathering phase allowed to deeply understand the relevant process and to make hypothesis on which parts of the process can be improved.

4.2.1 *The planning and surgical scheduling process at the Meyer hospital*

The Meyer hospital OT consists of seven ORs: five of these are partially interchangeable and host 15 surgical specialties (urology, otorhinolaryngology, paediatric surgery, neonatal surgery, ophthalmology, orthopaedic surgery, gynaecology and obstetrics, trauma centre, hand and microsurgery, oral and maxillofacial surgery, orthopaedic oncology, cardiothoracic surgery, gastroenterology, burns and plastic surgery); the remaining two ORs are dedicated almost entirely to specific surgical specialties or treatments (hemodynamics and bronchial endoscopy) and partially to emergencies and urgencies. At the Meyer hospital, emergencies and urgencies are managed by allocating them a fixed amount of operating room sessions and a fixed number of beds. The Meyer hospital actually allocates 42 beds to elective patients. However this number can change during the year due to unexpected urgencies and emergency. The beds are organised into three physically distinct units. One unit accommodates patients with short expected LoS, i.e. day surgeries, which occupy a bed for a single day. The other two units accommodate patients with longer expected LoS, i.e. ordinary surgeries, which occupy a bed for more than one day.

The hospital waiting lists, i.e. the set of patients needing a surgery, are populated on the basis of surgery request forms that are filled out by surgeons after visiting the patients. The form clearly indicates, for each case: (i) the surgical specialty, (ii) the diagnosis, (iii) the procedure that the patient is expected to undergo and (iv) a priority class. The priority class determines the maximum number of days within which the case should be scheduled and, thus, the case's due date. There are four possible priority classes. The first three classes are associated with 30, 60 and 180 days of waiting time, respectively. The fourth class indicates that the patient does not have a due date. In addition, the form indicates: (v) the expected duration of the procedure (surgery duration); and (vi) the expected LoS. There are three possible time ranges for procedure duration: less than one hour (short duration), between one and two hours (medium duration), and more than two hours (long duration). With respect to the expected LoS, the form indicates if the patient is a day surgery or an ordinary surgery.

Presently, at the Meyer hospital, the activities of the surgical specialties that use the five aforementioned interchangeable ORs are planned by the planning department. The activities of the remaining specialties are managed by their respective departments.

At the Meyer hospital the surgical planning and scheduling process works as the typical process reported in literature (Beliën and Demeulemeester, 2007). It is a three-phase process, consisting of case mix planning, MSS and patients selection and sequencing.

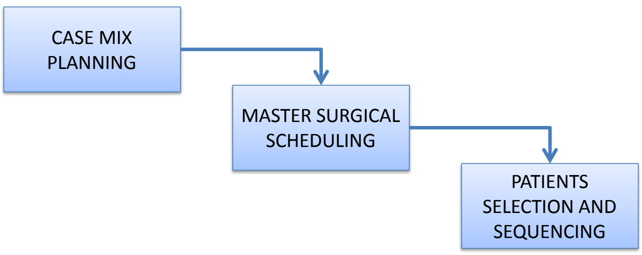


Figure 6 The surgical planning and scheduling process

As can be noticed from Figure 6, it is a cascade process in the sense that the output of the upstream sub-process represents the input for the downstream one. Each sub-process has different frequency and planning horizon. Moving from the case mix to the patients selection and sequencing, frequency increases and planning horizon decreases.

The case mix planning is performed once a year. It consists of a negotiation process between the hospital management and the surgeons responsible for the different surgical specialties. The decisions taken at this step concern the number of OR hours that each specialty is assigned with on an annual basis.

The MSS is performed once a month and consists of the assignment of the specific OR sessions to the different surgical specialties, for a number of sessions that depends on the output of the case mix planning. Basically the MSS is a timetable that specifies what specialty will operate on a given day, at a given time, in a given OR. In addition, it provides a rough indication of the number of day surgeries and ordinary surgeries that should be performed in each session. At this stage the availability of the OR anaesthetists, nurses and electro medical equipment is considered. The MSS is performed by the OR and the beds manager in concert with the surgeons. These latter are asked to indicate particular requirements and preferences. It is worth noticing that, due to logistical reasons, the MSS tends to remain constant during the year. In fact keeping the MSS constant makes the surgeons easier to coordinate their surgical activities with the other ones outside the hospital. However each month little changes in the MSS are usually made.

The patients selection and sequencing is performed once a week. In this phase the surgical planning department personnel call the patients and recruit them for undergoing the surgeries, thus populating the OR sessions indicated in the MSS. Patients are chosen such that: (i) the sum of the expected surgery duration of the cases assigned to each session does not exceed the duration of the session itself; (ii) the expected number of hospitalised patients for each day does not exceed the number of expected available beds; and (iii) the percentage of short-, medium- and long-lasting surgeries scheduled in the weekly plan reflects approximately the percentage on the waiting list. At the Meyer hospital, the demand for short-, medium- and long-lasting surgeries has proven to be fairly constant all year round. Hence, by scheduling a constant mix of short-, medium- and long-lasting surgeries, the hospital avoids leaving an excessive amount of long-lasting surgeries on waiting lists, which would make the scheduling process more complex in the following weeks or months. Lastly, if possible, patients with closer due dates are given higher priority.

During the patients schedule execution, the planning the department is also asked to manage the variations to the schedule that may occur as a consequence of:

- a cancellation due to a patient no-show (need to replace the cancelled patient);
- a patient with higher priority that must be scheduled in the place of a yet scheduled patient (need to cancel one or more surgery to schedule this patient);

- any other unpredictable event that can give rise to a schedule disruption, e.g. the lack of beds because some patients needed to stay in their beds more than expected.

4.2.2 *Information systems analysis*

From a "surgical request" perspective, the surgical process can be represented as in Figure 7.

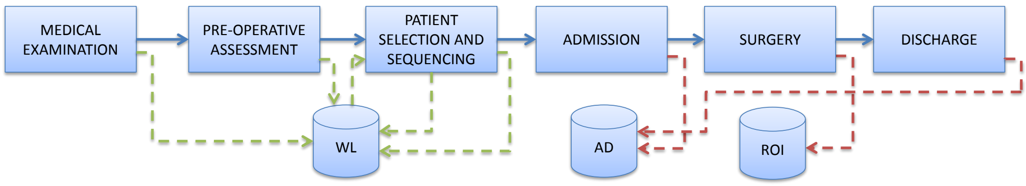


Figure 7 The surgical process at the Meyer hospital

The process starts with the medical examination. The surgeon sees the patient and make a surgery request whether the patient needs a surgery. The request is sent to the planning department and, together the other requests, constitute the waiting lists. Subsequently, the patient is called to make a pre-operative assessment with the anaesthetist. If the anaesthetist gives positive feedback than the patient can be scheduled. After a certain amount of time, that depends on different factors (see the previous paragraph), the patient is selected from the waiting lists to be operated. Then the patient is hospitalised and a surgeon, that is not necessarily the one who prescribed the surgery, operates the patient. After the surgery, the patient stays in a bed for a certain period and after that he is discharged. During this process several data about the patient/request are transferred and stored. These data concern:

1. information ante-surgery (green dotted line in Figure 7), that are information stored before the surgery is performed, mainly request's details, data about anaesthetist assessment, fixed date of the surgery, contacts with the patient's;
2. information post-surgery (red dotted line in Figure 7), that are information stored during the patient hospitalization concerning data about the surgery, such as, actual surgical date and times, name of the surgeon, actual diagnosis and surgical procedure performed and data about the hospitalization, such as admission date or discharge date.

All the data belonging to the second category were digitalised and stored in a two database called respectively ROI ("Registro operatorio informatizzato") and AD ("Ammesso-

Dimesso”). Instead data belonging the first category were not digitalised and the requests were managed “on paper”. However there was an on-going project aiming at implementing a new information system software. Such a new software, on the one hand, aims to digitalize the waiting list information, i.e. the information belonging to the aforementioned first category, on the other hand, aims to integrate the ante and post-surgeries’ information.

4.2.3 *Data analysis and process criticalities*

The collected data and information were then analysed. The fact that the waiting lists are managed on paper did not allow to make an initial quantitative assessment about the situation of the patients waiting times. However the interviews with the personnel of the planning department reveal how the lists were growing. They argued that each month the number of incoming surgical requests were higher than the number of surgeries they are able to schedule. Moreover they claimed other facts:

“It’s difficult is to select patients in order to respect their priorities and due dates”;

“The indications reported in the MSS about the number of patients to be scheduled are too rough. In order to prevent OR overtime and beds overbooking it happens that we schedule less surgeries than the indicated number”

“We are not able to adequately cope with the short notice cancellations. These patients must be replaced with other ones that have already undergone the anaesthetist examination and that are expected have the same ST and LoS. It is difficult to decide what patients to prepare because the indications about ST and LoS on the requests are too rough. Hence it happens that the replacement of a cancelled patient is made with a patient that requires a different amount of resources”

“Surgeons requests about scheduling specific patients in specific sessions are difficult to manage because often are not compatible with the indications reported in the MSS”

Instead, data about the STs and patients’ LoSs were available. The former have been collected from the ROI database, the latter from the AD database. The elaboration of these data allowed us to calculate respectively OR and beds’ utilisation. The analysis of the medical records showed how OR utilisation can be significantly improved. Beds’ utilisation on the contrary was high. From this first analysis beds seemed to be the bottleneck of the surgical process. Such a consideration was confirmed by the interviews with the OT and beds manager. This latter claimed that:

"It often happens that there are too many patients to accommodate thereby leading to open the day surgery unit at night or to cancel surgical patients, due to bed shortages"

These cancellations are probably the reason why OT manager claimed that:

"ORs are often underutilised"

Finally, we realised that the other resources involved in the surgical process, i.e. OR and nurses and electro medical equipment, were not scarce and their availability did not represent a constraint for the OT performance.

In summary, the direct observation of the surgical scheduling process, the interviews and the data analysis revealed different facts:

- besides the ORs, also surgical teams and beds were critical resources and must be necessarily considered in the surgical scheduling process;
- with respect to these critical resources, there was the need to report a more clear indication on the surgical requests about the critical resources that the patient is expected to require. The requests, in fact, reported only a clear indication about what surgical specialty the patient requires but not about the ST and LoS; this indications would help the planning department to
    - schedule an adequate number of patients while preventing operating room overtime and post-surgical beds overbooking and
    - prepare an adequate set of patients for replacing the short notice cancellations;
- the decision on how to assign the ORs to the different specialties is taken separately from the decision on how to assign the beds. The first decision is taken in the MSS and the second one in the patients selection and sequencing, after the first one has been already taken. Such a scarce coordination, which is due to the cascade approach, might lead to underutilise the ORs, because the beds might constraint the system more than the case in which they are considered in the MSS phase.

For these reasons the surgical scheduling process needed to be redesigned. The next paragraph illustrates how actions were planned to improve the process.

Chapter 5 Planning action

After the diagnosing phase, in which the members of the action research team collaborate to understand what is going on in the organisation under study, actions to solve the identified problems have to be planned. Since action research is “an unfolding series of unpredictable events” it was difficult to know exactly what actions would have been taken during the whole project. As a consequence it was difficult to make a specific plan and to expect how long the project would have taken to provide the pursued results.

In the previous chapter some criticalities of the actual surgical planning process identified by the AR team have been pointed out. In order to solve such criticalities the AR team decided to develop a MSS tool after redesigning the scheduling process. Despite the impossibility to make a detailed plan, we figure out to take different actions, following four different lines of intervention:

- patients classification;
- process redesign;
- tool development;
- tool fine-tuning.

5.1 Patients classification

In order to give the possibility to accurately compute how much resources are utilised as a consequence of a certain patients schedule, patients must be categorised according to their resources consumption. Since we identified three critical resources, namely the surgical teams, the ORs and the beds, we decided to classify the patients in the waiting list in the so-called *surgery groups*. The patients pertaining to a same surgery group require a surgical team belonging to the same specialty, are expected to require a similar amount of OR time, i.e. are characterised by a similar expected ST, and are expected to occupy a bed for the same number of days they are expected, i.e. are characterised by the same expected LoS. The information about the surgery group, besides being useful to the planning department personnel both in the phases of patients selection and replacement, would have been useful in the new scheduling process to integrate the decision of the assignment of the ORs and the beds to the specialties in the MSS phase.

5.2 Process redesign

As pointed out in the diagnosis section, the classic surgical planning framework, that is the one adopted at the Meyer hospital, does not allow for the downstream resources coordination. These resources are to some extent considered only in the patient selection and sequencing stage, after the MSS has been already performed. As a consequence, downstream resources, that are considered only in the patients selection and sequencing phase, might constraint the scheduling process more in the case that they are considered in the MSS. For this reason we decided to integrate the assignment of the surgical specialties to the beds and the assignment of the surgical specialties to the OR sessions in the MSS phase. We called this new approach enhanced MSS. In this new MSS process, besides deciding what specialties will occupy the different OR sessions, it is established also how many cases, belonging to the different surgery group, will be scheduled in the different OR sessions. Adding the surgery group information at this stage guarantees a higher coordination of the upstream and downstream resources, thus leading to higher utilisation of the OT. At this stage the selection of the surgery group is made also considering the due dates of the relevant patients. Since the planning horizon in the MSS is longer than in the patient selection and sequencing phase, considering the due date at this stage improves their fulfilment rate.

Besides improving the OT performance and the due date fulfilment, such a new process helps the planning department personnel in the patients selection and sequencing phase, because they have not to decide the number and the type of patients to schedule.

5.3 Tool development and implementation

The tool supporting the new MSS process should enable the end-user to:

1. produce the MSS following the enhanced MSS approach;
2. assess the feasibility of a modification of the current MSS as a consequence of certain events.

With regard to the first point, we decided to integrate in the tool a mixed integer programming model to create optimised MSSs. The model must take into account of different aspects that are considered as fundamental for the MSS. Specifically it must take into account:

- the limited availability of the three critical resources (ORs, beds, surgical teams);
- the patients characteristics, in terms of the critical resources consumption, i.e. the surgery groups, and the priorities/ urgency of the patients, i.e. the patients' due dates;
- the uncertainty affecting the surgical times and the LoSs;
- other characteristics reflecting the process stakeholders' priorities and needs.

The aim was to obtain a model able to produce schedules that satisfy the stakeholders' expectations.

In order to cope with the second point we decided to give the possibility to the end-user to visualise the utilisation of the critical resources of the current MSS and to calculate the impact that a modification may have on them. In fact it may happen that the MSS needs to be changed at the last minute because of

- a cancellation due to a scheduled patient that cannot undergo the surgery, in this case there is the need to replace the cancelled patient to avoid unused capacity;
- a patient with higher priority that must be scheduled, in this case there is the need to cancel one or more surgeries to schedule this patient without exceeding the available capacity.

In both cases it is necessary to assess the feasibility of the change with respect to the actual utilisation of the resources. In fact

- in the waiting lists it is not always possible to find a patient belonging to the same surgery group of the cancelled scheduled patient;
- in the schedule it is not always possible to find a patient belonging to the same surgery group of the patient with higher priority that needs to be scheduled.

In these cases the replacement must be done with a patient belonging to a different surgery group, thus requiring a different amount of resources. A tool that enables the end-user to assess the impact of any modification is essential to avoid significant under or over-utilisation of the available resources capacity.

5.4 Data requirement

In order to use the aforementioned scheduling tool several digitalised data were required. The tool in fact should be easily fed with updated data each time a new schedule is produced. These data are mainly relevant to the characteristics of the patients in the waiting list, e.g. how many patients are available to be scheduled, what are their priorities, what are the diagnoses and the surgical procedures they must undergo. As pointed out in the diagnosis section, all these data were available but they were not digitalised at that moment. However, as said, there was an on-going project aiming at introducing a software for the waiting list management through which the waiting list data would have been digitalised. Hence, even if data were not available for the prompt implementation of a scheduling tool, the on-going project for the introduction of the new waiting lists management system seemed to represent the best conditions for the introduction of a new scheduling tool, which might require some information that are not considered in standard waiting list management software. The fact that a new information system was in course of implementation, in fact, could allow us to guide the definition of the information about the patient that should be editable through the waiting list management software and that are required to the scheduling tool to work. From the discussions within the action research team, it emerged what information about the patients in the waiting list were required by the model and thus by the tool. Specifically the following information were needed:

- an indication of the expected ST, an indication of the expected pre and post-surgical LoS and the surgical specialty, i.e. the surgery group. This latter depends on the pathology affecting the patient and the procedure he/she must undergo and on the surgical specialty of the surgical team he/she requires. At that moment the surgery group information about the ST and the LoS were not reported on the surgery requests, however there was the need to give the possibility to associate this information to the patient when the request is inserted in the system through the new waiting list management system;
- indications about the day on which the surgery should be performed. These indications were usually reported on the surgery requests. However every time they

were written in different ways. Sometimes there was indicated only the priority class, sometimes there was a priority class and an extended indication, e.g. “to be operated before” and a date, sometimes no indication were provided. For the tool purposes, it was essential to standardise the way this information are provided on the surgery requests. For this reason it was decided to give the possibility to enter the following information in the new waiting list management system when prescribing a surgery:

- a latest due date, that can be explicitly specified by the surgeon on the surgery request or can be calculated adding the number of days within the surgery should be performed, 30, 60 or 180 days, given by the patient priority class, to the arrival date of the surgery request to the planning department;
- an earliest programmable date. In fact because despite the fact that the patient is in the waiting list, it may happen that the patient should not undergo the surgery before a certain date, e.g. the patients’ family may have problems to organise for the surgery before a certain day;
- a specific date, indicating the date the patient must necessarily undergo the surgery. It often happens when the patient hospitalization is difficult to organise from a logistic point of view, e.g. the patient lives far from the hospital.

The model will use these information when producing the MSS.

Chapter 6 Taking action

The previous chapter identifies what actions were required to implement the scheduling tool. Even if the contents of the actions to conduct were quite clear, it was difficult to predict how the project would have unfolded and thus how many cycles would have been needed to achieve the tool implementation. However we expected to undertake two types of cycles, depending on the relevant aim. A first type of cycle, i.e. the test cycles, would have aimed to develop a model able to produce schedules that are easily implementable A second type of cycle, i.e. the implementation cycle, aimed to transfer of the tool to the Meyer hospital. Each cycle would have represented a specific MSS production.

6.1 Action research test cycles

As pointed out in the methodology chapter, some of the smaller cycles of the project, i.e. the test cycles, have been addressed through the axiomatic model based methodology. However each models' development and the test is undertaken following an action research approach. In Figure 8 is reported a scheme of a test cycle.

In the diagnosing phase the relevant information are gathered and used to conceptualise the problem to solve. In the planning phase a model is formulated based on the conceptualisation at the previous step. In the taking action phase the model is fed with real data coming from the Meyer hospital and solved.

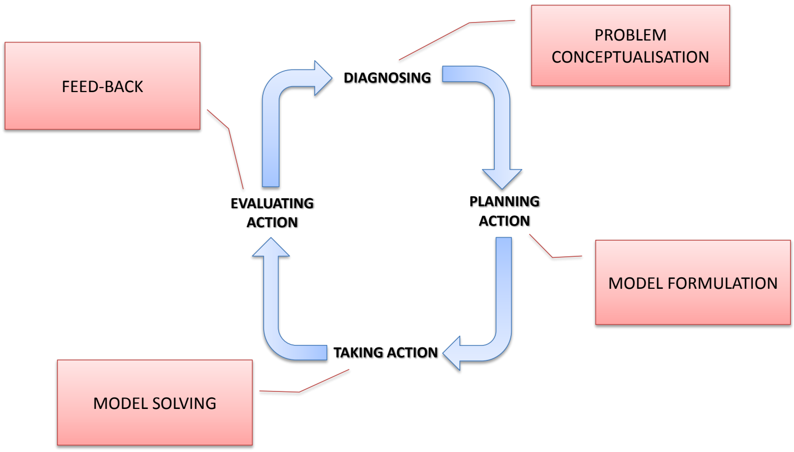


Figure 8 Action research test cycle scheme

The obtained solution is then discussed within the action research team to understand to what extent the schedule is satisfactory or not. In this latter case the feedback about the solution is used to start another cycle with the aim to modify the model and improve the quality of the solution. The self-reflection process on the outcomes of these cycles allowed to understand what characteristics of the MSS model make it easier to implement. During the unfolding of these cycle it was also defined the information that were required by the model and thus what data about the patients would have been needed to be editable through the new waiting list management system.

An implementable version of the MSS model was obtained after four action research test cycles. In each of these cycles a different version of the model was formulated and tested. The contents of the first three cycles have been the object of three studies, in which the results have been useful to answers to the research questions pointed out in the literature review chapter. Since the aim of this chapter is to show why and how the models have been changed during the project, it reports only a summary of the contents of these studies. Specifically each paragraph corresponds to a cycle, in which it is highlighted what are the characteristics of the developed model, why these characteristics have been chosen , what are the results of the tests, what findings emerged from these tests and how these results influenced the following cycles. The paragraphs do not report the mathematical formulation of the model, details about the experimental campaigns and numerical results. The interested reader is referred to the Appendix in which the complete versions of the articles concerning these studies are included.

All the models were developed using the surgery groups information and are based on the enhanced MSS approach. Each model is characterised by an objective function, which gives the criterion/criteria to optimise, and a set of constraints. Referring to the constraints, two types are considered: the first refers to the availability of the critical resources, the second refers to some quality requisites of the solution, mainly related to the patients' due dates and to the scheduled cases mix with respect to the surgery groups. Despite all the models consider these issues, each one differs in the way these issues are mathematically modelled. Since data about the waiting lists were not digitalised at moment of the these cycles, some ad-hoc instances were created to test the models. Details about the data analysis and the instances generation are reported in the original articles.

6.1.1 *Test cycle I*

Diagnosing – problem conceptualisation

The interviews with the members of the action research team allowed to define a first version of the MSS problem. Specifically, from the interviews emerged that:

- since resources availability (and hence costs) are fixed in the short term, in order to maximise efficiency, the solution must be produced maximising the number of scheduled surgeries; in this sense an efficient solution allows for the maximisation of revenues, for the containment of the costs and for an increase in the patient satisfaction as a consequence of the reduction of their waiting times;
- the solution must be robust against the variability of the ST and the LoS, in the sense that it should give rise to few or no disruptions when a surgery lasts more than its expected ST or a patient stays in his bed more than his expected LoS. In these cases we may have, respectively, overtime and overbooking, that entail costs and patients dissatisfaction;
- OR daily availability is organised in two sessions, morning and afternoon, for which the surgical specialties guarantee the availability of a certain number of surgical teams;
- beds are organised in three units, each accommodating cases characterised by the same LoS, e.g., day surgery unit for cases with a LoS equal to one day, week hospital unit for a LoS equal to two days and ordinary unit for LoSs longer than two days. When possible, each case type should be accommodated in the appropriate unit. However, bed mismatches, e.g., long-stay case types accommodated in day surgery unit, may be tolerated if they allow for an increase in the throughput;
- as many as possible patients' must be scheduled before the relevant due dates;

- leaving an excessive number of surgeries with approaching due date or long duration on the waiting list should be avoided. Otherwise, the maximisation of the throughput in the current planning horizon would lead to criticalities in the long run.

Action planning – model formulation

In order to deal with the aforementioned requirements, a mixed integer programming model was formulated. Such a model allowed the production of the MSS and exhibits the following constraints and objective function:

- at most, one surgical specialty can be assigned to a given OR session, i.e. a given OR, on a given session, on a given day; in a certain OR session, however, the model is free to schedule cases belonging to each surgery group within that specialty;
- surgical specialties guarantee the availability of a certain number of surgical teams (0 or 1) for each for each session, for each day;
- bed mismatches are allowed but their number is penalised;
- with respect to the patients' due dates, the model imposes the strict fulfilment for those patients whose due date expires in the planning horizon; the missed scheduling of the other patients, i.e. those whose due date expires in the following periods, is penalised according to the relevant due date, i.e. the more the patient's due date is forthcoming, the higher the penalty;
- the solution must be characterised by a certain mix with respect to the surgical time. Specifically we classified the surgery groups in three categories depending on the ST, short-, medium- and long-lasting surgery groups and imposes that the percentage of cases belonging to these time ranges must fall into a range;
- the objective function is composed of three terms and the importance of each term is given by a weight; specifically the weights are chosen in the way that the tree criteria are hierarchically ordered. The first and most important term is the maximisation of the number of scheduled surgeries. The second term is the minimization of the penalties relevant to the patients' due dates. The third and less important term is the minimization of the penalties associated with the bed mismatches.

In order to test the robustness of the solution produced by the optimisation model, we created also a discrete-event simulation model through which it was possible to assess the impact of the variability of the ST and the LoS on the deterministic solution, in terms of overtime and overbooking cancellations.

Action taking – model solving

The mixed-integer programming model formulated at the previous step has been used to produce the MSS. This schedule has been then simulated through a simulation model to assess its robustness. Since the schedule was not robust, i.e. it gave rise to an unacceptable amount of OR overtime and beds overbooking cancellations, we decided to introduce the resource slacks in the optimisation phase. Specifically the MSS is produced considering a smaller amount of OR time and beds than what is actually available. By introducing resource slacks, the optimisation model schedules fewer surgeries and, consequently, the obtained solution likely gives rise to fewer cancellations and to less overtime.

Evaluating action – feed-back

The experimental campaign performed at this cycle shown how a MSS that is robust against variability of ST and LoS can be achieved through the adoption of the resource slacks. A trade-off between efficiency and robustness does exist: the higher the efficiency, i.e. the number of scheduled surgeries, the lower the robustness and vice versa. Referring to the case of the Meyer hospital we demonstrate how adopting slacks of 10% and 12% respectively for ORs and beds, it is possible to schedule a higher number of surgeries with respect to the actual planning process, i.e. 582 vs. 495, experiencing at maximum 9 overbooking cancellations over a planning horizon of 4 weeks. The obtained solutions were discussed within the action research team. Despite the positive feedbacks about the results of the study, the proposed solutions did not satisfy the expectation of the stakeholders, specifically:

- in the stakeholders' opinion, the MSS exhibited too many day surgeries with respect to the number of day surgeries that were usually performed at the Meyer hospital;
- daily resources utilisations, especially the OR ones but also beds' ones, were unbalanced, e.g. in some cases OR sessions utilisation were under the 50%.

Another fact that we noticed during the experimental campaign was that the due dates' strict fulfilment can strongly limit the potentiality of the model and in some cases it might also give rise to the model infeasibility.

6.1.2 *Test cycle II*

Diagnosing – problem conceptualisation

Since the solution obtained at the previous cycle was not satisfactory, the model needed to be modified:

- the number of day surgeries was too high, and this was probably due to the possibility of the model to allow bed mismatches. In fact, if the model had not the possibility to make bed mismatches, the number of daily day surgery would be limited by the number of available beds in the day surgery unit. If bed mismatched are allowed, another bounding strategy should be adopted;
- the solution should have been more balanced with respect to the daily resources utilisations. This because a fair distribution of the workload positively affects the satisfaction of OR personnel, e.g. surgical teams, nurses. Moreover, in general, if the daily utilisation profiles of ORs and beds are nicely balanced, there should be some idle resources to absorb the unexpected peaks caused by ST and LoS variability (Beliën et al., 2009). In other terms, a higher balancing should lead to a higher robustness, especially when average resource utilisation is high;
- the due dates' strict fulfilment can limit the number of scheduled surgeries. The due dates' respect is an important issue for a hospital, however the hospital members of the action research team recognised that the strict respect of the due dates (via hard constraint) can be meaningless, especially for those patients with low priority and that are expected to stay in the waiting list for long time. For this reason we decided to relax the due dates' constraints with the aim to improve the objective function's value;
- in order to further improve the performance of the model we decided to change the way the model deals with the availability of the surgical teams.

Action planning – model formulation

At this stage some constraints and the objective function have been changed. Specifically:

- referring to the ORs, at most, one surgical specialty can be assigned to a given OR session; in a certain OR session, the model is still free to schedule cases belonging to each surgery group within that specialty;

- surgical teams availability is here expressed in terms of maximum weekly number of OR sessions; however, surgical teams belonging to the same surgical specialty cannot occupy more than one OR at the same time;
- mix and patients due dates' are regulated through the combination of two mix constraints; the first impose that the solution must be characterised by a certain mix with respect to the LoS, the second impose a mix with respect to the ST. Specifically, with regard to the LoS, the cases are subdivided in two classes: the day surgeries and the ordinary surgeries; on the other side, with regard to the ST, the cases are subdivided into the following two classes: short lasting surgeries (less than 1 hour) and long lasting surgeries (more than 1 hour). Then, the two mix constraints specify that for each class, the scheduled surgeries in that class fall within a minimum and maximum percentage of the overall scheduled surgeries. These minimum and maximum percentages are defined based on the composition of the current waiting list, i.e. mix and patients due dates';
- bed units are pooled, in the sense that bed mismatches can still occur; however, the number of day surgeries is now bounded by the mix constraints; hence there is not the need to penalise them in the objective function;
- the objective function still has three terms; however the terms about the penalisation of the due date's missing and the bed mismatches are now substituted by two terms aiming respectively at balancing the OR and the beds' daily utilisations. Since the aim is to obtain balanced solutions, the balancing terms were prioritised with respect to the efficiency term. This latter term allows to obtain, among the balanced solutions, the one characterised by the highest number of surgeries. Furthermore, since overbooking was considered more undesirable than overtime at the Meyer hospital, the beds' balancing was prioritised with respect to OR balancing.

Action taking – model solving

As in the first cycle, at this stage the MIP model was used to produce the MSS and the simulation model to test the robustness of the MSS against the ST and LoS variability. The experimental campaign here aimed to assess the impact on efficiency and robustness of three different objective functions, each of which incorporates a different criterion for the balancing of the resource utilisations. Specifically these criteria are (i) the minimisation of the maximum daily utilisation, (ii) the minimisation of the range between the maximum and minimum utilisation (minRng) and (iii) the minimisation of the *overrun (*minOvrn), i.e. the positive deviation between the actual resource utilisations and target utilisation values. By using the Meyer hospital's data, three different schedules were produced through the three models and their robustness were assessed by means of the simulation model. In order to give generalizability to the results of the study, these objective functions were tested in correspondence with several realistic hospital settings, that were created starting from the Meyer hospital's one. In particular these settings were characterised by different values of:

- minimum and maximum percentage values for the mix constraints;
- available beds/OR time ratio;
- minimum and maximum percentage values constraining the overall OR utilisation.

Evaluating action – feed-back

The study performed at this cycle revealed that none of the investigated policies allows superior performance in terms of efficiency, balancing and robustness to be achieved concurrently. However, depending on the hospital management's priorities and needs, it is always possible to identify a policy that allows for a reasonable trade-off among these performance criteria. In the case of the Meyer hospital, minOvrn seemed to be the best balancing policy because it represents somehow an intermediate case with respect to the other two policies. In fact minRng is at least effective as the other policies in balancing the beds. In addition, it allows for a better OR balancing and thus for a smaller overtime. Moreover, it leads to higher bed saturation and to a larger number of scheduled surgeries. However, it also causes a higher overbooking. The properties of minMax are quite the opposite than minRng. Finally looking at the numbers, consistently with the literature (Proudlove et al., 2007) and with the results obtained at the previous cycle, a target daily resource utilisation of around 85% seemed to be an adequate value for guarantying the robustness of the schedules.

With respect to the solution obtained, however, the stakeholders were not still satisfied: despite the fact that the number of day surgery was now adequate, the solution exhibited too many of surgeries with LoS equal to two days. In addition the combination of the two percentage mix constraints would still lead to the model infeasibility. Finally the solution also evidenced how the way with which the surgical teams were managed was too flexible and thus scarcely implementable: some assignment surgical specialty-day were not implementable because surgeons are not always available for surgical activities. They in fact need to coordinate them with the other activities inside and outside the hospital.

6.1.3 *Test cycle III*

Diagnosing – problem conceptualisation

As a consequence of the feedback of the previous cycle, the model required some changes as follows:

- we decided not to utilise percentages constraints to deal with the patients' due dates and the mix of the solutions because, besides causing problems in terms of model feasibility, they can strongly limit the potentiality of the model;
- it was necessary to modify again the way that surgical teams' availability is managed, making it less flexible;
- it should be useful to understand what is the impact of certain choices about how the resources are management and thus modelled.

Action planning – model formulation

Based on the emerged problems, the model was modified as follows:

- instead of percentage mix constraints, we added to the model some constraints that guarantee a minimum number of surgery to be scheduled for each surgery group in the waiting lists; each target number is established according to the due dates of the patients in the waiting list and to the mix desired by the process stakeholders;
- it was decided to fix the assignment of the specialties to the OR sessions
- with respect to ORs it was decided to model them in such a way that a given OR session is dedicated, i.e. it can host either day surgeries or ordinary surgeries;
- beds are here dedicated and not pooled as in the previous cycles;
- the objective function maximises the number of scheduled surgeries, i.e. the efficiency; robustness is pursued by means of resource slacks equal to 15% both for ORs and beds.

However, besides this base model, we decided to create several different version of the model in order to investigate what is the impact on the performance of the process of a more flexible management policy of the critical resources. Specifically, in the "flexible version" of the model, resources are modelled as follows:

- the assignment of the surgical specialties to the OR sessions is not fixed but can change every time the MSS is produced; the number of changes with respect to the original assignment can be set by through the model;
- bed units are pooled;
- OR session are mixed and can host both day and ordinary surgeries.

Action taking – model solving

The base model was solved to obtain the solution relevant to the Meyer hospital. However we wanted to show to the action research team what it would have been the increase in the efficiency if resource were managed more flexibly. To do so, and to give generalizability to the study, we carried out an experimental campaign based on a 2^3 experimental design (Montgomery and Runger, 2003). In detail, we consider the way the three critical resources are managed as factors and we assume two possible levels for each factor: "high" when the resource is managed in a flexible way and "low" otherwise. Combining factors and factor levels, we obtained eight ($=2^3$) configurations. For each of them we ran the optimisation model in correspondence with 30 randomly generated instances, that were obtained starting from real data coming from the Meyer hospital.

Evaluating action – feed-back

The analysis revealed that the best results in terms of efficiency can be achieved by managing flexibly both surgical teams and ORs. Moreover, the analysis showed that, if a hospital cannot manage flexibly the surgical teams, then it can still improve its efficiency by managing flexibly the ORs and vice versa. However, the analysis revealed that if both surgical teams and ORs are managed flexibly, pooling surgical units has no significant impact, while if only one of these two resources (or none) are managed flexibly, then pooling surgical units produces significant benefits. However, even if the flexible management could have improved the efficiency of the process, it was decided to not implement any of these practices. In fact the benefits arising from the flexible management of the resources were considered by the stakeholders too low to justify the organisational costs emerging to implement it. With respect to the solution relevant to the Meyer hospital, the stakeholders were still unsatisfied. The solution exhibited too few long ordinary surgeries in terms of both ST and LoS. In fact, in order to prevent infeasibilities, the target number of surgeries for each surgery group was set at a low level. Hence, the model, after satisfying these constraints, in order to maximise efficiency, chase surgeries that "consume" few resources to fill the remaining room. Finally, imposing a strict upper bound on the maximum utilisation of the resources, i.e. though the introduction of the resource slack, can make to exclude solutions that are more efficient than the optimal one and that are still robust. For example a solution with the daily utilisation of one OR equal to 85.5% is unfeasible for a model with a resource slack equal to 15%. However it could be more efficient and still robust than optimal solution of the model.

6.1.4 *Test cycle IV*

Diagnosing – problem conceptualisation

After three action research test cycles some evidences emerged:

- hard constraints for the quality requisites of the solutions, i.e. patients' due dates fulfilment and the mix of the scheduled surgeries, can make the model infeasible, depending on the input data; a model whose feasibility is too sensitive with respect to the data is not implementable; infeasibilities are in fact difficult to manage for people without expertise in the modelling field;
- there is the need to deal with patients' due dates and the mix, but neither the percentage range constraint nor the coverage constraint seemed to suitable to make the model able to produce satisfactory solutions;
- it is better to pursue robustness integrating the resource utilisations balancing in the objective function; utilising resource slacks can make to not consider some solutions that are still robust and more efficient.

In order to deal with these issues we decided to modify the model, including also the patients' due date fulfilment and the mix respect in the objective function. Hence the objective function comprises five criteria:

1. maximisation of the number of scheduled surgeries;
2. balancing of the daily utilisations of the ORs;
3. balancing of the daily utilisations of the beds;
4. patients' due dates fulfilment;
5. surgeries' mix respect.

We decided to adopt a goal programming approach in which the importance of each objective can be set depending on the preferences of the decision maker. This flexibility seemed to make the goal programming approach the most suitable to address the our problem. In fact the importance of the different criteria can be easily fine-tuned by the model user though a set of weights, thus allowing to alternatively focus on different objectives depending on the stakeholders' preferences and needs.

With respect to the resources management, the stakeholders decided to not implement any flexible practice. Surgical teams availability is fixed through an allocation grid which indicates for each day, for each OR and for each session the specialty and the type of surgery, i.e. ordinary or day surgeries, that can be performed. Hence each OR session can host only ordinary or day surgeries. Finally post-surgical units are dedicated, i.e. they can host exclusively patients belonging to certain surgery groups and bed mismatches are not allowed.

Action planning – model formulation

Based on the previous conceptualization, a new MIP model was developed. In this paragraph the mathematical formulation of the new model is given.

Let us define the following sets and parameters

S	the set of specialties
K	the set of surgery groups
O	the set of ORs
D	the set of days in the planning horizon
T	the set of time slots
B	the set of post-surgical beds units
P	the set of patients' priority classes
F	the set of surgery types, i.e. ordinary or day surgery
I	the set of the criteria
G_{sfodt}	the allocation grid, equal to 1 if surgeries of type f and belonging to the specialty s can be performed in the OR o, on day d, in session t, 0 otherwise
s_k	the specialty of surgery group k
f_k	the surgery type of surgery group k
H_{odt}	the available time of OR o on day d, time slot t
γ_k	the expected ST of surgery group k
α_k	the expected LoS after surgery required by group k
β_k	the expected LoS before surgery required by group k
R_{bd}	the number of beds in unit b available on day d
E_{kd}	the number of cases in surgery group k, whose earliest programmable date is on day d
e	the number of time periods preceding and following the planning horizon
L_{pc}	the number of cases of priority p whose due date is on day c, $\forall p \in P, \forall c \in D \cup D_1 \cup \ldots \cup D_e$
$\hat{q}$	the target value for the OR utilisation rate
$\hat{r}$	the target value for the beds utilisation rate
w_k	the weight associated with surgery group k

$\hat{y}_k$ the target number of surgeries belonging to surgery group k

w_{pj} the penalty associated with cases of priority p with due date in D_j not scheduled in the planning horizon, $\forall p \in P, \forall j \in 1,..,e$

$\hat{n}_i$ the target value for the objective i

W_i the weight associated with criteria i

In addition, let us define the following variables:

y_{kpodt} the number of procedures of surgery group k assigned to OR o on day d in time slot t

z_{bd} the number of beds of type b occupied on day d

q_{odt} the utilisation rate of the OR o, on day d, in session t

r_{bd} the utilisation rate of the beds in unit b on day d

u_{ph} the number of cases with priority p with due date in time period D_j not scheduled in the planning horizon, $\forall p \in P, \forall j \in 1,..,e$

q^+_{odt} the positive deviation of the utilisation rate of the OR o, on day d, in session t from the fixed target

q^-_{odt} the negative deviation of the utilisation rate of the OR o, on day d, in session t from the fixed target

r^+_{bd} the positive deviation of the utilisation rate of the beds in unit b on day d from the fixed target

r^-_{bd} the negative deviation of the utilisation rate of the beds in unit b on day d from the fixed target

y^-_k the negative deviation of the number of scheduled surgeries belonging to the group k the fixed target

n_i the value associated with the objective i

n^+_i the positive deviation of the objective i from the fixed target

n^-_i the negative deviation of the objective i from the fixed target

Since the allocation grid is fixed in input through the parameter G_{sfodt}, the surgery groups *k(S)* that can be scheduled in a given OR o, on a given day d in the time slot t are restricted to the ones belonging to the surgical specialty s and type f for which the parameter G_{sfodt} is equal to 1. For this reason variables y are defined on a subset of the set $(K \times P \times O \times D \times T)$. Specifically, we introduce, for each surgery group k, the set A_k that is a collection of triples *(o,d,t)* indicating, the OR sessions, in which the surgery group k can be scheduled. More formally, if s_k and f_k denote respectively the specialty and the type of surgery group k, A_k are defined as follows:

$$A_k = \{(o,d,t) \text{ s.t. } o \in O, d \in D, t \in T \text{ and } G_{s_k f_k odt} = 1\}$$

Variables y_{kpodt} are thus defined $\forall k \in K, \forall p \in P, \forall (o,d,t) \in A_k$. Given these sets, parameters and variables, let us define the following constraints:

$$\sum_{\substack{k \in K, p \in P: \\ (o,d,t) \in A_k}} \gamma_k y_{kpodt} \le H_{odt} \qquad \forall o \in O, \forall d \in D, \forall t \in T \tag{1}$$

$$\sum_{\substack{k \in K, p \in P, o \in O, \\ t \in T: (o,d,t) \in A_k}} \sum_{d'=\max(1, d-\alpha_k)}^{\min(|D|, d+\beta_k)} y_{kpod't} = z_{bd} \qquad \forall b \in B, \forall d \in D \tag{2}$$

$$z_{bd} \le R_{bd} \qquad \forall b \in B, \forall d \in D \tag{3}$$

$$q_{odt} = \frac{\sum_{\substack{k \in K, p \in P: \\ (o,d,t) \in A_k}} \gamma_k y_{kpodt}}{H_{odt}} \qquad \forall o \in O, \forall d \in D, \forall t \in T \tag{4}$$

$$r_{bd} = \frac{z_{bd}}{R_{bd}} \qquad \forall b \in B, \forall d \in D \tag{5}$$

$$\sum_{\substack{p \in P, o \in O, d' \in D: d' \le d, \\ t \in T: (o,d',t) \in A_k}} y_{kpod't} \le \sum_{d \in D: d' \le d} E_{kd'} \qquad \forall k \in K, \forall d \in D \tag{6}$$

$$\sum_{\substack{k \in K, o \in O, d \in D, \\ t \in T: (o,d,t) \in A_k}} y_{kpodt} + \sum_{h=1}^{j} u_{ph} \ge \sum_{c \in D \cup D_1 \cup \ldots \cup D_j} L_{pc} \qquad \forall p \in P, \forall j \in 1..e \tag{7}$$

$$q^+_{odt} \ge q_{odt} - \hat{q} \qquad \forall o \in O, \forall d \in D, \forall t \in T \tag{8}$$

$$q^-_{odt} \ge \hat{q} - q_{odt} \qquad \forall o \in O, \forall d \in D, \forall t \in T \tag{9}$$

$$r^+_{bd} \ge r_{bd} - \hat{r} \qquad \forall b \in B, \forall d \in D \tag{10}$$

$$r^-_{bd} \ge \hat{r} - r_{bd} \qquad \forall b \in B, \forall d \in D \tag{11}$$

$$\sum_{\substack{p \in P, o \in O, d \in D, \\ t \in T: (o,d,t) \in A_k}} y_{kpodt} + y^-_k \ge \hat{y}_k \qquad \forall k \in K \tag{12}$$

$$n_1 = \sum_{\substack{k \in K, p \in P, o \in O, \\ d \in D, t \in T: (o,d,t) \in A_k}} y_{kpodt} \tag{13}$$

$$n_2 = \sum_{\substack{o \in O, d \in D, \\ t \in T}} q^+_{odt} + q^-_{odt} \tag{14}$$

$$n_3 = \sum_{b \in B, d \in D} r^+_{bd} + r^-_{bd} \tag{15}$$

$$n_4 = \sum_{y \in K} w_k y_k^- \qquad (16)$$

$$n_5 = \sum_{p \in P, j \in 1..e} w_{pj} u_{pj} \qquad (17)$$

$$n_i^+ \geq n_i - \hat{n}_i \quad \forall i \in I \qquad (18)$$

$$n_i^- \geq \hat{n}_i - n_i \quad \forall i \in I \qquad (19)$$

$$y_{kpodt} \in \mathbb{N} \quad \forall k \in K, \forall p \in P, \forall (o,d,t) \in A_k \qquad (20)$$

$$u_{pj} \geq 0 \quad \forall p \in P, \forall j \in 1..e \qquad (21)$$

$$z_{bd} \geq 0 \quad \forall b \in B, \forall d \in D \qquad (22)$$

$$q_{odt}^+ \geq 0 \quad \forall o \in O, \forall d \in D, \forall t \in T \qquad (23)$$

$$q_{odt}^- \geq 0 \quad \forall o \in O, \forall d \in D, \forall t \in T \qquad (24)$$

$$r_{bd}^+ \geq 0 \quad \forall b \in B, \forall d \in D \qquad (25)$$

$$r_{bd}^- \geq 0 \quad \forall b \in B, \forall d \in D \qquad (26)$$

$$y_k^- \geq 0 \quad \forall k \in K \qquad (27)$$

$$n_i^+ \geq 0 \quad \forall i \in I \qquad (28)$$

$$n_i^- \geq 0 \quad \forall i \in I \qquad (29)$$

$$\min \sum_{i \in I} \frac{W_i (n_i^+ + n_i^-)}{\hat{n}_i} \qquad (30)$$

Constraints (1) assure that for each OR session, the sum of the surgical times of the scheduled surgeries does not exceed the available time. Constraints (2) compute the number of utilised beds for each unit and for each day of the planning horizon. Constraints (3) limit the number of occupied beds. Constraints (4) and constraints (5) compute respectively the daily utilisation of the OR sessions and of the different units. Constraints (6) assure that the number of scheduled surgeries for each group does not exceed the number of cases that are available, depending on the relevant earliest programmable dates.

