Considering patients' prioritization in the operating room planning

Mariana Bettencourt Medeiros Oliveira

Thesis to obtain the Master of Science Degree in

Biomedical Engineering

Supervisor: Professor Inês Marques Proença

Co-Supervisor: Professor Valérie Bélanger

Examination Committee

Chairperson: Professor Mónica Duarte Correia de Oliveira

Supervisor: Professor Inês Marques Proença

Members of the Committee: Professor Tânia Rute Xavier de Matos Pinto Varela

October 2018

Acknowledgements

I would like to express my deepest gratitude to Professor Inês Marques for all the availability and encouragement provided during this journey. It has been a pleasure to discover the area of operational research with you.

I would also like to thank Professor Valérie Bélanger at HEC Montreal and Professor Angel Ruiz at Université Laval for all the support and for sharing your vast knowledge with me. Although it is formally not possible to register more than two mentors, I have actually been fortunate enough to be counselled and challenged by three outstanding supervisors.

A special thanks to my mother, who provided me the opportunity and guidance to be where I am today. I absolutely love you! To my family, thank you for all the encouragement words in the last 5 years.

Eduardo, thank you for all your patient during these rough times. You are my soulmate.

Thanks to the friends I made during my academic path: it would not be the same without you!

Abstract

It is increasingly necessary to have the patient as the main concern in the scheduling of elective surgeries. Waiting time is often higher than the clinically recommended maximum response time, a reality that occurs in several contexts such as the Portuguese and Canadian ones, where the patient is given a clinical priority level at the time of the surgical proposal, which is not reviewed during the waiting time for the surgery. On the other hand, the guidelines for scheduling are summarized by criteria of priority and seniority, without reflecting the actual need of the patient and the evolution of their health condition during the waiting period. In order to prevent patient health and well-being from being affected during waiting, it is important to create a surgery scheduling system that considers the dynamics of patients' real needs and analyze its impact.

In this dissertation, it is proposed a scheduling model in integer linear programming that integrates a new prioritization system. Based on a case study from a Canadian hospital, waiting lists are generated to validate the proposed approach and test the quality of the solutions. The discussion of results includes a sensitivity analysis and analysis on patient selection, waiting times, the use of overtime and the number of scheduled surgeries. One formula for updating patients' condition is also proposed in a rolling horizon approach, which allow the model to select patients considering the dynamics of their real needs and creating equitable schedules.

Keywords: Operating room, Surgery scheduling, Utility, Patients’ needs, Optimization, Integer linear programming

Resumo

É cada vez mais necessário ter o paciente como principal preocupação no agendamento das cirurgias eletivas. O tempo de espera é geralmente superior ao tempo máximo de resposta clinicamente recomendado, realidade que se verifica em diversos contextos como o Português e o Canadiano, onde é atribuído um nível de prioridade clínica ao paciente no momento da proposta cirúrgica, que não é revisto durante o tempo em espera pela cirurgia. Contrariamente, as recomendações para o agendamento resumem-se a critérios de prioridade e antiguidade, sem refletir a necessidade real do paciente e a evolução do seu estado de saúde durante o período em espera. De forma a evitar que a saúde e o bem-estar do paciente sejam afetados durante a espera, é importante criar um sistema de agendamento de cirurgias que considere a dinâmica das necessidades reais dos pacientes e perceber o seu impacto.

Nesta dissertação, é proposto um modelo de escalonamento em programação linear inteira que integra um novo sistema de priorização. Baseado num caso de estudo de um hospital Canadiano, são geradas listas de espera para validar a abordagem proposta e testar a qualidade das soluções. A discussão de resultados inclui uma análise de sensibilidade e sobre a seleção de pacientes, os tempos de espera, a utilização de tempo extraordinário e o número de cirurgias marcadas. É ainda proposta, numa abordagem *rolling horizon,* uma fórmula de atualização da condição dos pacientes, que permite que o modelo selecione pacientes considerando a dinâmica das suas necessidades e que crie escalonamentos equitativos.