Constraints (7) allows for the respect of the due dates' of the patients in the waiting list. Specifically, these covering constraints impose that the number of schedules surgeries of a given priority p should be greater or equal to the number of cases in the waiting lists belonging to the same priority. If this cannot happen the corresponding variable u, which measures the number of not scheduled surgeries of priority p and with a due date falling in the time period D or D_j, assumes a value greater than zero. The summation of the u variables is penalised in the objective function according to the priority and the due date time period.

Constraints (8) and (9) compute the OR session daily utilisation positive and negative deviations from the fixed target, for each triple (o,d,t). Constraints (10) and (11) are their counterparts for the units utilisation. Constraints (12) calculate the negative deviation of the number of scheduled surgeries belonging to a given surgery group from the relevant fixed target. Constraints (13)-(17) compute the values of the five objectives, specifically we have: Constraint (13) computes the value of the number of scheduled surgeries, Constraints (14) and (15) calculate the sum of the deviations from the fixed targets of the OR session daily utilisations and of the units daily utilisations respectively. Constraint (16) calculate the weighted sum of the penalties associated with the missing achievement of the target levels for the number of scheduled surgeries of the different groups. Constraint (17) instead computes the weighted sum of the penalties associated with the scheduling of the patients with certain due dates and priorities. Constraints (18)-(19) compute, for each objective the positive and the negative deviation from the fixed target. The remaining constraints impose that y variables are positive, integer and defined as previously mentioned and the other variables are positive. The weighted sum of the deviations calculated in constraints (18) and (19) is minimised in the objective function (30).

Action taking – model solving

Since at that moment information about the waiting lists were digitalised, the model was tested with the actual data about the patients in the waiting lists.

With respect to the model's parameters we considered:

- a planning horizon of 4 weeks;
- 14 surgical specialties and 126 surgery groups;
- 5 ORs, whose availability was established by the allocation grid defined by the OR manager;
- 3 surgical units, i.e. day surgery unit, which accommodates the day surgery patients of each specialty, week hospital and ordinary unit, accommodating the ordinary patients of different specialties;
- a target value for the daily utilisations of OR sessions and units of 85%;

- for the cases whose earliest programmable date is before the first day of the planning horizon we consider the first day itself as earliest programmable date;
- the target number of cases for each surgery group is fixed according to stakeholders' preferences on the mix; the weights w_k are set in the way that the model prioritises the surgery groups with longer waiting lists;
- in order to penalise the missing scheduling of a patient having a certain priority class and due date, the patients are clustered in three time periods ($e=2$), according to the relevant due dates:
    - patients whose due date is expired D_1;
    - patients whose due date expires in the planning horizon D;
    - patients whose due date expires beyond the planning horizon D_2;

  In order to take into account of the due dates of the patients before and after the planning horizon, we extended the planning horizon at right and at left of a number of days that is three times the planning horizon itself. When a patient did not report a due date, we assumed its value equal to the last day of the right extension of the planning horizon. The weights w_{pj} in the fifth objective are set in the way that the model prioritises (i) cases belonging to different due dates' time clusters; and (ii) cases with higher priority. With respect to the time clusters the weights are set in the way that the model priorities the patients whose due date expires in the planning horizon, then it chooses among patients whose due date is expired and finally among patients whose due dates expire beyond the planning horizon;
- the weights in the objective function W_i are set to give a decreasing importance to the objectives following this order: OR balancing, bed balancing, mix, due dates and number of scheduled surgeries;
- the targets for the different objectives are represented by the ideal values that these objectives can assume; specifically they are calculated by solving the models considering only one objective at a time. The values obtained solving the five sub-problems are then used to solve the goal programming model. Since the target values are ideal values, depending on the typology of objective that is considered, i.e. minimisation or maximisation, respectively negative and positive deviation variables with respect to the objective can be eliminated. For example, for an objective that is maximised, i.e. the number of scheduled surgeries, the ideal value will be always higher or equal to the value assumed by the objective and positive deviation variable can be eliminated.

Evaluating action – feed-back

The MSS produced by the model was judged as satisfactory by the stakeholders: it allowed for a quite high number of scheduled surgeries (considering the available resources), i.e. 462, it presented a fairly balanced daily utilisation profile of the ORs and the beds' units, i.e. with daily utilisation values ranging from 79% to 90%, and it was satisfactory with respect to the due dates fulfilment and the mix, i.e. it allowed for the fulfilment of the 93% of the patients with a due date expiring in the planning horizon (100% of the patients with higher priority) and it reflected the target number of surgeries fixed by the stakeholders for each surgery group. This version of the model was then chosen to be integrated in the scheduling tool.

6.2 Action research implementation cycles

In this paragraph the implementation process of the model is described. From the first implementation, several research implementation cycles have been conducted, each of which represents the production and the implementation of a MSS relevant to a given period. The results of the steps of these cycles are summarised in the following paragraphs, organised as the phases of a single cycle. In each phase the most valuable findings are highlighted.

Diagnosing

Once an implementable version of the model was available, it would have been integrated in a scheduling tool, to make it usable by the personnel of the Meyer hospital. As pointed out in Chapter 5, the scheduling tool should enable the end-user to:

1. produce a new MSS, through the integration with the MIP model
2. manage the produced MSS, giving the possibility to change it after it was created.

In the first step the tool should enable the end-user to entry all the parameters needed by the model, e.g. the allocation grid, the daily availability of ORs and beds in the units, the weights of the objective function. This data, together with the information on the current waiting list, would have been used by the model to produce the MSS. Another information required for the MSS production were the requests that surgeons usually made to the planning department personnel about the patients to schedule. During the implementation cycles the model was modified to keep into consideration of these requests.

With respect to the time when the MSS is produced, since the planning department personnel start selecting and calling patients one week before the schedule is executed, in each cycle the model has been run one week in advance with respect to the period it refers to.

Planning action

Before the implementation, the scheduling tool needed to be created. Based on the model and the stakeholders' requirements the scheduling tool was developed as an excel-based tool. The tool integrates via VBA the production of the MSS (the MIP model and the solver) and its management during the its execution as shown in Figure 9.

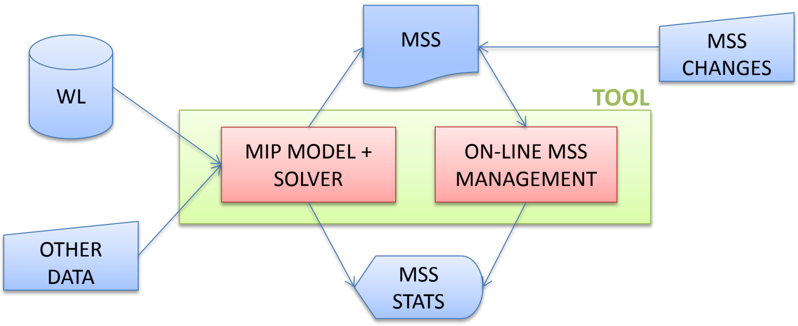


Figure 9 The surgical scheduling tool

In the production phase the tool is fed by the waiting lists and from other data that are manually entered by the end-user, e.g. the daily availability of ORs and units. The produced MSS is visualised on a excel sheet together with the relevant statistics, mainly the resource utilisations in the different days of the planning horizon. Each time the schedule on the excel sheet is changed, the end-user can visualise the impact of these changes on the statistics sheet.

In each cycle, actions were planned according to the results of the previous cycles and the feedbacks given by the stakeholders. Several changes to the model have been made during the implementation project. Details about these modifications are given in the next paragraphs.

Taking action

The first MSS was created on October 25th, 2013 and referred to the period November 4th – December 1st, with a planning horizon of four weeks. After the first implementation more than 20 MSSs has been produced, entailing more than 20 action research implementation cycles. In this thesis the last implemented MSS considered is the one relevant to July 2014.

During the whole implementation process the hospital members were trained to use the tool for both the MSS production and management. The training activities concerned not only the technical aspects of the scheduling tool usage, but also on how to identify the criticalities of a specific MSS through the tool and what are the possible ways to solve them. The personnel of the hospital has been thus trained to help themselves, that is fundamental for the implementation purposes.

In the first three cycles the MSS were produced and managed through the scheduling tool by the IBIS Lab members of the action research team. The hospital members of the team were called just to give the indications about the schedule requisites and to ask for making the changes to the schedule in order to verify their feasibility. Meanwhile they were able to become familiar with the tool, understanding better how to use it and what kind of analysis it allows to perform. After three cycles, the hospital members were able to use the tool for the MSS managing purposes. At that moment the MSS production phase was still demanded to the IBIS Lab researchers, but the tool was transferred to the hospital and used by the bed manager. At present the production of the MSS is made by the IBIS Lab and hospital members together and entirely managed by the beds manager.

Evaluating action

The results of the actions taken in the implementation cycles were assessed each time by the stakeholders. The feedbacks were used in the following cycles to modify the model and make it able to produce better solutions. Specifically:

1. the planning horizon was reduced from 4 weeks to 2 weeks. The MSS needs to be produced with updated data about the cases that are in the waiting lists. The shorter the planning horizon, the higher the frequency with which the MSS is performed and the more updated are the input waiting lists. However the planning horizon cannot be too short because that would make difficult to manage the patients' due dates (resources are less coordinated). Hence, when deciding about the length of the planning horizon, the need of updated data about the waiting list must be traded off with the need to respect the patients' due dates;
2. some coverage constraints about the number of cases to schedule belonging to the different surgery groups were added to the model and it was given the possibility to modify relevant parameters in the scheduling tool. Each time the MSS was created, in fact, surgeons made some requests about the patients or the typologies of patients to schedule. Some requests examples are:

- specific/minimum number of patients belonging to a specific specialty in a specific OR session;
- specific/minimum number of patients belonging to certain surgery type and specialty in a specific OR session;
- specific/minimum number of patients belonging to a certain surgery group in a specific OR session;
- specific/minimum number of patients belonging to a specific specialty in a specific week (for those specialties with more than one OR session per week);
- specific/minimum number of patients belonging to certain surgery type and specialty in a specific week (for those specialties with more than one OR session per week);
- specific/minimum number of patients belonging to a certain surgery group in a specific week (for those specialties with more than one OR session per week).

As hard constraints, these requests can cause model infeasibilities and jeopardise the effectiveness of the model. Moreover, since resources are shared by the different specialties, satisfying the requests of some surgeons may disadvantages the specialties whose surgeons do not make requests, making the process unfair. These requests were often justified by the clinical conditions of the patients, but sometimes they were not. For this reason it was decided to make the OT manager responsible for judging these issues and thus to establish which requests had to be considered when creating the MSS. The aim was to limit the number of these requests as much as possible;

3. a criterion concerning the scheduling of the patients with longer waiting times was added to the objective function. In fact, some surgical specialties, i.e. those specialties in which the most of patients are not assigned with a due date, had a significant number of patients with very long waiting times. Hence there was the need to prioritise these patients;
4. In order to improve the efficiency of the MSS, some changes to the allocation grid were made. The ordinary unit utilisation profile resulting from the starting allocation grid in fact was characterised by lower utilisation on the beginning of the week and higher in the end. For this reason it was decided to move a OR session of a specialty whose cases are hospitalised in that unit from Thursday to Monday. As a result, the number of scheduled surgeries improved and the unit utilisation were smoothed. However, it is worth to point out that only few changes to the allocation grid have

been made, because, as consequence of such changes, surgeons usually needed to reorganise their activities during the week;

5. the values of the weights assigned with the different criteria in the objective function were changed during the implementation process, mainly for two reasons:
    a. the priorities of the stakeholders changed, causing a different ranking of the objectives;
    b. we gained a knowledge about the relationships existing among the objectives. In general establishing what are the relationships among the weights in an objective function is not an easy issue. Given a certain weights configuration, the same variation of two different weights may have a totally different impact on the solution of the model, depending on the sensitivity of the solution to the weights. It not only depends on the structure of the model but also on the considered data set. For some objectives, the same value can be reached with different surgery combinations and thus the variation of the relevant weight does not impact on the value of the objective function. However the solutions can be very different for the other objectives perspective. For example, the same OR and units utilisation balancing can be reached with very different values of due dates' fulfilment and mix. The variation of the objectives corresponding to these latter criteria have a higher impact on the solution with respect to the same variation on the former two, i.e. the objective function is more sensitive to the weights of the due dates and the mix with respect to the one associated with the OR and units utilisation balancing.

Chapter 7 Evaluating Action

In this chapter the results of the action research study are presented. Specifically the chapter is organised in two parts: the contribution to practice and the contribution to knowledge. The first part illustrates the qualitative and quantitative results of the scheduling tool implementation process at the Meyer hospital. The second part highlights what is the concrete contribution to knowledge of the study, i.e. what insights and understanding have been developed about a MSS tool implementation.

7.1 Contribution to practice

7.1.1 *Qualitative results*

This study has effectively brought to the implementation of the developed scheduling tool, and thus the main objective of the study has been achieved. The enhanced MSS approach is now integral part of the Meyer hospital processes, influencing the activities of all the people involved in the MSS process. The new concept of surgery group is now familiar to everyone. This concept has made people more sensitive to the impact and the consequences on the utilisation of shared resources of the scheduling of the surgeries. Everyone agree with the fact that the scheduling process now is more structured than before and that every decision is taken with more awareness. For example, the OR manager said that:

"Surgeons now complete the assigned surgical sessions in time"

meaning that OR overtime, that cause extra-costs, is less likely to occur. The beds manager said that

"Bed units occupation is now more under control"

and this means that overbooking, which is one of the major cause of patients cancellations and postponements, is less likely to occur, increasing the patients satisfaction. The members of the planning department said that

"It is easier to select the patients to schedule"

In fact, they no more have to decide the number and the kind of patients to call to populate the MSS, because this already contains the information about the number of patients to schedule and the surgery group they must belong to. The surgeons were not very enthusiastic about the project at the beginning because they feared that the enhanced MSS approach would have reduced their autonomy in the choice of the patients to operate. For this reason the changes have been introduced gradually, allowing the surgeons to propose what patients to schedule and putting the OT manager judgement as a filter. This decision revealed to be successful: surgeons are now convinced about the benefits of the project. They recognised that the assignment of the beds to specialties is now more fair, as it is guided by a mathematical model. The general and medical directors, whose commitment has been fundamental to win the resistance to change in the organisation, are satisfied of the obtained results (details are given in the next paragraph) and decided to continue the collaboration with the researchers of the IBIS Lab.

7.1.2 *Quantitative results*

In order to demonstrate the effectiveness of the implemented scheduling tool, hereafter it is made a comparison of the relevant performance before and after the implementation. Specifically the comparison of the performance is made between the first months on the 2013, i.e. January 2013-October 2013, and the period from the first implementation and July 2014, i.e. November 2013-July 2014. The results are reported in Table 7. As shown in the first part of such a table, despite between the two periods the ORs available time were reduced of around the 13%, passing from 201.9 to 176.4 mean weekly hours, the overall workload (in terms of OR room utilised hours) has not significantly decreased, passing from 140 to 137.7 mean weekly hours. As a result we observed an increase of around the 9% of the utilisation rate of the ORs: in fact, the overall workload decrease has been less than proportional and thus lower than the ORs available time reduction. Such a little decrease in the OR workload reflected on the mean number of surgery performed each week, whose difference before and after the implementation is around 0.5 surgeries, i.e. 96.6 vs. 96.1.

Table 7 Tool implementation results

	BEFORE	AFTER
OR workload		
- OR mean utilisation [%]	69.3	78.1
- OR mean weekly utilisation [hours]	140	137.7
- OR mean weekly availability [hours]	201.9	176.4
Executed surgeries		
- Mean weekly number [patient]	96.6	96.1
- Percentage of day surgeries [%]	48.1	55.5
- Patients per OR hour [patient/hours]	0.48	0.54
Number of patients requiring an ordinary bed [patient]		
- Daily mean	20.9	20.3
- StDev	6.1	4.3
- Median	22	21
- 95th percentile	29.3	27
- Maximum value	35	30
Number of patients requiring a DS bed [patient]		
- Daily mean	9.3	10.8
- StDev	3.8	3.1
- Median	9	12
- 95th percentile	15.3	14
- Maximum value	24	16

The percentage of day surgeries has grown from the 48.1% to the 55.5%. This happened because between the two periods, some typologies of surgery that before the implementation were typically characterised by a LoS of 2 days are now performed as day surgeries. This fact and the little decrease in the OR workload have led to an increase in the mean daily number of patients requiring a day surgery bed, i.e. from 9.3 to 10.8 patients/day, and to a decrease in the mean daily number of patients requiring an ordinary bed, i.e. from 20.9 to 20.3. However, observing the other measures, it emerges how daily units utilisations are now more balanced: for both units in fact the standard deviation of the number of patients requiring a bed is now smaller. Moreover the value corresponding to the 95th percentile is smaller or equal to the number of available beds, i.e. respectively, for the day surgery unit 14 and 14, for the ordinary units 27 and 28., which more likely lead to less patients cancellations.

7.2 Contribution to knowledge

This paragraph reports the outcomes of the self-reflection process undertaken during project. Specifically it describes what lessons about the implementation of a MSS tool have been learned during the unfolding of the action research cycles. Lessons have been categorised in two typologies: the first is what are the characteristics of a MSS model that make it easier to implement. The second is what conditions facilitate the implementation process of a MSS tool. Hence, the first category refers to specific aspects of the mathematical model, e.g. how to model variables, constraints and objectives. The second type of lessons is more general, and refers to the insights and to the understanding about how to manage the implementation process.

7.2.1 *MSS model characteristics*

1. Importance of considering the surgery groups when creating the MSS: considering the surgery group at this stage entails considering the downstream resources. This would lead to better performance of the process in terms of throughput, due dates fulfilment and robustness. In addition surgery groups makes easier to replace scheduled patients, which is a particularly important issue for children's hospital.
2. If assignment of specialties to sessions can't significantly vary during the year, in order to optimise the units utilisation, it is important to choose the best allocation for those specialties whose surgery groups are almost characterised by the same post-surgical LoS (e.g. avoid to assign two specialties whose surgery groups are characterised by two days of post-surgical LoS in consecutive days).
3. Cases in the waiting list should be characterised by two dates: the earliest programmable date and the latest due date, i.e. the due date, which together define the interval in which the patient should undergo the surgery.
4. The tool should allow for fixing the date of the surgery for certain patients, i.e. those whose scheduling is complex from a logistic point of view, e.g. patients that live very far from the hospital and need to know the date of the surgery very far in advance to organise the trip.
5. Hard constraints on the quality of the schedule, e.g. due dates, mix constraints, should be avoided. Depending on the instance, in fact they can lead to model infeasibilities that are not easily manageable by the end-users. If hard constraints are re-

quired, e.g. the ones at point 4, train adequately the end-user to help him/herself in detecting what constraints may cause infeasibilities.

6. Constraints should be tuneable as much as possible. For this reason it is important to understand what are the operative conditions that may change most frequently and give the possibility to the scheduler to utilise the these constraints in the tool, e.g. the most frequent requests made by the surgeon about the quality of the schedule.
7. The tool should be easily tuneable with respect to the objectives to optimise, because the stakeholders priorities and needs may change in the course of time, e.g. multi-criteria approaches.

7.2.2 *MSS implementation process*

1. Consistently with Carvalho et al. (2014), regardless the stakeholders' importance, sometimes characteristics and expectations are explicit and easy to understand. Sometimes instead they can be tacit and difficult to deduce. The best way to obtain a model able to produce satisfactory schedules is to create a solution and asking the stakeholders to comment on it. This feedback can give information that are often more valuable that the one obtained with generic interviews.
2. In order to overcome the resistance to change of the members of the hospital, the strong commitment of the top management is essential.
3. Before introducing a change in a practice, assess the resistance to change of the involved members. Sometimes resistance to change can make the project fail. For this reason it is advisable to gradually introduce the changes.
4. In order to keep high the hospital members' participation, during the implementation process clearly show the practical benefits given by the introduction of the tool.

Chapter 8 Conclusions and future research

8.1 Conclusions

This thesis concerns a study aiming at developing and implementing a scheduling tool for the master surgical scheduling process. The project has been addressed combining the action research and the model based research methodologies. The former has guided the whole project, while the latter has been specifically used to develop and test the mathematical models which the tool is based on. As an action research, the project has contributed to practice, solving a practical problem, and to knowledge, answering to several research questions.

In particular, the project allowed the development and the successful implementation of a master surgical scheduling tool at the Meyer hospital, which is the organisation that has inspired the study. The implementation has led to an improvement of the surgical process performance as discussed in Chapter 7.

From the contribution to knowledge perspective, the study has allowed to answer to several research questions.

RQ1: What factors and conditions can facilitate the implementation of a master surgical scheduling tool?

During the unfolding of the action research cycles, the self-reflection process typical of the action research approach has led to rationalise what experienced and to develop understanding about the scheduling tool implementation process. The emerged insights have been categorised in two typologies: the characteristics which make the MSS model easier to implement and more general indications about how to manage the implementation process. Detailed indications are given in Chapter 7.

Four novel mixed integer programming models for the MSS production are proposed in this thesis. Each model was created in a specific cycle and represents an improved version of the model developed at the preceding cycle, which implements the feedbacks of the stakeholders on the solutions obtained in the relevant testing phases. These cycles are addressed through a model based approach. The studies conducted in these cycle has allowed to answer to different research questions that emerged during the project.

RQ2: How is it possible to obtain efficient and robust master surgical schedules?

The study relevant to this research question has been published in (Banditori et al., 2013). The relevant article, which is article n. 1, is reported in the Appendix. In the study a novel MSS MIP model was developed. The schedules produced by this model were then simulated, via a discrete event simulation model, in order to assess their robustness against the variability of surgical times and LoS. In order to trade-off efficiency and robustness, a resource-slack strategy, i.e. instantiating the optimization model considering an amount of resources lower than the one actually available, is tested. The performed experimental campaign shown how a MSS that is robust against variability of the surgical times and the length of stay can be achieved through the adoption of the resource slacks. A trade-off between efficiency and robustness does exist: the higher the efficiency, i.e. the number of scheduled surgeries, the lower the robustness and vice versa. Referring to the considered setting, it has been demonstrated how adopting slacks of 10% and 12% respectively for ORs and beds, it is possible to schedule a higher number of surgeries with respect to the actual planning process

RQ3: Is it possible to obtain efficient and robust master surgical schedules through the resources utilisation balancing?

The study relevant to this research question has been published in (Cappanera et al., 2014). The relevant article, which is article n. 2, is reported in the Appendix. In this study the relationships between efficiency, robustness and balancing are investigated through a combined optimization-simulation approach. Three well-known different balancing criteria were tested in correspondence with different hospital settings. The criteria are the following: the minimisation of the maximum value, i.e. minMax, the minimisation of the difference between the maximum and the minimum values, i.e. minRng, the minimisation of the squared positive deviation a the values from a fixed threshold, i.e. minOvrn. The experimental campaign reveals that none of the investigated policies allows superior performance in terms of efficiency, balancing and robustness to be achieved concurrently. However, depending on the hospital management's priorities and needs, it is always possible to identify a policy that allows for a reasonable trade-off among these performance criteria. Specifically minOvrn seemed to be the best balancing policy because it represents somehow an intermediate case with respect to the other two policies. In fact minRng is at least effective as the other policies in balancing the beds. In addition, it allows for a better OR balancing and thus for a smaller overtime. Moreover, it leads to higher bed saturation and to a larger number of scheduled surgeries. However, it also causes a higher overbooking. The properties of minMax are quite the opposite than minRng.

RQ4: What is the impact on the master surgical schedule of a flexible management of the critical resources?

The study relevant to this research question has been published in (Visintin et al., 2014). The relevant article, which is article n. 3, is reported in the Appendix. In the study a new MSS was developed. The model allows to consider two different level of flexibility in the management of surgical teams, ORs and post-surgical beds. A design of experiment was then conducted to assess to what extent the flexible management of one or more of these resources can improve the surgical process throughput. The analysis revealed that the best results can be achieved by managing flexibly both surgical teams and ORs. Moreover, the analysis showed that, if a hospital cannot manage flexibly the surgical teams, then it can still improve its efficiency by managing flexibly the ORs and vice versa.

However, the analysis revealed that if both surgical teams and ORs are managed flexibly, pooling surgical units has no significant impact, while if only one of these two resources (or none) are managed flexibly, then pooling surgical units produces significant benefits. However, even if the flexible management could improve the efficiency of the process, it was decided to not implement any of these practices. In fact the benefits arising from the flexible management of the resources were considered by the stakeholders too low to justify the organisational cost needed to implement it.

8.2 Limitations and future research

From the practice perspective, even if the tool have been implemented, the project is not over yet. The obtained results are quite satisfactory, but the OT performance can be further improved. In fact, changes were introduced gradually, removing only partially the actual practices at the Meyer hospital, because we realised that radical changes could make the project fail. Removing these practices, and thus further improving the operating theatre performance, will be certainly one of the aim of the future actions of the project.

From a research perspective, in this thesis, the findings emerging from the development and the implementation of a MSS tool have been presented. Specifically these findings are mainly relevant to (i) the relationships between efficiency, resources utilisation balancing and robustness and to the flexible management of the critical resources in the MSS context and to (ii) the lessons learned about the implementation of a MSS model/tool. All these findings have been empirically obtained stating from a single setting, i.e. the Meyer hospital. This fact reduce the generalizability of the study. In order to increase the external validity of the former findings the computational campaigns have been extended generating several instances and considering different hospital settings. The lessons learned instead, that are the findings emerging from the reflection process of the action research project, are based on that single setting. However, in the light of the literature and of the experience gained on the health care field, it is reasonable to assume that these lessons can be applied to other hospital settings, since the characteristics of the Meyer hospital MSS process are shared by the most of the hospital. Despite these facts, hospitals whose setting significantly differ from the Meyer's one, e.g. hospitals in which other resources are critical, hospitals in which emergencies are prevalent and are not managed through dedicated resources, may not take advantage of these findings. For these reasons, studying the development and the implementation of the scheduling models in very different settings would be useful to validate the empirical findings of this thesis.

References

ADAN, I., BEKKERS, J., DELLAERT, N., VISSERS, J. & YU, X. 2009. Patient mix optimisation and stochastic resource requirements: A case study in cardiothoracic surgery planning. *Health Care Management Science,* 12, 129-141.

AGNETIS, A., COPPI, A., CORSINI, M., DELLINO, G., MELONI, C. & PRANZO, M. 2012. Long term evaluation of operating theater planning policies. *Operations Research for Health Care,* 1, 95-104.

BANDITORI, C., CAPPANERA, P. & VISINTIN, F. 2013. A combined optimization–simulation approach to the master surgical scheduling problem. *IMA Journal of Management Mathematics,* 24, 155-187.

BANDITORI, C., CAPPANERA, P. & VISINTIN, F. Investigating the Relationship Between Resources Balancing and Robustness in Master Surgical Scheduling. Proceedings of the International Conference on Health Care Systems Engineering, 2014. Springer, 149-162.

BELIËN, J. & DEMEULEMEESTER, E. 2007. Building cyclic master surgery schedules with leveled resulting bed occupancy. *European Journal of Operational Research,* 176, 1185-1204.

BELIËN, J., DEMEULEMEESTER, E. & CARDOEN, B. 2006. Visualizing the demand for various resources as a function of the master surgery schedule: a case study. *Journal of Medical Systems,* 30, 343-350.

BELIËN, J., DEMEULEMEESTER, E. & CARDOEN, B. 2009. A decision support system for cyclic master surgery scheduling with multiple objectives. *Journal of Scheduling,* 12, 147-161.

BENNETT, D. & LEE, S. 2000. Total productive maintenance implementation in the newspaper printing industry: an action research approach.

BERTRAND, J. W. M. & FRANSOO, J. C. 2002. Operations management research methodologies using quantitative modeling. *International Journal of Operations & Production Management,* 22, 241-264.

BLAKE, J. T., DEXTER, F. & DONALD, J. 2002. Operating room managers' use of integer programming for assigning block time to surgical groups: A case study. *Anesthesia & Analgesia,* 94, 143-148.

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

BRAILSFORD, S., HARPER, P., PATEL, B. & PITT, M. 2009. An analysis of the academic literature on simulation and modelling in health care. *Journal of Simulation,* 3, 130-140.

BRAILSFORD, S. & VISSERS, J. 2011. OR in healthcare: A European perspective. *European Journal of Operational Research,* 212, 223-234.

CAPPANERA, P., VISINTIN, F. & BANDITORI, C. 2014. Comparing resource balancing criteria in master surgical scheduling: A combined optimisation-simulation approach. *International Journal of Production Economics,* 158, 179-196.

CARDOEN, B., DEMEULEMEESTER, E. & BELIËN, J. 2010. Operating room planning and scheduling: a literature review. *European Journal of Operational Research,* 201, 921-932.

CARTER, M. W. & KETABI, S. 2012. Bed Balancing in Surgical Wards via Block Scheduling. *Journal of Minimally Invasive Surgical Sciences,* 1, 129-137.

CARVALHO, A. N., SCAVARDA, L. F. & LUSTOSA, L. J. 2014. Implementing finite capacity production scheduling: lessons from a practical case. *International Journal of Production Research,* 52, 1215-1230.

CHOW, V. S., PUTERMAN, M. L., SALEHIRAD, N., HUANG, W. & ATKINS, D. 2011. Reducing surgical ward congestion through improved surgical scheduling and uncapacitated simulation. *Production and Operations Management,* 20, 418-430.

COGHLAN, D. & BRANNICK, T. 2005. *Doing action research in your own organization,* London, Sage.

COUGHLAN, P. & COGHLAN, D. 2002. Action research for operations management. *International Journal of Operations & Production Management,* 22, 220-240.

DENTON, B., VIAPIANO, J. & VOGL, A. 2007. Optimization of surgery sequencing and scheduling decisions under uncertainty. *Health Care Management Science,* 10, 13-24.

ELDABI, T. 2009. Implementation issues of modeling healthcare problems: misconceptions and lessons. Winter Simulation Conference, 1831-1839.

FATTORE, G. & TORBICA, A. 2006. Inpatient reimbursement system in Italy: How do tariffs relate to costs? *Health Care Management Science,* 9, 251-258.

GLOUBERMAN, S. & MINTZBERG, H. 2001. Managing the care of health and the cure of disease-Part I: Differentiation. *Health care management review,* 26, 56-69.

GUERRIERO, F. & GUIDO, R. 2011. Operational research in the management of the operating theatre: a survey. *Health Care Management Science,* 14, 89-114.

GUINET, A. & CHAABANE, S. 2003. Operating theatre planning. *International Journal of Production Economics,* 85, 69-81.

GUMMESSON, E. 2000. *Qualitative methods in management research*, Sage.

HALES, D. N. & CHAKRAVORTY, S. S. 2006. Implementation of Deming's style of quality management: An action research study in a plastics company. *International Journal of Production Economics,* 103, 131-148.
HANS, E., WULLINK, G., VAN HOUDENHOVEN, M. & KAZEMIER, G. 2008. Robust surgery loading. *European Journal of Operational Research,* 185, 1038-1050.
HARPER, P. & PITT, M. 2004. On the challenges of healthcare modelling and a proposed project life cycle for successful implementation. *Journal of the Operational Research Society,* 55, 657-661.
KARLSSON, C. & ÅHLSTRÖM, P. 1996. The difficult path to lean product development. *Journal of Product Innovation Management,* 13, 283-295.
LAGANGA, L. R. 2011. Lean service operations: reflections and new directions for capacity expansion in outpatient clinics. *Journal of Operations Management,* 29, 422-433.
LITVAK, E. & LONG, M. C. 2000. Cost and quality under managed care: Irreconcilable differences. *Am J Manag Care,* 6, 305-12.
MAHDAVI, M., MALMSTRÖM, T., VAN DE KLUNDERT, J., ELKHUIZEN, S. & VISSERS, J. 2013. Generic operational models in health service operations management: A systematic review. *Socio-Economic Planning Sciences,* 47, 271-280.
MAY, J. H., SPANGLER, W. E., STRUM, D. P. & VARGAS, L. G. 2011. The surgical scheduling problem: current research and future opportunities. *Production and Operations Management,* 20, 392-405.
MAY, J. H., STRUM, D. P. & VARGAS, L. G. 2000. Fitting the lognormal distribution to surgical procedure times. *Decision Sciences,* 31, 129-148.
MEREDITH, J. R., RATURI, A., AMOAKO-GYAMPAH, K. & KAPLAN, B. 1989. Alternative research paradigms in operations. *Journal of Operations Management,* 8, 297-326.
MITROFF, I. I., BETZ, F., PONDY, L. R. & SAGASTI, F. 1974. On managing science in the systems age: two schemas for the study of science as a whole systems phenomenon. *Interfaces,* 4, 46-58.
MONTGOMERY, D. C. & RUNGER, G. C. 2003. *Applied statistics and probability for engineers*, John Wiley and Sons.
PROUDLOVE, N., BLACK, S. & FLETCHER, A. 2007. OR and the challenge to improve the NHS: modelling for insight and improvement in in-patient flows. *Journal of the Operational Research Society,* 58, 145-158.
RITTEL, H. W. & WEBBER, M. M. 1973. Dilemmas in a general theory of planning. *Policy sciences,* 4, 155-169.
SAID, K., PASCAL, A. & SONDES, C. 2006. Block scheduling: toward a master surgical schedule. International Conference on Service Systems and Service Management, 429-435.
SANTIBANEZ, P., BEGEN, M. & ATKINS, D. 2007. Surgical block scheduling in a system of hospitals: an application to resource and wait list management in a British Columbia health authority. *Health Care Manag Sci,* 10, 269-82.
SANTIBÁÑEZ, P., BEGEN, M. & ATKINS, D. 2007. Surgical block scheduling in a system of hospitals: an application to resource and wait list management in a British Columbia health authority. *Health Care Management Science,* 10, 269-282.

SEN, T., DILEEPAN, P. & LIND, M. R. 1996. Minimizing a weighted quadratic function of job lateness in the single machine system. *International Journal of Production Economics,* 42, 237-243.

SHANI, A. B. & PASMORE, W. A. 1985. Organization inquiry: Towards a new model of the action research process. *Contemporary organization development: Current thinking and applications,* 1, 438-48.

SODHI, M. S. & TANG, C. S. 2014. Guiding the next generation of doctoral students in operations management. *International Journal of Production Economics,* 150, 28-36.

TÀNFANI, E. & TESTI, A. 2010. A pre-assignment heuristic algorithm for the Master Surgical Schedule Problem (MSSP). *Annals of Operations Research,* 178, 105-119.

TESTI, A., TANFANI, E. & TORRE, G. 2007. A three-phase approach for operating theatre schedules. *Health Care Management Scencei,* 10, 163-72.

VAN OOSTRUM, J. M., BREDENHOFF, E. & HANS, E. W. 2010. Suitability and managerial implications of a master surgical scheduling approach. *Annals of Operations Research,* 178, 91-104.

VAN OOSTRUM, J. M., VAN HOUDENHOVEN, M., HURINK, J., HANS, E., WULLINK, G. & KAZEMIER, G. 2008. A master surgical scheduling approach for cyclic scheduling in operating room departments. *OR Spectrum,* 30, 355-374.

VANBERKEL, P. T. & BLAKE, J. T. 2007. A comprehensive simulation for wait time reduction and capacity planning applied in general surgery. *Health Care Management Science,* 10, 373-385.

VINCENT, C., TAYLOR-ADAMS, S. & STANHOPE, N. 1998. Framework for analysing risk and safety in clinical medicine. *BMJ: British Medical Journal,* 316, 1154.

VIRTUE, A., CHAUSSALET, T. & KELLY, J. 2013. Healthcare planning and its potential role increasing operational efficiency in the health sector: A viewpoint. *Journal of Enterprise Information Management,* 26, 8-20.

VISINTIN, F., CAPPANERA, P. & BANDITORI, C. 2014. Evaluating the impact of flexible practices on the master surgical scheduling process: an empirical analysis. *Flexible Services and Manufacturing Journal*, 1-24, DOI: 10.1007/s10696-014-9208-9.

VISSERS, J., ADAN, I. & BEKKERS, J. 2005. Patient mix optimization in tactical cardiothoracic surgery planning: a case study. *IMA Journal of Management Mathematics,* 16, 281-304.

WESTBROOK, R. 1993. Orderbook models for priority management: a taxonomy of data structures. *Journal of Operations Management,* 11, 123-142.

WESTBROOK, R. 1995. Action research: a new paradigm for research in production and operations management. *International Journal of Operations & Production Management,* 15, 6-20.

WILSON, J. T. 1981. Implementation of computer simulation projects in health care. *Journal of the Operational Research Society*, 825-832.

ZHANG, B., MURALI, P., DESSOUKY, M. M. & BELSON, D. 2008. A mixed integer programming approach for allocating operating room capacity. *Journal of the Operational Research Society,* 60, 663-673.

Appendix – Article I

This is a pre-copy-editing, author-produced PDF of an article accepted for publication in IMA JOURNAL OF MANAGEMENT MATHEMATICS following peer review. The definitive publisher-authenticated version

BANDITORI, C., CAPPANERA, P. & VISINTIN, F. 2013. A combined optimization–simulation approach to the master surgical scheduling problem. *IMA Journal of Management Mathematics,* 24, 155-187

is available online at

http://imaman.oxfordjournals.org/content/early/2013/01/17/imaman.dps033.full.pdf+html

1 Introduction

The operating theatre is considered to be the 'engine that drives the hospital' (Beliën *et al.*, 2006). In fact, its activities are tightly interconnected with those of other departments and, consequently, its performance dramatically influences hospital performance as a whole (Cardoen *et al.*, 2010). In addition to being one of the most costly functional areas of the hospital (Denton *et al.* 2007; May *et al.* 2011), the operating theatre is the principal reason for almost 70% of all hospital admissions (Denton *et al.*, 2007). Hospital managers are, therefore, urged to maximise the patient throughput and associated revenues, and to optimise the use of medical resources to reduce costs. In this regard, surgical scheduling is of paramount importance. However, solving a surgical scheduling problem is remarkably complex. It requires, in general, consideration of: (i) many different types of cases, characterised by different priority levels and requiring different procedures; (ii) many different types of resources, such as operating rooms (ORs), OR personnel (e.g., surgeons, anaesthetists and nurses), surgical and electro-medical equipment, postsurgical resources (e.g., ICU, wards); (iii) the randomness associated with patients' arrival, surgeries' duration and patients' length of stay (LoS) (May *et al.*, 2000); and (iv) the conflicting priorities and preferences of the scheduling process stakeholders (Glouberman & Mintzberg, 2001). For these reasons, the use of quantitative techniques, such as mathematical modelling and simulation, seems necessary (Utley *et al.*, 2009). In the literature, the surgical scheduling process is typically seen as entailing three stages: (i) determination of the OR time to assign to each surgical specialty (the case mix planning problem); (ii) determination of the specialty (or specialties) to assign to each OR on each day of the planning horizon (the so-called Master Surgical Scheduling problem, or MSS); and (iii) selection and sequencing of patients who have to undergo surgery, according to the MSS. This process is usually studied in cascade, i.e., by considering the output of the upstream stage as the input for the downstream stage (Beliën & Demeulemeester, 2007). In this regard, it is worth pointing out that the literature offers slightly different definitions of MSS (e.g., Blake *et al.* 2002; Van Oostrum *et al.* 2008). However, there is unanimous consensus that MSS construction does not entail the selection and sequencing of the actual patients to undergo surgery.