Palavras-chave: Bloco operatório, Escalonamento de cirurgias, Utilidade, Necessidades dos pacientes, Otimização, Programação linear inteira

Table of Contents

List of tables

List of figures

1. Introduction

Health care must be provided quickly and effectively. A good quality of life is increasingly demanded by the population in general and for this it is necessary to ensure access to health care in an efficient and equitable manner, which can be a major challenge in the management of health systems. The main barriers are the high costs, the lack of availability of resources and the increasing demand motivated by the ageing of the population and the development of new and sophisticated technology. Although subjective, measurement of access to health care can be quantified, for example, through the amount of waiting time for some appointment.

The health system in Canada is managed by territorial and provincial governments and its main objective is to provide health services to the entire population in an equitable manner. There is no national health plan, but there are 13 health insurance plans distributed across the various provinces. According to the Canada Health Act, each territorial and provincial insurance plan must be comprehensive, accessible, portable, universal and managed by a public authority (https://www.canada.ca/). In the specific case of the province of Québec, the responsible authority is the *Régie de l'Assurance Maladie du Québec* (RAMQ), which should ensure that the high costs of healthcare and the low availability of resources and services are not an obstacle for the population to achieve a good quality of life.

However, serious access problems are known regarding, for example, elective surgeries. For most medical specialties, the waiting time for treatment is higher than recommended, which can have serious consequences for the health status of the patient. Some causes of the long waiting times are the insufficient response to the increasing demand for surgical services, shortages in personnel and inefficient use of operating room resources (WTA, 2014). Often, the long waiting time causes the patients to have increased pain, anxiety, irreversible aggravation of the situation or even lead to their death (Barua, 2017). It can also lead to a poorer health system performance and to high and unexpected expenses to the individual and for the government due to less productivity, less earned income and less tax revenues ("The Federal Government Role in Reducing Health Care Wait Times," 2014).

Patients who need surgery are registered in waiting lists. When there is a life-threatening condition, urgent patients are ordered according to the priority of their condition, regardless on the time they entered the waiting list. On the other hand, for cases where the patient's life is not under some risk – elective patients – they are generally served under the motto First-Come-First-Served (FCFS), meaning that the first patients requiring surgery are those who are scheduled first (Hall, 2013). This is also the guideline given by Portuguese and Canadian legislation which state that priority level and seniority as the criteria to elective patient scheduling. Urgent patients are very hard to predict and thus they are usually incorporated to the schedule when they arrive or there are dedicated capacity for those situations (Wullink *et al.*, 2007).

Generally, waiting lists can be managed in many ways. In most cases, the guidelines to manage waiting lists are according to patients' priority and waiting time. Patients with more severe conditions should be treated ahead of the ones that are less urgent. Under the same level of priority, the patients that enter the waiting list first need to be served first (FCFS strategy). To evaluate patients' priority, standardized measures must be defined to allow an equitable patient assessment. Some used criteria are, for example, the extent of pain, limits to the performance of basic daily routine activities, risk of death, social factors, need and expected benefit (Hadorn, 2000). In Canada, as in Portugal, a family doctor refers the patient to a specialist; after the appointment and medical tests, if a surgery is required and the patient agrees, a priority level is assigned. In both countries, there are four priority levels, being level 1 the most urgent and level 4 the least urgent. However, the patient condition is considered as static: when the patient enters the waiting list, the health status is evaluated and a priority level is assigned, and during all the waiting time the condition and severity are considered to be the same. Moreover, about 90-95% of the patients in the waiting list have usually priority level 4 which means that waiting time is the only scheduling criterion for a large number of patients (Marques *et al.*, 2015).

Accordingly, the difficulty in accessing health care services and the high waiting times are the two main obstacles to overcome. To address these two shortcomings, it is important to achieve several goals: to utilize human and material resources optimally (efficiency), to perform surgeries in a timely manner (effectiveness), to maximize patients' flow, to maximize patient and medical staff satisfaction (quality), to minimize waiting times and to minimize costs. Usually this management is done in an empirical way by the doctors and because of that it can lead to an inefficient strategy. For this reason, there is an increasing need to create accurate and systematic procedures for the planning of surgeries (Guerriero and Guido, 2011), so that patients can be evaluated and ordered according to the chosen criteria in the most appropriate and equitable possible ways.