In this paper, we propose a Mixed Integer Programming (MIP) model to address the MSS problem on a one-month planning horizon, where the time is split into time slots. We assume a block scheduling approach, i.e., MSS is produced by assigning on a monthly basis a number of OR blocks to each specialty (Van Oostrum *et al.*, 2010).

Specifically, our model assumes that each case on the hospital waiting list: (i) is characterised by a due date and (ii) can be assigned to a *surgery group*. Each surgery group includes procedures that belong to the same specialty and are expected to require a similar amount of resources, i.e., they are characterised by similar expected durations and LoS. We also hypothesise that different case types should be accommodated in different bed types, according to their LoS. The proposed model produces a solution that indicates, for each day of the month and for each time slot (block) of the day, the number of cases to be treated in each surgery group. Consequently, the model focuses on the second stage of the surgical scheduling process while supporting also the third one. It does not, in fact, select and sequence the patients who must undergo surgery during a given month, but it does indicate *pools* of cases (those needing surgeries that fall within the specified surgery groups) that should be given higher priority during the selection and sequencing process. In particular, the model assigns surgery groups to time slots in a way that (i) patient throughput is maximised, (ii) surgery groups with a higher number of cases with closer due dates are given higher priority, and (iii) the number of bed mismatches (e.g., long-stay patients temporarily accommodated in short-stay beds) is minimised.

In this study, through discrete event simulation, we also test the MSS's robustness against the variability of both surgery duration and patient LoS. Moreover, we show that, by combining optimisation and simulation, it is possible to trade off efficiency and robustness (Bertsimas & Sim, 2004), i.e., it is possible to find solutions that allow for the execution of a satisfactory number of surgeries without incurring undue overtime and/or excessive overbooking cancellations.

Our study is inspired by the Meyer University Children's Hospital (hereafter Meyer Hospital), one of the most renowned children's hospitals in Italy, and both the optimisation and the simulation models presented here are tested on empirical data from this hospital.

Children's hospitals represent very interesting settings for the study of surgical scheduling problems (Crowe *et al.*, 2011), specifically to address issues related to schedule robustness. These hospitals present two main peculiarities. First, they need to manage a higher number of patient-driven cancellations, as children are more likely to fall ill than adults (Bathla *et al.*, 2010). Consequently, while on the one hand, it is impractical to fix the day of surgery too far in advance, on the other hand, even if patients are scheduled with short notice (e.g., one week), some of them will probably fall ill between the day the appointment is made and the day the surgery is scheduled. Second, last-minute hospital-driven cancellations, i.e., cancellation on the day of surgery due to a bed shortage, must be prevented or minimised. A cancellation can actually lead to psychological trauma to both the patient and her/his family (Tait *et al.*, 1997), especially if the child has been obligated to fast for a prolonged time before the cancellation (Bathla *et al.*, 2010). Furthermore, cancellations represent a significant inconvenience to the child's parents, who may miss additional workdays (Bathla *et al.*, 2010; Tait *et al.*, 1997). For a children's hospital, thus, being able to rely on a robust schedule is of paramount importance.

It is worth noting that schedule disruptions and cancellations are also a major concern even outside the paediatric setting (Beliën & Demeulemeester, 2007; Hans *et al.*, 2008). Indeed, our study is based on features and requirements that are shared by many types of hospitals (Cardoen *et al.*, 2010; Guerriero & Guido, 2011; May *et al.*, 2011) including: the adoption of a block scheduling process; the subdivision of the procedures into surgery groups; the presence of ORs, beds and surgical teams as critical resources; incompatibility between specialties and ORs, as well as between case types and bed types; the need to meet certain due dates; and the wish to maximise the patient throughput. As a result, the MIP model, as well as the combined approach presented in this paper, can be applied in many different real-life settings and is of interest to a wide audience of scholars and practitioners.

2 Literature review

The problem of surgical planning and scheduling has been the subject of a sizeable number of contributions, especially over the last decade (Cardoen *et al.*, 2010). Given that excellent updated reviews on this topic have recently been published (Cardoen *et al.* 2010; Guerriero & Guido 2011; May *et al.* 2011), a broad review of the literature is outside the scope of this work. Instead, we narrow the scope of our review to papers that have presented mathematical models that support the MSS process. With respect to this subset, the review covers, to the best of our knowledge, all of the most relevant contributions that have been published in peer-reviewed journals.

Building on the taxonomy/dimensions proposed by Cardoen *et al.* (2010), we provide a thorough description of the mathematical models available in the literature and identify the gaps that we aim to fill with our model. The papers reviewed here are analysed according to the following dimensions: (i) *patient characteristics*, i.e., the typology of the patients scheduled (elective vs. non-elective, inpatient vs. outpatient); (ii) *performance measures*, i.e., the optimised utility function (throughput, resource utilisation and so on); (iii) the *decision delineation*, which identifies the entity (specialty, patient, etc.) to which/whom the decision applies and the type of decision to support (e.g., the assignment of a specialty to a day vs. the assignment of a specific patient to a time slot); (iv) *research methodology*, which refers to the type of analysis (e.g., heuristic vs. exact optimisation) and to the solution techniques adopted (e.g., mathematical programming vs. simulation); (v) *type of constraints*, particularly the hard constraints that are considered (e.g., resource availability, demand, release/due date); (vi) *uncertainty*, which indicates if and how data randomness is managed; (vii) *applicability of the research*, which explains how the models have been tested (i.e., with real data, with realistic data, or not tested); and (viii) a *planning horizon* indicating the time horizon on which the models have been applied. Dimensions (i), (v) and (vii) are taken as-is from Cardoen *et al.* (2010), while dimensions (ii), (iii), (iv) and (vi) have been adapted through the addition of more details in order to better position our contribution with respect to the literature. Finally, dimension (viii) has been introduced ex-novo. The review is organised and presented into tabular form (see Table 1 and Table 2), where rows represent the aforementioned dimensions, and each column represents one paper. Hence, each cell provides a brief description of a particular paper from a specific perspective.

TABLE 1. *MSS literature review: part 1/2*

		Our work	Blake *et al.* (2002)	Vissers *et al.* (2005)	Said *et al.* (2006)	Santibàñez *et al.* (2007)
Patient characteristics		Elective Inpatients	Elective Not specified	Elective Inpatients	Elective Not specified	Elective Not specified
Performance criteria		- Throughput maximisation - Appropriate waiting lists consumption - Proper bed allocation	Minimisation, for each specialty, of the OR time undersupply with respect to fixed targets	Minimisation of the deviation between realised and target resource utilisation	Minimisation, for each specialty (or surgeon), of the gap between OR time demand and supply	- Minimisation of the deviation among scheduled and target throughput - Minimisation of bed utilisation
Decision delineation	**Scheduled 'object'**	Specialties + procedure typologies	Specialties	Procedure typologies	Specialties/ surgeons + procedure typologies	Specialties + procedure typologies
	Decision details	Date, time slot, OR	Date, OR	Date	Date, time, OR	Date, hospital, OR
Research methodology	**Type of analysis**	- Multi-criteria hierarchical exact optimisation - Scenario analysis	Single criterion heuristic optimisation	- Single criterion exact optimisation - Scenario analysis	Single criterion exact optimisation	- Single criterion exact optimisation - Scenario analysis
	Solution technique	- Mixed integer programming - Discrete event simulation	- Mixed integer programming - Constructive heuristic	Mixed integer programming	Mixed integer programming	Mixed integer programming
Type of constraints	**Resource**	Wards, surgical staff, equipment, regular OR time	Surgical staff, equipment, regular OR time	Wards, ICUs, nursing staff, regular OR time	Surgical staff, regular OR time	Wards, ICUs, surgical staff, equipment, regular OR time
	Others	- Procedures' due dates - Procedures mix	Max and min n° of OR blocks per week to specialties	- Throughput target - Additional restrictions	Max and min n° of OR blocks per week to specialties	- Throughput target - Schedule cyclicity
Uncertainty		Deterministic (optimisation), stochastic (robustness test) surgery duration and LoS	Deterministic	Deterministic	Deterministic	Deterministic
Applicability		Tested on real and realistic data	Tested on real data	Tested on real data	Randomly generated surgery duration and specialty/surgeon demand	Tested on real data
Planning horizon		1 month	1 week	1 month	1 week	1 month

TABLE 2. *MSS literature review: part 2/2*

		Testi *et al.* (2007)	**Van Oostrum *et al.* (2008)**	**Zhang *et al.* (2008)**	**Beliën *et al.* (2009)**	**Tànfani & Testi (2010)**
Patient characteristics		Elective Inpatients	Elective Not specified	-Elective In&outpatients -Non-elective Emergency cases	Elective Inpatients	Elective Inpatients
Performance criteria		- Minimisation of the gap between specialty demand and supply - Fulfilment of the surgeons' preferences - OR overtime, resource utilisation, n° of shifted cases	- Minimisation of the required ORs - Bed occupancy levelling	- Minimisation of the patients' LoS - Minimisation of OR time undersupply to specialties	- Bed occupancy levelling - Schedule cyclicity - Minimisation of OR sharing among different specialties	Minimisation of patients' waiting time
Decision delineation	**Scheduled 'object'**	Specialties, surgeons, patients	Procedure typologies	Specialties	Surgeons	Patients
	Decision details	Date, time, OR	Date, OR	Date, time, OR	Date, time, OR	Date, OR
Research methodology	**Type of analysis**	- Single criterion exact optimisation - Scenario analysis	- Multi-criteria exact optimisation - Multi-criteria heuristic optimisation	- Multi-criteria hierarchical exact optimisation - Scenario analysis	- Multi-criteria exact optimisation - Multi-criteria heuristic optimisation	Single criterion heuristic optimisation
	Solution technique	- Mixed integer programming - Discrete event simulation	- Mixed integer programming - Column generation -Decomposition approach	- Mixed integer programming - Discrete event simulation	- Goal programming - Simulated annealing	Constructive heuristic
Type of constraints	**Resource**	Surgical staff, regular OR time, OR overtime	Wards, ICUs, OR overtime	Surgical staff, equipment, regular OR time	Regular OR time	Wards, ICUs, surgical staff, regular OR time, OR overtime
	Others	Max and min n° of OR blocks per week to specialties	Throughput target	Specialty demand (elective, non-elective)	Surgeon demand	Additional restrictions
Uncertainty		Deterministic (optimisation), stochastic (scenario analysis) surgery duration and arrivals	Deterministic LoS, stochastic surgery duration	Deterministic (optimisation), stochastic (scenario analysis) surgery duration and arrivals	Deterministic (multinomial distribution for the number of patients per OR block and patient LoS)	Deterministic
Applicability		Tested on real data	Tested on real data	Tested on real data	Tested on real data	Tested on realistic data
Planning horizon		1 week	1–2 weeks, 1 month	1 week	1–2 weeks	1 week

The first column of Table 1 refers to our work. Comparing Tables 1 and 2, it can be noted that our model exhibits decision variables that are similar to those used in Santibanez *et al.* (2007). However, the two models differ in several aspects. The most important is that our model (as will be thoroughly explained in Sections 4 and 7) takes into account the cases' due dates and, consequently, allows—to a certain extent—the exertion of control over the hospital's waiting list. Another important feature of our model is that it actually schedules *procedure typologies* (which are referred to as *surgery groups*) instead of cases. Such a characteristic is shared by half of the reviewed papers. However, none of these deals explicitly with cases' due dates. While due dates are, indeed, considered in Tànfani & Testi (2010), their model assigns time slots to actual patients (instead of to procedure typologies) and assumes a planning horizon of one week. As such, their model is unsuitable for monthly planning, especially in contexts like children's hospitals, where a high rate of patient-driven cancellations makes it impossible to schedule patients too far in advance.

Finally, another important contribution of our study is that it addresses uncertainty. Several other authors have incorporated LoS or surgery duration uncertainty into their models (see Cardoen *et al.*, 2010, p. 928). For example, Van Oostrum *et al.* (2008) proposed an optimisation model where a constraint is inserted to keep the probability of realising an OR overtime from exceeding a defined threshold. Specifically, they exploited portfolio optimisation theory (Hans *et al.*, 2008) to reduce the time required to complete a surgical session. In addition, they mitigated the effects of LoS variability through the proper balancing of bed usage. Other authors (e.g., Testi *et al.* (2007), VanBerkel & Blake (2007), Zhang *et al.* (2008)) have instead utilised simulation to evaluate, ex-post, the robustness of schedules produced by optimisation models.

In our study, we propose also a novel optimisation-simulation approach to the MSS problem. Our approach allows for the evaluation of the robustness of the MSS produced by the MIP model, and permits the fine-tuning of the optimisation model to trade off robustness and efficiency. While the use of simulation in health care settings is by no means a new topic (Sobolev *et al.*, 2011), to the best of our knowledge, simulation has never before been used to fine tune optimisation models.

In sum, our study offers two major contributions to the MSS literature. First, it presents a new MIP model to support the surgical scheduling process. This model maximises patient throughput, assigns procedure typologies to OR time slots and presents three new features, allowing us to (i) take into account cases' due-dates, (ii) minimise bed mismatches and (iii) produce solutions characterised by a desired mix of surgeries. Second, this study proposes an optimisation-simulation approach that allows for the fine-tuning of the MIP model to obtain robust and easy-to-implement solutions.

In the following sections, we will present the problem addressed, the MIP model developed, the results of our study and, finally, our conclusions and suggestions for future research.

3 Problem addressed

In this section, we describe the main features that characterise our problem. Given a planning horizon where each day is organised as a set of time slots, a set of resources and a set of elective cases, each of which is characterised by a *due date* and by a *surgery group*, we jointly address the following two problems:

1. the assignment of specialties to each OR and time slot of the planning horizon;
2. the determination of the number of procedures that belong to each surgery group to be scheduled in each OR and time slot of the planning horizon,

with the objective of maximising the number of scheduled surgeries.

We consider three resources: ORs, surgical teams and beds. Moreover, we assume that beds are organised into a certain number of wards each accommodating cases characterised by the same LoS, e.g., short-stay beds for cases with a LoS equal to one day, medium-stay beds for a LoS equal to two days and long-stay beds for LoSs longer than two days.

The problem's solution must respect several feasibility constraints. Specifically, it has to: (i) be compatible with the daily availability of the aforementioned resources; and (ii) respect the compatibility between ORs and specialties, as well as between bed types and surgery groups. We assume, in fact, that certain ORs/time slots may not be suitable for certain specialties/surgery groups and that certain bed types are not compatible with certain surgery groups. Finally, the solution must (iii) guarantee the fulfilment of the cases' due dates.

The solution should respect quality requirements as well. In particular, when possible, each case type should be accommodated in an appropriate bed type. However, bed mismatches, e.g., long-stay case types accommodated in short-stay beds, may be tolerated if they allow for an increase in the throughput. Additionally, leaving an excessive number of surgeries with approaching due date, long duration and/or long LoS on the waiting list should be avoided. Otherwise, the maximisation of the throughput in the current planning horizon will lead to criticalities in the long run.

In the next section, we present the mathematical formulation of the model we have developed to address the problem described here.

4 Optimisation model

The proposed model satisfies the aforementioned feasibility and quality requirements. In order to limit the emergence of critical situations after the planning horizon, the horizon is extended to the right with e consecutive and not overlapping time periods (as an example, we might consider a one-month planning horizon and e=2 extra time periods of seven days each). Moreover, we assume that each surgery group is also characterised by a *surgery duration range* that classifies the procedures of the group as short-, medium-, or long-lasting procedures, according to their duration. Hence, surgery durations are categorised within a set of surgery duration ranges, thereby allowing for the specification of the mix of short-, medium- and long-lasting surgeries to be scheduled in the planning horizon. The mix constraints represent a means to avoid leaving an excessive number of long-lasting surgeries on the waiting list. It is worth noting that mix constraints could also be formulated in terms of LoS ranges to control the mix of short-, medium- and long-stay case types scheduled in the planning horizon.

Let us define the following sets and parameters:

e the number of time periods following the planning horizon

D the set of days in the planning horizon indexed by d

$\tilde{D}_j$ the j-th time period following the planning horizon, $\forall j = 1,..,e$

T the set of time slots indexed by t

O the set of ORs indexed by o

S the set of specialties indexed by s

K the set of surgery groups indexed by k

B the set of bed types indexed by b

G the set of surgery duration ranges indexed by g

M a suitably big constant

$W_0 >> W_1 >> W_2$ the weights used in the objective function

w_{kj} the penalty associated with cases in surgery group k with due date in $\tilde{D}_j$ not scheduled in the planning horizon, $\forall k \in K, \forall j \in 1,..,e$

V_{sdt} the maximum number of surgical teams available for specialty s, on day

d in time slot t, $\forall s \in S, \forall d \in D, \forall t \in T$

H_{odt} the available time of OR o on day d, time slot t, $\forall o \in O, \forall d \in D, \forall t \in T$

L_{kd} the number of cases in surgery group k needing a surgical procedure within day d, $\forall k \in K, \forall d \in D \cup \tilde{D}_1 \cup ... \cup \tilde{D}_e$

B_{bd} the number of beds of type b available on day d, $\forall b \in B, \forall d \in D$

$\underline{\underline{T}}_g$ the minimum percentage of procedures with a surgery duration in range g that has to be scheduled, $\forall g \in G$

$\overline{\overline{T}}_g$ the maximum percentage of procedures with a surgery duration in range g that can be scheduled, $\forall g \in G$

s_k the specialty of surgery group k, $\forall k \in K$

r_k the bed type required by surgery group k, $\forall k \in K$

$\bar{\gamma}_k$ the average surgery duration, expressed in minutes, of surgery group k, $\forall k \in K$

g_k the range to which the surgery duration of surgery group k belongs, $\forall k \in K$

$\bar{\beta}_k$ the average number of days of hospitalisation before surgery required by group k, $\forall k \in K$

$\bar{\alpha}_k$ the average number of days of hospitalisation after surgery required by group k, $\forall k \in K$.

Then let us define the following four families of variables:

$$x_{sodt} = \begin{cases} 1 & \text{if specialty } s \text{ is assigned to OR } o \text{ on day } d \text{ and time slot } t \\ 0 & \text{otherwise} \end{cases}$$

$\forall s \in S, \forall o \in O, \forall d \in D, \forall t \in T$, such that specialty s can be performed in OR o

y_{kodt} the number of procedures of surgery group k assigned to OR o on day d in time slot t, $\forall k \in K, \forall o \in O, \forall d \in D, \forall t \in T$

u_{kj} the number of cases of surgery group k with due date in time period $\tilde{D}_j$ not scheduled in the planning horizon, $\forall k \in K, \forall j \in 1,..,e$

$v_{bb'd}$ the number of beds of type b' used in place of beds of type b on day d, $\forall b, b' \in B, \forall d \in D$.

Furthermore, let us define the following auxiliary variables:

z_{bd} the number of beds of type b occupied on day d, $\forall b \in B, \forall d \in D$.

Observe that compatibility constraints between specialties/surgical groups and ORs/time slots are defined implicitly in the statement of variables x and y.

Using these variables and parameters, we can state the model formally as follows:

$$\max \; W_0 \sum_{\substack{k \in K, o \in O, \\ d \in D, t \in T}} y_{kodt} - W_1 \sum_{\substack{k \in K, \\ j=1,..,e}} w_{kj} u_{kj} - W_2 \sum_{\substack{b,b' \in B: b \neq b', \\ d \in D}} v_{bb'd} \qquad (4.1)$$

$$\sum_{s\in S} x_{sodt} \le 1 \qquad \forall o \in O, \forall d \in D, \forall t \in T \qquad (4.2)$$

$$\sum_{o\in O} x_{sodt} \le V_{sdt} \qquad \forall s \in S, \forall d \in D, \forall t \in T \qquad (4.3)$$

$$\sum_{k\in K: s_k = s} y_{kodt} \le M x_{sodt} \qquad \forall s \in S, \forall o \in O, \forall d \in D, \forall t \in T \qquad (4.4)$$

$$\sum_{k\in K} \bar{\gamma}_k y_{kodt} \le H_{odt} \qquad \forall o \in O, \forall d \in D, \forall t \in T \qquad (4.5)$$

$$\sum_{\substack{o\in O,\\ d'\in D: d'\le d, t\in T}} y_{kod't} \ge \sum_{d'\in D: d'\le d} L_{kd'} \qquad \forall k \in K, \forall d \in D : L_{kd} > 0 \qquad (4.6)$$

$$\sum_{\substack{o\in O,\\ d\in D, t\in T}} y_{kodt} + \sum_{h=1}^{j} u_{kh} \ge \sum_{d\in D\cup \tilde{D}_1 \cup \ldots \cup \tilde{D}_j} L_{kd} \qquad \forall k \in K, \forall j \in 1,\ldots,e \qquad (4.7)$$

$$\sum_{\substack{k\in K: r_k = b,\\ o\in O, t\in T}} \sum_{d'=\max(1, d-\bar{\alpha}_k)}^{\min(|D|, d+\bar{\beta}_k)} y_{kod't} = z_{bd} \qquad \forall b \in B, \forall d \in D \qquad (4.8)$$

$$z_{bd} \le B_{bd} + \sum_{b'\in B: b'\ne b} v_{bb'd} \qquad \forall b \in B, \forall d \in D \qquad (4.9)$$

$$\sum_{b\in B} z_{bd} + \sum_{b,b'\in B: b\ne b'} v_{bb'd} \le \sum_{b\in B} B_{bd} \qquad \forall d \in D \qquad (4.10)$$

$$\underline{\underline{T}}_g \sum_{\substack{k\in K, o\in O,\\ d\in D, t\in T}} y_{kodt} \le \sum_{\substack{k\in K: g_k = g,\\ o\in O, d\in D, t\in T}} y_{kodt} \le \overline{\overline{T}}_g \sum_{\substack{k\in K, o\in O\\ d\in D, t\in T}} y_{kodt} \qquad \forall g \in G \qquad (4.11)$$

$$x_{sodt} \in \{0,1\} \qquad \forall s \in S, \forall o \in O, \forall d \in D, \forall t \in T$$

$$y_{kodt} \in \mathbb{N} \qquad \forall k \in K, \forall o \in O, \forall d \in D, \forall t \in T$$

$$z_{bd} \ge 0 \qquad \forall b \in B, \forall d \in D$$

$$u_{kj} \ge 0 \qquad \forall k \in K, \forall j \in 1,\ldots,e$$

$$v_{bb'd} \ge 0 \qquad \forall b, b' \in B : b \ne b', \forall d \in D$$

The objective function (4.1) involves three criteria that are hierarchically ordered. Specifically, the first criterion is the maximisation of the number of surgeries planned (and consequently, of patient throughput), whereas the second and third criteria are, respectively, the minimisation of penalties resulting from missing due dates and the minimisation of bed mismatches.

In particular, the model guarantees that all of the cases with a due date within the planning horizon will be planned via a hard constraint (see Constraints (4.6)). In addition, if there are enough resources, the model also schedules cases with a due date outside the planning horizon. The second criterion of the objective function determines how these latter cases are selected. Specifically, weights w_{kj} depend both on the surgery group k and on the time period j. For example, the closer the due date is, the higher the weight, while the shorter the waiting list relative to a given surgery group, the lower its weight. W_0, W_1 and W_2 are weights that are used to reflect the priority of the three criteria.

In this model, each day of the planning horizon is characterised by $|T|$ time slots; in many hospitals (and also in our case study), each day is organised into two time slots, morning and afternoon. Constraints (4.2) assure that, at most, one specialty can be assigned to an OR in a given time slot, each day. Constraints (4.3) guarantee that the number of ORs assigned during the same time slot for a given specialty s does not exceed the maximum number V_{sdt} of surgical teams available for specialty s, on day d in time slot t. A pre-processing phase is performed to eliminate all variables x_{sodt} for which the corresponding parameter V_{sdt} is zero. Constraints (4.4) bind the x and y variables together; specifically, they state that, given specialty s, OR o, day d and time slot t, no procedure of that specialty can be planned ($\sum_{k \in K: s_k = s} y_{kodt} = 0$) if that specialty has not been assigned to OR o ($x_{sodt} = 0$). On the other hand, Constraints (4.4) are redundant when specialty s is assigned to OR o on a given day d, time slot t (i.e., $x_{sodt} = 1$); in that case, the constraints state that the maximum number of procedures performed cannot exceed a big-M, which represents an upper bound on the number of workable procedures. Constraints (4.5) guarantee that the time consumed by all of the procedures planned in OR o on a given day d during time slot t will not exceed the OR available time H_{odt}. Constraints (4.6) and (4.7) are covering constraints used to manage the waiting lists. Specifically, the number L_{kd} of patients who must undergo a procedure of surgery group k by day d (the so-called *due date*) is supposed to be known. Constraints (4.6) assure that such procedures take place in the planning horizon within day d, i.e., on any of the days preceding d or on d itself. Additionally, Constraints (4.7) also allow the potential for scheduling patients with due dates outside the planning horizon, so as to limit criticality in the following planning horizons. Constraints (4.7) exhibit a similar structure to Constraints (4.6) and count the number u_{kj} of patients requiring a surgery in group k, not scheduled in the planning horizon and whose due dates fall in time period $\tilde{D}_j$. For example, when the first time period outside the planning horizon is considered (i.e., j=1), the constraints assure that, for each surgery group, the sum of the number of procedures scheduled in the planning horizon and the number of procedures with a due date within $\tilde{D}_1$ not scheduled is at least equal to the number of patients who must undergo a procedure within the last day in $\tilde{D}_1$. The objective function minimises the unscheduled procedures by giving priority to surgery groups *k*, for which the due date falls within a closer time period and for which the waiting list is longer. The idea is that the closer the due date is, the higher the priority becomes; additionally, all of the surgery groups with a due date within the same time period that are not eventually planned in the time horizon are penalised in the same way. Constraints (4.8)–(4.10) control bed occupancy. As discussed in Section 3, several bed types are considered. Constraints (4.8) compute the number of beds of type *b* occupied on day *d* by taking into account all of those procedures planned on day *d'* that require a bed of type *b* and whose LoS comprises *d*. This number is stored in the auxiliary variable z_{bd}. Constraints (4.9), for each bed type *b* and day *d,* impose an upper limit on the number of occupied beds and, at the same time, state that a bed mismatch may occur at the cost of paying a penalty $v_{bb'd}$, which measures the number of beds of type *b'* used in place of beds of type *b* on day *d*. Note that if beds of type *b'* cannot be used instead of beds of type *b* on a given day *d*, the corresponding variable $v_{bb'd}$ is fixed to zero. Finally, Constraints (4.10) state that the number of beds occupied, either properly or not, must not exceed the total number of beds available, regardless of the type. Constraints (4.11), guarantee that, for each surgery duration range $g \in G$, the number of procedures belonging to *g* that are performed within the current planning horizon is between a lower ($\underline{\underline{T}}_g$) and an upper ($\overline{\overline{T}}_g$) threshold percentage of the total number of procedures performed. The remaining constraints impose that *x* variables are binary, *y* variables are non-negative integers and other variables are non-negative.

5 Scope of application

In this section, we illustrate a computational analysis where the model's performance has been tested in correspondence with different hospital settings.

The settings presented in this section were based upon real empirical data, i.e., the Meyer Hospital, which is thoroughly described in Section 6.1. The settings were then modified through changing the value of the most significant parameters.

In particular, to test the applicability of our MIP model in different contexts and to assess the generalisability of our study, we have considered:

- three different hospital dimensions (Dim1, Dim2, Dim3);
- three different planning horizon lengths (7, 14, 28 days); and
- two different ways of organising the surgical sessions (1 time slot, 2 time slots).

Specifically, we consider a relatively ‘small’ hospital (Dim1), like Meyer Hospital, which is characterised by 5 ORs, 47 beds and $V_{sdt} \leq 1$ surgical teams available for each specialty *s*, day *d,* time slot *t.* We then performed analyses with two further hospital dimensions settings, in which the ORs, beds and number of surgical teams were doubled (Dim2) and tripled (Dim3), respectively.

For all of these types of hospitals, we have considered short ($|D|$=7), medium ($|D|$=14) and long ($|D|$=28) planning horizons, and we hypothesised that the surgical activities are organised either into one time slot ($|T|$=1) or into two time slots ($|T|$=2).

Finally, we have considered that the ORs can be utilised 5 days a week with an available time of 690 min/day, and we have assumed full compatibility between ORs/time slots and specialties/surgical groups. These latter hypotheses make the problem less constrained than the real case presented in Section 6.

We have thus investigated, in total, 3x3x2=18 different settings. The values of the parameters utilised to define these settings are justified by the literature (see Section 2 and Cardoen *et al.* (2010)). For each of the aforementioned settings, we have analysed the model’s behaviour in correspondence with 10 different realistic waiting lists. These waiting lists, in fact, were based upon a real one containing 2,391 cases organised into 54 surgical groups *k* ($|K|$=54), each of which was characterised by an average surgery duration ($\overline{\gamma}_k$) and LoS $(\overline{\alpha}_k + \overline{\beta}_k)$. In line with the dimensions Dim2 and Dim3, the number of patients on the list for each surgery group was doubled and tripled, respectively. The randomisation was carried out by adding a number of days to the real cases’ due dates; this number was obtained by sampling a discrete uniform distribution that ranged from -15 to +15.

Therefore, this computational analysis is based on 180 (=18x10) different and randomly generated problem instances. The model was coded using AMPL and solved with the IBM ILOG CPLEX solver (version 11.2) running on a PC equipped with an Intel iCore 5 processor and 4 GB of RAM. For each instance, we analysed the solutions obtained by bounding the computational time to 10 and 30 minutes.

Table 3 shows the average number of variables and constraints characterising the problem in correspondence with each setting after the CPLEX pre-solving. Each average value presented in this section refers to one setting and it is computed over ten instances.

TABLE 3. *Number of variables and constraints after pre-solving*

	Variables						**Constraints**					
	|*T*|=1			**|*T*|=2**			**|*T*|=1**			**|*T*|=2**		
	Dim1	**Dim2**	**Dim3**	**Dim1**	**Dim2**	**Dim3**	**Dim1**	**Dim2**	**Dim3**	**Dim1**	**Dim2**	**Dim3**
|*D*|=7	978.9	1813.9	2648.9	1801.4	3471.4	5141.4	986.9	1752.4	2517.4	1775.3	3305.7	4835.7
|*D*|=14	1873.7	3543.2	5212.7	3535.8	6875.8	10215.8	1860.3	3389.8	4919.3	3442.6	6502.6	9562.6
|*D*|=28	3632.6	6968.6	10304.6	6973.8	13653.8	20333.8	3585.7	6641.7	9697.7	6760.6	12880.6	19000.6

Figure 1, instead, shows the *average relaxation time* for each setting.

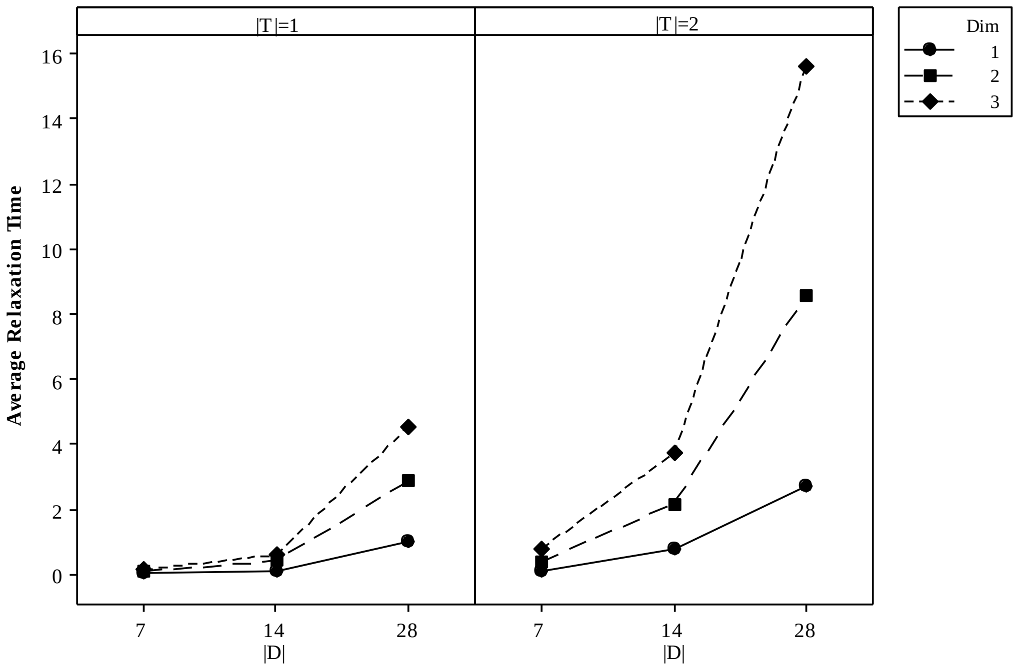


FIG. 1. Average relaxation time (sec) for different settings.

As can be noted, the average relaxation time, i.e., the time required to solve the linear relaxation of the MIP problem, grows as the dimension of the instances increases. Nonetheless, it remains acceptable (less than 16 seconds) even for the biggest instances in the computational campaign. It is worth pointing out that, even with a 10-minute time limit, we found at least one feasible solution for all the 180 instances. However, we found optimal solutions in only a small number of cases. Indeed, with a time limit of 10 minutes, we found no optimal solutions for $|D|$=28. Similarly, for $|D|$

=14, we found only one optimal solution. The number of optimal solutions does not increase significantly when the time limit is increased to 30 minutes. However, the *Relative Mipgap* reported in Table 4 reveals that solutions of very good quality were found in most cases (gaps<1.5%). Unfortunately, the quality of the solution degrades as the dimension of the instances grows. Nevertheless, solutions of tolerable quality (i.e., gap<12%) were found within 10 minutes even for quite big instances, e.g., $|D|$ =28 or $|T|$=2. These gaps decrease when the time limit is increased to 30 minutes (see Table 4). On the other hand, no reduction of the gap occurs for the biggest instances, even if the time limit is increased (see bold entries in Table 4).

TABLE 4. *Relative Optimality Gaps*

	Time limit = 10 min						Time limit = 30 min					
	|*T*|=1			|*T*|=2			|*T*|=1			|*T*|=2		
	Dim1	**Dim2**	**Dim3**	**Dim1**	**Dim2**	**Dim3**	**Dim1**	**Dim2**	**Dim3**	**Dim1**	**Dim2**	**Dim3**
|*D*|=7	0.09%	0.06%	0.13%	0.15%	0.12%	0.29%	0.07%	0.06%	0.11%	0.15%	0.06%	0.15%
|*D*|=14	0.42%	0.59%	0.60%	0.57%	7.46%	33.99%	0.34%	0.44%	0.48%	0.57%	1.30%	19.33%
|*D*|=28	1.31%	6.80%	9.19%	11.62%	**40.56%**	**38.97%**	1.25%	5.81%	9.11%	3.06%	**40.56%**	**38.97%**

In sum, it is possible to state that the model allows us to find satisfactory solutions in a reasonable amount of time for: (i) relatively small hospitals (5 ORs/47 beds), regardless of the planning horizon and of the number of time slots; and (ii) short planning horizons (7 days), regardless of the hospital dimensions (up to 15 ORs and 141 beds) and time slots (up to 2 time slots).

6 Case description: the surgical scheduling process at Meyer Hospital

The surgical unit at Meyer Hospital consists of seven ORs: five of these are partially interchangeable and host 15 surgical specialties (urology, otorhinolaryngology, paediatric surgery, neonatal surgery, ophthalmology, orthopaedic surgery, gynaecology and obstetrics, trauma centre, hand and microsurgery, oral and maxillofacial surgery, orthopaedic oncology, cardiothoracic surgery, gastroenterology, burns and plastic surgery); the remaining two ORs are dedicated almost entirely to specific surgical specialties (neurosurgery) or treatments (hemodynamics and bronchial endoscopy) and partially to emergencies and urgencies. At Meyer Hospital, emergencies and urgencies are managed by allocating them a fixed amount of OR time-slots and a fixed number of beds. Meyer Hospital allocates 47 beds to elective patients. These beds are organised into three wards according to the patients' expected LoS, i.e., long, medium and short. The hospital waiting lists are populated on the basis of surgery request forms that are filled out by surgeons. The form clearly indicates, for each case: (i) the diagnosis; (ii) the procedure that the patient is expected to undergo; and (iii) a priority class. The priority class determines the maximum number of days within which the case should be scheduled and, thus, the case's due date. There are three possible priority classes, and these are associated with 30, 60 and 90 days of waiting time, respectively. Even though the hospital is not obliged by law to meet the promised due date, there is a strong commitment to increase hospital performance in terms of due date fulfilment. In addition, the form indicates: (iv) the expected duration of the procedure (surgery duration); and (v) the expected LoS. There are three possible time ranges for procedure duration: less than one hour (short duration), between one and two hours (medium duration), and more than two hours (long duration). With respect to the expected LoS, a distinction is made between *day surgeries* (which occupy a bed for a single day) and *ordinary surgeries* (which occupy a bed for more than one day). Presently, at Meyer Hospital, the activities of the 15 surgical specialties that use the five aforementioned interchangeable ORs are planned by a Planning Department, which is headed by a bed manager. The activities of the remaining specialties are managed by their respective departments and are not considered in this paper. The entire planning process is performed manually, and it is organised into two stages. In the first stage, on a monthly basis, a timetable is produced that indicates the specialties assigned to each OR and to each time slot (morning or afternoon) on each day. In addition, it provides a rough indication of the number of day surgeries and ordinary surgeries that should be performed in each slot. Such a timetable is produced by taking into account the fact that each specialty can ensure the availability of a surgical team only during certain days/time slots within the planning horizon. It is worth pointing out that, given the need to coordinate surgeons' activities within and outside the surgical department, the specialties tend to hold these availabilities constant year round.

In the second stage, Planning Department personnel compile the timetable on a weekly basis, assigning cases to each OR for each time slot. Cases are chosen such that: (i) the sum of the expected surgery duration of the cases assigned to each time slot does not exceed the duration of the time slot itself; (ii) the expected number of hospitalised patients for each day does not exceed the number of expected available beds; and (iii) the percentage of short-, medium- and long-lasting surgeries scheduled in the weekly plan reflects approximately the percentage on the waiting list. At

Meyer Hospital, the demand for short-, medium- and long-lasting surgeries has proven to be fairly constant year round. Hence, by scheduling a constant mix of short-, medium- and long-lasting surgeries, the hospital avoids leaving an excessive amount of long-lasting surgeries on waiting lists, which would make the scheduling process more complex in the following weeks or months. Lastly, if possible, patients with closer due dates are given higher priority. The selected patients are then called to be operated on and a recovery date is given. If a patient is not available to be scheduled (i.e., because s/he is ill), then another case within the same specialty and with a similar (or shorter) expected surgery duration and LoS is called in her/his place. To prevent overtime and cancellations, each week the Planning Department schedules a fewer number of surgeries (by almost 15%) than what is suggested in the monthly plan. This practice leads to the underutilisation of both the ORs and the beds. Unfortunately, such underutilisation is further exacerbated by last-minute patient-driven cancellations, which account for almost 10% of the cases scheduled.

In the next subsection, we present the data we used to test our model, all of which originated from the planning period between 5 September 2011 and 2 October 2011.

7 Computational results

In this section, we present the computational results of our study. In the following subsections, we illustrate; (i) the data we used to test both the optimisation and the simulation model; (ii) the results of the optimisation model and a scenario analysis; (iii) the results of the simulation analysis; and (iv) the combined optimisation–simulation approach.

For all of the analysed scenarios, we limited the computational time to 10 minutes. For each, we found a solution with an optimality gap of at most 0.5%.

Hereinafter, we will denote random variables with capital Greek letters, such as A, B, Γ, the values that random variables take on with lowercase Greek letters, such as α, β, γ, and the random variables' mean values with lowercase Greek letters with a bar on top, such as $\overline{\alpha}$, $\overline{\beta}$, $\bar{\gamma}$.

7.1 **Input data**

7.1.1 *OR available time and beds*

The model considers 47 beds: 14 in the short-stay ward (day surgery), 19 in the medium-stay ward and 14 in the long-stay ward. As mentioned, it considers five ORs that can be utilised by 15 specialties. These ORs are not available on all the days and time slots in the planning horizon. The total monthly OR *available time* is equal to 819.5 hours.

7.1.2 *Surgery duration, LoS and surgery groups*

To calculate the values of the surgery durations and the LoS to use in the model, we analysed two years of surgical records. Each record indicated: i) the estimated surgery duration (short, medium, or long) and LoS (expressed in days), as indicated by the surgeon when s/he prescribed the surgery (see Section 6); and ii) the actual values of both the surgery duration and the LoS. Combining, for each specialty, the different values of the estimated surgery duration and LoS, we created 54 homogeneous surgery groups. For example, a group labelled *Urology-Short-2* includes procedures (e.g., varicocelectomy, orchidopexy, etc.) that, ex-ante, were expected to require a urology surgical team, occupy the OR for a small amount of time (less than one hour) and give rise to a post-surgical LoS of two days (i.e., the patient was expected to occupy one bed for two days: the day of the surgery and the following day). For each group *k*, the surgery duration (Γ_k) and the post-surgical LoS (A_k) are random variables. Hence, we determined their empirical distribution and calculated their respective mean values $\bar{\gamma}_k$ and $\bar{\alpha}_k$. Finally, since in our setting, patients do not occupy beds in the surgical department in the days preceding surgery, we have set $\bar{\beta}_k$=0. All of the solutions of the optimisation model presented in this paper are based on these mean values.

7.1.3 *Availability of surgical teams*

We considered the actual number of available surgical teams that each specialty could ensure for each day and time slot of the period under investigation.

7.1.4 *Waiting list*

We analysed the hospital waiting list as follows. For each surgery group, we created a dedicated waiting list that contained all of the cases that needed a procedure that fell within the surgery group itself. Then, surgery groups were clustered based on the number of cases on their waiting lists. With this method, we identified five clusters. The first includes all of the surgery groups with more than 200 cases on their lists; the second includes the surgery groups with 151 to 200 cases on their lists, and so on (see the rows in Table 5). The cases within each cluster were then subdivided based on their due dates. We identified six time intervals into which each case's due date could fall: within 28 days, from 29 to 36 days and so on (see the columns in Table 5).

The resulting breakdown of the waiting lists on 5 September 2011 is shown in Table 5 and will be commented on hereafter.

TABLE 5. *Actual waiting list on 5 September 2011*

Clusters	N° of cases in list for each surgery group included in the cluster	Time intervals (days)						Tot N° of cases in list
		< 28	29-36	37-44	45-52	53-60	> 60	
1	**> 200**	106	28	4	17	90	565	**810**
2	**151-200**	26	18	6	1	28	98	**177**
3	**101-150**	11	2	0	0	8	108	**129**
4	**51-100**	51	21	11	8	37	394	**522**
5	**0-50**	120	61	9	13	123	427	**753**
	Total N° of cases in list	**314**	**130**	**30**	**39**	**286**	**1592**	

For example, cluster 1 contained 810 cases, 106 of which had due dates expiring within 28 days. More specifically, these cases belonged to two surgery groups: *Urology-Short-1* (331 cases, with 28 expiring within 28 days) and *Paediatric Surgery-Short-1* (479 cases, with 78 expiring within 28 days). Looking at Table 5, it is possible to observe that: i) many cases (1,592) had due dates that were still somewhat remote (more than 60 days from the beginning of the planning horizon); and ii) 314 cases had due dates that expired within 28 days (these also include cases that were already late on 5 September 2011). Both of these facts were due to a lack of control over the waiting lists. On the one hand, many cases had been placed on the waiting list with no due date or with very remote due dates (this is typical for surgeries that, sooner or later, children are expected to undergo, but that are relevant to pathologies that don't cause any problem in the short term). On the other hand, a consistent number of surgeries with approaching (or expired) due dates had not been scheduled in the preceding weeks. Unfortunately, the accommodation of such a high number of cases (314) with due dates that fell within the planning horizon was not compatible with the availability of surgical teams that each specialty had ensured for the same period. For example, a certain number of cases had a due date *d'*, but required a surgical team that was only available later than *d'*. Thus, it wasn't possible to meet the expected due date and, consequently, the model wasn't able to find any feasible solution. To handle such a criticality, we agreed with the hospital management to postpone all of the due dates for 15 days. The composition of the resulting *adjusted* waiting lists is shown in Table 6.

TABLE 6. *Adjusted waiting lists on 5 September 2011*

Cluster	N° of cases in list for each surgery group included in the cluster	Time intervals (days)						Tot N° of cases in list
		< 28	29-36	37-44	45-52	53-60	> 60	
1	**> 200**	21	11	74	30	11	663	**810**
2	**151-200**	11	6	10	17	7	126	**177**
3	**101-150**	1	1	9	2	0	116	**129**
4	**51-100**	21	10	26	21	18	488	**584**
5	**0-50**	25	14	79	60	11	502	**691**
	Total N° of cases in list	**79**	**42**	**198**	**130**	**47**	**1895**	

All of the solutions presented in this paper are based on the adjusted waiting lists (Table 6). As will be shown in Section 7.2.1, despite using the adjusted waiting lists, the solutions presented allow us to respect most of the real due dates (i.e., before postponement). The adjusted lists were also used to create the realistic instances presented in Section 5.

7.1.5 *Objective function penalties*

The weights w_{kj} of the objective function are set in such a way that the model prioritises (i) the groups whose waiting lists include a higher number of cases with closer due dates; and (ii) groups with longer waiting lists. The weights are set to give higher priority to the due dates. Hence, the length of the waiting list will discriminate between two solutions only when these solutions are equivalent in terms of due dates.

7.1.6 *Allowed mix variation.*

We have set the lower bound ($\underline{\underline{T}}_g$) and the upper bound ($\overline{\overline{T}}_g$) in Constraints (4.11) so that the model is allowed to schedule a certain percentage of short-, medium- and long-lasting surgeries, which will differ at most by 3% from the percentage on the waiting list. The actual mix on the waiting list is 74% short-, 21% medium-, 5% long-lasting surgeries.

7.2 **Optimisation results**

7.2.1 *Model output*

Table 7 summarises the model's output relative to the base scenario.

TABLE 7. *Base scenario: output summary*

	Model output
Planned surgeries	651
- short duration [%]	74.3
- medium duration [%]	20.6
- long duration [%]	5.1
OR utilisation rate [%]	83.2
Bed utilisation rate [%]	81.8
Bed mismatch rate [%]	3.3

The number of planned surgeries (651) is remarkably higher than the number planned by the hospital for the same period (495). In addition, it allows for obtaining high OR and bed utilisation rates (which rise to 95% on weekdays) and a low bed mismatch rate. Figure 2 illustrates the daily profile of the OR (left) and bed utilisation (right).

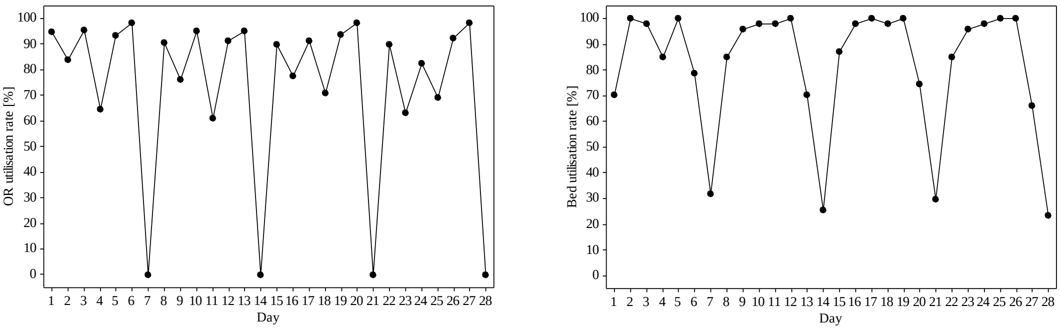


FIG. 2. Base scenario: OR and bed utilisation daily rates.

As can be noted, bed utilisation has a similar profile over the last three weeks (days 8 to 28), whereas it denotes a different pattern in the first week. This is because the number of time slots that are allotted, each day, to the 15 specialties considered by the model is not constant, but varies from a maximum of nine to a minimum of zero. In particular, on the third day of the first week, more than 30% of the available time slots are allotted to specialties that are not considered by the model (see Fig.3). This explains why the bed utilisation rate (which refers only to the specialties considered by the model) sharply decreases on the fourth day.

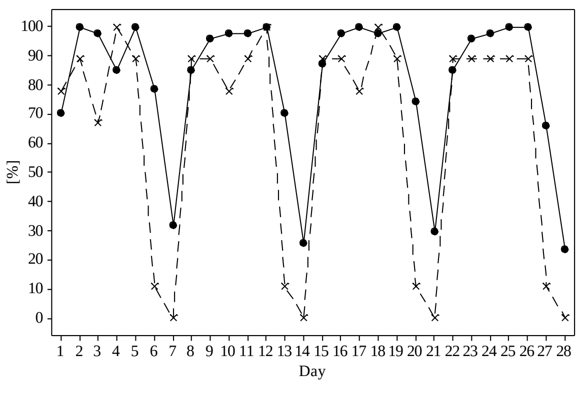


FIG. 3. Base scenario: bed utilisation (●) and time-slot availability (×) daily rates.

Table 8 shows how waiting lists are *consumed*, indicating, in each cell, the number of surgeries planned and, in round brackets, the percentage of surgeries planned that were on the waiting list.

TABLE 8. *Waiting list consumption*

Cluster	Time intervals (days)						TOT
	< 28	29-36	37-44	45-52	53-60	> 60	
1	21 (100%)	11 (100%)	74 (100%)	21 (70%)	4 (36.4%)	0 (0%)	**131 (16.2%)**
2	11 (100%)	6 (100%)	10 (100%)	17 (100%)	4 (57.1%)	0 (0%)	**48 (27.1%)**
3	1 (100%)	1 (100%)	5 (55.6%)	0 (0%)	0 (N.A.)	0 (0%)	**7 (5.4%)**
4	21 (100%)	7 (70%)	16 (61.5%)	13 (61.9%)	7 (38.9%)	32 (6.6%)	**96 (16.4%)**
5	25 (100%)	11 (78.6%)	59 (74.7%)	42 (70%)	3 (27.3%)	229 (45.6%)	**369 (53.4%)**
	79 (100%)	**36 (85.7%)**	**164 (82.8%)**	**93 (71.5%)**	**18 (38.3%)**	**261 (13.8%)**	

As expected, the model allows for planning a higher percentage of surgeries with closer due dates. To obtain a higher consumption of the waiting lists of the groups belonging to the first clusters, it would be sufficient to appropriately set the weights (due to space constraints, such a setting is not presented here). In addition, it is worth noting that, although the due dates had been postponed by 15 days (see Section 7.1.4), 279 (=79+36+164) of the 314 surgeries (see Table 5) that were due in the first 28 days (and whose adjusted due dates fell between the first and the 44th day of the planning horizon), were in fact scheduled. This means that postponing the due dates allowed us to find a solution that scheduled 89% of the interventions that were actually due for the current month. The remaining 11% of interventions could not be scheduled due to a lack of resources.

7.2.2 *Scenario analysis.*

In this section, we illustrate the results of a scenario analysis. In particular, we want to test how the model solution, specifically the number of planned surgeries, would change as a consequence of: (i) an increase in the number of beds and/or the OR available time; (ii) an increase in the availability of surgical teams; or (iii) variation in the composition of the mix of short-, medium- and long-lasting surgeries that the model is allowed to plan.

7.2.2.1 Impact of an increase in the available number of beds and/or OR time

To investigate the impact of an increase in the number of beds and/or OR time on the model output, we analysed three scenarios. In the first, the number of beds is increased by 10% (ΔBed=10%). In the second, the total OR available time is increased by 10% (ΔOR=10%). In the third, both the number of beds and the total OR available time are increased by 10% (ΔOR=ΔBed=10%). The results are presented in Table 9.

TABLE 9. *Scenario 1 to 3: increase of the OR available time and/or the number of beds*

	Base scenario	Scenario 1 (ΔOR=0% ΔBed=10%)	Scenario 2 (ΔOR=10% ΔBed=0%)	Scenario 3 (ΔOR=10% ΔBed=10%)
Planned surgeries	651	694	677	737
- short duration [%]	74.3	74.4	74	74.9
- medium duration [%]	20.6	20.6	21	20.1
- long duration [%]	5.1	5	5	5
OR utilisation rate [%]	83.2	89.1	78.2	85.3
Bed utilisation rate [%]	81.8	78.2	82.8	80.7
Bed mismatch rate [%]	3.3	0.7	4.7	1.1

Starting from the base scenario (Table 7), it is possible to increase the number of planned surgeries both by only adding beds (scenario 1, +43 surgeries) and by only increasing the OR available time (scenario 2, +26 surgeries). The marginal benefits of increasing one resource once the other has already been increased leads to substantial benefits as well. In fact, in a move from scenario 1 to scenario 3 or from scenario 2 to scenario 3, it is possible to plan an additional 43 or 60 surgeries, respectively.

7.2.2.2 Impact of an increase of surgeons' availability

One feature of our setting is that the MSS must be produced by taking into account that, for each specialty, the availability of a surgical team is ensured only on pre-determined days and time slots within the planning horizon (see Section 6). Considering this fact, however, does not allow for the full exploitation of the model. On the contrary, by allowing the model to freely assign specialties to sessions, it is possible to determine *when* each specialty should ensure the availability of surgical teams, in order to maximise the value of the objective function. In Table 10, for example, we show the results if each specialty were potentially able to ensure the availability of one surgical team for every day and session within the planning horizon, and if the model were completely free to assign specialties to ORs/sessions.

TABLE 10. *Scenario 4: increase in the surgical teams' availability*

	Scenario 4 (Higher surgical teams availability)
Planned surgeries	685
- short duration [%]	74.9
- medium duration [%]	20
- long duration [%]	5.1
OR utilisation rate [%]	86.6
Bed utilisation rate [%]	82.0
Bed mismatch rate [%]	3.2

With respect to the base scenario (Table 7), it can be noted that an increase in the surgical teams' availability leads to a substantial increase in the number of planned surgeries (685 vs. 651) and in OR utilisation (86.6% vs. 83.2%). The surgery mix, the bed utilisation and the bed mismatch rate, however, remain constant. Hence, *ceteris paribus*, a higher availability of surgical teams allows for a substantial increase in the number of planned surgeries.

7.2.2.3 Impact of mix variation

Another feature of our setting is that the MSS must contain a constant mix of short-, medium- and long-lasting procedures. Such a result is obtained by limiting the allowed mix variation (Constraints (4.11)) to a maximum value of 3%. In this subsection, we illustrate how the solution would change if we were to provide a different setting for this constraint. In particular, we test what happens if: i) the maximum allowed mix variation is set to 0% (no mix variation allowed); and ii) the mix constraint is removed (100% mix variation allowed). The results of these analyses are summarised in Table 11 and commented upon hereafter.

TABLE 11. *Scenarios 5 and 6: change in the allowed mix variation*

	Scenario 5 (Allowed mix variation=0%)	**Scenario 6** (Allowed mix variation=100%)
Planned surgeries	600	661
- short duration [%]	74.0	73.1
- medium duration [%]	21.0	25.3
- long duration [%]	5.0	1.7
OR utilisation rate [%]	77.7	83.3
Bed utilisation rate [%]	78.4	81.8
Bed mismatch rate [%]	2.1	3.4

When the allowed mix variation is set to zero, the solution is remarkably worse than the one in the base scenario. In fact, in this setting, the model is able to schedule only 600 surgeries instead of 651. On the contrary, if we remove the mix constraints, the number of planned surgeries increases from 651 to 661. Such a solution, however, is characterised by a surgery mix that is particularly lacking in long-lasting surgeries (11 vs. 33). This means that a higher percentage of long-lasting (and resource-consuming) procedures will be left on the waiting list, thereby making the scheduling process more complex in the following months. By allowing a mix variation of 3% (base scenario) instead, we obtain a better trade-off between planned surgeries and actual surgery mix.

7.3 **Simulation analysis**

In this section, we illustrate the results of the simulation analysis. This section is organised into two parts. First, we provide a short description of the simulation model, which will also be used in Section 7.4. Second, we use the simulation model to test the robustness of the solution produced by the optimisation model against the variability of the surgery duration and of post-surgical LoS. The simulation model described in this section was created with Rockwell Arena (version 13.9) and integrated with AMPL via VBA.

7.3.1 *Simulation model*

The analyses presented in the next sections are based on a discrete event simulation model that works as follows. Based on the optimisation model's solution, for each simulated day, the simulator generates a number of entities equal to the number of surgeries planned for the day and assigns a surgery group, OR and time slot to each entity. Hence, every day, before the beginning of the morning session, the model verifies whether there are enough beds to accommodate the number of cases planned in the MSS for that day. If the number of surgeries planned exceeds the number of beds available, than the model randomly deletes a number of entities that is equal to the number of surgeries planned minus the number of beds available. The remaining entities seize a bed each for a time that is randomly sampled from the empirical distribution of A_k. In addition, by following the MSS, each of these entities seizes an OR for a time that is randomly sampled from the empirical distribution of the surgery duration Γ_k. Then, the model records the number of surgeries actually *executed* (ε), the number of surgeries *cancelled* (ω), the total *overtime* (θ), the *OR utilisation rate* (τ), and the *bed utilisation rate* (σ). In the next section, we present the results of the simulation with respect to the base scenario.

7.3.2 *Simulation model results: base scenario*

Table 12 summarises the output of 20 simulation runs relevant to the base scenario. For each performance, we indicate the value obtained with the optimisation model and the mean values, standard errors (SE) and minimum and maximum obtained with the simulation.

TABLE 12. *Simulation results: base scenario*

	Optimisation output	Simulation output			
		Mean	SE	Min	Max
Planned surgeries	651	651	-	651	651
OR utilisation rate [%]	83.2	76.8	0.2	75.4	78.2
Bed utilisation rate [%]	81.8	77.5	0.1	76.4	78.6
Overtime [h]	-	16.3	0.9	9.4	23.3
Cancelled surgeries	-	46.9	1.4	35	59
Executed surgeries	-	604.1	1.4	592	616

As can be noted, because of the variability of the post-surgical LoS, almost 47 surgeries (7.2%) of the planned 651 could not be executed due to a bed shortage.

Such a circumstance, in turn, led to lower rates of bed and OR utilisation. In Fig. 4, we show the boxplot of the daily OR and bed utilisation rates, respectively.

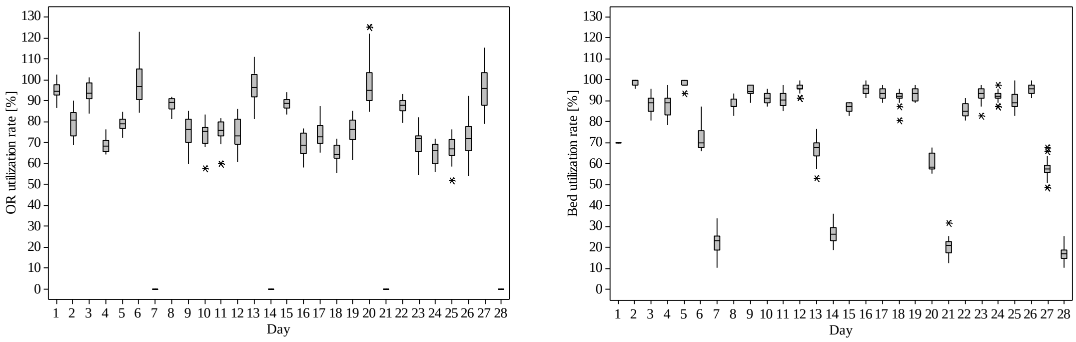


FIG. 4. Boxplot of the daily OR (left) and bed utilisation (right) rates.

Looking at the boxplots (Fig. 4), it is possible to observe that the OR utilisation rate is characterised by a high variability and is sometimes larger than 100%. In fact, despite the 47 cancellations, the variability of the surgery duration relevant to the remaining 604 surgeries caused an overtime of almost 16 h/month. The range for daily overtime, in turn, is [0; 6.75] hours, where 6.75 is calculated, as suggested by Kelton *et al.* (2002, p. 39), as the maximum of the individual replication maxima. The mean value of daily overtime, instead, is 0.58 hours (or 0.68, if Sundays are excluded). On the contrary, it is possible to observe that the first quartile of daily bed utilisation is always above 80% (excluding weekends), and the inter-quartile range is rather small. In sum, the simulation suggests that if the MSS obtained with the optimisation model were implemented, then the hospital would need to cancel 7.2% of the surgeries planned, and it would experience an overtime of 16 h/month. Such a high number of cancellations would surely lead to patient dissatisfaction and, to a certain extent, undermine the hospital's ability to control the waiting list (i.e., a patient with an approaching due date might be cancelled even though it would be impossible to schedule her/him again in the current month). In the next section, we illustrate how these issues can be fixed by using a combined optimisation–simulation approach.

7.4 Combined optimisation-simulation approach

In the previous section, we have shown that, because of the randomness of surgery duration and post-surgical LoS, the implementation of the solution of the optimisation model leads to cancellations and overtime. In many cases, cancellations and overtime can be highly undesirable (which is surely the case at Meyer Hospital). Given a certain set of resources (e.g., beds, ORs and surgical teams), a way to reduce the number of cancellations and obtain a more robust solution consists of running the optimisation model while considering a smaller amount of resources than what is actually available (Hans *et al.*, 2008). By introducing resource slacks, the optimisation model schedules fewer surgeries and, consequently, the obtained solution likely gives rise to fewer cancellations and to less overtime.

However if the decrease of cancelled surgeries ($\Delta\omega<0$), is smaller than the decrease of the planned surgeries ($\Delta N<0$), then also the number of executed surgeries decreases ($\Delta\varepsilon=\Delta N-\Delta\omega<0$). In other words, utilising a more robust schedule likely leads to the execution of fewer surgeries. Consequently, the benefits in terms of patient satisfaction arising from the reduction in cancellations need to be balanced against the (opportunity) costs of executing fewer surgeries. Hereafter, we illustrate how the simulation and optimisation models can be used jointly to manage this trade-off between robustness and efficiency. The combined optimisation–simulation approach is based upon three steps.

1. *Optimisation*. The optimisation model runs in several different configurations. In each configuration, the model finds a solution that is based on an OR available time and on the availability of a certain number of beds, which are reduced by a percentage equal to $\%_h$ and $\%_b$, respectively, with respect to real values. The number of surgeries that the optimisation model will plan (N), as well as the *total planned surgery duration* (P) (left-hand-side of Constraints (4.5)), obviously depend on $\%_h$ and $\%_b$, i.e., $N=N(\%_h,\%_b)$ and $P=P(\%_h,\%_b)$.

2. *Simulation*. Each solution obtained in the previous step undergoes several simulation runs. The simulation is carried out by considering the real number of beds and the real total OR available time. The number of cancellations (Ω), the number of surgeries executed (E), and overtime (Θ) will depend upon $\%_h$ and $\%_b$ as well, i.e., $\Omega=\Omega(\%_h,\%_b)$, $E=E(\%_h,\%_b)$, $\Theta=\Theta(\%_h,\%_b)$.

3. *Analysis*. The results of the simulations are analysed to identify solutions with an *acceptable* robustness. To identify the acceptable robustness levels, we proceed as follows. For each configuration ($\%_h,\%_b$) we define a *cancellation threshold* T_ω and an *overtime threshold* T_θ. The former is calculated as a percentage $\%_\omega$ of the number of planned surgeries (i.e., $T_\omega=\%_\omega N(\%_h,\%_b)$). The latter is calculated as a percentage $\%_\theta$ of the total planned surgery duration (i.e., $T_\theta=\%_\theta P(\%_h,\%_b)$). The solutions for which the probability of exceeding at least one of

these two thresholds is smaller than 0.05 are considered 'robust'. Among the robust solutions, we choose the one allowing, on average, for the execution of the highest number of surgeries.

The results of the application of the combined approach to the Meyer Hospital case are illustrated in the next subsection.

7.4.1 *Combined approach results*

We defined 11 different possible values for $\%_h$ and $\%_b$ (i.e., 0%, 2%, 4%, 6%, 8%, 10%, 12%, 14%, 16%, 18% and 20%), thereby obtaining 121 different configurations. Hence, for each configuration, we solved the optimisation model and carried out 20 simulation runs. To identify the *acceptable* solutions, we have set $\%_\omega$= $\%_\theta$=1%. Then, for each configuration, we performed two one-tailed *t*-tests and calculated the relevant *p*-values (p_1, p_2) (Montgomery & Runger, 2002, p. 337). With the first *t*-test, we tested whether the mean value of the number of cancelled surgeries ($\overline{\omega}$) was significantly smaller than T_ω. In statistical terms, we tested the null hypothesis, $H_0 : \overline{\omega} = T_\omega$, against the alternative hypothesis, $H_1 : \overline{\omega} < T_\omega$.With the second *t*-test, instead, we tested whether the mean value of the overtime ($\overline{\theta}$) was significantly smaller than T_θ (in this case, $H_0 : \overline{\theta} = T_\theta$ and $H_1 : \overline{\theta} < T_\theta$). The configurations for which $\overline{\omega}$ or $\overline{\theta}$ were not significantly smaller than their threshold values (p>0.05) were discarded. For all the remaining configurations, we carried out a Pearson correlation analysis revealing that Ω and Θ were not significantly correlated (p>0.05). Hence, for each of these configurations, we calculated the probability (p_{12}) of considering the configuration acceptable when it actually is not. To do so, we operated as follows. For each of the aforementioned *t*-test *i*, the probability of no Type I error is $(1-p_i)$. Hence, since the variables Ω and Θ were not correlated, the overall probability of no Type I error is $(1-p_{12}) = (1-p_1)(1-p_2)$ and, consequently, the probability of at least one Type I error is equal to $p_{12} = [1-(1-p_1)(1-p_2)]$. Therefore, we considered as acceptable those configurations for which p_{12}<0.05. Fig. 5 shows the contour maps plotted against $\%_h$ and $\%_b$, respectively, of the planned surgeries (*N*) and of the mean values of the executed surgeries ($\overline{\varepsilon}$), the cancellations ($\overline{\omega}$) and overtime ($\overline{\theta}$).

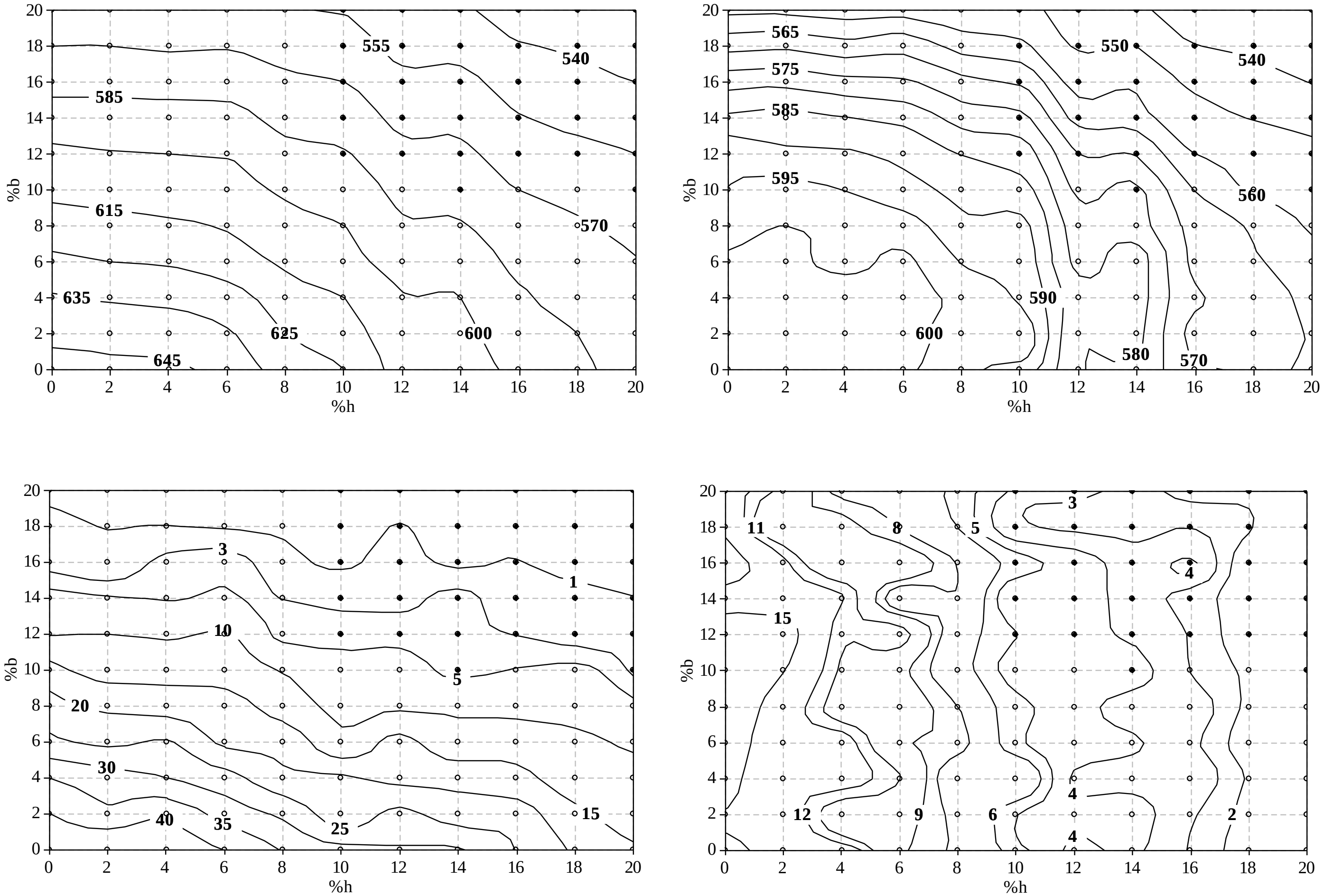


FIG. 5. Contour maps of N (upper left graph), $\bar{\boldsymbol{\varepsilon}}$ (upper right graph), $\bar{\boldsymbol{\omega}}$ (lower left graph) and $\bar{\boldsymbol{\theta}}$ (lower right graph)

The dots in each graph of Fig. 5 represent the configurations that were investigated. The black dots represent the acceptable configurations, whereas the white dots represent the unacceptable (non-robust) ones. Table 13, instead, shows the simulation results (mean, Standard Error SE, threshold value T_ω, *t*-value with the relevant degrees of freedom $t(19)$ and *p*-value p_i) in correspondence with the acceptable configurations. Due to space constraints, here we only show the results relative to the robust configurations allowing the execution of at least 565 surgeries.

TABLE 13. *Hospital performance in correspondence with the acceptable configurations*

Config.	***N***	***P***	**Executed surgeries**		**Cancelled surgeries**					**Overtime (h)**					p_{12}
			Mean	**SE**	**Mean**	**SE**	T_ω	**t(19)**	p_1	**Mean**	**SE**	T_θ	**t(19)**	p_2	
(10,12)	586	661.2	582.4	0.6	3.6	0.6	5.9	-3.9	0.001	5	0.5	6.6	-3	0.004	0.004
(10,14)	579	648.8	576.3	0.5	2.7	0.5	5.8	-6	0.000	4.1	0.4	6.5	-6.8	0.000	0.000
(10,16)	570	640.1	569.2	0.3	0.8	0.3	5.7	-17.7	0.000	5.5	0.5	6.4	-1.8	0.046	0.046
(12,12)	574	648.5	570	0.6	4	0.6	5.7	-2.9	0.005	4.3	0.4	6.5	-5.9	0.000	0.005
(14,10)	580	656.8	575.8	0.7	4.2	0.7	5.8	-2.2	0.018	4.5	0.3	6.6	-6.5	0.000	0.018
(14,12)	573	651.9	569.6	0.6	3.4	0.6	5.7	-4	0.000	3.9	0.4	6.5	-7.7	0.000	0.000

Looking at Fig. 5 and at Table 13, it is possible to draw several conclusions. First, a trade-off between robustness and efficiency does exist. As a matter of fact, the configurations characterised by high robustness are clearly separated from those characterised by high numbers of planned and executed surgeries. The former are situated in the upper-right part of the contour maps, whereas the latter are in the lower-left corner. Second, the configurations associated with robust solutions are characterised by moderately high values of *both* $\%_h$ (e.g., $\%_h \geq 10\%$) and $\%_b$ (e.g., $\%_b \geq 10\%$). Hence, acting on only one type of resource slack brings about unsatisfactory solutions. In fact, increasing only $\%_b$ keeping $\%_h=0$, while leading to a small number of cancellations, also gives rise to high amounts of overtime. Instead, by increasing $\%_h$ keeping $\%_b=0$, it is possible to reduce overtime, but the number of cancellations remains high**.** Third, when the values of $\%_h$ and $\%_b$ are too high (e.g., $\%_h \geq 14\%$ or $\%_b \geq 16\%$, or both) the number of executed surgeries decreases too much. Fourth, the best configuration is the one characterised by $\%_h$ =10% and $\%_b$=12%. Among the acceptable solutions, this is, in fact, the configuration that has, on average, the highest value of executed surgeries. Table 14 compares the performance obtained utilising only the optimisation model with the outcomes associated with the combined approach. Both solutions refer to the base scenario (see Section 6).

TABLE 14. *Combined approach results: base scenario*

	Base Configuration ($\%_h$=0; $\%_b$=0)					Best Configuration ($\%_h$=10; $\%_b$=12)				
	Optim. output	Simulation output				Optim. output	Simulation output			
		Mean	SE	Min	Max		Mean	SE	Min	Max
Planned surgeries	651	651	-	651	651	586	586	-	586	586
OR utilisation rate [%]	83.2	76.8	0.2	75.4	78.2	74.2	73.7	0.2	72.4	75.5
Bed utilisation rate [%]	81.8	77.5	0.1	76.4	78.6	72	73.8	0.2	72.6	75.4
Overtime [h]	-	16.3	0.9	9.4	23.3	-	5	0.5	1.3	9.7
Cancelled surgeries	-	46.9	1.4	35	59	-	3.6	0.6	0	9
Executed surgeries	-	604.1	1.4	592	616	-	582.4	0.6	577	586

If compared with the solution of the optimisation model, the one obtained with the combined approach is characterised by a number of cancellations that is significantly smaller ($t(25)$=28.25, p=0.000) and by a significantly smaller amount of overtime ($t(31)$=10.87, p=0.000), i.e., it is more robust. Such a solution will, therefore, lead to higher patient satisfaction, better control of the waiting list and lower costs. These benefits, however, need to be balanced against the opportunity cost of executing significantly fewer surgeries ($t(25)$=14.16, p=0.000).

7.4.2 *Scenario analysis*

In this section, we illustrate an example of the application of the combined approach to a scenario analysis. In particular, we investigate scenarios 1 (ΔBed=10%), 2 (ΔOR time=10%) and 3 (ΔBed=ΔOR=10%), presented in Section 7.2.2.1.

To apply the combined approach, for each scenario analysis we investigated 121 configurations (obtained with the same criteria that were described in Section 7.2.2). After we identified the robust configurations, we chose the best one. The results of the scenario analyses are summarised in Table 15 and commented upon hereafter.

TABLE 15. *Combined approach results: scenario analysis*

	Scenario 1 (ΔOR=0% ΔBed=10%) Best configuration ($\%_h$=10; $\%_b$=12)				Scenario 2 (ΔOR=10% ΔBed=0%) Best configuration ($\%_h$=10; $\%_b$=14)				Scenario 3 (ΔOR=10% ΔBed=10%) Best configuration ($\%_h$=14; $\%_b$=16)			
	Mean	SE	Min	Max	Mean	SE	Min	Max	Mean	SE	Min	Max
Planned	625	-	625	625	592	-	592	592	633	-	633	633
OR utilisation rate [%]	78.7	0.2	77	80.1	67.7	0.1	66.5	68.6	72.1	0.2	70.3	73.2
Bed utilisation rate [%]	70.7	0.2	69.1	71.6	72.1	0.2	70.9	73,4	69.6	0.2	68.6	71.4
Overtime [h]	4.9	0.5	1.4	9.8	5.2	0.4	1.8	9.2	5.5	0.4	2.6	9.4
Cancelled surgeries	2.5	0.5	0	8	4.6	0.6	0	8	4.2	0.7	0	10
Executed surgeries	622.5	0.5	617	625	587.4	0.6	584	592	628.8	0.7	623	633

By comparing the results presented in Table 15 with those relevant to the base scenario (Table 14), it is possible to observe that: first, adding 10% more beds (scenario 1) leads to an increase in the number of executed surgeries (+40.1%/+6.9%). This is, on average, relatively higher than the number that can be obtained by increasing the OR available time by 10% (+5%/+0.8%, scenario 2). Second, increasing the OR available time by 10% once 10% more beds have already been added (i.e., moving from scenario 1 to scenario 3) allows, on average, the execution of only 6.3 additional surgeries (+1%).

It is worth noting that scenario analysis based on the combined approach allows us to take decisions based on accurate estimates of the performance that the hospital will actually achieve if the investigated solutions are implemented. On the contrary, if we utilise only the optimisation model, the comparison of scenarios would be based on solutions that neglect the randomness of the data. These solutions, unfortunately, can differ a great deal from those that will actually be implemented. Consequently, the resulting comparison can be misleading. In fact, in Section 7.2.2.1, we concluded that increasing the OR available time by 10%, after having added 10% more beds, would allow for the planning of 43 more surgeries. Unfortunately, however, the combined approach has revealed that such a solution is not acceptable and that the best acceptable solution associated with the third scenario ($\%_h$=14%, $\%_b$=16%) is characterised, on average, by a mere 6.3 more executed surgeries than the best acceptable solution associated with the first scenario ($\%_h$=4%, $\%_b$=14%). Therefore, investments to increase the OR available time, if undertaken exclusively on the basis of the optimisation model results, will probably fail to produce the expected returns.

8 Conclusions and future research

This paper offers two main contributions to the research on the master surgical scheduling problem. It presents both an original MIP model and an original combined optimisation–simulation approach. The MIP model allows us to obtain a MSS that permits the maximisation of patient throughput and control of hospital waiting lists. The combined optimisation–simulation approach, instead, allows us to obtain a robust MSS while effectively trading off robustness and efficiency. Our study has, therefore, notable practical implications as well. By applying the presented combined approach, hospital managers can obtain a MSS that is characterised by the desired degree of robustness, and will thus be better able to manage waiting lists and make more informed investment decisions. Moreover, the adoption of surgery groups can simplify the short-term assignment of cases to OR time slots (the third step of the surgical scheduling process). In fact, the combined approach, if applied to the Meyer Hospital case, would help improve the hospital's performance in three ways. First, by selecting the best robust configuration (i.e., $\%_h$=10, $\%_b$=12), the Planning Department would have scheduled 582 surgeries instead of 495. Since the MSS suggests a pool of cases, i.e., those needing surgeries that fall within specific surgery groups, eligible to be scheduled, it would have been easier for the Planning Department to fill in the schedule. In addition, when necessary, it would have made it easier to replace unavailable patients on short notice without significant schedule disruptions. Third, by executing the surgeries planned in the MSS, the hospital would have appropriately utilised the waiting lists, reduced its backlog and experienced a maximum of nine overbooking cancellations.

However, the scenario analysis presented here does not suffice to support the hospital's investment decisions. These decisions would also require taking into account the costs that adding resources (e.g., surgical teams, beds, or OR time) would incur, and they should not be based solely on one problem instance. Nonetheless, our study demonstrates that the combined approach can offer more accurate scenario analyses.

This study, of course, is not without its limitations. First, as we illustrated in Section 5, the MIP model might be difficult to apply to support a master surgical scheduling process with a planning horizon of one month in large hospitals (e.g., a hospital with more than 10 ORs and 94 beds). In these settings, in fact, the applicability of the presented approach can be undermined by excessive computational time. To address these problems, ad-hoc algorithms should be purposely developed. Second, we have not investigated how different ways of grouping the procedures into surgical groups might affect the model's solution. Third, we have neglected to factor in certain hospital resources (e.g., ICU, electro-medical devices) that were not considered critical in our setting, but that in other hospitals may be highly critical. Finally, since in our setting non-elective patients are handled with dedicated resources (i.e., dedicated OR, time slots and beds), we only considered elective patients. Nonetheless, the presented combined approach would also seem promising for the study of the impact of the randomness caused by emergencies, urgencies and no-shows on hospital performance.

The development of ad-hoc algorithms to solve large problem instances, the addition of new resource constraints, the investigation of different ways of clustering procedures into surgery groups and the incorporation of emergencies, urgencies and no-shows into the simulation model will certainly be the object of our future research efforts.

9 Acknowledgements

The authors are grateful to Meyer University Children's Hospital, and in particular to General Director Dr. Tommaso Langiano, Medical Director Dr. Monica Frassineti and Bed Manager Dr. Mirco Gregorini for their constant support throughout the research project.

The authors would **also** like to express their sincere gratitude to the anonymous referees for their insightful critical comments and suggestions.

10 References

BATHLA, S., MOHTA, A., GUPTA, A. & KAMAL, G. (2010) Cancellation of elective cases in pediatric surgery: an audit. *J Indian Assoc Pediatr Surg,* **15,** 90-92.

BELIËN, J. & DEMEULEMEESTER, E. (2007) Building cyclic master surgery schedules with leveled resulting bed occupancy. *Eur. J. Oper. Res.,* **176,** 1185-1204.

BELIËN, J., DEMEULEMEESTER, E. & CARDOEN, B. (2009) A decision support system for cyclic master surgery scheduling with multiple objectives. *J. Sched.,* **12,** 147-161.

BELIËN, J., DEMEULEMEESTER, E. & CARDOEN, B. (2006) Visualizing the demand for various resources as a function of the master surgery schedule: a case study. *J. Med. Syst.,* **30,** 343-350.

BERTSIMAS, D. & SIM, M. (2004) The Price of Robustness. *Oper. Res.,* **52,** 35-53.

BLAKE, J. T., DEXTER, F. & DONALD, J. (2002) Operating room managers' use of integer programming for assigning block time to surgical groups: a case study. *Anesth. Analg.,* **94,** 143-148.

CARDOEN, B., DEMEULEMEESTER, E. & BELIËN, J. (2010) Operating room planning and scheduling: a literature review. *Eur. J. Oper. Res.,* **201,** 921-932.

CROWE, S., PAGEL, C., BULL, K., FENTON, M., VASILAKIS, C., GALLIVAN, S. & UTLEY, M. (2011) Using a mathematical model to assist with the management of paediatric heart transplant waiting lists: a case study. *IMA J. Manage. Math.,* **23,** 99-116.

DENTON, B., VIAPIANO, J. & VOGL, A. (2007) Optimization of surgery sequencing and scheduling decisions under uncertainty. *Health Care Manage. Sci.,* **10,** 13-24.