Often operating rooms are pointed out as the largest consumer of costs of a hospital, so it is extremely important not only for the patients but also for the hospital administration to use this service as efficiently as possible (HFMA, 2003). For this reason, the number of published articles proposing algorithms for operating room scheduling has increased progressively over the last few years (Samudra *et al.*, 2016). In fact, there are several solutions in the field of operational research that can be brought to different types of situations in the operating room planning and scheduling. This discipline is based on optimization procedures and can be used as a decision-making tool. Therefore, it allows the management, monitoring and optimization of the hospital resources, namely time, money and staff resources, while considering the specific constraints and goals of the hospital under study.

As shown in Figure 1, when addressing the problem of operating room planning and scheduling, the level of decisions can be: strategic, tactical or operational (Hans *et al.*, 2012). In the case of the strategic level (also called case mix planning), decisions are made about the mission of the hospital: with a longer planning horizon issues as distribution of time by the several hospital specialties and what capacity is used by each of these specialties can be taken within this level of decision. For an intermediate length of the planning horizon, the tactical level (or master surgery scheduling) deals with

the implementation of strategic missions: e.g. issues related to the fitting of the time blocks of each medical specialty in all the available operating rooms. Finally, for the operational level there is almost none flexibility since it deals with a short planning horizon: it may decide e.g. on the patients' assignment to operating rooms and time blocks (Marques *et al*., 2012). A distinction can also be made between two types of operational decisions: advance scheduling and allocation scheduling. Advance scheduling defines the date of the surgery for each patient while allocation scheduling defines not only the starting time of the surgery on some day but also the operating room in which the surgery is be performed (Magerlein and Martin, 1978). Two strategies for scheduling are open scheduling and close scheduling. Open scheduling is a process in which surgeons can plan the operations for any operating room according to the demand. On the other hand, close scheduling assigns slots of time and rooms to each type of surgical groups (van Oostrum *et al*., 2010) – master surgery schedule – and the surgeries are scheduled to the timeslots assigned to the corresponding surgical group. All these decisions allow constructing a complete schedule for the elective surgeries performed in a hospital.

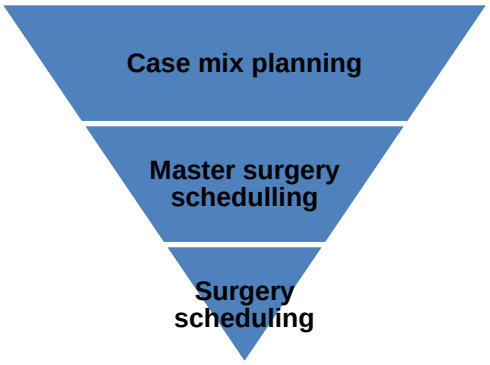

Figure 1: Hierarchical levelling of the planning process.

So, to generate appropriate and optimized schedules, it is necessary to gather all relevant data to the problem and to construct a mathematical model that is a sufficiently precise representation of the situation that is being studied. A model should be created to mimic the hospital characteristics to predict the best way to allocate specialties and patients. The main barrier for researchers when modelling is to define a trade-off between the robustness of the model and an acceptable complexity. It is also necessary to consider the visions of the various stakeholders: patients, medical staff and administration. The different stakeholders are driven by different interests: for example, the main interest for patients will be to reduce the waiting time; for the hospital staff the greatest concern is related to the fairness of workload distributions; while administration keeps the focus of managing costs and all the financial issues.

This work is part of a project carried out in collaboration with the Operations and Logistics Management Department of the *Hautes Études Commerciales* (HEC), Montréal, and a Canadian university hospital. The main objective of this project is to propose a more accurate prioritization system, establish a link between the prioritization system of surgical patients and the scheduling of elective surgeries, and compare the new approach with the existing one (Figure 2).

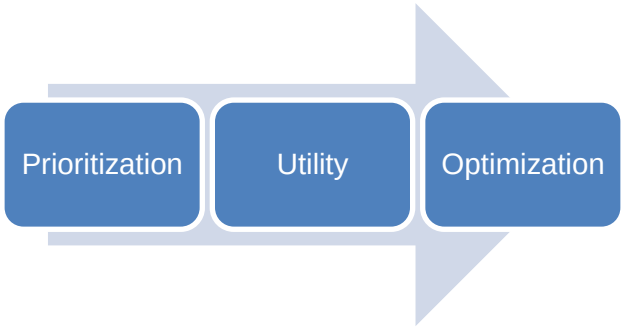

Figure 2: Overall scheme of the project work.

This dissertation aims to explore the potential advantages that can be extracted from linking the new prioritization system and the optimization model. The prioritization system gives a utility value for each patient. In addition to patients' medical situation, the prioritization system also considers the clinical and social situation and the strategic value of the surgeries. This utility value can then be used as an input of the scheduling problem.

The introduction of this utility value in the planning process is justified by the fact that, most times, even patients with the same biomedical factors can be in more need of some surgical procedure and the scheduling of surgeries should be done with the intention of satisfying the prioritized needs of patients as effectively as possible, avoiding doing it so with goals such as generally reducing waiting times (as often suggested in legislation) or hospital costs (also considered in the literature) that often do not comply with the trend for patient-centered care. The use of a utility value specific for each patient that includes several criteria instead of just clinical priority and waiting time allow for a patient-centered operating room scheduling. Moreover, if we are able to capture the evolution of the patient health state along the waiting time and reflect this evolution in a dynamic utility value, the operating room schedule is able to capture the real situation and need of each patient waiting for surgery and to increase equity in surgical access for elective patients, thus better fitting the ideals of most of health care systems.

In collaboration with the staff and administration of the hospital under study, it is possible to identify preferences and restrictions for the surgery scheduling. Several constraints are thus introduced in the optimization model. Besides the usual capacity and time constraints, some limits are defined regarding skills and availability of each surgeon and even to guarantee workload balance among these surgeons. Patients' interests and satisfaction are considered in a rolling horizon approach that updates the health status of the patients waiting for surgery.

One of the objectives of the project is the link of the new prioritization system and the scheduling method, considering that the priority is assigned according to a new utility system that evaluates the physical, psychological and social condition of the patient. This dissertation focuses on the part of the optimization part of the project and aims to propose an integer linear programming model for elective surgery scheduling of the hospital under study that integrates, among other objectives, the compliance of dynamic utility values of the patients in the waiting list. This work intends to fill the literature gap

regarding the fact that the management of waiting list for surgery is done considering the waiting time and the clinical priority at the moment of patient registration in the waiting list which does not consider the dynamic of the health status and needs of the patients during the waiting time. Thus, this dissertation aims to answer the research question: “What is the impact of considering a utility function representing the real needs of the patients and its evolution in an optimization system for surgery scheduling, on the selection of patients and waiting times?”.

This dissertation is organized as follows. Chapter 2 develops a literature review, which reports on the approaches proposed by several authors for operating room scheduling. There are many studies related to the topic under study, but a selection of some papers is made with the aim of showing that the different chosen approaches may be consequence of the type of patients that is considered, the integration of other hospital facilities, the incorporation of uncertainty and its forms, the prioritization system or even of the main performance measures or interests taken into account. In this chapter, this dissertation is framed in the field and the literature gap that this dissertation aims to fill is highlighted. Chapter 3 describes the case study and the specific problem that motivates this work, highlighting its most important characteristics. All the rules and restrictions imposed by the hospital are explained. In Chapter 4, an integer linear programming model is proposed to cover the most relevant aspects of the case study although being generalized for other cases. Chapter 5 presents the data provided by the hospital and the instance generation process used for computational experiments. The data is used to validate the model and perform sensitivity analysis evaluating how some changes in the data or in the mathematical formulation impact the results, as described in Chapter 6. The main results are presented and analyzed in this chapter with an extended discussion on all the decisions taken and their impact on the results. Two rolling horizon approaches are also proposed and tested. These results intend to answer the research question of this dissertation. Finally, Chapter 7 concludes this dissertation, highlights the limitations of the project and outlines suggestions for future work.