GLOUBERMAN, S. & MINTZBERG, H. (2001) Managing the care of health and the cure of disease-part I: differentiation. *Health Care Manage. R.,* **26,** 56-69.

GUERRIERO, F. & GUIDO, R. (2011) Operational research in the management of the operating theatre: a survey. *Health Care Manage. Sci.,* **14,** 89-114.

HANS, E., WULLINK, G., VAN HOUDENHOVEN, M. & KAZEMIER, G. (2008) Robust surgery loading. *Eur. J. Oper. Res.,* **185,** 1038-1050.

KELTON, W. D., SADOWSKI, R. P. & SADOWSKI, D. A. (2002) *Simulation with Arena*, McGraw-Hill.

MAY, J. H., SPANGLER, W. E., STRUM, D. P. & VARGAS, L. G. (2011) The surgical scheduling problem: current research and future opportunities. *Prod. Oper. Manage.,* **20,** 392-405.

MAY, J. H., STRUM, D. P. & VARGAS, L. G. (2000) Fitting the lognormal distribution to surgical procedure times. *Dec. Sc.,* **31,** 129-148.

MONTGOMERY, D. C. & RUNGER, G. C. (2002) *Applied Statistics and Probability for Engineers*, John Wiley & Sons.

SAID, K., PASCAL, A. & SONDES, C. (2006) Block scheduling: toward a master surgical schedule. *International Conference on Service Systems and Service Management, 2006.*

SANTIBÀÑEZ, P., BEGEN, M. & ATKINS, D. (2007) Surgical block scheduling in a system of hospitals: an application to resource and wait list management in a British Columbia health authority. *Health Care Manag Sci,* **10,** 269-282.

SOBOLEV, B. G., SANCHEZ, V. & VASILAKIS, C. (2011) Systematic review of the use of computer simulation modeling of patient flow in surgical care. *J Med Syst,* **35,** 1-16.

TAIT, A. R., VOEPEL-LEWIS, T., MUNRO, H. M., GUTSTEIN, H. B. & REYNOLDS, P. I. (1997) Cancellation of pediatric outpatient surgery: Economic and emotional implications for patients and their families. *J. Clin. Anesth.,* **9,** 213-219.

TÀNFANI, E. & TESTI, A. (2010) A pre-assignment heuristic algorithm for the master surgical schedule problem (MSSP). *Ann. Oper. Res.,* **178,** 105-119.

TESTI, A., TÀNFANI, E. & TORRE, G. (2007) A three-phase approach for operating theatre schedules. *Health Care Manag Sci,* **10,** 163-172.

UTLEY, M., CHAUSSALET, T. & BAKER, R. (2009) Editorial. *IMA J. Manage. Math.,* **20,** 323-325.

VAN OOSTRUM, J. M., BREDENHOFF, E. & HANS, E. W. (2010) Suitability and managerial implications of a master surgical scheduling approach. *Ann. Oper. Res.,* **178,** 91-104.

VAN OOSTRUM, J. M., VAN HOUDENHOVEN, M., HURINK, J., HANS, E., WULLINK, G. & KAZEMIER, G. (2008) A master surgical scheduling approach for cyclic scheduling in operating room departments. *OR Spectrum,* **30,** 355-374.

VANBERKEL, P. T. & BLAKE, J. T. (2007) A comprehensive simulation for wait time reduction and capacity planning applied in general surgery. *Health Care Manage. Sci.,* **10,** 373-385.

VISSERS, J. M. H., ADAN, I. J. B. F. & BEKKERS, J. A. (2005) Patient mix optimization in tactical cardiothoracic surgery planning: a case study. *IMA J. Manage. Math.*

ZHANG, B., MURALI, P., DESSOUKY, M. M. & BELSON, D. (2008) A mixed integer programming approach for allocating operating room capacity. *J. Oper. Res. Soc.,* **60,** 663-673.

Appendix – Article II

This is a pre-copy-editing, author-produced PDF of an article accepted for publication in INTERNATIONAL JOURNAL OF PRODUCTION ECONOMICS following peer review. The definitive publisher-authenticated version

CAPPANERA, P., VISINTIN, F. & BANDITORI, C. 2014. Comparing resource balancing criteria in master surgical scheduling: A combined optimisation-simulation approach. *International Journal of Production Economics*, 158, 179-196

is available online at

http://www.sciencedirect.com/science/article/pii/S0925527314002606

1 Introduction

The Operating Theatre (OT) is one of the most critical functional areas in a hospital. It drives almost 70% of the hospital's admissions and determines most of its costs (Denton et al., 2007). Improving the OT performance, thus, represents a strategic objective for a growing number of hospitals. In this regard, hospital managers have widely recognised that the performance of the OT largely depends on the way the surgical activities are scheduled (Litvak and Long, 2000, Guinet and Chaabane, 2003). This challenging topic has encouraged the development of a significant number of mathematical models that support the surgical planning and scheduling process (Cardoen et al., 2010, Guerriero and Guido, 2011, May et al., 2011, Dobrzykowski et al., 2013).

In the literature, such a process is considered to entail three stages i.e., *case mix planning*, *master surgical scheduling* (MSS) and *patients scheduling*, where the output of the upstream stage is the input of the downstream one (Beliën and Demeulemeester, 2007).

In the *case mix planning* stage, each specialty (e.g. urology, orthopaedic surgery, etc.) is assigned with a *total* OR time, which is usually expressed in terms of sessions per week/month. The *master surgical scheduling* stage, instead, consists in producing a timetable (the MSS) where a specialty is assigned to each OR session for each day of the planning horizon. Finally, in the *patients scheduling* stage, patients who have to undergo surgery are selected and sequenced within each session of the MSS.

This study focuses on the second stage, i.e. the MSS problem. Coherently with Banditori et al. (2013 and 2014), in this study we consider a situation where the *case mix planning*, has already been performed and we address the problem of determining: (i) the specialty (or specialties) to assign to each operating room (OR) and session of each day of the planning horizon; (ii) the number and type of surgeries that should be performed in each OR session (van Oostrum et al., 2008). Such a plan serves as an input for the *patient scheduling* stage. Solving a MSS problem has been proven to be extremely complex. Indeed, it requires taking into account many resources (ORs, post-surgical beds, surgical teams, ICU) and dealing with the randomness of surgical times (ST) and patients' length of stay (LoS) (Cardoen et al., 2010, Guerriero and Guido, 2011). In addition, it necessitates to take into consideration the conflicting priorities of different stakeholders, e.g. hospital managers, surgeons, nurses, patients (Glouberman and Mintzberg, 2001, Marcon et al., 2003).

In general, to fulfil the expectations of these stakeholders a MSS should be *efficient* (Cardoen et al., 2010, Guerriero and Guido, 2011), *balanced* (Litvak and Long, 2000) and *robust* (Banditori et al., 2013, 2014). In fact, it should allow for the increase of revenues and for the reduction of waiting times by maximising the number of patients scheduled (*efficiency*). In addition, it should determine a fair allocation of the workload among the people (doctors, nurses, etc.) working in the OT and in the post-surgical wards (*balancing*). Finally, it should prevent schedule disruptions, i.e. it should prevent OR overtime and/or bed overbooking that are usually caused by the variability of both ST and LoS (*robustness*).

This study is based on a combined optimisation-simulation approach and has a twofold aim:

i. compare three different scheduling policies and identify the one that under given operational conditions allows for the trade-off between efficiency, balancing and robustness best fitting the hospital priorities and its needs,

ii. explain *why* in certain conditions certain scheduling policies are superior to the others.

All the investigated policies aim to maximise the number of scheduled surgeries and to balance the utilisation of both post-surgical beds (hereafter beds) and ORs. However, these policies adopt different balancing criteria. The first policy (hereafter referred to as minMax) minimises the maximum daily utilisation of beds and ORs. The second one (hereafter referred to as minRng), instead, minimises the range between the maximum and minimum utilisation of these resources. Finally, the third policy (hereafter referred to as minOvrn), minimises the *overrun*, i.e. the positive deviation between the actual resource utilisations and target utilisation values.

In this study, we develop a mixed-integer programming (MIP) model, which is based on the models presented in Banditori et al. (2013 and 2014) and compare three alternative objective functions.

Each objective function corresponds to one of the aforementioned scheduling policies. The model variables and constraints do not vary across policies. We assume that the cases in a hospital’s waiting list can be classified into homogeneous surgery groups that are based on the resources (e.g. ORs, beds) that they are expected to require. Hence, the model produces a solution (the MSS) indicating the number of cases to treat and the surgery group these cases must belong to for each day of the planning horizon, for each OR, and for each session of the day. Such a solution also has to satisfy Quality of Service (QoS) requisites, i.e. it should allow the desired case-mix and the desired level of OR utilisation to be obtained.

The MIP model assumes deterministic values for the parameters ST and LoS. Thus, to assess the impact that the variability of these parameters has on the MSS robustness, we use a discrete-event simulation model. Such a model samples the values of ST and LoS from suitable probability distributions. By combining optimisation and simulation, we are able to calculate the overtime and the overbooking that would emerge as a consequence of the implementation of a given MSS.

The underlying conjecture of this study is that, in general, if the daily utilisation profiles of ORs and beds are nicely balanced, there should be some idle resources to absorb the unexpected peaks caused by ST and LoS variability (Beliën et al., 2009). In other terms, a higher balancing should lead to a higher robustness, especially when average resource utilisation is high. However, resource balancing can be achieved by using different scheduling policies, where each policy allows for the scheduling of a different number of surgeries (efficiency). In this study, the trade-off between efficiency, balancing and robustness is empirically investigated.

The main contribution of this work is to offer fresh insights into the relationship between efficiency, balancing and robustness in the surgical scheduling field, and to provide a thorough assessment of the pros and cons associated with the utilisation of three alternative scheduling policies. This work is based on real data from the Meyer University Children's Hospital (hereinafter Meyer Hospital) a leading Italian hospital. Starting from this data, we create 26 additional "realistic" hospital settings, thus to compare the scheduling policies in different scenarios. Moreover, to increase the external validity of our findings, the schedules produced in the optimisation phase have been simulated using both empirical distributions and theoretical (lognormal) distributions, for both ST and LoS.

The major findings of our work are that:

(i) a scheduling policy that allows achieving for a given hospital setting, superior performances in terms of efficiency *and* balancing *and* robustness, does not exist;

(ii) in general, when the focus is on efficiency, the best policy is the one that minimises the resources utilisation range (minRng). This policy allows for the containment of the overtime and for a good balancing of both beds and ORs. On the contrary, when the focus is on how to avoid overbooking, other policies (minMax, minOvrn) should be preferred.

(iii) these results are consistent across different distributional models

Another important contribution of this study is to explain the causal mechanisms that make some scheduling policies outperform the others.

The empirical results of this work are organised in tables (Table 5 to Table 12) that can help managers in choosing the scheduling policies that best fit their own hospital settings and priorities.

The remainder of the paper is organised as follows: in Section 2, we provide a review of the literature. In Section 3, we present the characteristics of the addressed MSS problem. In Section 4, we describe the optimisation and simulation models. In Section 5, we illustrate the experimental campaign we have carried out, whose results are presented in Section 6. Subsequently, in Section 7, we draw the conclusions and outline the direction of our future research efforts.

2 Literature review

Balancing/levelling issues emerge from different fields of application, i.e. machine scheduling (Sen et al., 1995, Caramia and Dell'Olmo, 2003), crew scheduling (Cappanera and Scutellà, 2011), project scheduling (Neumann and Zimmerman, 1999), surgical scheduling (Banditori et al 2014) and have been the object of a large number of contributions.

In this review, we primarily focus on works studying the workload balancing problem in the MSS context, i.e. the problem of equally distributing a certain workload among a given set of resources (e.g. beds, ORs). The papers reviewed here are analysed according to the following seven dimensions: (i) *balancing criteria*, i.e. the criterion adopted to balance resource utilisation; (ii) *balanced resources*, i.e. the resources whose utilisation is balanced; (iii) *solution technique*, i.e. the typologies of model/s adopted to solve the problem addressed; (iv) *type of analysis*, i.e. the approach followed to solve the problem; (v) *uncertainty*, that indicates if the parameters used in the model/s are deterministic or stochastic and, in this latter case, if the effect of randomness is assessed ex-post via simulation; (vi) *types of distributions*, i.e. empirical, theoretical or both, used to model the stochasticity of ST and/or LoS; (vii) *investigated setting*, i.e. the number and the type (real and/or realistic) of hospital settings where the proposed models are tested, and the number of dimensions (experimental factors) used to differentiate the settings from each other. Dimensions (iii) and (iv) are taken as is from the review scheme given by Cardoen et al. (2010). Dimensions (v) and (vii) has been adapted by adding some details. Dimensions (i), (ii) and (vi) have been developed ex-novo. The review is organised in tabular form and presented in Table 1. Each column of the table represents one dimension, while each row represents a paper. In order to emphasise the differences between our study and the related literature we have added one row representing our study.

As it can be observed in Table 1, the *minimisation of the maximum utilisation* is the most common balancing criterion. It can involve one or more resources, thus respectively entailing the minimisation of a single maximum value of daily utilisation or the sum of the maximum daily utilisation values.

Referring to the balanced resources, beds (belonging to a single or multiple wards/hospitals) are considered in all of the examined papers. In these papers, the major aim of this balancing is to reduce the bed utilisation variability thus to prevent schedule disruptions and patient cancellations. In addition, some authors also consider other resources (e.g. IC beds, IC nurses). In the literature, OR balancing is only addressed by Adan et al. (2009) and Banditori et al. (2014). Most of the examined papers deal with the LoS randomness. Instead, ST randomness is only considered by van Oostrum et al. (2008) and Banditori et al (2014).

Table 1- The MSS balancing literature review

	Balancing criteria	Balanced resources	Solution technique	Type of analysis	Uncertainty	Type of distr.	Investigated settings
Our study	Minimisation of the maximum daily utilisation Minimisation of the difference between the maximum and the minimum daily utilisations Minimisation of the sum of the quadratic overrun	ORs Beds of a single ward	Mixed integer programming Quadratic programming Discrete-event simulation	Single criterion exact optimisation Scenario analysis	Stochastic ST (ex-post) Stochastic LoS (ex-post)	Empirical Theoretical	1 real setting 26 realistic settings 3 experimental factors (Beds/ORs ratio, OR utilisation rate and case MIX)
Santibáñez et al. (2007)	Minimisation of the sum of the maximum daily utilisations	Beds of different hospitals	Mixed integer programming	Single criterion exact optimisation	Deterministic ST Deterministic LoS	None	1 real setting
van Oostrum et al. (2008)	Minimisation of the maximum daily utilisation	Beds of different wards	Mixed integer programming Column generation Decomposition approach	Multi-criteria exact/heuristic optimisation	Stochastic ST Deterministic LoS	Empirical	1 real setting 35 realistic settings 3 experimental factors (Planning horizon, N° of ORs. N° of bed types)
Adan et al. (2009)	Minimisation of the deviation from a target utilisation	ORs Medium care beds IC beds IC nurses	Mixed integer programming	Single criterion exact optimisation	Deterministic ST Deterministic IC nursing load Stochastic LoS	Empirical Theoretical	1 real setting
Beliën et al. (2009)	Minimisation of the weighted sum of the quadratic mean and variance of the utilisations	Beds of different wards	Mixed integer programming Quadratic programming Goal programming Simulated annealing	Multi-criteria exact/heuristic optimisation	Stochastic Los	Empirical	1 real setting 1 realistic setting 1 experimental factor (Planning horizon)
Chow et al. (2011)	Minimisation of the sum of the maximum daily utilisations	Beds of different wards	Mixed integer programming	Single criterion exact optimisation Scenario analysis	Deterministic ST Stochastic LoS (ex-post)	Empirical	1 real setting
Carter and Ketabi (2012)	Minimisation of the sum of the maximum daily utilisations	Beds of different wards	Integer programming	Single criterion exact optimisation	Deterministic ST Stochastic LoS	Theoretical	1 real setting
Banditori et al. (2014)	Minimisation of the maximum daily utilisation Minimisation of the difference between the maximum and the minimum daily utilisations	ORs Beds of a single ward	Mixed integer programming Discrete-event simulation	Single criterion exact optimisation Scenario analysis	Stochastic ST (ex-post) Stochastic LoS (ex-post)	Empirical	1 real setting 4 realistic setting 1 experimental factor (OR utilisation rate)

It is worth pointing out, however, that the papers in Table 1 are not the only works addressing the robustness issues in the surgical scheduling field. There are, indeed, studies that address these issues, also considering different sources of randomness. However, they do not investigate the relationship between robustness and resources balancing. For instance, Mannino et al. (2012) propose a pattern based MIP model for the MSS and apply a light robustness approach (Fischetti and Monaci, 2009) to cope with the uncertainty associated with the surgery demand. Hans et al. (2008) instead address the patient scheduling problem by proposing different heuristics, in which one of the objectives is to minimise the risk of overtime. They consider stochastic STs and exploit the portfolio effect, thus to minimise the required OR slacks. Resource slacks are also used in Banditori et al. (2013) both for beds and ORs. There an optimisation-simulation approach is used to determine the resource slacks that best fit the hospital needs. A combined optimisation-simulation approach is also used by Lamiri et al. (2009), who combine Monte Carlo simulation and mixed integer programming to address problems where OR capacity is shared by elective and emergency patients, and by Zhang et al. (2008), who propose a MIP model and then test the robustness of the model' solutions against the randomness of surgery demand, via simulation. Choi and Wilhelm (2014), instead, solve a block surgical schedule problem following a newsvendor approach. Specifically they assume normally distributed ST and determine the duration and the sequence of the OR time blocks in order to minimise the costs associated with the expected early or late completion of the OR activities.

In sum, within the MSS literature, only the study of Banditori et al. (2014) explores how to obtain MSSs that are robust against *both* ST and LoS variability, by balancing *both* beds and ORs. However, such a study presents a number of shortcomings. First and foremost, it compares two balancing criteria, namely minMax and minRng, but does not explain *why* in certain conditions and for certain performances one criterion performs better than the others. Second, the study's computational campaign includes only a limited number of very similar hospital settings and it is based on empirical distributions only. These facts clearly hamper the external validity of the study findings that are, indeed, very context-specific. Third, the study of Banditori et al. (2014) does not consider a fairly well known balancing criterion, i.e. the minOvrn one (Sen et al., 1995). Our study overcomes all these shortcomings. Indeed, we compare three balancing criteria, through an extensive computational campaign that combines 27 hospital settings and both empirical and theoretical distributions. Such experimental campaign allowed us to explain why some criteria outmatch the others and to obtain results that are generalizable to a wide set of hospital settings.

In the following section, we will present the models used to make these comparisons and to address the aforementioned literature gaps.

3 Problem addressed

In this study, we consider a planning horizon expressed in days and three critical resources: surgeons, ORs and beds. We assume that the hospital can always rely on a sufficient number of OR nurses and anaesthetists, thus, these resources are not included in the model. Consistently with Banditori et al. (2013 and 2014), we organise the elective cases in the waiting list in specialties and within the same specialty, in surgery groups, including surgeries with similar ST and LoS (see Section 5.2). Each specialty is assigned with a certain number of time slots per week, which depends on the number of surgeons the specialty relies on. We hypothesise that once a specialty is assigned to a certain time slot it will always be able to deploy a surgeon team suitable to execute the surgeries scheduled in that time slot. ORs are characterised by an available time, which is subdivided into a set of time slots. Beds accommodate patients after the surgery. Based on its surgery group, a case occupies one OR for a time equal to ST and one bed for a time equal to LoS.

The problem we address consists in determining, for each OR, for each time slot and for each day of the planning horizon:

(i) the specialty to schedule, and
(ii) the number of surgeries belonging to each surgery group that should be performed

with the aim of maximising the number of surgeries scheduled and balancing the beds and OR daily workloads.

In addition, the solution must also comply with some of the hospital's management requirements (QoS requisites). These requirements pertain to the case-mix and the OR target utilisation. The former specify that the mix of surgeries in the MSS has to reflect on the mix of the surgeries on the waiting list. To this aim, the cases in the waiting list are organised into classes according to their LoS and ST. Specifically, with regard to the LoS, the cases are subdivided in two classes: the ones with a short LoS (*short-stay surgeries*) and the ones with a long LoS (*long-stay surgery*); on the other side, with regard to the ST, the cases are subdivided into the following two classes: short lasting surgeries and long lasting surgeries. Then, the case mix constraints specify that for each class, the scheduled surgeries in that class fall within a minimum and maximum percentage of the overall scheduled surgeries; the range is consistently defined with the dimension of the class in the waiting list. The latter QoS requisite sets a range, where the average OR utilisation should fall. By fixing the case mix, on the one hand, it is possible to avoid leaving an excessive number of complex cases (i.e. with long LoS and/or ST) on the waiting list, which would make the scheduling process more complex in the following periods. On the other hand, it is possible to avoid hospitalising an excessive number of complex cases at the same time, thereby reducing the clinical risk (Vincent et al., 1998). Similarly, by setting a range for the OR utilisation, the management ensures that the solution complies with the given efficiency targets (lower bound) and, at the same time, it avoids an excessive OR utilisation (upper bound), which could result in an excessive OT personnel workload.

The optimisation and simulation models we developed in this study are thoroughly described in the next section.

4 Models

4.1 Optimisation model

Let us define the following sets and parameters:

D the set of days of the planning horizon, indexed by d

$\tilde{D}$ the set of days in D in which the ORs are open

T the set of time slots, indexed by t

O the set of ORs, indexed by o

S the set of surgical specialties, indexed by s

K the set of surgery groups, indexed by k

G the set of the ST classes, indexed by g (short-lasting vs. long-lasting surgeries)

J the set of the LoS classes, indexed by j (short-stay vs. long-stay surgeries)

H_{odt} the available time of OR o, on day d and time slot t

F_d the number of beds available on day d

s_k the specialty of surgery group k

g_k the ST class that the surgery group k belongs to

j_k the LoS class that the surgery group k belongs to

c_k the average surgery duration of surgery group k

$\overline{\overline{G}}_g, \underline{\underline{G}}_g$ the maximum and the minimum percentage of schedulable surgeries belonging to the *g-th* ST class

$\overline{\overline{J}}_j, \underline{\underline{J}}_j$ the maximum and the minimum percentage of schedulable surgeries belonging to the *j-th* LoS class

$\overline{\overline{U}}, \underline{\underline{U}}$ the upper and the lower threshold on the total OR utilisation

W_1, W_2, W_3 the weights used in the objective functions.

The problem mathematical formulation involves two families of main variables each of them related to a specific kind of decision: the binary *assignment* variables x and the non-negative integer *scheduling* variables y. The former define which specialty is assigned to each OR in each day and in each time slot of the planning horizon. The latter define the number of surgeries scheduled in each time slot for each surgery group. Specifically,

x_{sodt} binary, 1 if specialty s is assigned to OR o on day d and time slot t, 0 otherwise

y_{kodt} the number of surgeries in surgery group k assigned to OR o on day d in time slot t.

Hereafter we discuss the feasibility set these variables belong to, splitting it in two blocks: the first block refers to constraints that are quite common in any hospital setting and that have already been widely discussed and used in the literature. The second block of constraints, instead refers to *quality of service* (QoS) constraints which reflect the peculiarities of the specific settings addressed, though quite widespread, and to the definition of balancing criteria.

The feasibility constraints belonging to the first block are described informally without giving their mathematical formulation. The interested reader is referred to Banditori et al. (2014) for a detailed description. They include constraints that control the following issues: (i) at most one surgical specialty can be assigned to a given OR in each time slot of the planning horizon; (ii) in each time slot, a given specialty cannot occupy more than one OR; (iii) the correct binding of the assignment variables *x* and scheduling variables *y*: specifically, these constraints guarantee that no surgery of a given specialty is scheduled in a given OR and time slot, unless that specialty has been assigned to that OR in that time slot; (iv) the total time consumed by all the surgeries scheduled in a given OR, in each time slot, cannot exceed the available OR time; (v) the computation of the number of beds occupied in each day properly keeping into account the average number of days of hospitalisation, before and after surgeries; and (vi) the surgeons availability for each week, i.e. the number of slots assigned to a given specialty in a given week cannot exceed the number of slots that such a specialty can cover with the surgeons available. In the following the feasibility set defined by the constraints (i) to (vi) is referred to as set *E*.

Conversely, the second block of feasibility constraints is defined in a more formal way. In addition to the assignment (*x*) and scheduling variables (*y*), let us define the following auxiliary variables:

z_d the number of beds occupied on day *d*

u_{odt} the utilisation rate of OR *o*, on the day *d* and time slot *t*

v_d the utilisation rate of beds on day *d*.

Using these variables and the parameters listed at the beginning of this section, we can complete the definition of the feasibility set as follows:

$$z_d \leq F_d \quad \forall d \in D \tag{4.1.1}$$

$$\underline{\underline{G}}_g \sum_{\substack{k\in K, o\in O \\ d\in D, t\in T}} y_{kodt} \leq \sum_{\substack{k\in K: g_k=g \\ o\in O, d\in D, t\in T}} y_{kodt} \leq \overline{\overline{G}}_g \sum_{\substack{k\in K, o\in O \\ d\in D, t\in T}} y_{kodt} \qquad \forall g \in G \tag{4.1.2}$$

$$\underline{\underline{J}}_j \sum_{\substack{k\in K, o\in O \\ d\in D, t\in T}} y_{kodt} \leq \sum_{\substack{k\in K: j_k=j \\ o\in O, d\in D, t\in T}} y_{kodt} \leq \overline{\overline{J}}_j \sum_{\substack{k\in K, o\in O \\ d\in D, t\in T}} y_{kodt} \qquad \forall j \in J \tag{4.1.3}$$

$$u_{odt} = \frac{\sum_{k \in K} c_k y_{kodt}}{H_{odt}} \qquad \forall o \in O, \forall d \in \tilde{D}, \forall t \in T \tag{4.1.4}$$

$$v_d = \frac{z_d}{F_d} \qquad \forall d \in \tilde{D} \tag{4.1.5}$$

$$\underline{\underline{U}} \leq \frac{\sum_{\substack{o \in O, d \in \tilde{D}, \\ t \in T}} u_{odt}}{|O||\tilde{D}||T|} \leq \overline{\overline{U}} \tag{4.1.6}$$

$$x_{sodt} \in E \qquad \forall s \in S, \forall o \in O, \forall d \in D, \forall t \in T \tag{4.1.7}$$

$$y_{kodt} \in E \qquad \forall k \in K, \forall o \in O, \forall d \in D, \forall t \in T \tag{4.1.8}$$

A brief description of the constraints follows. Constraints (4.1.1) guarantee that the number of beds occupied in each day of the planning horizon does not exceed the bed availability. Constraints (4.1.2) and (4.1.3) are the case mix constraints. The first control the composition of the mix of surgeries in terms of ST; for each ST class, they state that the number of scheduled surgeries in the class falls inside a pre-defined range. The second are the counterparts for the LoS classes and assure that the number of scheduled surgeries belonging to each LoS class falls in a specified range. Constraints (4.1.4) and (4.1.5) respectively compute the daily utilisation rate of OR time slots and beds. These auxiliary variables are only inserted for matter of clarity. Constraint (4.1.6) pertains to the OR utilisation target and imposes that the average OR utilisation falls between the pre-defined lower and upper bounds, $\underline{\underline{U}}$ and $\overline{\overline{U}}$. Finally, constraints (4.1.7) and (4.1.8) assure that the assignment and scheduling variables satisfy also the feasibility set denoted by E.

Three alternative objective functions are considered, and each of them implements a different scheduling policy, as discussed in the introduction. All the objective functions are composed of three terms, whose relative importance is given by the weights W_1, W_2, W_3. The first and the second term of the three objective functions are the balancing terms. The former acts on OR utilisations, while the latter acts on the beds' ones. Finally, the third term of the objective functions maximises the number of scheduled surgeries. For each objective function, specific variables and constraints are defined.

The first objective function (4.1.9), referred to as minMax, minimises the maximum ORs ($\bar{\bar{u}}$) and beds ($\bar{\bar{v}}$) daily utilisations, i.e.:

$$\min \quad W_1\bar{\bar{u}} + W_2\bar{\bar{v}} - W_3 \sum_{\substack{k\in K, o\in O \\ d\in D, t\in T}} y_{kodt} \tag{4.1.9}$$

$$u_{odt} \le \bar{\bar{u}} \qquad \forall o \in O, \forall d \in \tilde{D}, \forall t \in T \tag{4.1.10}$$

$$v_d \le \bar{\bar{v}} \qquad \forall d \in \tilde{D} \tag{4.1.11}$$

Constraints (4.1.10) and (4.1.11), used in combination with the objective function (4.1.9), respectively assure that daily utilisation of ORs and beds does not exceed the corresponding maximum daily utilisation.

The second objective function (4.1.12), referred to as minRng, minimises the gaps between the maximum and the minimum values of ORs and beds' daily utilisations: as before, $\bar{\bar{u}}$ and $\bar{\bar{v}}$ represent the maximum daily utilisations of ORs and beds, whereas $\underline{\underline{u}}$ and $\underline{\underline{v}}$ represent the minimum values of such utilisations.

$$\min \quad W_1(\bar{\bar{u}} - \underline{\underline{u}}) + W_2(\bar{\bar{v}} - \underline{\underline{v}}) - W_3 \sum_{\substack{k\in K, o\in O \\ d\in D, t\in T}} y_{kodt} \tag{4.1.12}$$

$$\underline{\underline{u}} \le u_{odt} \le \bar{\bar{u}} \qquad \forall o \in O, \forall d \in \tilde{D}, \forall t \in T \tag{4.1.13}$$

$$\underline{\underline{v}} \le v_d \le \bar{\bar{v}} \qquad \forall d \in \tilde{D} \tag{4.1.14}$$

The third objective function (4.1.15), referred to as minOvrn, minimises the sum of the quadratic positive deviations (overrun) of the ORs and the beds daily utilisations from a fixed threshold. Specifically, $\hat{u}_{odt}$ represents, for the OR *o*, on the day *d* and time slot *t*, the positive deviation of the total operating time scheduled from the fixed percentage threshold *U* of the available time for that OR, day, and time slot (see constraints (4.1.16)); on the other hand, $\hat{v}_d$ represents, for each day *d*, the positive deviation of the number of occupied beds from the fixed percentage threshold *V* of the number of available beds for that day (see constraints (4.1.17)). In order to penalise the bigger deviations more than the smaller ones, the objective function is quadratic. As a consequence, here the resource utilisations (and the introduced overrun variables) are expressed in terms of absolute values instead of relative values ranging from 0 to 1. Constraints (4.1.18) and (4.1.19) define the non-negativity of the variables involved.

$$\min \quad W_1 \sum_{\substack{o \in O, d \in \tilde{D} \\ t \in T}} \hat{u}_{odt}^2 + W_2 \sum_{d \in D} \hat{v}_d^2 - W_3 \sum_{\substack{k \in K, o \in O \\ d \in D, t \in T}} y_{kodt} \tag{4.1.15}$$

$$\hat{u}_{odt} \geq \sum_{k \in K} c_k y_{kodt} - UH_{odt} \qquad \forall o \in O, \forall d \in \tilde{D}, \forall t \in T \tag{4.1.16}$$

$$\hat{v}_d \geq z_d - VF_d \qquad \forall d \in \tilde{D} \tag{4.1.17}$$

$$\hat{u}_{odt} \geq 0 \qquad \forall o \in O, \forall d \in \tilde{D}, \forall t \in T \tag{4.1.18}$$

$$\hat{v}_d \geq 0 \qquad \forall d \in \tilde{D} \tag{4.1.19}$$

4.2 Simulation model

The simulation model used in this study works as follows. The model reads the schedule produced in the optimisation phase, generates a number of entities equal to the number of surgeries planned for the planning horizon and assigns a surgery group to each entity. Hence, for each simulated day, the model creates a number of entities equal to those planned for the day. These entities seize the ORs and the beds they are assigned to and release them after a time that is randomly sampled from a suitable probability distribution (a thorough discussion of the probability distributions we use in the model is presented in Section 5.2). The model, thus, keeps track of the actual duration of the surgical sessions and of the number of beds that would actually be needed to accommodate the scheduled patient. If the duration of a session exceeds the OR time allotted to the session itself, then the model registers the number of overtime minutes worked. Similarly, if the number of beds occupied on a given day exceeds the number of beds that are actually available, the model then keeps track of the overbooking.

5 Experimental campaign

5.1 Input data

As we pointed out in the introduction, our experimental campaign was inspired by Meyer Hospital. Such a hospital is characterised by:

- *12 surgical specialties*. Each surgical specialty is associated with surgeon teams that can cover a certain number of time slots per week.

- *39 surgery groups*. Each case in the Meyer Hospital waiting list is characterised by a surgery group. Surgery groups are assigned by surgeons when they prescribe a surgery and indicate the specialty (e.g. urology), an estimate of the surgery duration and an estimate of the LoS. More specifically, regarding to the ST, cases are labelled as *type A* (0<ST≤60 minutes), *type B* (60<ST≤120 minutes) and *type C* (ST>120 minutes). Instead, regarding the LoS, cases are labelled according to the days the patient is expected to occupy a bed (1,2,3,… days). For example, a group labelled *ORL-A-2* includes cases that are expected to require an *othorinolaringoiatry* surgeon team, occupy the OR for less than one hour and give rise to a LoS of two days.
- *A planning horizon of 2 weeks.*
- *47 beds and 4 ORs* dedicated to elective patients. Each OR is open 11.5 hours a day, 5 days per week. This leads to a *Beds/OR Ratio* being equal to 1.02 [beds/hour]. Such a ratio is calculated by dividing the number of beds for the daily OR available time, this latter being calculated as the product of the number of ORs and the OR daily available time, i.e. *Beds/OR time Ratio* =47/(4x11.5) =1.02. Additional OR time-slots and beds are allocated to non-elective patients (emergencies and urgencies).
- A target *case-mix* composed of
    - 40% of *short-stay surgeries* (SLoS, LoS<2 days) and 60% of *long-stay surgeries* (LLoS, LoS≥2 days),
    - 64% of *short-lasting surgeries* (SST, *type* A, ST≤60 minutes) and 36% of and *long-lasting surgeries* (LST, *type* B and C, ST>60 minutes).
- A target OR utilisation range equal to 80-85%.
- A strong focus on preventing bed shortage.

5.2 Data analysis

To calculate the values of ST and LoS to use in the optimisation and simulation models, we analysed two years of surgical records. Each record corresponded to a case and indicated:

- the surgery group the case was assigned to when surgeon prescribed the surgery,
- the actual duration of the surgery (ST) and the actual patient's LoS.

For each surgery group, we calculated the descriptive statistics of ST and LoS. The *mean values* of ST and LoS were used to run the optimisation model. To perform the simulation analysis, instead, we needed to identify, for each surgery group, suitable probability distributions. To this aim, we first extracted the empirical distributions associated with ST and LoS. Using empirical distributions, instead of theoretical ones, carries two main advantages (Law and Kelton, 2000, pag. 296): (i) it allows avoiding the occurrence of fairly large (or small) values that might not practically occur in realty; (ii) it allows for better capturing the characteristics of the data when theoretical distributions display poor fit. On the other hand, the aim of our simulation model is not to reproduce very accurately a specific hospital setting, rather, its aim is to derive conclusions that can be extended to a variety of hospitals. To this purpose, the utilisation of empirical distributions represents a limitation. To extend the generalizability of the results it is advisable to use theoretical distributions (Law and Kelton, 2000, p. 296). In fact, it is reasonable to assume that ST and LoS may follow similar distributions across a variety of hospitals, even though the distributions' parameters may change from one hospital to another. Since our data are positively skewed and non-negative, we decided to fit several non-negative continuous theoretical distributions, namely, lognormal, 3-parameters lognormal, gamma and Weibull.

More specifically, for each surgery group, for each variable (ST, LoS) and for each distribution, we carried out an Anderson Darling (AD) goodness-of-fit test and tested the null hypothesis of the data being distributed according to the investigated distribution. Looking at the test's statistics and p-values and by visually inspecting the relevant probability plots, we found that the lognormal models fitted our data better than the other ones, both for ST and LOS. Such a finding is consistent with the literature. There is, in fact, consensus that lognormal models are suitable to represent both ST (May et al., 2000, Stepaniak et al., 2009) and LoS (Marazzi et al., 1998, Carter and Ketabi, 2012).

However, even if the lognormal distributions fitted better than other distributions for certain surgery groups, they do not showed, in absolute terms, a good fit, especially for LoS. In Table 2, we cluster the surgery groups according with the number of occurrences (No) in the data set (the dimensional classes are taken from (Stepaniak et al., 2009)). Hence, for each cluster we show the number of times we fail to reject the null hypotheses ($p>0.05$) of the data being lognormally distributed.

Table 2 –Results of the Anderson Darling (AD) goodness-of-fit test for ST (left side) and LoS (right side)

ST			
Number of occurrences	**Groups in the class**	**Groups where p>0.05**	**%**
No≥200	9	0	0%
30≤No<200	16	7	44%
No<30	13	12	92%
Total	38	19	50%

LoS			
Number of occurrences	**Groups in the class**	**Groups where p>0.05**	**%**
No≥200	9	0	0%
30≤No<200	16	2	13%
No<30	13	5	38%
Total	38	7	18%

Referring to the 2-parameters lognormal model, we rejected the null hypotheses 50% of the time for ST, and 82% of the time for LoS. It is necessary to point out, however, that when samples are large, e.g. No>200, the power of goodness-of-fit tests increases, the confidence intervals shrink and consequently it is very likely to reject the null hypotheses (p<0.05) even for small and practically not relevant deviations from the investigated distributions. For that reason, it is always necessary to plot the data in order to make an informed decision about the extent of the deviation (Field, 2005 p. 93). Indeed, in several cases even when the p-value was smaller than 0.05, the histogram and probability plots revealed a quite satisfactory fit. In Figure 1 there is the example of the surgery group URO-B-1 for which ST displayed a satisfactory fit but the AD test returned p<0.05.

Referring to the 3-parameters lognormal model, instead, there is no established method for calculating the p-value (D'Agostino and Stephens, 1986). Hence, we performed a Likelihood Ratio Test (LRT) and compared the 3-parameter model with its 2-parameter counterparts. The LRT is a statistical test of the goodness-of-fit between two models. In our case, a LRT's p-value (which is referred to as LRT-p) smaller than 0.05, implies that the 3-parameters distribution fits significantly better than the 2-parameter one.

Looking at the LRT p-values, we found that only for a limited number of surgery groups the 3-parameters lognormal distribution improved the fit (LRT-P<0.05, 11 times for ST and 7 times for LoS). In most of the other cases, we obtained LRT-p> 0.05 and lower values of the AD test statistic. In addition, also in those cases where LRT-p<0.05, the AD test statistic relevant to the 2- and 3-parameters distributions were rather similar. In these situations, it is advisable to use the distribution that has a calculated p-value, i.e. the 2-parameters lognormal one. In sum, combining a visual inspection of the histograms and probability plots with an analysis of p-values, LRT-p values and AD test statistics, we concluded that for both ST and LoS the theoretical distribution best fitting our data was the 2-parameters lognormal one.

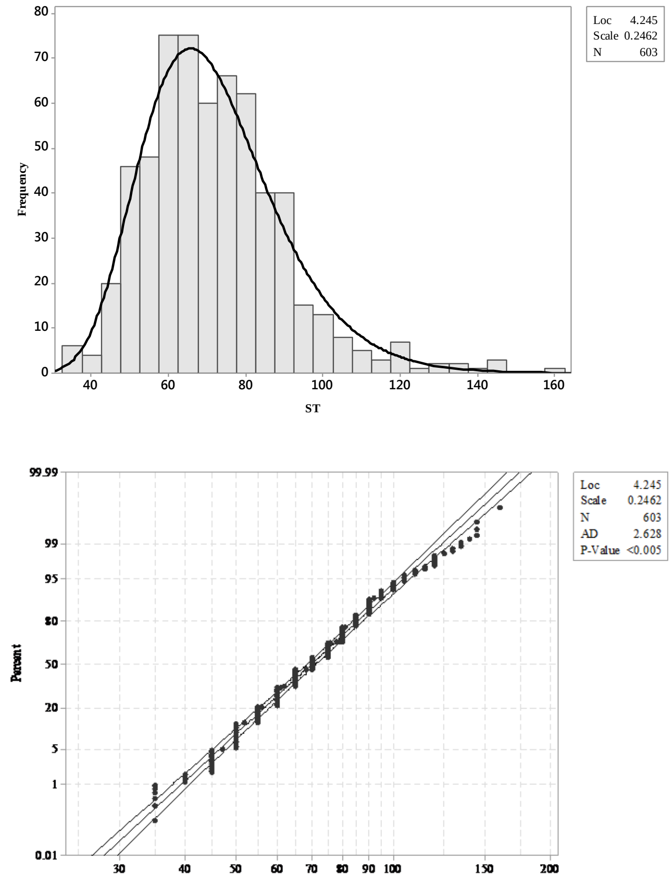


Figure 1 - Histogram with fitted lognormal distribution (top) and probability plot (bottom) of ST for surgery group URO-B-1

5.3 Scenario analysed

In this study, we investigated 27 hospital settings. For each setting, we run the optimisation model in correspondence with all the objective functions. For each obtained solution we run the simulation model using four different types of probability distributions. The hospital settings we investigated and the distributions we used are described in the next sections.

5.3.1 Hospital settings

In this study, we investigate 27 settings. These settings have been obtained by combining different values of the following parameters:

(i) case mix complexity (MIX);

(ii) hospital *Beds/OR time Ratio* (BOR);

(iii) target OR utilisation range (OUR).

Specifically, we started from the Meyer Hospital values of MIX, BOR and OUR, and we increased (decreased) them by a fixed percentage (25% for MIX, 10% for BOR, 5% for the OUR boundaries), thus, to obtain three different levels (low, medium, high) for each parameter. Parameters and levels are combined together, thus summing up to the 27 scenarios, the Meyer Hospital one plus 26 additional realistic hospital settings.