2. Literature Review

The literature regarding the scheduling of the operating room is very wide since to be able to manage such a complex space it is necessary to consider several aspects and the interests of many stakeholders. This dissertation focuses in operational decisions and intends to generate a schedule of elective surgeries in such a way that the patients' interests are protected and the use of the hospital resources are optimized. For that reason, this literature review is centralized in the operating room scheduling which is inherently dependent on the results from the previous hierarchical levels of decisions (case mix planning and master surgery scheduling): after the assignment of time blocks among specialties within a hospital and the assignment of the operating rooms to each group of surgeons, it is finally possible to assign each patient's case to certain time and place according to the previous decisions. The papers identified as the most important to this work are highlighted, namely to the definition of the problem and in the followed approach. The Web of Knowledge database was used to find the most relevant papers in the literature about the topics "operating room" and "scheduling". The search was narrowed to documents published between 1900 and 2018, written in English and in the categories of Operations Research & Management Science and Health Care Sciences Services. Some other papers considered relevant found while checking reference lists are also included in this review.

There are many considerations in those papers that can influence how the scheduling is done and how the results are obtained. In the analysis of problems related to operating room scheduling, many factors must be considered as shown in Figure 3: distinct type of decisions, several types of patients, many objectives, various types of prioritization, isolation of the operating room as the only hospital service or its incorporation in other up and downstream facilities, uncertainty integration and several approaches to deal with non-elective patients. All these decisions must be made in accordance with the goal to be achieved, which also varies from hospital to hospital as well as from stakeholder to stakeholder.

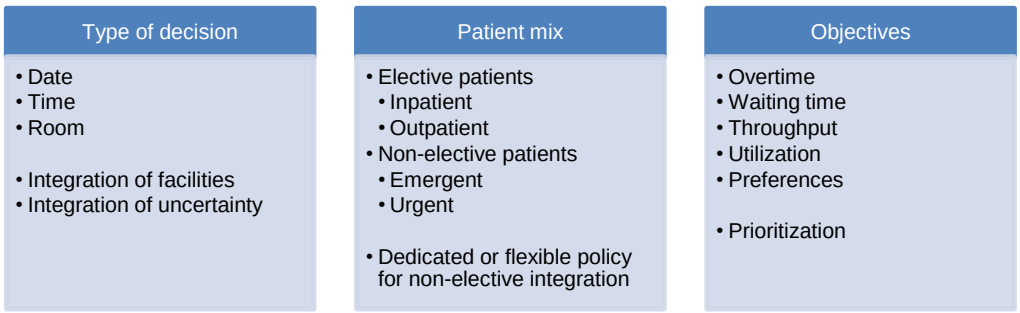

Figure 3: Framework used to describe the main findings in the literature review.

This chapter is organized according to Figure 3: Section 2.1. describes the types of decisions that can be taken in solving the problem of operating room scheduling and explains how complex it can be by integrating facilities and integrating uncertainty; Section 2.2. enumerates the type of patients that can be considered for the operating room scheduling namely elective and non-elective and list the approaches that can be followed to deal with the arrival of non-elective patients to a hospital; Section 2.3. specify what are the main objectives to be considered in the existent models and how the management of patients in the waiting list is done.

2.1. Type of decision

The procedure of scheduling surgeries can be based on several types of decisions. For example, decisions such as date, time and operating room are often taken during this process. In fact, the results of the operating room scheduling are highly influenced by the way that the problem is defined, namely regarding the integration of other facilities besides the operating room or the integration of uncertainty in some parameters.

Intuitively, it is intended to result in a schedule in which each patient is assigned with a date, a starting time and room in which the surgery must be performed. Gomes *et al*. (2012) use data mining techniques to predict the duration of surgeries to be scheduled. This method is created so that this process is not performed by surgeons themselves who generally incur in over-planning to protect their interests which lead to delays, cancellations and a total lack of control over schedules. Adopting the computed surgery duration, the authors use optimization methods to assign each patient a room, a day and a shift where the surgery must be performed. However, they do not order surgeries within each shift, but they point out that this part of the problem is easily solved by random methods. On the other hand, Marques *et al*. (2012) use an integer linear programming model to assign not only a room and a day to each patient but also the starting time of the surgery. They use Boolean starting time decision variables that dictate if a certain surgery starts on a certain day, room and time period.

These scheduling decisions can be made not only from the point of view of the patient but also from the point of view of the hospital staff. They can affect each individual surgeon or some specialty surgeons as a whole. For example, Van Houdenhoven *et al.* (2007) propose a model to calculate the optimal utilization rate of operating rooms depending on the patient mix and on the willingness to accept overtime. This study highlights the idea that utilization rates between studies can only be compared considering the various characteristics of each patient mix and the management choices of the hospital under study.

Besides the decisions regarding date, time and room of a surgery, some other decisions need to be taken. Sometimes some assumptions need to be made to make the simplify the problem of patients' assignment. Although many papers consider the operating room as an isolated unit, stakeholder interests may include issues related to other parts of the hospital such as upstream or

downstream facilities. In this case, the complexity of the problem is increased and an integrated model must be considered. Upstream facilities include services wards like outpatient clinic and downstream facilities include intensive care units or post-anesthesia care units (Samudra *et al.*, 2016). For example, Kharraja *et al*. (2006) develop two integer linear programming methods to allocate time blocks to the surgeons performing the surgeries. To model this problem no other facilities besides operating room are considered, although in reality a patient passes through more than one place during their stay in the hospital. After modeling the operating room characteristics, the authors compare the two approaches in which the first consider every surgeon as an individual and the second one considers surgical groups. The results show that the second method has better results because there is more flexibility within a group to perform all the assigned surgeries. On the other hand, Vanberkel *et al*. (2011) propose an analytical approach using an already constructed master surgery schedule to calculate the workload of downstream departments regarding ward occupancy, patient admission, patient discharge and ongoing interventions distributions. The integration of the operating room with other departments allows evaluating the impact that its management has on the entire hospital operation. In the same paper, it is stated that all the activities that happen after the surgery highly depend on the activities of the operating room, so the workload of downstream departments is described as a function of the master surgery schedule. It is also mentioned that upstream facilities are not so sensitive to the operating room management as the downstream facilities are. However, Gul *et al*. (2011) for instance present a simulation model in which they deal with both the upstream and downstream facilities in order to test some heuristics for an outpatient clinical center and evaluate the average patient waiting time and the operating room overtime. The authors conclude that the waiting time and overtime are highly determined by the arrival time of patients.

In fact, some uncertain factors can disrupt the operating room schedules. It is found that stochasticity, i.e. incorporation of uncertainty, is often introduced in mathematical formulations regarding the durations of surgeries and the time of arrival of elective and non-elective patients (Samudra *et al.*, 2016). This incorporation adds some complexity to the model and to the problem solving. The durations of surgeries may suffer many deviations from the expected time due to several factors, such as the characteristics of the patient, the surgeon and even the surgical team. They are usually modeled through a lognormal distribution (Van Riet and Demeulemeester, 2015). Stuart and Kozan (2012) deal with re-scheduling problems using a branch-and-bound approach and considering some variability regarding the surgery durations. Branch-and-bound approach performs an implicit enumeration of the feasible solutions, according to upper and lower bounds on the optimal value calculated iteratively. The surgery durations are assumed to be lognormal random variables and independent from one surgery to another. However, it is noted that lognormal distributions are realistic enough to represent a surgery duration but when there is a disruption in the schedule doctors can try to compensate and perform the next surgery faster which makes the surgeries duration dependent of each other in reality. In this case, two decisions should be made along this process: the operating room in which the current surgery will be performed and the next surgery to be scheduled. Re-scheduling is done to a list of elective patients when an emergency patient arrives or another disruptive factor occurs. Regarding the arrival times, although it can suffer some variability, the arrival

of elective patients is often considered deterministic, meaning that uncertainty is ignored. A deterministic approach, while making the problem less realistic, makes it simpler, since it is usually hard to predict any kind variability (Van Riet and Demeulemeester, 2015). On the other hand, the arrival of non-elective patients is more unpredictable and when modeled should be considered with a certain level of uncertainty. The model can be created using the arrival rate or the interarrivals time. For the expected number of arrivals per unit time, a Poisson distribution is generally assumed. Adan *et al.* (2011) propose a mixed integer programming model to obtain the daily schedule of surgeries minimizing the difference between the actual resources' utilization and the target values. In addition to elective patients, it also considers emergency cases to whom some capacity is reserved as a buffer to accommodate the arrival of uncertain cases. Arrivals of patients are modeled with a Poisson process. One approach used by these authors to adapt to variations of patient arrivals is over-planning, which increases patients' throughput goal value to reach a greater number of patients admitted. The time between arrivals is often represented by an exponential distribution (Van Riet and Demeulemeester, 2015), since exponential inter-arrival times described by parameter λ are equivalent to Poisson arrival rates described by λt (with t representing the time interval).