The different values of BOR were obtained by increasing (decreasing) the number of beds and by keeping the OR time constant. Table 3 shows the values associated with each level.

Table 3 – Hospital settings parameters

Level	MIX (X-axes)	BOR (Y-axes)	OUR (Z-axes)
Low	SLoS (LLoS)=30% (70%) SST (LST)=52%(48%)	0.9	$\underline{\underline{U}}=75\%$ $\overline{\overline{U}}=80\%$
Medium	SLoS (LLoS)=40% (60%) SST (LST)=64%(36%)	1.0	$\underline{\underline{U}}=80\%$ $\overline{\overline{U}}=85\%$
High	SLoS (LLoS)=50% (50%) SST (LST)=80%(20%)	1.1	$\underline{\underline{U}}=85\%$ $\overline{\overline{U}}=90\%$

In the additional 26 hospital settings, all the other model parameters are either constant and equal to their counterparts in the Meyer Hospital setting or depend on the parameters used to calculate MIX, BOR and OUR. More specifically, we used the same weights for all the objective functions. These weights do not vary across scenarios and are set so that $W_2 >> W_1 >> W_3$.

In fact, since we aimed to obtain balanced solutions, the balancing terms were prioritised with respect to the efficiency term. This latter term allows us to obtain, among the balanced solutions, the one characterised by the highest number of surgeries. Furthermore, since overbooking is considered more undesirable than overtime at Meyer Hospital, the beds' balancing is prioritised with respect to OR balancing. In addition, we linked the parameters ($\underline{\underline{G}}_g, \overline{\overline{G}}_g$) and ($\underline{\underline{J}}_j, \overline{\overline{J}}_j$) with the target MIX and we allowed a maximum deviation of 10%. Finally, for all the scenarios, we set the thresholds *U* and *V* equal to $\underline{\underline{U}}$.

Figure 2 graphically represents the hospital settings investigated in our experimental campaign. The grey dot identifies the Meyer Hospital setting.

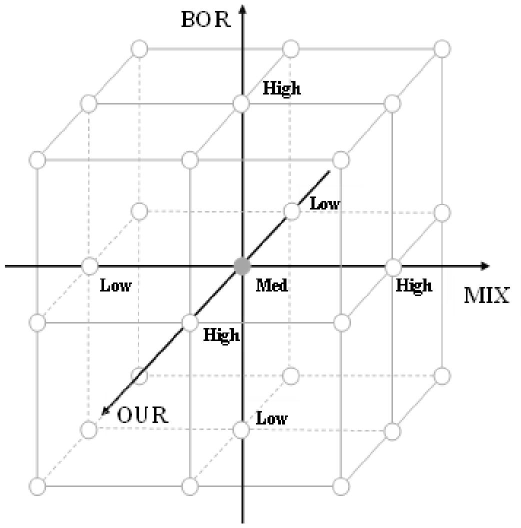


Figure 2–Investigated settings

The rationale inspiring the generation of these settings is threefold.

First, we wanted to use *easy-to-calculate* and *relative* parameters (indeed, MIX, BOR and OUR are percentages). By doing so, it is possible to obtain study's findings that are less dependent from the absolute hospital dimensions (in terms of beds and OR available time) and thus easier to extend to many different hospitals settings.

Second, we were interested in combining different levels of MIX and BOR. As Bowers (2013) argues, in fact, the benefits of having more beds given a certain number of ORs - or vice-versa - is "greater if the resources are reasonably well matched, relative to the mean theatre and bed requirements per patient" i.e. these benefits are greater for some coherent combination of MIX and BOR.

Third, we wanted to assess whether the ranking of the different scheduling policies, for any given combination of MIX and BOR, remains the same across different OUR.

In addition to test our scheduling policies in different settings, we wanted also to obtain findings tenable under different, yet realistic, distributional assumptions. To this aim, we simulated each optimisation model's solution using four types of distributions. These distributions are described in the next section.

5.3.2 Distributions

The simulation analysis was conducted using, for each surgery group, and for both ST and LoS, four types of distributions. The basic idea was to verify whether the relative performance of the three scheduling policies, in terms of overbooking and overtime, are somehow influenced by the distributions adopted to our data. To this aim, we considered, for each surgery group, and for both ST and LoS, their *empirical distributions* (D1), the *best fitting (lognormal) distributions* (D2), as well as "extreme" examples of artificially created lognormal distributions (D3, D4). These latter distributions are characterised by the same expected value of D2 but, respectively, by an extremely *skewed and platykurtic* shape (D3) and by an extremely *symmetric and leptokurtic* shape (D4).

To define the parameters of D3 and D4, we proceeded as follows. First, for each surgery group we used the parameters of the fitted lognormal distributions (D2) to calculate the expected values of both ST and LoS. If a variable X is lognormally distributed its probability density function is:

$$f(x:\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}} \qquad (5.3.2.1)$$

and its expected value can be calculated as:

$$E[X] = e^{\mu+\frac{1}{2}\sigma^2} \qquad (5.3.2.2)$$

Second, for each surgery group we set the values of $E[ST], E[LoS]$ equal to the ones relevant to D2, we fixed the values of the scale parameter (σ) and, finally, we used equation (5.3.2.2) again to calculate the location parameter μ ($\mu=ln(E[x])-0.5\sigma^2$). For all the surgery groups, the values of σ used in D3 and D4 are the reported in Table 4.

Table 4 - Scale parameters of the lognormal distributions D3 and D4

	D3	D4
σ(ST)	0.5	0.1
σ(LoS)	1.5	0.1

The values of σ used in D3 and D4 are extreme yet reasonable. In fact they represent, respectively, the upper and lower bound of σ(ST) and σ(LoS) across surgery groups in our data-set. As an example in Figure 3, we show the shapes and the parameters of the distributions D2, D3, and D4, relevant to the surgery group URO-B-1 (the shape of D1 is represented in the histogram in Figure 1). For all the distributions E[ST]=71.94 minutes.

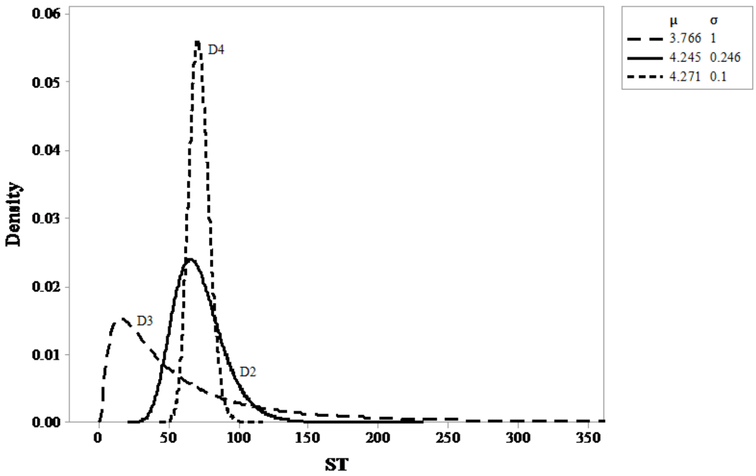


Figure 3 Lognormal probability density functions for ST of surgery group URO-B-1 (D2-continuous line, D3-dashed line, D4-dotted line)

5.4 Analysis performed

For each setting, we instantiated the optimisation model using the mean value of ST and LoS, and tested the three scheduling policies described in Section 4. We obtained 81 (=3x27) solutions. For each solution, we calculated the *number of scheduled surgeries* (N) and the *mean* (M), the *standard deviation* (Sd), the *maximum* (Max), the *range* (Rng) and the *sum of the quadratic overrun* (Ovrn), of both the beds and OR daily utilisations. Each solution was subsequently simulated performing 30 simulation runs and using the empirical distributions of ST and LoS. For each solution we calculated the *mean* (M), *standard error of the mean* (SE) and the *maximum* (Max) across the replications of the overtime (OVT) and overbooking (OVB). To test whether there was a significant effect of the scheduling policy on the M(OVB), for each setting, we carried out a one-way ANOVA using the scheduling policies as factors. The ANOVA revealed a significant effect ($p<0.05$) for all the settings for which we found a feasible solution. Secondly, for each setting, we carried out a Games-Howell post-hoc test, to compare all policies with each other, rank them and control the family error rate (Field, 2005, p.310) to a 0.05 level. For each pairwise comparison, we assigned the same rank to those policies for which the post-hoc test did not allow for the identification of a significant ($p>0.05$) difference between the relevant M(OVB). The same procedure was applied for M(OVT). The result of these tests, and thus the policy rankings are presented in Table 8 and Table 9.

Subsequently, the simulation experiments and the follow-up analysis were replicated using the remaining three distributions (D2, D3, D4). As we will discuss in detail in Section 6.2.4, when distributions are positively skewed and platykurtic it is more likely to obtain very high values of ST and LoS that could lead to unacceptable values of OVT and OVB. In this situation, even well balanced solutions might not be robust. In this cases, a possible way to achieve robustness - at the expenses of the overall efficiency- is to use values greater than the mean to run the optimisation model. To explore this alternative way to obtain robustness, we have replicated the whole analysis, using, in the optimisation phase, the third quartiles of ST and LoS.

It is worth mentioning that ANOVA is a parametric test and, as such, it requires the data to meet the hypotheses of independence, normality and homoscedasticity. We checked these assumptions, but unfortunately, for some settings, we found out that our data violated the assumption of homoscedasticity. In fact, in some cases the Levene's test of homogeneity of variance revealed that it was not possible to reject the test's null hypothesis that the sample variances were unequal, at a 0.05 level of significance. However, as pointed out by Field (2005, p.354), even if the assumption of homoscedasticity is violated, ANOVA is still robust when: (i) sample sizes are equal and (ii) variances are proportional to the means (Budescu, 1982, Budescu and Appelbaum, 1981). Since our data respect both these conditions, we can trust the results of our ANOVAs. To deal with the fact that the homogeneity of variances hypothesis was violated, we carried out the post-hoc comparison using the Games–Howell procedure. Such a test is one of the post-hoc tests specially designed for situations in which variances differ. In particular, it has been proven to be the most powerful when, as in our case, the sample sizes are not too small (Field, 2005, p.355). We can thus trust our post-hoc test results as well.

We coded the optimisation models in AMPL and solved them through the IBM ILOG Cplex Solver (version 12.4) running on a PC equipped with an Intel iCore 7 processor and 8 GB of RAM. For each optimisation run, we bound the computational time to 1 hour. The Cplex options were set so as to emphasise feasibility over optimality, perform an aggressive level of probing (Savelsbergh, 1994) and limit the maximum size for the Branch and Bound node file (mipemphasis=1, probe=3, nodefile=2, workfilelim=1028). These settings allowed us to find a feasible solution for a higher number of scenarios with respect to the default Cplex options. The simulation model, instead, was created using Rockwell Arena (version 13.9) and integrated with AMPL via VBA.

6 Empirical results

Due to space constraints, it is not possible to present the results relevant to all the 27 settings in full. Consequently, at first, we will thoroughly discuss the results associated with three base settings. One of these setting refers to the Meyer Hospital case, the other two settings are obtained starting from the Meyer Hospital case and varying the values of OUR. Presenting these three base settings in detail, allows us to explain the causal mechanisms that make some scheduling policies outperform the others. The simulation experiments used in this section 6.1, use empirical distributions for both ST and LoS.

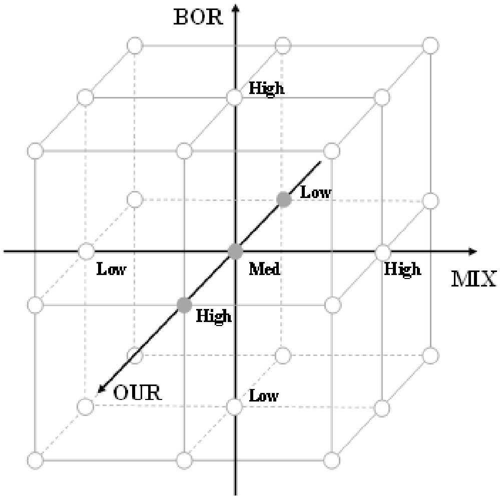


Figure 4 - Base settings

Subsequently, we will present the results relevant to all the other settings and distributional models in a more aggregated manner.

6.1 Base settings

6.1.1 Optimisation phase

Figure 5 and Figure 6 present the results of the base settings. Each figure includes four graphs.

The x-entry of each graph represents a different level of OUR. The y-entry, instead, represents an output of the optimisation phase. The four graphs in Figure 5 report the statistics (Max, M, Rng and Sd, respectively) relevant to the bed utilisation, whereas those in Figure 6 report the same statistics that are relevant to the OR utilisation. Each symbol on the graphs is associated with a different objective function, the lines connecting the symbols are drawn for the sake of clarity, but do not represent any experimental point.

Looking at Figure 5, it is possible to notice that the beds are pretty well balanced. In fact, Sd(Beds)=0, Rng(Beds)=0 and thus M(Beds)=Max(Beds) for all the solutions, except for the one corresponding to a high OUR value and the minOvrn policy. In addition, minRng obtains the highest M(Beds) across OUR levels. Such a high beds utilisation, however, leads to high Ovrn(Beds).

Looking at Figure 6, instead, we can observe that: (i) no policy leads to perfect OR balancing (Sd(ORs)>0 and Rng(ORs)>0), (ii) across OUR levels, minRng performs better than minMax and minOvrn. In fact, it allows us to obtain the highest M(ORs) and, at the same time, the lowest Max(ORs), Rng(ORs) and Ovrn(ORs). Indeed, minRng leads to a lower Max(ORs) than minMax, and to a lower Ovrn(ORs) than minOvrn. This fact can be the explained as follows: all the objective functions are hierarchical and the bed balancing term is more important than the OR balancing term. Hence, for both minRng and minMax, the best solutions are characterised by Sd(Beds)=0 i.e. by perfect bed balancing. When the beds are perfectly balanced, Rng(Beds)=0, Max(Beds)=M(Beds)=Min(Beds) and Ovrn(Beds)=0. A perfect bed balance, can indeed be obtained in correspondence with different levels of bed utilisation. However, minMax is only optimised in those cases where Max(Beds) =M(Beds) is at a minimum. Instead, minRng is also optimised in the cases where Max(Beds) differs from the minimum. It means that with minRng, it is possible to explore a higher number of solutions in order to find the one that leads to a better balancing of the ORs. Our experimental campaign reveals that among the solutions for which Rng(Beds)=0 and Max(Beds) is not optimised, it is possible to find solutions characterised by the smallest possible values of Rng(ORs), Max(ORs) and Ovrn(ORs).

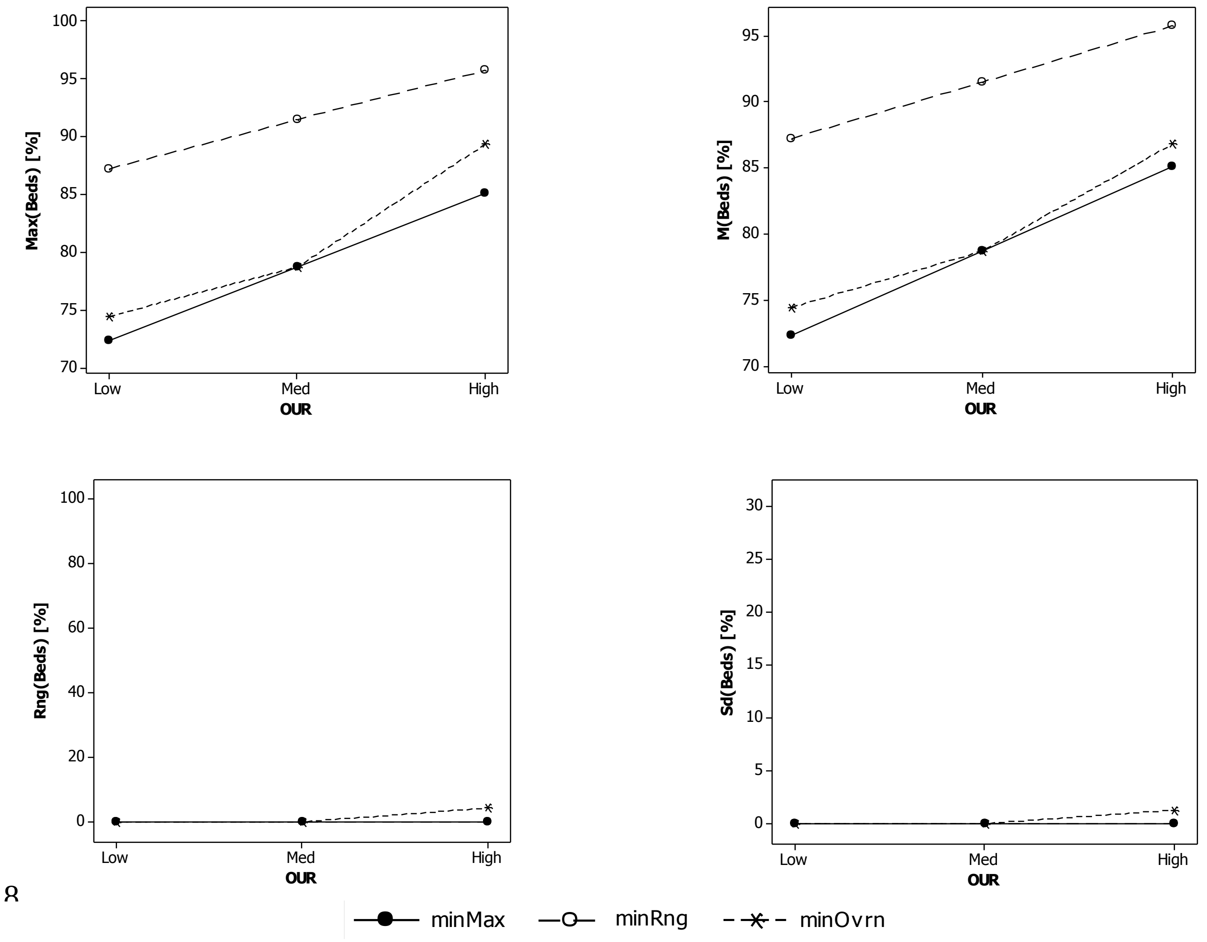


Figure 5 - Base settings optimisation results – bed utilisation

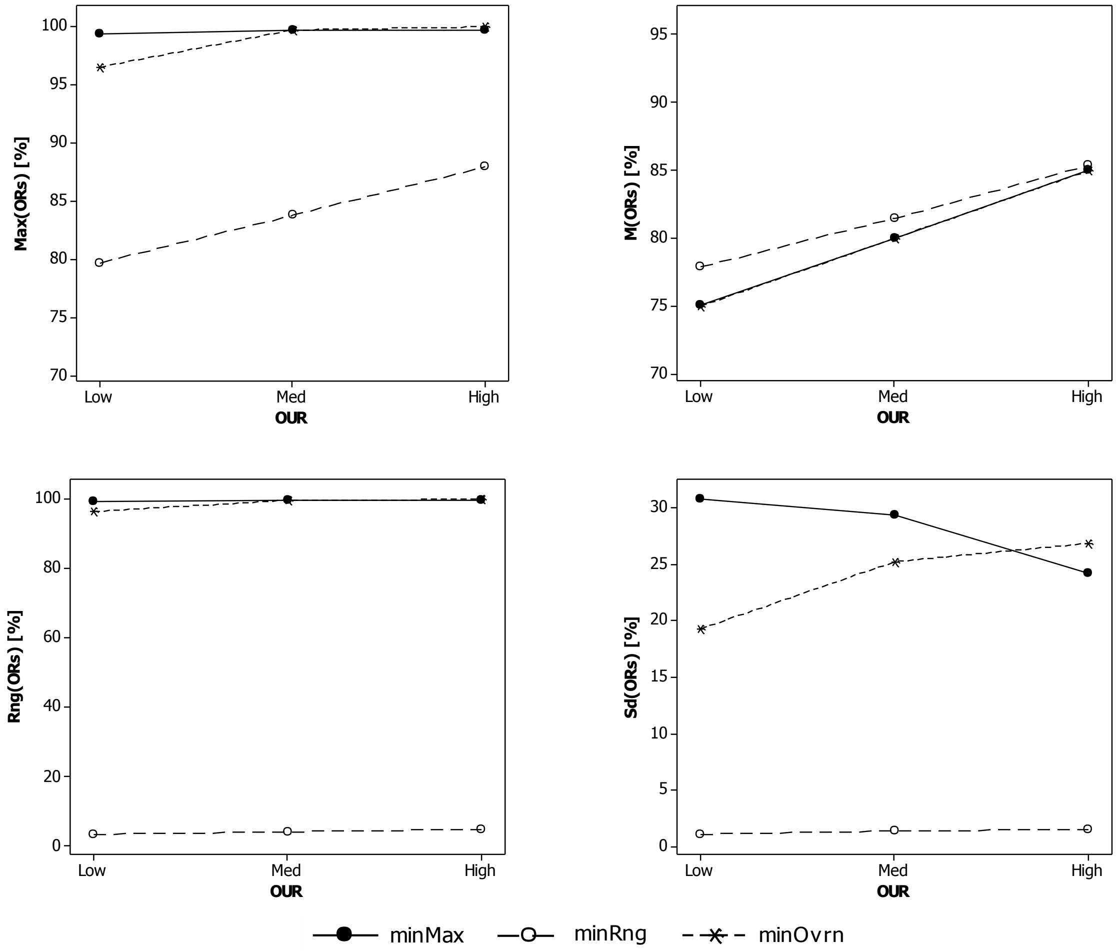


Figure 6 - Base settings optimisation results – OR utilisation

Figure 7 reports the number of scheduled surgeries associated with each solution. As it can be noticed, for all the policies the number of scheduled surgeries increases as OUR increases.

By comparing the policies to each other, it emerges that minRng is the one that allows for the largest number of surgeries to be scheduled. This is not surprising; minMax and minOvrn, in fact, are associated, for each OUR level, with a lower M(Beds) and Ovrn(Beds). It implies that in order to achieve a given OUR, they schedule a smaller number of surgeries characterised by a higher ST.

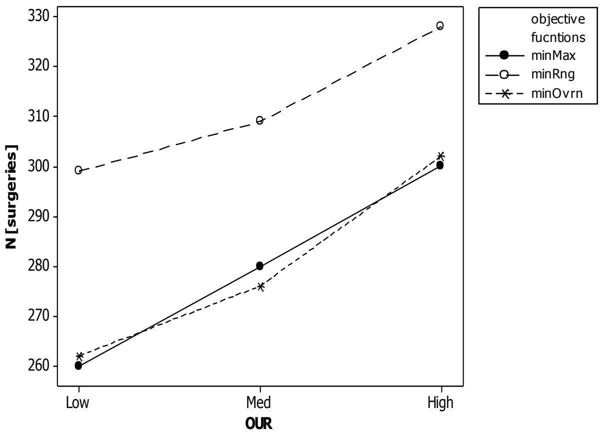


Figure 7 - Base settings number of scheduled surgeries

6.1.2 Simulation phase

Figure 8 presents the results of the simulation phase. In particular, it reports the 95% confidence intervals for the M(OVB) and M(OVT) (left- and right-hand graph, respectively). The results presented in this section refers to empirical distributions (D1).

As it can be noticed, both OVT and OVB increases as the OUR increases.

From statistical analysis, it emerges that, for each scenario, minRng leads to a M(OVB) that is significantly larger than the one associated with minMax and minOvrn. The difference between the values associated with these latter policies is indeed statistically significant ($p<0.05$) only for high values of OUR. However, also in this latter case, such a difference is not practically relevant. M(OVT) increases with OUR, as well. However, in this case, for all the scenarios, minRng leads to an overtime that is significantly lower than the one associated with minMax and minOrvn.

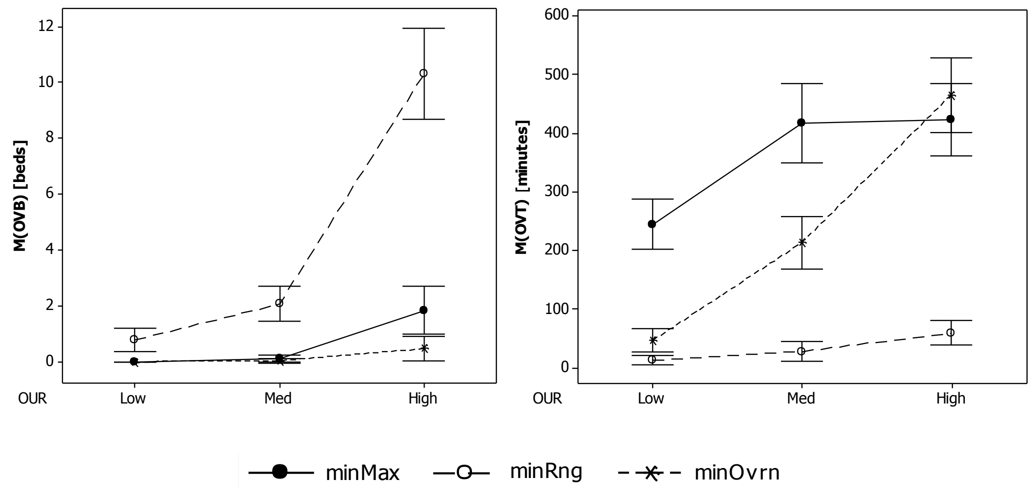


Figure 8 - Base settings simulation results

In summary, from the analysis of these base settings it emerges that MinRng is at least effective as the other policies in balancing the beds. In addition, it allows for a better OR balancing and thus for a smaller overtime. Moreover, it leads to higher bed saturation and to a larger number of scheduled surgeries. However, it also causes a higher overbooking, especially when OUR is high. The properties of minMax are quite the opposite, while minOvrn somehow represents an intermediate case.

6.2 Generalisation

In this section, we discuss the generalisability of the previous findings exploring different hospital settings and different types of distributions. Furthermore, we will present the insight emerging from the additional scenarios.

6.2.1 Optimisation phase

Table 5, Table 6 and Table 7 report the results of the optimisation phase, where each table is associated with a different performance (number of scheduled surgeries N, standard deviation of the bed utilisation Sd(Beds) and of OR utilisation Sd(ORs), respectively). Each cell of the tables represents a different setting. Each cell indicates, for each scheduling policy, the value of the performance under investigation. Within each cell, the policies are ranked from the best (rank=1) to the worst one (rank=3). Due to space constraints, the data relevant to the maximum utilisations, utilisation ranges and overruns (that we presented for the base settings) is not presented here.

Looking at the experimental results, it can be noticed that there is a setting, i.e. MIX=high, BOR=low and OUR=high, for which it was not possible to find a feasible solution. The explanation for this is that when the number of beds is low (BOR=low), to achieve a high OR utilisation (OUR=high), the model, regardless of the objective function, has to schedule a high number of surgeries characterised by low LoS and high ST. It, obviously, does not allow to obtain solutions that are characterised by a high percentage of short surgeries (MIX=high).

Table 5 - Experimental campaign optimisation results – number of scheduled surgeries

N		*Low OUR*	*Medium OUR*	*High OUR*
High MIX	*High BOR*	1.minRng (330) 2.minOvrn (293) 3.minMax (286)	1.minRng (327) 2.minOvrn (315) 3.minMax (312)	1.minRng (354) 2.minMax (340) 3.minOvrn (329)
	Med BOR	1.minRng (309) 2.minMax (286) 3.minOvrn (281)	1.minRng (334) 2.minMax (313) 3.minOvrn (304)	1.minRng (340) 2.minMax (330) 3.minOvrn (322)
	Low BOR	1.minRng (312) 2.minMax (286) 3.minOvrn (284)	1.minRng (318) 2.minMax (312) 3.minOvrn (305)	infeasible
Med MIX	*High BOR*	1.minRng (300) 2.minOvrn (270) 3.minMax (260)	1.minRng (314) 2.minOvrn (286) 3.minMax (280)	1.minRng (341) 2.minMax (301) 3.minOvrn (300)
	Med BOR	1.minRng (299) 2.minOvrn (262) 3.minMax (260)	1.minRng (309) 2.minMax (280) 3.minOvrn (276)	1.minRng (328) 2.minOvrn (302) 3.minMax (300)
	Low BOR	1.minRng (300) 2.minOvrn (260) 3.minMax (259)	1.minRng (302) 2.minMax (280) 2.minOvrn (280)	1.minMax (301) 1.minRng (301) 3.minOvrn (300)
Low MIX	*High BOR*	1.minRng (280) 2.minOvrn (252) 3.minMax (241)	1.minRng (293) 2.minOvrn (265) 3.minMax (261)	1.minRng (297) 2.minOvrn (282) 3.minMax (279)
	Med BOR	1.minRng (286) 2.minOvrn (241) 3.minMax (240)	1.minRng (293) 2.minMax (258) 2.minOvrn (258)	1.minRng (292) 2.minMax (280) 3.minOvrn (277)
	Low BOR	1.minRng (279) 2.minMax (243) 3.minOvrn (236)	1.minRng (282) 2.minMax (261) 3.minOvrn (258)	1.minRng (282) 2.minMax (280) 3.minOvrn (276)

Table 6 - Experimental campaign optimisation results - standard deviation of the bed utilisation [%]

Sd(Beds)		*Low OUR*	*Medium OUR*	*High OUR*
High MIX	*High BOR*	1.minMax (0) 1.minRng (0) 1.minOvrn (0)	1.minMax (0) 1.minRng (0) 1.minOvrn (0)	1.minMax (0) 1.minRng (0) 1.minOvrn (0)
	Med BOR	1.minMax (0) 1.minRng (0) 2.minOvrn (8.3)	1.minMax (0) 1.minRng (0) 2.minOvrn (0.9)	1.minMax (0) 1.minRng (0) 2.minOvrn (1.1)
	Low BOR	1.minMax (0) 1.minRng (0) 2.minOvrn (1.9)	1.minMax (0) 1.minRng (0) 2.minOvrn (1)	infeasible
Med MIX	*High BOR*	1.minMax (0) 1.minRng (0) 2.minOvrn (0.6)	1.minMax (0) 1.minRng (0) 1.minOvrn (0)	1.minMax (0) 1.minRng (0) 2.minOvrn (2.4)
	Med BOR	1.minMax (0) 1.minRng (0) 1.minOvrn (0)	1.minMax (0) 1.minRng (0) 1.minOvrn (0)	1.minMax (0) 1.minRng (0) 2.minOvrn (1.3)
	Low BOR	1.minMax (0) 1.minRng (0) 2.minOvrn (1.7)	1.minMax (0) 1.minRng (0) 2.minOvrn (2.8)	1.minMax (0) 1.minRng (0) 2.minOvrn (2)
Low MIX	*High BOR*	1.minMax (0) 1.minRng (0) 2.minOvrn (0.6)	1.minMax (0) 1.minRng (0) 2.minOvrn (0.6)	1.minMax (0) 1.minRng (0) 2.minOvrn (0.8)
	Med BOR	1.minMax (0) 1.minRng (0) 1.minOvrn (0)	1.minMax (0) 1.minRng (0) 1.minOvrn (0)	1.minMax (0) 1.minRng (0) 1.minOvrn (0)
	Low BOR	1.minMax (0) 1.minRng (0) 2.minOvrn (2)	1.minMax (0) 1.minRng (0) 2.minOvrn (3.3)	1.minMax (0) 1.minRng (0) 2.minOvrn (2.5)

Table 7 - Experimental campaign optimisation results - standard deviation of the OR utilisation [%]

Sd(ORs)		*Low OUR*	*Medium OUR*	*High OUR*
High MIX	*High BOR*	1.minRng (2.2) 2.minOvrn (4) 3.minMax (19.4)	1.minRng (1.4) 2.minOvrn (8.9) 3.minMax (15.2)	1.minRng (2.7) 2.minOvrn (9) 3.minMax (15.3)
	Med BOR	1.minRng (1.6) 2.minMax (23) 3.minOvrn (26.8)	1.minRng (1.6) 2.minMax (15.4) 3.minOvrn (18.9)	1.minRng (2.4) 2.minMax (11.8) 3.minOvrn (16.2)
	Low BOR	1.minRng (1.4) 2.minMax (23.1) 3.minOvrn (24.3)	1.minRng (3) 2.minMax (15.4) 3.minOvrn (18.8)	infeasible
Med MIX	*High BOR*	1.minRng (1.1) 2.minOvrn (3.2) 3.minMax (32.2)	1.minRng (1.3) 2.minOvrn (8) 3.minMax (29.5)	1.minRng (1.7) 2.minOvrn (3.5) 3.minMax (25.5)
	Med BOR	1.minRng (1.1) 2.minOvrn (19.3) 3.minMax (30.8)	1.minRng (1.4) 2.minOvrn (25.2) 3.minMax (29.4)	1.minRng (1.5) 2.minMax (24.2) 3.minOvrn (26.9)
	Low BOR	1.minRng (1) 2.minMax (30.7) 3.minOvrn (32.1)	1.minRng (1.8) 2.minOvrn (24.7) 3.minMax (29.4)	1.minRng (8.9) 2.minOvrn (21.9) 3.minMax (25.4)
Low MIX	*High BOR*	1.minRng (1) 2.minOvrn (3.4) 3.minMax (34.2)	1.minRng (1.2) 2.minOvrn (4.1) 3.minMax (31.4)	1.minRng (0.9) 2.minOvrn (3.7) 3.minMax (28.4)
	Med BOR	1.minRng (1.5) 2.minOvrn (21) 3.minMax (33.3)	1.minRng (1.3) 2.minOvrn (27.5) 3.minMax (32.1)	1.minRng (0.9) 2.minOvrn (26.6) 3.minMax (27.8)
	Low BOR	1.minRng (1) 2.minMax (35.5) 3.minOvrn (36.6)	1.minRng (2.1) 2.minMax (31.7) 3.minOvrn (33.2)	1.minRng (11.7) 2.minMax (27.2) 3.minOvrn (28.2)

Moreover, the results clearly show that some of the findings emerged from the analysis of the base settings, are indeed generalisable to the other settings. In fact, regardless of BOR and MIX, whenever it was possible to find feasible solutions, we observed that: (i) minRng is the policy that allows for the scheduling of the largest number of surgeries; (ii) for each policy, the beds are more balanced than ORs; (iii) the solutions obtained by minMax and minRng are characterised by perfect bed balancing (Sd(Beds)=0). Instead, most of the solutions obtained with minOvrn are pretty unbalanced (yet, also for this policy, Sd(Beds) is always lower than 10%); (iv) no policy leads to perfect ORs balancing, in fact, Sd(ORs) is always larger than zero; (v) minRng performs better than the other policies in terms of OR balancing and utilisation. In fact it allows the smallest values of Max(ORs), Rng(ORs) and Ovrn(ORs) to be obtained and at the same time the highest value of M(ORs) (due to space constraints are not reported here). The explanation we gave in Section 6.1.1 for this unobvious phenomenon seems to hold across the scenarios.

Furthermore, the analysis of these additional settings showed that regardless of the level of MIX and OUR, when BOR increases, minOvrn allows for a fairly good OR balancing. In fact, with BOR=low or BOR=medium, minOvrn led to values of Sd(ORs) similar to the ones of minMax. On the contrary, when BOR=high, Sd(ORs) decreases till the values are similar to the ones of minRng. We have not observed any significant change in the ranking of scheduling policies across MIX levels.

6.2.2 Simulation phase, empirical distributions

Table 8 and Table 9 report the results of the simulation phase. As for the base settings, the results presented in this section refer to empirical distributions (D1). These tables are organised as the ones relevant to the optimisation phase and report, respectively, the mean values (M) and the standard deviations (Sd) of OVB and OVT. Since OVB and OVT are stochastic variables, to rank the different scheduling policies within each cell, we proceeded as explained in Section 5.4.

Table 8 - Experimental campaign simulation results, empirical distributions (D1) – mean values (M) and standard deviations (Sd) of overbooking [beds]

Overbooking		*Low OUR*	*Medium OUR*	*High OUR*
High MIX	*High BOR*	1.minMax (M=0, Sd=0) 1.minRng (M=0, Sd=0) 1.minOvrn (M=0.1, Sd=0.3)	1.minMax (M=0, Sd=0) 1.minOvrn (M=0, Sd=0) 2.minRng (M=1.2, Sd=1.8)	1.minOvrn (M=0.1, Sd=0.3) 2.minMax (M=0.5, Sd=1) 3.minRng (M=6.2, Sd=3.9)
	Med BOR	1.minMax (M=0, Sd=0) 1.minRng (M=0.2, Sd=0.7) 2.minOvrn (M=1.5, Sd=1.4)	1.minMax (M=0.2, Sd=0.5) 1.minOvrn (M=0.2, Sd=0.6) 2.minRng (M=6.1, Sd=4.1)	1.minOvrn (M=2.5, Sd=2.1) 2.minMax (M=4.5, Sd=3) 3.minRng (M=22.1, Sd=6.3)
	Low BOR	1.minOvrn (M=0.8, Sd=1.2) 1.minMax (M=1, Sd=1.5) 2.minRng (M=20.3, Sd=5.8)	1.minOvrn (M=10.7, Sd=4.5) 2.minMax (M=14.6, Sd=5.5) 3.minRng (M=22.6, Sd=5.1)	infeasible
Med MIX	*High BOR*	1.minMax (M=0, Sd=0) 1.minRng (M=0, Sd=0) 1.minOvrn (M=0, Sd=0)	1.minMax (M=0, Sd=0) 1.minOvrn (M=0, Sd=0) 1.minRng (M=0.1, Sd=0.3)	1.minMax (M=0, Sd=0) 1.minOvrn (M=0.1, Sd=0.3) 2.minRng (M=3.9, Sd=3)
	Med BOR	1.minMax (M=0, Sd=0) 1.minOvrn (M=0, Sd=0) 2.minRng (M=0.8, Sd=1.1)	1.minOvrn (M=0, Sd=0.2) 1.minMax (M=0.1, Sd=0.4) 2.minRng (M=2.1, Sd=1.7)	1.minOvrn (M=0.5, Sd=1.2) 2.minMax (M=1.8, Sd=2.3) 3.minRng (M=10.3, Sd=4.4)
	Low BOR	1.minMax (M=0.5, Sd=0.8) 1.minOvrn (M=0.6, Sd=1.3) 2.minRng (M=22.3, Sd=7.5)	1.minOvrn (M=5.2, Sd=2.7) 1.minMax (M=6.4, Sd=3.5) 2.minRng (M=20, Sd=4.9)	1.minMax (M=19.9, Sd=6.5) 1-2.minOvrn (M=23.5, Sd=7) 2.minRng (M=26.8, Sd=6.3)
Low MIX	*High BOR*	1.minMax (M=0,Sd=0) 1.minRng (M=0, Sd=0) 1.minOvrn (M=0, Sd=0)	1.minMax (M=0, Sd=0) 1.minRng (M=0, Sd=0) 1.minOvrn (M=0, Sd=0)	1.minMax (M=0, Sd=0.2) 1.minOvrn (M=0.1, Sd=0.3) 1.minRng (M=0.2, Sd=0.4)
	Med BOR	1.minMax (M=0, Sd=0) 1.minOvrn (M=0, Sd=0) 2.minRng (M=1, Sd=1.2)	1.minMax (M=0, Sd=0) 1.minOvrn (M=0, Sd=0) 2.minRng (M=1.9, Sd=2.2)	1.minOvrn (M=0.4, Sd=0.9) 1.minMax (M=1.6, Sd=2.7) 2.minRng (M=5, Sd=3.2)
	Low BOR	1.minMax (M=0.5, Sd=1) 1.minOvrn (M=0.9, Sd=1.4) 2.minRng (M=18.5, Sd=5.6)	1.minMax (M=3.7, Sd=2.2) 2.minOvrn (M=5.3, Sd=2.8) 3.minRng (M=17.9, Sd=5.1)	1.minOvrn (M=15.4, Sd=5.9) 1.minMax (M=16.6, Sd=4.4) 2.minRng (M=19.2, Sd=5.9)

Table 9 - Experimental campaign simulation results, empirical distributions (D1) – mean values (M) and standard deviations (Sd) of overtime [min]

Overtime		*Low OUR*	*Medium OUR*	*High OUR*
High MIX	*High BOR*	1.minOvrn (M=0.5, Sd=2.7) 2.minRng (M=8.9, Sd=16.8) 3.minMax (M=99.9, Sd=93.3)	1.minOvrn (M=44.5, Sd=53.7) 1.minRng (M=52.7, Sd=100.3) 2.minMax (M=116.3, Sd=64)	1.minRng (M=83.3, Sd=66.1) 1.minOvrn (M=102.7, Sd=99.5) 2.minMax (M=226.4, Sd=107.8)
	Med BOR	1.minRng (M=8.2, Sd=19.3) 2.minMax (M=118.3, Sd=81.1) 3.minOvrn (M=203.2, Sd=121.9)	1.minRng (M=22.5, Sd=34.9) 2.minMax (M=101.8, Sd=67.6) 3.minOvrn (M=243.4, Sd=128.7)	1.minRng (M=102.3, Sd=64.5) 2.minMax (M=158.5, Sd=77.8) 3.minOvrn (M=296.4, Sd=130.5)
	Low BOR	1.minRng (M=2.7, Sd=8.8) 2.minMax (M=115.4, Sd=75.2) 3.minOvrn (M=243.3, Sd=108.1)	1.minRng (M=22.6, Sd=30.2) 2.minMax (M=84.9, Sd=74.7) 3.minOvrn (M=247.6, Sd=137.4)	infeasible
Med MIX	*High BOR*	1.minRng (M=1.2, Sd=3.9) 1.minOvrn (M=3, Sd=10.5) 2.minMax (M=363.9, Sd=138.9)	1.minRng (M=27.2, Sd=42.9) 1.minOvrn (M=31.6, Sd=40.9) 2.minMax (M=459.9, Sd=180.9)	1.minOvrn (M=63.4, Sd=78.6) 2.minRng (M=130, Sd=97.6) 3.minMax (M=457.6, Sd=169)
	Med BOR	1.minRng (M=13.2, Sd=21.2) 2.minOvrn (M=47.2, Sd=54.5) 3.minMax (M=245.4, Sd=115.6)	1.minRng (M=28.5, Sd=43.5) 2.minOvrn (M=214.1, Sd=121) 3.minMax (M=417.6, Sd=182.8)	1.minRng (M=59.8, Sd=55.8) 2.minMax (M=423.5, Sd=166.4) 2.minOvrn (M=466, Sd=170.5)
	Low BOR	1.minRng (M=6.8, Sd=14.7) 2.minMax (M=268.4, Sd=144) 2.minOvrn (M=282.9, Sd=126.4)	1.minRng (M=19.1, Sd=32.1) 2.minOvrn (M=240.1, Sd=120.3) 2.minMax (M=299.6, Sd=144.4)	1.minRng (M=259.2, Sd=142.6) 2.minOvrn (M=415.9, Sd=164.7) 2.minMax (M=425.1, Sd=128.3)
Low MIX	*High BOR*	1.minOvrn (M=4.2, Sd=14.3) 1.minRng (M=8.4, Sd=22.3) 2.minMax (M=303.2, Sd=136.2)	1.minOvrn (M=19.4, Sd=40.4) 1.minRng (M=26.4, Sd=34.6) 2.minMax (M=362.9, Sd=135.3)	1.minOvrn (M=42.5, Sd=49) 1.minRng (M=77.1, Sd=85) 2.minMax (M=465.8, Sd=163.4)
	Med BOR	1.minRng (M=14.2, Sd=19.5) 2.minOvrn (M=53.3, Sd=67.6) 3.minMax (M=222.6, Sd=116.9)	1.minRng (M=28.8, Sd=47.9) 2.minOvrn (M=248.1, Sd=131.3) 3.minMax (M=508.4, Sd=168.9)	1.minRng (M=67.9, Sd=55.7) 2.minOvrn (M=448.2, Sd=160.6) 2.minMax (M=515.8, Sd=156.1)
	Low BOR	1.minRng (M=2.7, Sd=7.7) 2.minOvrn (M=389.8, Sd=156.4) 2.minMax (M=455.5, Sd=192.5)	1.minRng (M=20.8, Sd=31) 2.minMax (M=382.9, Sd=143.6) 3.minOvrn (M=552.4, Sd=215)	1.minRng (M=276.6, Sd=129.7) 2.minOvrn (M=494.6, Sd=164.7) 2.minMax (M=526.6, Sd=203.9)

Referring to the relative rank of the scheduling policies, the simulation analysis confirms most of the findings found for the base settings. In particular: (i) minRng leads to overbookings that (if different from zero) are significantly higher than those associated with minMax and minOvrn, regardless of the levels of MIX and BOR. These latter policies, instead, do not significantly differ in terms of overbooking. There is only one setting (MIX=high, BOR=med, OUR=low) where minOvrn leads to the highest values of overbooking. This is probably due to the fact that this is the only scenario where the solver was unable to find a feasible solution where beds were well balanced within the given time limit (Table 6). (ii) minRng leads to the best results in terms of overtimes in most of the scenarios. However, when BOR is high minOvrn and minRng lead to overtimes that do not significantly differ from each other. Moreover, in the two scenarios (MIX=med, BOR=high, OUR=high and MIX=high, BOR=high, OUR=low) minOvrn leads to overtimes that are significantly smaller than those of minRng. However, the difference between the results associated with the two policies is only practically relevant in the first case (in the second case is just 8.4 minutes in two weeks).