2.2. Type of patients

In fact, as mentioned on the previous section, patients can be divided into elective and non-elective. These two classes are divided as follows: elective patients are those for whom the surgery can be planned, with no immediate urgency for its treatment, and non-elective are the ones for whom the surgery cannot be planned in advance (Cardoen *et al.*, 2010b). In addition, sub-classifications of the two groups of patients are defined: elective patients may be inpatient or outpatient, with the former referring to patients who stay overnight in the hospital, while the latter only go to the hospital the very day of the surgery and leave the hospital on the same day; on the other hand, non-elective patients can be subdivided into emergent when surgery has to be performed as quickly as possible and urgent when surgery, although also unexpected, may be delayed for a short time due to patient stability (Cardoen *et al.*, 2010a). For the sake of simplicity, in this dissertation both emergent and urgent patients are treated indistinguishable. The choice of the type of patients to consider in the analysis of the problem entails some implications, for example, outpatient surgeries are usually more standardized procedures which may lead to a greater level of certainty in the duration of the surgery but at the same their time of arrival at the hospital is very uncertain because these patients are not hospitalized (Samudra *et al.*, 2016).

Although some papers do not establish differences between the various types of patients, a very common discussion topic is whether operating rooms should be organized by different types of patients with dedicated rooms, or if an integrated approach should be adopted. Ferrand *et al.* (2010) discuss the trade-off between the efficiency of performing elective surgeries and responding to emergencies arriving at the hospital from the point of view of the patient's waiting time and the surgeon's overtime. In this case study, two policies are considered: a focused policy in which a certain

number of operating rooms are dedicated to elective cases and other operating rooms are dedicated to emergency cases, and a flexible policy in which emergency cases are inserted in the schedules of elective cases. This work shows that a focused policy leads to less overtime and less waiting time for elective cases but more waiting time for emergency cases and less operating room access for elective patients. According to Heng and Wright (2013), the use of the operating room, waiting times, percentage of cases answered within the time limit, overtime, cancellations, overruns and length of stay are evaluated. From this analysis it turn out that having an operating room dedicated only to emergency cases leads to a decrease in the number of cancellations and overruns in elective cases. The number of patients with the lowest priority level having access to surgery within the initially defined time and with a shorter waiting time is also increased. The calculated use of the operating room is 53% and there is a decrease of the overtime and of the average length of stay in the hospital after assigning an operating room just for emergency cases. On the other hand, Wullink *et al.* (2007) show that the flexible policy leads to a decrease in the waiting time for emergency surgery and in the overtime. In addition, the utilization of the operating room is increased, reflecting higher levels of quality of service, staff satisfaction and cost-effectiveness. These different conclusions may be justified by the hospital case mix, namely the proportion of emergency cases with respect to elective ones and the number of operating rooms. A different approach is to consider a mixed version of the two policies. Erdem *et al.* (2012) present this hybrid policy and suggest the ideal number of operating rooms to assign to both flexible and dedicated policy. This work shows that this method outperforms both isolated policies regarding patient waiting time and overtime.

2.3. Type of objectives

As previously discussed, the preferences of various stakeholders should be considered along with other operational performance measures, such as overtime, waiting time, throughput or utilization. Each type of stakeholder may have its own interests which can be contradictory to others'. In most cases, several performance measures are combined in the form of a weighted sum so that the interests of as many stakeholders as possible are met. After a good definition of the objectives to be addressed and according to its purpose and restrictions, it is necessary to choose which approach to follow. Mathematical programming, branch-and-price, dynamic programing, heuristics, simulation and scenario analysis are some of the many tools that operations research can offer to deal with scheduling problems.