From the simulation analysis, it also emerges that if BOR increases, then M(OVB) decreases (and vice-versa). In fact, for each target OR utilisation (OUR) and MIX, if the number of available beds (and thus BOR) increases, the probability of having additional beds to accommodate patients whose actual LoS lasts more than its expected value increases as well.

Finally, we can observe that, if the ORs are not well balanced then M(OVT) can also be high when OUR is low. On the contrary, if the ORs are well balanced then M(OVT) is only high for high level of OUR. This fact confirms our initial conjecture.

6.2.3 Simulation phase, theoretical distributions

In this section we discuss whether the findings relevant to D1 hold also for the other distributions. In Table 10 and Table 11 we report the results associated with the use of fitted lognormal distributions (D2). Comparing these tables with Tables 8 and 9, it clearly emerges that the relative rankings of the different policies do not change for most of the settings. In fact, as in the previous case: (i) minRng leads to the largest M(OVB) even if in two settings (MIX=low, BOR=low, OUR=high and MIX=med, BOR=low, OUR=high) M(OVB) associated with minRng are not significantly larger than those of minOvrn: (ii) minRng leads to the smallest M(OVT), with the only exception of scenario (MIX=med, BOR=high, OUR=high). Comparing the results of D2 and D1 it can also be noticed that, on average, M(OVT) and M(OVB) of D2 are larger than those of D1. As we pointed out on Section 5.2, this is due to the fact that contrary to the empirical distributions the lognormal ones are unbounded, as such, they periodically return fairly large values of ST and LoS during the run.

Due to space constraints, we do not report the tables relevant to D3 and D4. However, the results associated with these distributions are consistent with those presented so far.

In particular, with D3 (leptokurtic and symmetric distributions - small values of σ(ST) and σ(LoS)) we still obtain that minRng leads to the largest overbookings. There is only one case (MIX=high, BOR=med, OUR=low) where minOvrn is characterised by a M(OVB) that is significantly larger than the one of minRng but the difference between these values (+ 0.3 beds in two weeks) is not practically relevant. Looking at the overtimes, instead, minRng always leads to the smallest values.

Table 10 - Experimental campaign simulation results, fitted lognormal distributions (D2) – mean values (M) and standard deviations (Sd) of overbooking [beds]

Overbooking		*Low OUR*	*Medium OUR*	*High OUR*
High MIX	*High BOR*	1.minMax (M=0, Sd=0) 1.minOvrn (M=0, Sd=0.2) 2.minRng (M=2.8, Sd=3)	1.minMax (M=0.4, Sd=1) 1.minOvrn (M=0.5, Sd=1.4) 2.minRng (M=7.2, Sd=3.6)	1.minOvrn (M=3.6, Sd=2.7) 2.minMax (M=6.3, Sd=3.4) 3.minRng (M=20.6, Sd=6.6)
	Med BOR	1.minMax (M=0, Sd=0.2) 2.minOvrn (M=3.7, Sd=2.3) 2.minRng (M=5.7, Sd=4.1)	1.minOvrn (M=3.4, Sd=2.7) 1.minMax (M=3.6, Sd=2.8) 2.minRng (M=24.9, Sd=7.5)	1.minOvrn (M=14.6, Sd=6.5) 1.minMax (M=16.1, Sd=7.6) 2.minRng (M=44.8, Sd=8.8)
	Low BOR	1.minMax (M=7.5, Sd=3.7) 1.minOvrn (M=8.7, Sd=3.9) 2.minRng (M=43.6, Sd=10)	1.minOvrn (M=28.8, Sd=8) 1.minMax (M=32.6, Sd=7.8) 2.minRng (M=45.4, Sd=9.7)	infeasible
Med MIX	*High BOR*	1.minMax (M=0, Sd=0) 1.minOvrn (M=0, Sd=0) 1.minRng (M=0.5, Sd=1.5)	1.minMax (M=0, Sd=0.2) 1.minOvrn (M=0.2, Sd=0.5) 2.minRng (M=1.6, Sd=2)	1.minMax (M=0.8, Sd=1.3) 1.minOvrn (M=1.3, Sd=1.8) 2.minRng (M=15.3, Sd=6.9)
	Med BOR	1.minOvrn (M=0, Sd=0.2) 1.minMax (M=0.2, Sd=0.6) 2.minRng (M=6.2, Sd=5.1)	1.minMax (M=1, Sd=1.5) 1.minOvrn (M=1.2, Sd=1.9) 2.minRng (M=13, Sd=5.3)	1.minOvrn (M=5.4, Sd=3.1) 1.minMax (M=7.3, Sd=4.9) 2.minRng (M=24, Sd=8.6)
	Low BOR	1.minMax (M=3.9, Sd=2.9) 1.minOvrn (M=4, Sd=2.4) 2.minRng (M=38.9, Sd=9.2)	1.minMax (M=13.9, Sd=6.4) 1.minOvrn (M=14.3, Sd=5.1) 2.minRng (M=39.8, Sd=9.9)	1.minMax (M=36.5, Sd=7.8) 2.minRng (M=43.6, Sd=9.4) 2.minOvrn (M=46.2, Sd=7.4)
Low MIX	*High BOR*	1.minOvrn (M=0, Sd=0) 1.minMax (M=0, Sd=0) 1.minRng (M=0.1, Sd=0.3)	1.minMax (M=0, Sd=0) 1.minOvrn (M=0.1, Sd=0.3) 2.minRng (M=0.9, Sd=1.3)	1.minMax (M=0.1, Sd=0.3) 2.minOvrn (M=1.4, Sd=1.5) 3.minRng (M=3.8, Sd=3.3)\
	Med BOR	1.minMax (M=0, Sd=0) 1.minOvrn (M=0.1, Sd=0.3) 2.minRng (M=7.5, Sd=5)	1.minMax (M=0.3, Sd=0.6) 1.minOvrn (M=0.2, Sd=0.7) 2.minRng (M=11.4, Sd=7)	1.minOvrn (M=1.5, Sd=1.9) 2.minMax (M=4.6, Sd=2.8) 3.minRng (M=15.5, Sd=6.2)
	Low BOR	1.minOvrn (M=2.1, Sd=2.1) 1.minMax (M=2.9, Sd=2.2) 2.minRng (M=31.3, Sd=7.2)	1.minMax (M=9.5, Sd=4.3) 1.minOvrn (M=10.2, Sd=4.3) 2.minRng (M=28.6, Sd=6)	1.minOvrn (M=26, Sd=7.4) 1.minMax (M=28.7, Sd=8.1) 1.minRng (M=30, Sd=7.1)

Table 11 - Experimental campaign simulation results, fitted lognormal distributions (D2) – mean values (M) and standard deviations (Sd) of overtime [min]

Overtime		*Low OUR*	*Medium OUR*	*High OUR*
High MIX	*High BOR*	1.minRng (M=25.8, Sd=48.4) 1.minOvrn (M=42.5, Sd=65.7) 2.minMax (M=215.2, Sd=130.2)	1.minOvrn (M=93.3, Sd=79.7) 1.minRng (M=135.1, Sd=103.7) 2.minMax (M=228.7, Sd=102.4)	1.minOvrn (M=137.8, Sd=85.4) 1.minRng (M=153.4, Sd=92) 2.minMax (M=337, Sd=171.8)
	Med BOR	1.minRng (M=6.9, Sd=25.7) 2.minMax (M=233.2, Sd=133.8) 2.minOvrn (M=277.2, Sd=138.2)	1.minRng (M=35.5, Sd=43.3) 2.minMax (M=195.8, Sd=121.8) 3.minOvrn (M=307.3, Sd=138.2)	1.minRng (M=146.7, Sd=132.5) 2.minMax (M=295.4, Sd=201.8) 3.minOvrn (M=426, Sd=190.7)
	Low BOR	1.minRng (M=13.6, Sd=38.1) 2.minMax (M=253.8, Sd=143.4) 2.minOvrn (M=342.5, Sd=159.1)	1.minRng (M=66.7, Sd=91) 2.minMax (M=175.2, Sd=103.4) 3.minOvrn (M=330.8, Sd=138.9)	infeasible
Med MIX	*High BOR*	1.minOvrn (M=16.2, Sd=28.6) 1.minRng (M=38.1, Sd=53.1) 2.minMax (M=471.8, Sd=198.9)	1.minRng (M=41.6, Sd=48.4) 2.minOvrn (M=83, Sd=73) 3.minMax (M=627.9, Sd=197.2)	1.minOvrn (M=97.5, Sd=86.5) 2.minRng (M=200.6, Sd=138.9) 3.minMax (M=627.8, Sd=221.2)
	Med BOR	1.minRng (M=25.9, Sd=54.1) 2.minOvrn (M=122.3, Sd=118.4) 3.minMax (M=374, Sd=172)	1.minRng (M=56.5, Sd=59) 2.minOvrn (M=257, Sd=137.1) 3.minMax (M=562.3, Sd=127.3)	1.minRng (M=109.6, Sd=114) 2.minMax (M=538.4, Sd=222.2) 2.minOvrn (M=576.2, Sd=209.8)
	Low BOR	1.minRng (M=17.6, Sd=24.5) 2.minMax (M=352.6, Sd=186.7) 3.minOvrn (M=495.2, Sd=181.5)	1.minRng (M=38.1, Sd=54.9) 2.minOvrn (M=363, Sd=210.1) 2.minMax (M=454.8, Sd=195.8)	1.minRng (M=254.7, Sd=115.6) 2.minOvrn (M=470.8, Sd=164) 2.minMax (M=538.4, Sd=170.3)
Low MIX	*High BOR*	1.minOvrn (M=28.9, Sd=54.1) 1.minRng (M=37, Sd=49.4) 2.minMax (M=348.6, Sd=105.6)	1.minRng (M=26.6, Sd=31.1) 2.minOvrn (M=86.6, Sd=87.4) 3.minMax (M=510.8, Sd=204.8)	1.minRng (M=93.4 Sd=,71.3) 2.minOvrn (M=170.8, Sd=129.6) 3.minMax (M=698.8, Sd=254.3)
	Med BOR	1.minRng (M=23.9, Sd=34.2) 2.minOvrn (M=84.2, Sd=70.3) 3.minMax (M=386.5, Sd=193)	1.minRng (M=33.1, Sd=41.9) 2.minOvrn (M=384.1, Sd=148.8) 2.minMax (M=468.8, Sd=186.7)	1.minRng (M=141.9, Sd=121.6) 2.minOvrn (M=545.9, Sd=178.3) 3.minMax (M=679.1, Sd=203.1)
	Low BOR	1.minRng (M=31.8, Sd=74) 2.minOvrn (M=525.6, Sd=161.5) 2.minMax (M=568.6, Sd=226.7)	1.minRng (M=53.2, Sd=82.8) 2.minMax (M=520.2, Sd=228.3) 2.minOvrn (M=607.6, Sd=228.7)	1.minRng (M=392.9, Sd=146.5) 2.minMax (M=643.3, Sd=191.7) 2.minOvrn (M=667, Sd=223.5)

With D4 (platykurtic and positively skewed distributions - small values of σ(ST) and σ(LoS)), again, minRng is associated with the largest overbookings. In several scenarios, however, these values are not significantly larger than those of the other scheduling policies. There is also a scenario (MIX=low, BOR=med, OUR=high) where minMax causes a M(OVB) that is significantly larger than the one of minRng. The difference between these values, however, is very small (+ 2.3 beds in two weeks). Referring to the overtime, D4 shows results that are consistent with those of D2.

6.2.4 Optimisation and simulation approach using third quartiles values of ST and LoS

Looking at Tables 8 and 9 (or equivalently at Tables 10 and 11) it emerges that for certain settings (e.g. MIX=low, BOR=low, OUR=high) the values of both overtime and overbooking might not be acceptable. In these cases, implementing the well-balanced solutions obtained in the optimisation phase does not allow achieving robustness. In these cases, a way to achieve robustness is to use values of ST and LoS greater than the mean in the optimisation phase. By doing so, it is possible to obtain solutions that are still balanced but are characterised by a smaller number of surgeries and, consequently, more robust.

In Table 12, we show the results relevant to the “critical” setting (MIX=low, BOR=low, OUR=high) and compare the solutions obtained using: in the optimisation phase, the mean values (M) and the third quartiles (Q3) of ST and LoS respectively; in the simulation phase, the fitted lognormal distributions (D2) of ST and LoS.

Table 12 Simulation results using the mean values (M) and third quartiles (Q3) of ST and LoS in the optimisation phase (setting (MIX=low, BOR=low, OUR=high)).

	M			Q3		
	Scheduled surgeries	Overtime [min]	Overbooking [beds]	Scheduled surgeries	Overtime [min]	Overbooking [beds]
minMax	280	M=643.3, Sd=191.7	M=28.7, Sd=8.1	246	M=42.9, Sd=42.8	M=2.1, Sd=2.2
minRng	282	M=392.9, Sd=146.5	M=30, Sd=7.1	270	M=8.8, Sd=20.7	M=21.6, Sd=6.7
minOvrn	276	M=667, Sd=223.5	M=26, Sd=7.4	243	M=59.7, Sd=61.2	M=2.5, Sd=2.4

As can be noticed using Q3 in the optimisation phase leads to less efficient solutions, especially for minMax (ΔScheduled=36) and minOvrn and (ΔScheduled=32). However, by scheduling fewer surgeries it is possible to obtain solutions that are sufficiently robust with respect to both overtime and overbooking. Indeed, over a planning horizon of two weeks, M(OVT) and M(OVB) associated with minMax are respectively equal to 42.9 minutes and 2.1 beds and those associated with minOvrn are respectively equal to 59.7 minutes 2.5 beds. minRng, instead, still lead to unsatisfactory overbookings (21.6 beds). Finally, it can be observed that despite of the statistic used in the optimisation model (mean, Q3), the overtimes and overbookings associated with minMax and minOvrn do not significantly differ ($p>0.05$) from each other. Due to space constraints the results relevant to the other settings are not reported upon here.

7 Conclusions

The aim of this study was to compare three different scheduling policies, namely minRng, minMax and minOvrn. Specifically, we were interested in comparing the value of efficiency, balancing and robustness that can be obtained by implementing these policies in different hospital settings. To do so, we have utilised a combined optimisation-simulation approach, where the schedule produced by the optimisation model was tested via simulation in order to take into consideration the variability of surgical time and length of stay.

We generated 27 hospital settings, starting from the data of a real hospital (Meyer Hospital). Specifically, we combined three different levels of the following three parameters: (i) hospital's *Beds/OR time Ratio (BOR)*; (ii) case mix complexity (MIX); (iii) target OR utilisation range (OUR). The data pertaining to both surgical time and length of stay comes from real medical records.

The simulation analysis was performed using both empirical and theoretical lognormal distributions. In particular we use the lognormal distribution best fitting our data, as well as two "extreme" lognormal distributions, artificially created to explore distributions with highly skewed and platykurtic shape and situation with symmetric and leptokurtic shape.

From our experimental campaign, it clearly emerges that there is no policy dominating the others in terms of efficiency *and* balancing *and* robustness. In general, minRng leads to the highest bed utilisation and to the largest number of scheduled surgeries, to the best OR balancing and to the smallest overtime.

For each setting, minRng leads to a maximum OR utilisation that is lower than the one of minMax and to an OR overrun than is lower than the one of minOvrn. However, minRng also leads to the highest values of overbooking. In fact, even if allowing for a perfect bed balancing, this policy leads to very high levels of bed utilisation. As the simulation results clearly show, such a high bed utilisation likely leads to bed shortages when the patients' LoS lasts more than expected, and thus to overbooking.

In summary, in order to avoid overbooking, minMax and minOvrn should be preferred to minRng. The relative ranking of minMax and minOvrn, in terms of efficiency, varies according to the setting, but the number of scheduled surgeries (see Table 5) with these policies does not differ much from each other, and it is almost always smaller than the one associated with minRng. On the contrary, if the focus is on efficiency, the best choice is minRng, which also allows for low values of overtime.

For most of the investigated settings, our analysis reveals that there is also an unobvious trade-off between overbooking and overtime. In fact, only in a few cases it is possible to obtain low values for both of these performances. Nonetheless, extremely robust solutions can still be obtained when BOR is high. In these settings in fact, minOvrn allows for achieving solutions characterised by both low overtimes and overbookings. Indeed, in these cases, minOvrn leads to overtimes that either do not significantly differ from, or are significantly smaller than, those obtained with minRng. In addition, minOvrn leads to overbookings that are smaller than, or equal to, the ones of minRng and minMax. These solutions are however characterised by a small number of scheduled surgeries.

The analysis also reveals that for certain settings (e.g. those with BOR=low and OUR= high) using the mean values of ST and LoS in the optimisation phase may lead to large values of overtime and overbooking. In these situations, to obtain robust solutions it is advisable to run the optimisation model with values greater than the mean, e.g. the third quartile, and to use the minMax or minOvrn objective functions.

This study addresses a literature gap and, at the same time, has notable practical implications. Indeed, to the best of our knowledge, this is the first work in the vast literature on the MSS problem that:

i) compares three alternative scheduling policies investigating the efficiency, balancing and robustness that can be achieved implementing them;

ii) explains the causal mechanism that in given circumstances make certain polices outperform the others;

iii) assesses the generalizability of the proposed findings by means of a vast experimental campaign including a vast number (27) of realistic settings and four different types of distributions.

Referring to the practical relevance, this study can help hospital managers to understand the pros and cons associated with the use of different scheduling policies in different operational conditions. In this regard, Table 5 to Table 12 can be seen as tools that support the hospital managers in order to identify the scheduling policy best fitting their priorities and needs. Since the parameters of MIX, BOR and OUR are relative, i.e. they do not depend on the absolute hospital dimensions, and since our findings do not change according to the probability distributions used to model the stochasticity of ST and LoS, each studied setting can be considered as being representative of a wide set of hospitals.

This study, however, is not without its limitations. First, we investigated a vast, yet limited, number of hospital settings. Hospitals with MIX, BOR and OUR that are significantly different from the ones considered in this study might not take advantage of our research. Second, regardless of the values of MIX, BOR and OUR, we have not addressed how the actual hospital dimension, e.g. the number of beds and ORs, could affect the computational efficiency of the optimisation model. Third, we have not considered certain hospital resources (e.g. anaesthetists, ICU beds, nursing staff, medical devices) whose utilisation might need to be balanced as well. Finally, in this study, we assumed that non-elective patients are handled with dedicated resources (as it actually happens at the Meyer Hospital). The extension of the computational campaign to other hospital settings, the analysis of the optimisation model scalability and the evaluation of the impact of non-elective patients on the resource balancing will certainly be the object of our future research efforts. Another interesting avenue for future research includes the definition of alternative ways to address the robustness of the schedules, e.g. the formulation of a stochastic integer programming problem.

Acknowledgements

The authors are grateful to Meyer University Children's Hospital, and in particular to General Director Dr. Tommaso Langiano, and to the Medical Director Dr. Monica Frassineti for supporting the research project that has inspired this study.

The authors would also express their sincere gratitude to the anonymous referees for their insightful critical comments and suggestions.

References

ADAN, I., BEKKERS, J., DELLAERT, N., VISSERS, J. & YU, X. 2009. Patient mix optimisation and stochastic resource requirements: A case study in cardiothoracic surgery planning. Health Care Management Science, 12, 129-141.

BANDITORI, C., CAPPANERA, P. & VISINTIN, F. 2013. A combined optimization–simulation approach to the master surgical scheduling problem. IMA Journal of Management Mathematics, 24, 155-187.

BANDITORI, C., CAPPANERA, P. & VISINTIN, F. 2014. Investigating the relationship between resources balancing and robustness in master surgical scheduling. Proceedings of the international conference on health care systems engineering. Springer, 149-162.

BELIËN, J. & DEMEULEMEESTER, E. 2007. Building cyclic master surgery schedules with leveled resulting bed occupancy. European Journal of Operational Research, 176, 1185-1204.

BELIËN, J., DEMEULEMEESTER, E. & CARDOEN, B. 2009. A decision support system for cyclic master surgery scheduling with multiple objectives. Journal of Scheduling, 12, 147-161.

BOWERS, J. 2013. Balancing operating theatre and bed capacity in a cardiothoracic centre. Health Care Management Science, 16, 1-9.

BUDESCU, D. V. 1982. The power of the F test in normal populations with heterogeneous variances. Educational and Psychological Measurement, 42, 409-416.

BUDESCU, D. V. & APPELBAUM, M. I. 1981. Variance stabilizing transformations and the power of the F test. Journal of Educational and Behavioral Statistics, 6, 55-74.

CAPPANERA, P. & SCUTELLÀ, M. G. 2011. Color-coding algorithms to the balanced path problem: computational issues. INFORMS Journal on Computing, 23, 446-459.

CARAMIA, M. & DELL'OLMO, P. 2003. Assessing the resource usage in scheduling with incompatibilities. OR Spectrum, 25, 521-547.

CARDOEN, B., DEMEULEMEESTER, E. & BELIËN, J. 2010. Operating room planning and scheduling: A literature review. European Journal of Operational Research, 201, 921-932.

CARTER, M. W. & KETABI, S. 2012. Bed Balancing in Surgical Wards via Block Scheduling. Journal of Minimally Invasive Surgical Sciences, 1, 129-137.

CHOI, S. & WILHELM, W. E. 2014. An approach to optimize block surgical schedules. European Journal of Operational Research, 235, 138-148.

CHOW, V. S., PUTERMAN, M. L., SALEHIRAD, N., HUANG, W. & ATKINS, D. 2011. Reducing surgical ward congestion through improved surgical scheduling and uncapacitated simulation. Production and Operations Management, 20, 418-430.

D’AGOSTINO, R. B. & STEPHENS, M. A. 1986. Goodness-of-fit techniques. New York, Marcel Dekker.

DENTON, B., VIAPIANO, J. & VOGL, A. 2007. Optimization of surgery sequencing and scheduling decisions under uncertainty. Health Care Management Science, 10, 13-24.

DOBRZYKOWSKI, D., SABOORI DEILAMI, V., HONG, P. & KIM, S. 2014. A structured analysis of operations and supply chain management research in healthcare (1982–2011). International Journal of Production Economics, 147, 514-530.

FIELD, A. 2005. Discovering statistics using SPSS, London, SAGE Publications.

FISCHETTI, M. & MONACI, M. 2009. Light robustness. Robust and online large-scale optimization. Springer. Lecture Notes in Computer Science, 5868, 61-84.

GLOUBERMAN, S. & MINTZBERG, H. 2001. Managing the care of health and the cure of disease-Part I: Differentiation. Health care management review, 26, 56-69.

GUERRIERO, F. & GUIDO, R. 2011. Operational research in the management of the operating theatre: a survey. Health Care Management Science, 14, 89-114.

GUINET, A. & CHAABANE, S. 2003. Operating theatre planning. International Journal of Production Economics, 85, 69-81.

HANS, E., WULLINK, G., VAN HOUDENHOVEN, M. & KAZEMIER, G. 2008. Robust surgery loading. European Journal of Operational Research, 185, 1038-1050.

LAMIRI, M., GRIMAUD, F. & XIE, X. 2009. Optimization methods for a stochastic surgery planning problem. International Journal of Production Economics, 120, 400-410.

LAW, A. M. & KELTON W. D. 2000. Simulation modeling and analysis, New York, McGraw-Hill.

LITVAK, E. & LONG, M. C. 2000. Cost and quality under managed care: Irreconcilable differences. Am J Manag Care, 6, 305-12.

MANNINO, C., NILSSEN, E. J. & NORDLANDER, T. E. 2012. A pattern based, robust approach to cyclic master surgery scheduling. Journal of Scheduling, 15, 553-563.

MARAZZI, A., PACCAUD, F., RUFFIEUX, C. & BEGUIN, C. 1998. Fitting the distributions of lenght of stay by parametric model. Medical care, 36, 915-927.

MARCON, E., KHARRAJA, S. D. & SIMONNET, G. 2003. The operating theatre planning by the follow-up of the risk of no realization. International Journal of Production Economics, 85, 83-90.

MAY, J. H., W. E., STRUM, D. P. & VARGAS, L. G. 2000. Fitting the lognormal distribution to surgical procedure times. Decision sciences, 31, 129-148.

MAY, J. H., SPANGLER, W. E., STRUM, D. P. & VARGAS, L. G. 2011. The surgical scheduling problem: Current research and future opportunities. Production and Operations Management, 20, 392-405.

SANTIBÁÑEZ, P., BEGEN, M. & ATKINS, D. 2007. Surgical block scheduling in a system of hospitals: an application to resource and wait list management in a British Columbia health authority. Health Care Management Science, 10, 269-282.

SAVELSBERGH, M. W. 1994. Preprocessing and probing techniques for mixed integer programming problems. ORSA Journal on Computing, 6, 445-454.

SEN, T., DILEEPAN, P. & LIND, M. R. 1995. Minimizing a weighted quadratic function of job lateness in the single machine system. International Journal of Production Economics, 42, 237-243.

STEPANIAK, P. S., HEIJ, C., HURINK, J., MANNAERTS, G. H. H., DE QUELERIJ, M. & DE VRIES, G. 2009. Modeling procedure and surgical times for current procedural terminology-anesthesia-surgeon combinations and evaluation in terms of case-duration prediction and operating room efficiency: a multicenter study. Anesthesia & Analgesia, 109, 1232-1245.

VAN OOSTRUM, J. M., VAN HOUDENHOVEN, M., HURINK, J., HANS, E., WULLINK, G. & KAZEMIER, G. 2008. A master surgical scheduling approach for cyclic scheduling in operating room departments. OR Spectrum, 30, 355-374.

VINCENT, C., TAYLOR-ADAMS, S, & STANHOPE, N. 1998. Framework for analysing risk and safety in clinical medicine. BMJ, 316, 1154-1157.

ZHANG, B., MURALI, P., DESSOUKY, M. M. & BELSON, D. 2008. A mixed integer programming approach for allocating operating room capacity. Journal of the Operational Research Society, 60, 663-673.

Appendix – Article III

This is a pre-copy-editing, author-produced PDF of an article accepted for publication in FLEXIBLE SERVICES AND MANUFACTURING JOURNAL following peer review. The definitive publisher-authenticated version

VISINTIN, F., CAPPANERA, P. & BANDITORI, C. 2014. Evaluating the impact of flexible practices on the master surgical scheduling process: an empirical analysis. *Flexible Services and Manufacturing Journal*, 1-24, DOI: 10.1007/s10696-014-9208-9

is available online at

http://link.springer.com/article/10.1007/s10696-014-9208-9

1 Introduction

The Operating Theatre (OT) is widely acknowledged as the functional area driving most hospitals' costs and revenues (Denton et al., 2007). The surgical scheduling process, i.e. the process by which OT activities are planned, dramatically influences OT performance and, as such, it is the object of growing attention from hospital managers worldwide. Such a process, however, is extremely complex to manage. In fact, it requires the consideration of many resources (operating rooms (ORs), surgical teams, and nursing staff as well as downstream resources, such as surgical units and intensive care units (ICUs)) operating in a context affected by a high variability (Litvak and Long, 2000) and characterised by people - surgeons, patients, hospital managers - with conflicting priorities (Glouberman and Mintzberg, 2001). The complexity of the surgical scheduling process coupled with its significant economic and social impact has thus stimulated, in recent years, intensive research activities as well (Cardoen et al., 2010, Guerriero and Guido, 2011, May et al., 2011). The literature, indeed, abounds with models supporting the scheduling of surgical activities. In particular, the mainstream literature presents the consensus that solving a surgical scheduling problem requires addressing three intertwined sub-problems (Beliën and Demeulemeester, 2007): (i) the *case-mix planning*, i.e. the determination (usually on a yearly basis) of the total amount of OR time to assign to each surgical specialty, (ii) the *master surgical scheduling*, i.e. the determination of the specialty (or specialties) to assign to each OR on each day of the planning horizon (e.g. two weeks or one month) and, in certain cases, the specification of the number and typology of surgeries to be performed each day, and finally (iii) the *selection and sequencing* of patients who have to undergo surgery. Typically, these three sub-problems are solved in cascade; the case-mix determined at the first stage is used in the definition of the master surgical schedule (MSS). The MSS, in turn, is used as input for patient selection and sequencing.

This study focuses on the *master surgical scheduling* sub-problem. In the literature, the models supporting such a sub-problem consider slightly different sets of resources (ORs, surgical units, surgical teams, and the ICU) and make different assumptions about how *flexibly* these critical resources are managed. Some studies propose scenario analysis to assess the effects associated with the flexible management of certain resources, such as surgical teams or surgical units (Banditori et al., 2013, Agnetis et al., 2012). However, despite the fact that flexibility is by no means a new topic (Balasubramanian et al., 2012, Buzacott and Mandelbaum, 2008, Chou et al., 2008, Gupta and Shanthikumar, 2008), to date the literature lacks of contributions that have systematically studied the impact of flexibility on OT performance.

This study addresses this gap by adding two main contributions. First, it presents a novel mixed integer programming model to support MSS production. Second, it uses the model to investigate the main and interaction effects associated with the flexible management of three critical resources: *surgical teams*, *ORs* and *surgical units*.

The model assumes that surgical cases can be organised into homogeneous surgery groups (Santibáñez et al., 2007, Banditori et al., 2013) based on their specialty, their expected surgical time (ST) and their expected length of stay (LoS), that is, based on the extent to which these cases are expected to “consume” the previously mentioned critical resources. The model creates a solution indicating for each *OR session* (i.e. for each day, for each OR and for each session) in the planning horizon the number of surgeries to perform and the surgery group these cases must belong to. The model’s objective function is the maximisation of the number of scheduled surgeries.

In addition to presenting the model, we show how such a model can be modified by acting on its variables, parameters and constraints to incorporate a more or less flexible management of surgical teams, ORs and surgical units. The different versions of the model are then used to carry out an experimental campaign based on a 2^3 experimental design (Montgomery and Runger, 2003). In detail, we consider the way the three critical resources are managed as *factors* and we assume two possible levels for each factor: “high” when the resource is managed in a flexible way and “low” otherwise. More specifically:

1) With respect to surgical teams (“*Teams*” factor), we analyse the case where the assignment of surgical teams to sessions is fixed (fixed surgical teams assignment, low level) and the case where such an assignment can change every time the MSS is produced (variable surgical teams assignment, high level).
2) With respect to ORs (“*ORs*” factor), we distinguish the case where ORs are used to perform, within the same session, either long-stay (LoS>1 day) surgeries or short-stay (LoS=1 day) surgeries (dedicated sessions, low level) and the case where both types of surgeries can be performed within the same session (mixed sessions, high level).

3) With respect to surgical units *("Units"* factor), we distinguish the case where units characterised by the same care setting (in terms of nursing staff, equipment, etc.) are used to host cases of specific specialties only (dedicated units, low level) and the case where these units are pooled to host patients of all specialties (pooled units, high level).

In the remainder of the paper, when a resource is managed flexibly, we will say that the hospital implements a *flexible practice* with respect to such a resource. Combining factors and factor levels, we obtained eight (=2^3) configurations. For each of them we ran the optimisation model in correspondence with 30 randomly generated instances. These instances were obtained starting from real data coming from the Meyer University Children's Hospital (hereinafter Meyer Hospital) a leading Italian hospital. The remainder of the paper is organised as follows: in Section 2, we provide a brief review of the literature. In Section 3, we describe the optimisation models. In Section 4, we illustrate the experimental campaign. In Section 5, we present the empirical results and in Section 6 we discuss them. Subsequently, in Section 7, we draw the conclusions and outline the direction of our future research efforts.

2 Literature review

The master surgical scheduling problem has been the object of several studies (see the reviews of Cardoen et al. (2010), Guerriero and Guido (2011), May et al. (2011)). In Table 1, we review the most important mathematical models supporting the production of MSS that appeared in the literature. Each column of the table (except the last one) represents a resource, while each row represents a model. In each cell, we specify if and how the resource is modelled. When a resource is not explicitly considered in the model, the cell contains "NEC." In the last column of the table, instead, we report the methodology adopted in the relevant study.

In order to emphasise similarities and differences between our study and the related literature, we have added a row representing our model. When a study proposes both flexible and rigid approaches to manage a resource, we report all the alternatives in the table. Table 1 reveals that most of the authors considered three main critical resources in their models: *surgical teams, ORs* and *surgical units' beds*. Therefore, the remainder of this review will focus on these resources.

Table 1 – MSS literature review: resources modelled and operational assumptions

Paper	Surgical teams	OR	Surgical units' beds	Other resources	Type of analysis and solution technique
Our study	Number of sessions per surgical specialty bounded on a daily and on a weekly basis Session assignment performed: - Once and then considered as fixed (low flex) - Every time the MSS is produced, but only limited changes are allowed with respect to a predefined assignment (high flex)	Fully interchangeable ORs Two sessions per day/OR Sessions: - Dedicated (low flex) - Mixed (high flex)	Three types of surgical units (one day surgery unit and two regular units). - All units are dedicated to specific patient types, no mismatch allowed (low flex) - Regular units are pooled (high flex)	NEC	Single criterion exact optimisation, scenario analysis Mixed integer programming
Blake et al. (2002)	Number of sessions per surgical specialty bounded on a daily and on a weekly basis Session assignment performed once and then kept constant in the following period	Partially interchangeable ORs One session per day/OR Mixed sessions	NEC	Medical equipment	Single criterion heuristic optimisation, scenario analysis Mixed integer programming, constructive heuristic
Vissers et al. (2005)	NEC	Fully interchangeable ORs One session per day/OR Mixed sessions	Two types of surgical units (ICU and regular unit) Dedicated units, no mismatch allowed	ICU nursing staff	Single criterion exact optimisation Mixed integer programming
Santibáñez et al. (2007)	Number of sessions per surgical specialty bounded on a daily and on a monthly basis Session assignment performed once and then considered as fixed	Partially interchangeable ORs One or two sessions per day/OR Mixed sessions	Two types of surgical units (SCU and regular unit) Dedicated units, no mismatch allowed	NEC	Single criterion exact optimisation Mixed integer programming
van Oostrum et al. (2008)	NEC	Fully interchangeable ORs One session per day/OR Mixed sessions	Two types of surgical units (ICU and regular unit) Dedicated units, no mismatch allowed	NEC	Multi-criteria exact optimisation, multi-criteria heuristic optimisation Mixed integer programming, column generation, decomposition approach
Beliën et al. (2009)	Number of sessions per surgical specialty bounded on a weekly basis Session assignment performed once and then considered as fixed	Fully interchangeable ORs One or more sessions per day/OR Mixed sessions	Several types of surgical units Dedicated units, no mismatch allowed	NEC	Multi-criteria heuristic optimisation, goal programming Simulated annealing
Tànfani and Testi (2010)	Number of sessions per surgical specialty bounded on a weekly basis Session assignment performed every time MSS is produced	Fully interchangeable ORs One or two sessions per day/OR Mixed sessions	Two types of surgical units (ICU and regular unit) Dedicated units, no mismatch allowed	NEC	Single criterion heuristic optimisation Constructive heuristic
Banditori et al. (2013)	Number of sessions per surgical specialty bounded on a daily basis Session assignments performed every time MSS is produced	Partially interchangeable ORs Two sessions per day/OR Mixed sessions	Three types of surgical units Dedicated units, no mismatch allowed	NEC	Multi-criteria hierarchical exact optimisation, scenario analysis Mixed integer programming, discrete event simulation
Agnetis et al. (2012)	Number of sessions per surgical specialty bounded on a daily and on a weekly basis Session assignment performed: - Once and then considered as fixed (low flex) - Every time the MSS is produced, but only limited changes are allowed with respect to a predefined assignment (medium flex) - Every time the MSS is produced without limiting the changes allowed with respect to a predefined assignment (high flex)	Partially interchangeable ORs One or two sessions per day/OR Dedicated sessions	NEC	NEC	Single criterion exact optimisation, scenario analysis Mixed integer programming

Surgical teams, i.e. the teams of surgeons belonging to the same specialty that actually carry out surgeries are considered explicitly in all but two models (i.e. the model of Vissers et al. (2005) and van Oostrum et al. (2008)). In the remaining works, the availability of surgical teams is modelled by limiting the number of sessions that each surgical specialty can perform on a weekly basis and/or daily basis. Based on these constraints, almost all models assign sessions to specialties, thereby identifying *when* a surgical team will potentially operate in the planning horizon (*session assignment*). In addition, some models (Santibáñez et al., 2007, van Oostrum et al., 2008, Banditori et al., 2013) also determine the *type* and/or the *number* of surgeries that surgical teams will execute in each session (*surgery types assignment*). In (Agnetis et al., 2012), instead, one of the proposed models assumes that the *session assignment* has already been done and, consequently, supports the surgery types assignment only. Most studies suggest that the *session assignment* should be carried out once and should not be changed frequently (Guerriero and Guido, 2011). The underlying assumption of these studies is that it is not technically feasible to change the *session assignment* on a monthly (or more frequent) basis because it would make it very complex for surgeons to coordinate their activities inside and outside the OT (van Oostrum et al., 2010). Nonetheless, Agnetis et al. (2012) demonstrate that small and frequent changes in the *session assignment* can yield substantial benefits and that these benefits are higher than those associated with large yet less frequent changes. Therefore, the authors argue that a limited amount of flexibility in managing surgical teams can produce benefits that are higher than the organisational cost of implementing this solution. For that reason, we decided to include this latter case in our study and compare it with the case where the session assignment is considered as already having been performed.

Contrary to surgical teams, ORs are considered as critical in all the reviewed models. However, different authors model these resources in different ways. A first distinction is between *interchangeable* and *partially interchangeable ORs.* The former can host every type of surgery; the latter, instead, can host only a limited subset of surgeries and/or specialties. A second distinction pertains to how OR time is divided into sessions. Some authors consider one session per OR per day (van Oostrum et al., 2008), some consider two (Santibáñez et al., 2007) or more (Beliën et al., 2009) sessions per OR per day, while others allow both daily sessions and shorter sessions (Agnetis et al., 2012). A third distinction concerns the types of surgery that can be performed in the same OR session. For example, Agnetis et al. (2012) distinguishes two macro-types of surgeries: *general surgeries* and *day surgeries.* The former includes all the procedures leading to a LoS of at least two days (one night), and the latter includes those procedures associated with a LoS of just one day. Based on this distinction, Agnetis et al. (2012)'s model allows only *dedicated sessions*, meaning that within the same session it is not possible to execute both *day-surgeries and general surgeries*. Instead, other models (e.g. Banditori et al. (2013)) allow *mixed sessions* where these types of surgeries can coexist. While the interchangeability of an OR depends on the structural characteristics (e.g. the presence of certain equipment) of the OR itself, hospital managers have more degrees of freedom in deciding how to subdivide the OR time. Nonetheless, this decision is influenced by the actual number of surgical teams available for each specialty (Banditori et al., 2013). For example, all-day-long sessions cannot be planned for those specialties relying on less than two surgical teams per day (except in extraordinary cases, one team cannot operate for the entire day). The decision to organise dedicated or mixed sessions, instead, is generally free. The literature suggests that surgeons usually prefer dedicated sessions; surgeons, in fact, can reduce surgery time because of the repetitive nature of their work (Hans et al., 2008). On the other hand, a mixed session makes the scheduling process less constrained and as such, it potentially allows scheduling a greater number of surgeries. In this study, we explore both options.

Finally, *surgical units*, i.e. the facilities where patients are cared for following surgical procedures, are considered in six out of eight models. These units are usually classified based on the intensity of care required by the hospitalised patients: e.g. ICUs, day-surgery units, regular units. Moreover, these units are characterised by a given capacity that is expressed in terms of the number of beds. Certain hospitals (e.g. Meyer Hospital) allocate patients to the regular units based on the specialty. Such a practice makes it easier and faster for surgeons to control and visit their hospitalised patients. Different models assume different numbers of units and unit types. All the reviewed models except Banditori et al. (2013) constrain each type of patient to be hospitalised into a specific unit. In general, the literature (Vincent et al., 1998) suggests that it is risky to accommodate patients requiring thorough care in units characterised by reduced nursing staff or that are physically located far away from the intensive care unit. Thus, units should be pooled only if they are characterised by similar care settings, which is the flexible practice explored in this study. Banditori et al. (2013)'s model, instead, violates this recommendation and allows bed mismatches whenever they allow increasing the OT throughput.

According to Table 1, it can be argued that flexible practices are considered in several studies. However, no study proposes an analysis that investigates how different flexible practices can interact. With this study, we aim to address this literature gap. In sum, our study (i) proposes a model that considers critical resources that are included in the vast majority of the other studies; (ii) investigates three flexible practices that are reasonable and justified in light of the extant literature but that previous studies have considered only separately or by combining a very limited number of different scenarios (two at maximum); (iii) assesses, in statistical terms, the main and the interaction effects of the mentioned practices and to the best of our knowledge is the only study to do so. These facts ensure that the results presented in the next sections can be of value for a wide audience of practitioners and scholars and also that this study adds a significant contribution to the literature.

3 Model description

In this section, we present the mathematical models we have developed. Specifically, first we present a version of the model that does not implement any flexible practice (hereafter referred to as the "rigid model"). Then we show how such a model can be modified to incorporate flexibility with respect to the management of surgical teams, ORs and surgical units.

All the models presented in this work address a twofold problem: (i) determining the number of cases to assign to each OR session of the planning horizon; (ii) determining the surgery group these cases must belong to. The models consider three critical resources: (i) ORs, whose available time is organised in sessions; (ii) surgical units, which accommodate patients after the surgery; (iii) surgical teams, dedicated to one specialty each, whose availability is defined in terms of number of OR sessions. Cases belonging to the same surgery group require the same specialty, the same amount of OR time and will occupy a surgical unit for the same amount of time.

Let us define the following sets and parameters that are common to the rigid model and to its extensions as follows:

W	the set of weeks in the planning horizon, indexed by w
D	the set of days in the planning horizon, indexed by d
T	the set of sessions, indexed by t
O	the set of ORs, indexed by o

S the set of surgical specialties, indexed by s

K the set of surgery groups, indexed by k

M a suitably big constant

H_{odt} the available time of OR o, on day d and session t

F_{bd} the number of beds in the surgical unit b available on day d

L_{sw} the availability of surgical teams with specialty s for week w, expressed in number of OR sessions

s_k the specialty of surgery group k

r_k the typology of surgery group k (short-stay surgery – SS vs. long-stay surgery – LS)

c_k the average surgery duration of surgery group k

b_k, a_k the average number of days of hospitalisation, before and after surgery, required by surgery group k

Y_k the minimum number of procedures of surgery group k to be scheduled.

3.1 Rigid model

In this model, we assume that the *session assignment* has already been done. Consequently, we rely on an *allocation grid G* as an input. Specifically, for each specialty s, day d and session t, G_{sdt} is equal to 1 if specialty s is allocated on day d, session t, and 0 otherwise.

Grid G must respect the following feasibility constraints:

$$\sum_{s\in S} G_{sdt} \le |O| \qquad \forall d \in D, \forall t \in T \qquad (3.1.1)$$

$$\sum_{d=7w-6}^{7w} \sum_{t\in T} G_{sdt} = L_{sw} \qquad \forall s \in S, \forall w \in W \qquad (3.1.2)$$

Constraints (3.1.1) assure that on each day d and session t, the number of specialties assigned to an OR does not exceed the number of available ORs ($|O|$). Constraints (3.1.2) instead control that the number of sessions assigned weekly to a given s is exactly the value resulting from the upstream case-mix planning problem. Then, in the rigid model, an OR *o* has to be assigned to each triple *(s,d,t)* for which $G_{sdt} = 1$. For a matter of convenience, we denote this with the following:

$$\bar{G} = \{(s,d,t) \text{ s.t. } s \in S, d \in D, t \in T \text{ and } G_{sdt} = 1\} \quad .$$

The rigid scheduling model takes the following two main decisions:

1. Assign an OR to each triple *(s,d,t)* in $\bar{G}$
2. Determine, for each surgery group *k*, the number of procedures to schedule in correspondence with each triple *(s,d,t)* in $\bar{G}$ where *s* is the specialty associated with *k*.

Then let us define the following main and auxiliary variables:

q_{go}	binary, 1 if triple $g = (s,d,t)$ in $\bar{G}$ is assigned to OR *o*, 0 otherwise
y_{kdto}	the number of procedures in surgery group *k* assigned to OR *o* on day *d* in time slot *t*
z_{bd}	the number of beds belonging to surgical unit *b* occupied on day *d*
u_{odt}	binary, 1 if OR *o* on day *d* and session *t* is dedicated to short-stay surgeries, 0 otherwise.

Using these variables and parameters, we can state the rigid model as follows:

$$\max \sum_{\substack{k \in K, o \in O, \\ d \in D, t \in T}} y_{kodt} \tag{3.1.3}$$

$$\sum_{s \in S: g=(s,d,t) \in \bar{G}} q_{go} \leq 1 \qquad \forall o \in O, \forall d \in D, \forall t \in T \tag{3.1.4}$$

$$\sum_{o \in O} q_{go} = 1 \qquad \forall g \in \bar{G} \tag{3.1.5}$$

$$\sum_{k \in K: s_k = s} y_{kdto} \leq M q_{go} \qquad \forall g = (s,d,t) \in \bar{G}, \forall o \in O \tag{3.1.6}$$

$$\sum_{k\in K} c_k y_{kdto} \le H_{odt} \qquad \forall o \in O, \forall d \in D, \forall t \in T \qquad (3.1.7)$$

$$\sum_{\substack{k\in K, o\in O,\\ t\in T}} \sum_{d'=\max(1,d-a_k)}^{\min(|D|,d+b_k)} y_{kd'to} = z_{bd} \qquad \forall b \in B, \forall d \in D \qquad (3.1.8)$$

$$z_{bd} \le F_{bd} \qquad \forall b \in B, \forall d \in D \qquad (3.1.9)$$

$$\sum_{k\in K: \eta_k = "SS"} y_{kdto} \le M u_{odt} \qquad \forall o \in O, \forall d \in D, \forall t \in T \qquad (3.1.10)$$

$$\sum_{k\in K: \eta_k = "LS"} y_{kdto} \le M(1-u_{odt}) \qquad \forall o \in O, \forall d \in D, \forall t \in T \qquad (3.1.11)$$

$$\sum_{\substack{o\in O, d\in D,\\ t\in T}} y_{kdto} \ge Y_k \qquad \forall k \in K \qquad (3.1.12)$$

$$q_{go} \in \{0,1\} \qquad \forall g \in \overline{G}, \forall o \in O \qquad (3.1.13)$$

$$y_{kdto} \in \mathrm{N} \qquad \forall k \in K, \forall d \in D, \forall t \in T, \forall o \in O \qquad (3.1.14)$$

$$z_{bd} \in \mathrm{N} \qquad \forall b \in B, \forall d \in D \qquad (3.1.15)$$

$$u_{odt} \in \{0,1\} \qquad \forall o \in O, \forall d \in D, \forall t \in T \qquad (3.1.16)$$

The objective function (3.1.3) maximises the number of procedures scheduled in the planning horizon. Constraints (3.1.4) guarantee that each OR-session can host a specialty at most. Constraints (3.1.5) assure that all the triples (*s*,*d*,*t*) in $\overline{G}$ are assigned to some OR *o*. Constraints (3.1.6) bind together assignment variables *q* and variables *y*: specifically, they state that if the triple (*s*,*d*,*t*) in $\overline{G}$ has not been assigned to OR *o*, then no procedure belonging to a group characterised by specialty *s* can be performed in OR *o*, on day *d* and session *t*. In contrast, when the triple *g*=(*s*,*d*,*t*) in $\overline{G}$ is assigned to OR *o* (q_{go}=1), then the corresponding constraint is redundant since it imposes that the number of procedures of that specialty scheduled in that OR session does not exceed the suitably big constant *M*. Specifically, *M* is set equal to the maximum number of shortest procedures a session can host. Constraints (3.1.7) guarantee that the total duration of the procedures scheduled in an OR session does not exceed the available time of that OR session. Constraints (3.1.8) and (3.1.9) are used to properly manage beds; specifically, for each day *d* and surgical unit *b*, they respectively compute the number z_{bd} of beds occupied and limit such a number to the bed availability F_{bd}. To correctly determine the bed occupancy on a given day d, we have to consider all the patients whose stay in the surgical units, before (b_k) and after surgery (a_k), overlaps day *d*. More specifically, in a given day *d*, we have to consider the beds occupied by patients who have undergone a surgery before day *d* and who are still in the hospital on day *d* as well as all the patients who will undergo surgeries after day *d* and that have been pre-hospitalised, in addition to the patients that undergo a surgery exactly on day *d*. Constraints (3.1.10) and (3.1.11) refer to the management of dedicated sessions, and they assure that in a given OR session, long-stay and short-stay surgeries are mutually exclusive. In fact, the binary variable u_{odt} is equal to 1 if OR *o* on day *d* and session *t* is dedicated to short-stay surgeries. In this case, constraints (3.1.11) assure that in that OR session no long-stay surgery is performed. In contrast, when u_{odt} is equal to 0, the corresponding OR session can host only long-stay surgeries. Constraints (3.1.12) relate to target efficiency and they impose that for each surgery group *k* the number of procedures performed is not smaller than the target value Y_k. Indeed, the MSS must guarantee to schedule a minimum number of surgeries for each surgery group. Such a requisite is set to avoid solutions planning an excessive number of surgeries belonging to easy-to-schedule surgery groups (i.e. groups characterised by short ST and LoS). This method ensures a reasonable waiting time for patients of each group and allows distributing complex-to-schedule surgeries over time.

Finally, constraints (3.1.13), (3.1.14), (3.1.15), and (3.1.16) define the domain of the variables.

In the following section, we describe how to extend/modify the rigid model in order to take into account the flexible practices discussed in the previous section.

3.2 Flexibility with respect to surgical teams

In this scenario, differently from the rigid model, the allocation grid is not an input for the scheduling model. Instead, the grid is the output of the model that decides the specialty to assign to each OR, day and session in the planning horizon. However, only limited variations with respect to the original grid are allowed in order to guarantee that the new grid is still implementable. To this aim, the following variables are defined:

x_{sdto} binary, 1 if specialty *s* is assigned to OR *o* on day *d* and session *t*, 0 otherwise

x^{+}_{sdt} binary, 1 if a swap from 0 to 1 occurs with respect to G_{sdt}, 0 otherwise.

All the constraints that in the rigid model are implicitly satisfied by the pre-defined grid G have now to be explicitly guaranteed through the following set of constraints:

$$\sum_{s \in S} x_{sodt} \le 1 \qquad \forall o \in O, \forall d \in D, \forall t \in T \qquad (3.2.1)$$

$$\sum_{o \in O} x_{sodt} \le 1 \qquad \forall s \in S, \forall d \in D, \forall t \in T \qquad (3.2.2)$$

$$\sum_{o \in O, t \in T} \sum_{d=7w-6}^{7w} x_{sodt} = L_{sw} \qquad \forall s \in S, \forall w \in W \qquad (3.2.3)$$

$$\sum_{k \in K: s_k = s} y_{kodt} \le M x_{sodt} \qquad \forall s \in S, \forall o \in O, \forall d \in D, \forall t \in T \qquad (3.2.4)$$

$$x_{sodt} \in \{0,1\} \qquad \forall s \in S, \forall o \in O, \forall d \in D, \forall t \in T \qquad (3.2.5)$$

Specifically, constraints (3.2.1) assure that each OR on each day and in each session of the planning horizon is assigned to at most one specialty. Constraints (3.2.2) guarantee that each specialty is assigned to at most one OR in each day and session. Constraints (3.2.3) impose that the number of sessions assigned weekly to a given specialty *s* is exactly the value resulting from the upstream case-mix planning problem. Constraints (3.2.4) bind together assignment (*x*) and scheduling (*y*) variables; specifically, they assure that no procedure with specialty *s* is scheduled in OR *o*, on day *d* and session *t* unless specialty *s* has been assigned to that OR, on that day and session. Conversely, these constraints become redundant when x_{sodt}=1 since they impose that the number of procedures scheduled does not exceed a suitably defined big M. Finally, constraints (3.2.5) define the domain of the assignment variables.

Furthermore, we introduced the following constraints to control the variations with respect to the grid *G*:

$$\sum_{o \in O} x_{sodt} \le G_{sdt} + x^{+}_{sdt} \qquad \forall s \in S, \forall d \in D, \forall t \in T \qquad (3.2.6)$$

$$\sum_{d \in D, t \in T} x^{+}_{sdt} \le \overline{A} \qquad \forall s \in S \qquad (3.2.7)$$

Specifically, constraints (3.2.6) allow that any variation of element G_{sdt} can occur. In particular, if $G_{sdt} = 0$, i.e. if specialty s is not allocated to day d, session t, the new grid defined through variables x may allow that specialty s is assigned to some OR in that day and session. When this variation occurs, variable x^{+}_{sdt} takes value 1 and it accounts for a zero to one swap with respect to G. In addition, x_{sodt} specifies the OR o to which specialty s is assigned in day d, session t.

One to zero swaps, instead, do not need to be explicitly controlled. In fact, since we hypothesize that the number of sessions dedicated to each specialty in the planning period is constant, when a one to zero swap occurs also a zero to one swap takes place and this latter swap is controlled by x^{+}_{sdt} as well. The number of zero to one swaps affecting the specialty s cannot exceed the maximum number $\overline{A}$ of allowed variations (see constraints (3.2.7)).

3.3 Flexibility with respect to ORs

To implement this type of flexibility, it is sufficient to remove constraints (3.1.10) and (3.1.11), thus enlarging the feasibility region and allowing both short-stay and long-stay surgeries to be scheduled in the same session.

3.4 Flexibility with respect to surgical units

Each procedure is associated with a surgical unit. If surgical units are managed flexibly, then they are pooled. With this method, patients can be hospitalised in units that differ from the one originally assigned to them. To model this practice, we introduce the following variables:

$v_{bb'd}$ the number of beds of surgical unit b' used in place of beds of surgical unit b on day d.

Constraints (3.1.9) in the rigid model are then updated with constraints (3.4.1). These constraints allow that on a given day for a given surgical unit the number of beds occupied in that unit exceeds capacity. Moreover, we add constraints (3.4.2), which limit the overall number of beds occupied to the overall bed availability.

$$z_{bd} \le F_{bd} + \sum_{b' \in B: b' \ne b} v_{bb'd} \qquad \forall b \in B, \forall d \in D \qquad (3.4.1)$$

$$\sum_{b \in B} z_{bd} + \sum_{b,b' \in B: b \ne b'} v_{bb'd} \le \sum_{b \in B} F_{db} \qquad \forall d \in D \qquad (3.4.2)$$

4 Methodology

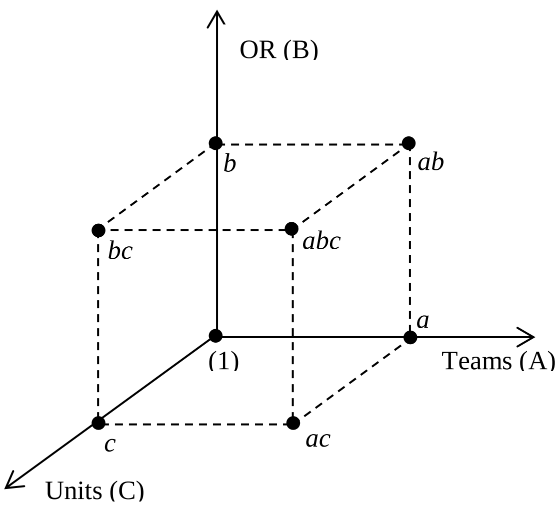
To assess the effects of the implementation of flexible practices on MSS efficiency, we use a 2^3 factorial design comprising the following:

- Three factors: *Teams, ORs, Units***.** Each factor corresponds to one of the critical resources incorporated in the model.
- Two possible levels for each factor: *high* when the resource is managed in a flexible way and *low* otherwise.
- One response variable, i.e. the number of surgeries scheduled.

Factors and factor levels are reported in Table 2, and the experimental design is illustrated in Figure 1.

Table 2 - Factorial design

Symbol	Factor Name	Low level	High level
A	Teams	Fixed surgical teams assignment. The allocation grid is fixed.	Variable surgical teams assignment. At maximum, one swap per specialty is allowed with respect to a predefined allocation grid.
B	ORs	Dedicated sessions. OR can host either short-stay or long-stay surgeries.	Mixed sessions. OR can host both short-stay and long-stay surgeries.
C	Units	Dedicated units. Unit 1 and Unit 2 are dedicated to different types of long-stay patients.	Pooled units. Unit 1 and Unit 2 are used interchangeably.



Treatment	Factors		
	Teams (A)	**ORs (B)**	**Units (C)**
(1)	Low	Low	Low
a	High	Low	Low
b	Low	High	Low
c	Low	Low	High
ab	High	High	Low
ac	High	Low	High
bc	Low	High	High
abc	High	High	High

Figure 1 Experimental design

Each factor is associated with an uppercase letter (A, B, C). Each vertex of the cube represents a treatment. Treatments are labelled according to the Montgomery and Runger's (2003, p.524) notation. According to this notation, a treatment combination is represented by a series of lowercase letters. If a letter is present, the corresponding factor is run at the high level in that treatment combination; if it is absent, the factor is run at its low level. The treatment combination with all the factors at the low level is represented by (1).

To implement the different treatments, the optimisation model is extended as described in Section 3. For each treatment, we have analysed the result of the optimisation model in correspondence of 30 randomly generated instances. These instances differ in each other's in terms of allocation grid G.

We coded the optimisation models in AMPL and solved them through the IBM ILOG Cplex Solver (version 12.4) running on a personal computer equipped with an Intel Core i7 processor and 8 GB of RAM. For each optimisation run, we bound the computational time to 1 hour. The results of our experimental campaign are presented in the next section.

5 Empirical results

In this section, we present the data we used to run the optimisation model(s) and the results of the experiments.

5.1 Input data

As we pointed out in the introduction, our study was inspired by Meyer Hospital. Such a hospital is characterised by the following features:

(i) *12 surgical specialties.* Each surgical specialty is associated with surgical teams that can cover a certain number of sessions per week.

(ii) *38 surgery groups*. Surgery groups have been created following Banditori et al.'s (2013) methodology. For each surgery group (k), we calculated the mean value of LoS and ST and used these values to set the parameters a_k and c_k of the optimisation models, respectively.

(iii) A *planning horizon* of two weeks.

(iv) Defined lower bounds (Y_k) for the number of surgeries to schedule within the planning horizon for each surgery group k. These lower bounds are fixed by the hospital's top management on a yearly basis.

(v) 3 *surgical units*: a day surgery unit and two regular units (Unit 1 and 2). The day surgery unit contains 14 beds, and Unit 1 and Unit 2 contain 19 and 14 beds, respectively.

(vi) The day surgery unit can host only short-stay patients, i.e. patients whose expected LoS is one day (no night), regardless of the speciality. In contrast, Units 1 and 2 can accommodate long-stay patients only for certain specialties. Long-stay patients can be hospitalised either in Unit 1 or in Unit 2, and mismatches are not allowed.

(vii) *4 interchangeable ORs* dedicated to elective patients. Each OR is open 10 hours a day, 5 days per week. The OR time is subdivided into two sessions, morning and afternoon. Additional OR sessions and beds are allocated to non-elective patients (emergencies and urgencies).

(viii) *OR* sessions are "dedicated," i.e. in a session where long-stay surgeries are performed, no short-stay surgery can be scheduled and vice versa. In addition, afternoon sessions can host only long-stay surgeries, while morning sessions can host both long-stay and short-stay surgeries.

(ix) An allocation grid *G* that fixes the specialty to assign to each OR session.

(x) No deviation from the allocation grid *G* is tolerated.

Features (i, ii, iii, iv, v, and vii) do not change across treatments and instances. Features (vi, viii and x) change depending on the treatment, as described in Table 2. Feature (ix) changes according to the instance, which is randomly generated. The Meyer Hospital case corresponds to the treatment (1) in Table 2.

5.2 Optimisation output

Table 3 shows the results of the optimisation models. It displays the mean values, calculated across instances, of the scheduled surgeries and of the optimality gap. In addition, for each treatment, the table shows the number of instances for which the optimisation model found the optimal solution and the minimum and the maximum optimality gaps across the 30 instances.

As can be seen for some treatments and instances, it was not possible to find an optimal solution within the fixed time limit. Nonetheless, the mean optimality gap associated with each treatment never exceeds 3.6%.

Table 3 Optimisation Output

Treatment	Mean of scheduled surgeries	Mean of optimality gap	Optimal solutions found	Min of optimality gap	Max of optimality gap
(1)	272.1	0.0%	30/30	0.0%	0.0%
a	280.2	2.7%	0/30	1.8%	3.6%
b	278.2	0.0%	30/30	0.0%	0.0%
c	274.7	0.0%	30/30	0.0%	0.0%
ab	286.7	0.5%	7/30	0.0%	1.4%
ac	281.9	2.6%	0/30	1.4%	3.6%
bc	279.2	0.0%	30/30	0.0%	0.0%
abc	287.0	0.5%	5/30	0.0%	1.4%

The table shows that when moving from treatment (1) to treatment *abc*, the number of surgeries scheduled increases by 14.9. Therefore, implementing all the mentioned flexible practices yields, on average, a monthly increase of around 30 surgeries. To interpret these results, we performed an analysis of variance (ANOVA) and assessed the statistical significance and the magnitude of all main and interaction effects. Moreover, we carried out several Tukey's post-hoc tests to compare treatments with each other and rank them in terms of scheduled surgeries while controlling the familywise error rate (Field, 2005, p.310) to a 0.05 level. These statistical analyses are presented in the next section.

5.3 Statistical analysis

Table 4 displays the complete ANOVA table including the magnitude of the estimated effects and their level of significance. The ANOVA analysis included an accurate check of the assumptions of normality of error terms and homogeneity of variance. More specifically, we carried out a Ryan-Joiner test and failed to reject (p=0.099) the null hypothesis of normally distributed errors. Similarly, we performed the Levene's test and failed to reject the null hypothesis of the variances being equal (p=0.087).

Table 4 Analysis of variance and effects summary table for scheduled surgeries

	DF	Sum of squares	Mean squares	F	Effect	p
Teams	1	3744.6	3744.6	2378.4	7.9	0.000 (*)
Ors	1	1837.1	1837.1	1166.8	5.5	0.000 (*)
Units	1	117.6	117.6	74.7	1.4	0.000 (*)
Teams*Ors	1	4.8	4.8	3.1	0.3	0.080
Teams*Units	1	8.8	8.8	5.6	-0.4	0.020 (*)
Ors*Units	1	33.8	33.8	21.4	-0.8	0.000 (*)
Teams*Ors*Units	1	0.1	0.1	0.0	0.0	0.837
Error	232.0	365.3	2.1			
Total	239.0	6112.0				

(*) significant at the $\alpha = 0.05$ level

Table 4 shows that, assuming an $\alpha = 0.05$ significance level, all the main effects are significant ($p<0.05$). Similarly there is a significant, yet negative, interaction effect between Teams factor (A) and the Units factor (C) ($p=0.020$) and between the ORs factor (B) and Units factor (C) ($p=0.000$). Other 2-way and 3-way interaction effects, instead, are not statistically significant ($p>0.05$). The main and interaction effects are plotted in Figure 2 and Figure 3, respectively.

Looking at the main effects (Figure 2), it can be noted that, on average, an increase in the level of each factor leads to an increase in the number of surgeries scheduled. For example, when the Teams factor (A) is run at a high level (i.e. treatments a, ab, ac, abc), the model schedules, on average, 283.9 surgeries. Instead, when the Teams factor (A) is run at its low level (i.e. treatments b, c, bc, (1)), the model schedules, on average, 276 surgeries (in fact, main effect (A) =283.9-276= 7.9)

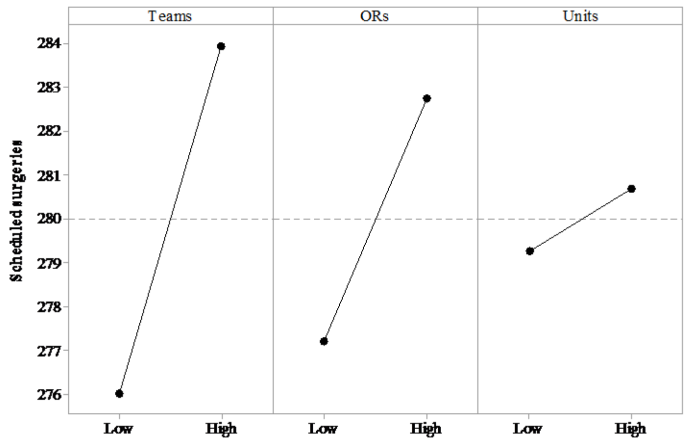


Figure 2 Main effects for scheduled surgeries, mean values

However, as in our case, when one or more significant interaction effects are present, the interpretation of the main effects can be incomplete or misleading. In fact, when an interaction factor is significant, the impact of one factor depends on the level of another factor. For example, in our case, the significant interaction between B and C factors implies that the effect on scheduling surgeries (dependent variable) of B depends on the level of C and vice versa. In particular, since the interaction effect is negative, increasing B when C is at a high level leads, on average, to a variation in terms of the number of scheduled surgeries that is significantly smaller than the variation obtained by increasing B when C is at a low level. If this latter variation were negative, i.e. if increasing B when C is low would lead to a decrease in the surgeries scheduled, then the interpretation of the main effects would be completely misleading. In this latter case, in fact, increasing B from low to high in the presence of a high level of C would determine a decrease of the surgeries scheduled, which is the opposite of what one would expect looking at the main effects of B and C. To prevent misleading interpretations of the main effects, however, it is sufficient to observe the interaction graphs in Figure 3.

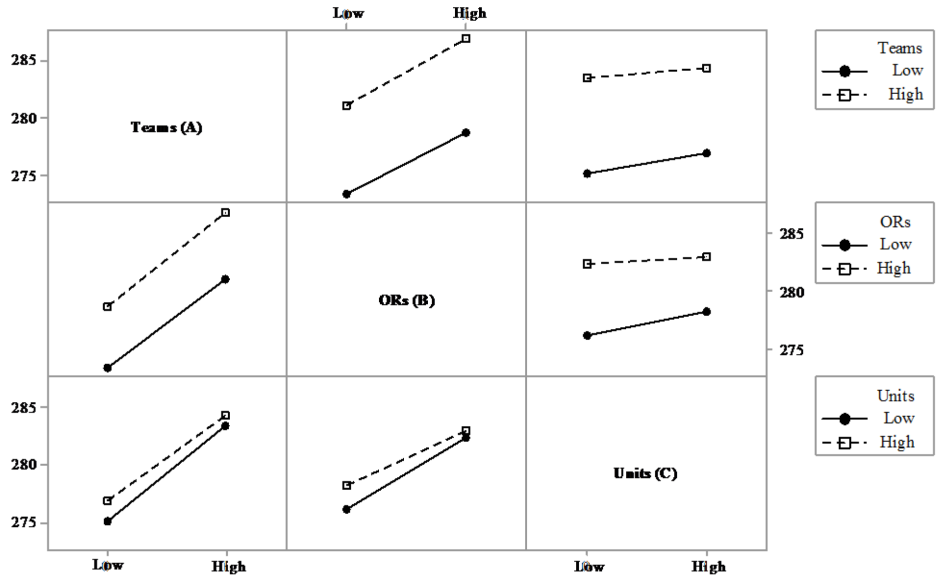


Figure 3 Interaction plot for scheduled surgeries, data means

In Figure 3, the lines in each cell do not cross. Therefore, for each factor, the number of surgeries scheduled is, on average, higher when the factor is high than when the factor is low, regardless of the level of the other factors. Therefore, moving a factor from low to high leads to a benefit in terms of scheduled surgeries, regardless of the levels of the other factors.

To compare treatments with each other and rank them, we used the Tukey's post hoc procedure. This procedure allows us to compare all different combinations of the treatment groups and to control the familywise error rate without sacrificing the statistical power. For each pairwise comparison, we assigned the same rank (1, 2, etc.) to those treatment groups for which the post-hoc test did not allow for the identification of a significant ($p>0.05$) difference between the number of scheduled surgeries. The results of these tests are shown in Table 5 and will be discussed in the next section along with their practical implications.

Table 5 Pairwise comparisons, grouping information using Tukey's method and 95.0% confidence level

Comparisons	Treatment group code	Treatment groups	N	Mean	Rank
1	(1.1)	a, ab, ac, abc	120	283.9	1
	(1.2)	b, c, bc, (1)	120	276	2
2	(2.1)	ab, abc, bc, b	120	282.8	1
	(2.2)	ac, a, c, (1)	120	277.2	2
3	(3.1)	abc, ac, bc, c	120	280.7	1
	(3.2)	ab, a , b, (1)	120	279.3	2
4	(4.1)	ab, abc	60	286.9	1
	(4.2)	ac, a	60	281	2
	(4.3)	bc, b	60	278.7	3
	(4.4)	c, (1)	60	273.4	4
5	(5.1)	abc, ac	60	284.5	1
	(5.2)	ab, a	60	283.4	2
	(5.3)	bc, c	60	276.9	3
	(5.4)	b, (1)	60	275.1	4
6	(6.1)	abc, bc	60	283.1	1
	(6.2)	ab, b	60	282.4	2
	(6.3)	ac, c	60	278.3	3
	(6.4)	a, (1)	60	276.1	4
7	(7.1)	abc	30	287	1
	(7.2)	ab	30	286.7	1
	(7.3)	ac	30	281.9	2
	(7.4)	a	30	280.2	3
	(7.5)	bc	30	279.2	4
	(7.6)	b	30	278.2	5
	(7.7)	c	30	274.7	6
	(7.8)	(1)	30	272.1	7

6 Discussion

Looking at pairwise comparisons 1 to 3 in Table 5, emerges that, between the three investigated flexible practices, the one that, *on average*, leads to the largest increase in the number of surgeries scheduled is the variable surgical teams assignment (groups 1.1 vs. 1.2). This practice is followed by the introduction of mixed session (groups 2.1 vs. 2.2) and by the surgical units pooling (groups 3.1 vs. 3.2).

The pairwise comparisons 4 in Table 5, in their turn, reveal that, *on average*, the introduction of a variable surgical teams assignment significantly increases the number of surgeries scheduled both when ORs can host mixed sessions (groups 4.1 vs. 4.3) and when ORs are organised into dedicated sessions (groups 4.2 vs. 4.4). Similarly, they reveal that introducing mixed sessions increases the number of surgeries scheduled in the presence of both a variable surgical teams assignment (groups 4.1 vs. 4.2) and fixed surgical teams assignment (groups 4.3 vs. 4.4). However, the increase that can be obtained introducing a variable surgical teams assignment is larger than the one that can be obtained introducing mixed sessions (groups 4.2 vs. 4.3).

Similarly, the pairwise comparisons 5 in Table 5, show that, *on average*, introducing a variable surgical teams assignment increases the number of surgeries scheduled both when surgical units are pooled (groups 5.1 vs. 5.3) and when they are not (groups 5.2 vs. 5.4). Similarly, pooling surgical units increases the number of surgeries scheduled both in presence of a variable (groups 5.1 vs. 5.2) and a fixed surgical teams assignment (groups 5.3 vs. 5.4). The increase that can be obtained introducing a variable surgical teams assignment is larger than the one that can be obtained by pooling surgical units (groups 5.2 vs. 5.3).

In the same way, the pairwise comparisons 6 in Table 5, show that, *on average*, introducing mixed sessions increases the number of surgeries scheduled, both when surgical units are pooled (groups 6.1 vs. 6.3) and when they are not pooled (groups 6.2 vs. 6.4). Similarly, pooling surgical units increases the number of surgeries scheduled, both in the presence of dedicated sessions (groups 6.1 vs. 6.2) and in presence of mixed sessions (groups 6.3 vs. 6.4). The increase that can be obtained introducing mixed sessions is larger than those that can be obtained by pooling surgical units (groups 6.2 vs. 6.3)

Finally, from the pairwise comparisons 7 in Table 5 emerges that for hospitals where no flexible practices are implemented, the best results in terms of surgeries scheduled can be achieved by introducing flexibility with respect to surgical teams and ORs (groups 7.2 vs. 7.8). In fact, once these two flexible practices are implemented, pooling surgical units does not yield any significant additional advantage (groups 7.1 vs. 7.2). On the other hand, if mixed session cannot be implemented, then pooling surgical units significantly increases the number of surgeries scheduled both when surgical teams are managed flexibly (groups 7.3 vs. 7.4) and when they are not (groups 7.7 vs. 7.8). Equivalently, if surgical teams cannot be managed flexibly, then pooling surgical units significantly increases the number of surgeries scheduled both when ORs are managed flexibly (groups 7.5 vs. 7.6) and when they are not (groups 7.6 vs. 7.8). Finally, for hospitals where no flexible practice is implemented, introducing flexibility with respect to surgical teams leads to an increase in the scheduled surgeries that is statistically larger than the one that can be obtained by introducing flexibility with respect to ORs and surgical units (groups 7.4 vs. 7.5).

As a final remark, it is worth mentioning that the *statistical significance* of an effect does not necessarily imply that such an effect is also *practically relevant*. The post-hoc test, in fact, reveals if the difference between the mean number of surgeries associated with two treatment groups is statistically different from zero. A difference greater than zero (say, one surgery in two weeks) is not necessarily practically relevant and does not necessarily imply that the associated flexible practice deserves to be implemented. Indeed, the benefits that are possible to obtain with a flexible practice should always be traded off with the costs of implementation. For example, the sessions assignment is often the output of a lengthy and complex negotiation process between stakeholders (surgeons, management, nursing staff) with different priorities and needs. Thus to avoid conflicts, a hospital could also decide to renounce the potential benefits of implementing a variable surgical teams assignment.

7 Conclusion and future research

In this study, we presented a novel mixed integer programming model to address the master surgical scheduling problem. In addition, we evaluated the impact in terms of scheduled surgeries of the implementation of different combinations of three flexible practices: (i) variable surgical teams assignment, (ii) mixed sessions and (iii) pooled surgical units.

Our analysis revealed that to maximise the number of scheduled surgeries it is sufficient to introduce a variable surgical teams assignment *and* mixed sessions. In fact, if both these practices are implemented, pooling surgical units carries no additional advantages. However, if only one of these flexible practices (or none) is implemented, then pooling surgical units produces significant benefits. Moreover, the analysis showed that, if a hospital cannot implement a variable surgical teams assignment, then it can still improve its efficiency by introducing mixed sessions and, similarly, if it cannot implement mixed sessions, it can improve its efficiency by introducing a variable surgical teams assignment.

This study considers hospital features that are included in the vast majority of the contributions available in the master surgical scheduling literature and explores flexible practices that are reasonable according to such a literature. Moreover, it is the first study to propose a systematic analysis of the effect of the implementation of these practices. As such, both the presented model and the implications of the analysis can be of interest for a wide audience of practitioners and scholars.

Of course, this study is not without limitations. First, we investigated only a limited number of hospital settings. For example, we neglected to factor in certain hospital resources (e.g. ICU, electro-medical devices) that are not considered critical at Meyer Hospital but that may be highly critical in other hospitals. Second, we have not investigated how the MIP model would perform in terms of computational time if the problem dimension increases, e.g. if the planning horizon is extended to one month or if the number of ORs and beds increases. Finally, we only considered elective patients. Nonetheless, it might be interesting to investigate how the implementation of flexible practices could help improve hospital performance in presence of emergencies, urgencies and no-shows (Stuart and Kozan, 2012). The extension of the computational campaign to other hospital settings, the analysis of the optimisation model scalability, the design of ad-hoc methodologies to cope with large scale instances and the incorporation of non-elective patients in the analysis will certainly be the object of our future research efforts.

Acknowledgements

The authors are grateful to Meyer University Children's Hospital, and in particular to General Director Dr. Tommaso Langiano, and to the Medical Director Dr. Monica Frassineti for supporting the research project that has inspired this study.

The authors would also express their sincere gratitude to the anonymous referees for their insightful critical comments and suggestions.

References

AGNETIS, A., COPPI, A., CORSINI, M., DELLINO, G., MELONI, C. & PRANZO, M. 2012. Long term evaluation of operating theater planning policies. *Operations Research for Health Care,* 1**,** 95-104.

BALASUBRAMANIAN, H., MURIEL, A. & WANG, L. 2012. The impact of provider flexibility and capacity allocation on the performance of primary care practices. *Flexible Services and Manufacturing Journal,* 24**,** 422-447.

BANDITORI, C., CAPPANERA, P. & VISINTIN, F. 2013. A combined optimization–simulation approach to the master surgical scheduling problem. *IMA Journal of Management Mathematics,* 24**,** 155-187.

BELIËN, J. & DEMEULEMEESTER, E. 2007. Building cyclic master surgery schedules with leveled resulting bed occupancy. *European Journal of Operational Research,* 176**,** 1185-1204.

BELIËN, J., DEMEULEMEESTER, E. & CARDOEN, B. 2009. A decision support system for cyclic master surgery scheduling with multiple objectives. *Journal of Scheduling,* 12**,** 147-161.

BLAKE, J. T., DEXTER, F. & DONALD, J. 2002. Operating room managers' use of integer programming for assigning block time to surgical groups: A case study. *Anesthesia & Analgesia,* 94**,** 143-148.

BUZACOTT, J. A. & MANDELBAUM, M. 2008. Flexibility in manufacturing and services: achievements, insights and challenges. *Flexible Services and Manufacturing Journal,* 20**,** 13-58.

CARDOEN, B., DEMEULEMEESTER, E. & BELIËN, J. 2010. Operating room planning and scheduling: A literature review. *European Journal of Operational Research,* 201**,** 921-932.

CHOU, M. C., TEO, C.-P. & ZHENG, H. 2008. Process flexibility: design, evaluation, and applications. *Flexible Services and Manufacturing Journal,* 20**,** 59-94.

DENTON, B., VIAPIANO, J. & VOGL, A. 2007. Optimization of surgery sequencing and scheduling decisions under uncertainty. *Health Care Management Science,* 10**,** 13-24.

FIELD, A. 2005. *Discovering statistics using SPSS,* London, SAGE Publications.

GLOUBERMAN, S. & MINTZBERG, H. 2001. Managing the care of health and the cure of disease-Part I: Differentiation. *Health care management review,* 26**,** 56-69.

GUERRIERO, F. & GUIDO, R. 2011. Operational research in the management of the operating theatre: a survey. *Health Care Management Science,* 14**,** 89-114.

GUPTA, D. & SHANTHIKUMAR, G. 2008. Editorial. *Flexible Services and Manufacturing Journal,* 20**,** 1-10.

HANS, E., WULLINK, G., VAN HOUDENHOVEN, M. & KAZEMIER, G. 2008. Robust surgery loading. *European Journal of Operational Research,* 185**,** 1038-1050.

LITVAK, E. & LONG, M. C. 2000. Cost and quality under managed care: Irreconcilable differences. *Am J Manag Care,* 6**,** 305-12.

MAY, J. H., SPANGLER, W. E., STRUM, D. P. & VARGAS, L. G. 2011. The surgical scheduling problem: Current research and future opportunities. *Production and Operations Management,* 20**,** 392-405.

MONTGOMERY, D. C. & RUNGER, G. C. 2003. *Applied statistics and probability for engineers*, John Wiley and Sons.

SANTIBÁÑEZ, P., BEGEN, M. & ATKINS, D. 2007. Surgical block scheduling in a system of hospitals: an application to resource and wait list management in a British Columbia health authority. *Health Care Management Science,* 10**,** 269-282.

STUART, K. & KOZAN, E. 2012. Reactive scheduling model for the operating theatre. *Flexible Services and Manufacturing Journal,* 24**,** 400-421.

TÀNFANI, E. & TESTI, A. 2010. A pre-assignment heuristic algorithm for the Master Surgical Schedule Problem (MSSP). *Annals of Operations Research,* 178**,** 105-119.

VAN OOSTRUM, J. M., BREDENHOFF, E. & HANS, E. W. 2010. Suitability and managerial implications of a master surgical scheduling approach. *Annals of Operations Research,* 178**,** 91-104.

VAN OOSTRUM, J. M., VAN HOUDENHOVEN, M., HURINK, J., HANS, E., WULLINK, G. & KAZEMIER, G. 2008. A master surgical scheduling approach for cyclic scheduling in operating room departments. *OR Spectrum,* 30**,** 355-374.

VINCENT, C., TAYLOR-ADAMS, S. & STANHOPE, N. 1998. Framework for analysing risk and safety in clinical medicine. *BMJ: British Medical Journal,* 316**,** 1154.

VISSERS, J., ADAN, I. & BEKKERS, J. 2005. Patient mix optimization in tactical cardiothoracic surgery planning: a case study. *IMA Journal of Management Mathematics,* 16**,** 281-304.