For the administration, one of the biggest interests is related to hospital costs and efficiency. By maintaining service quality levels and managing the interests of employees and patients, it is intended to optimize the schedules of surgeries in order to minimize the associated costs. Fei *et al.* (2006) developed a software for the block scheduling of elective surgeries through a deterministic approach with the objective of minimizing the cost of assignments by minimizing the cost of unexploited operating room time and overtime. Firstly, a mathematical model is presented to solve the weekly scheduling problem of the operating rooms. A column generation procedure is applied to determine

the list of surgeries to be done in each week. A column generation algorithm is a procedure in which the constrained variables are being added as the problem is being solved ("IBM Knowledge Center," n.d.). Considering the previous results, a hybrid genetic algorithm is performed to assign a starting time to each surgery considering performance measures such as utilization of the operating room and makespan. The proposed hybrid algorithm combines tabu search and genetic algorithms. Tabu search is a meta-heuristic in which a local search based on adaptive memory and responsive exploration is done (Glover and Marti, 2006). A genetic algorithm uses genetic processes to mimic the evolution of a population, selecting the best characteristics of each solution (Mallawaarachchi, 2017). Beliën and Demeulemeester (2008) formulate a deterministic and integrated branch-and-price approach where the objective is used to ensure ward levelling for elective inpatient surgeries, ensuring compliance with the restrictions related to the operating room and the intensive care and post-anesthesia care units. The ward levelling leads to a minimization of the required number of nurses, which decreases staffing costs. Branch-and-price methods are a combination of branch-and-bound and column generation. Several scenarios are then analyzed through a discrete event simulation. Van Essen *et al.* (2013) use a complex analytical model to calculate the ideal number of available beds so the hospital can reduce the number of beds to the ideal one and consequently decrease costs. This approach considers the variability of the length of stay of each patient in the hospital and it also presents a heuristic algorithm based on local search. The authors state that if the use of beds is leveled over time, for each moment, it is necessary to have a smaller number of beds available. For that reason, the cost of maintaining and cleaning the beds and labor and personal costs is minimized. This type of constraints is complex mainly because patients' length of stay can be influenced by many factors and can suffer great variability, so an attempt is made to simplify the problem to be solved simply as an integer linear programming problem. Moreover, Min and Yih (2014) establish a relationship between the scheduling of elective surgeries and waiting list management by calculating the cost of postponing the surgery and the likelihood of patients leaving the waiting list (bulking) which is time dependent. It also considers the cost of overtime, waiting time and the risk of adverse events. The objective is to select an optimal number of patients to be scheduled from the waiting list, given that assigning too many patients leads to overtime while assigning few patients causes delays in performing the surgeries because the capacity of the operating room is underused. To get more realistic results some uncertainty is applied to the patient arrivals time and to the duration of surgeries. It is intended to obtain a schedule of surgeries through a Markov decision process using a waiting list that considers the evolution of time and the adverse events that may arise.

On the other hand, the medical staff has as main priority good working conditions, namely the reduction of overtime, and the perception of fairness among staff. Denton *et al.* (2007) develop a two-stage stochastic programming model to schedule and sequence surgeries in which the objective function is the minimization of a weighted sum of surgeon waiting time, idle time and operating room overutilization. In this work, some heuristics are applied and tested to improve the patient sequencing because optimal sequencing and scheduled times highly influence the improvements on the operating room scheduling done by the heuristic methods. Day *et al.* (2012) present a model divided into 3 phases where it is intended to combine predictable physician access to the operating room and high

values of capacity utilization. In order to guarantee that all doctors have access to some operating room time, this time can be either exclusive or shared by a small group of surgeons. In the first place an individual or a group of surgeons is assigned the various time blocks available, providing regular access, maximizing utilization rates and reducing staff costs. Then, as a new surgery arises, for each of them a day and a time consistent with the time allocated to the surgeons are chosen and as soon as possible to avoid deferrals. Finally, the operating room is chosen by minimizing the excess downtime for surgeons and for the rooms. Marques and Captivo (2017) use mixed integer linear programming to combine the interests not only of surgeons but also of administration in the scheduling of elective surgeries. For a mixed version they model the administration interests for the morning shifts and the surgeons' interests for the afternoon shifts. In the morning shifts, equity in access, timely access and the use of the operating room available time are guaranteed. For the afternoon shifts, the use of available surgical resources, the access and the number of scheduled surgeries are maximized while the patients are scheduled by a Last-In-First-Out (LIFO) strategy mimicking the human nature of forgetting past events in a scenario of absence of a system to systematically select patients from the waiting list. A deterministic version and a robust optimization approach are created to deal with uncertainty in the duration of surgeries. This work concludes that despite of having the same number of optimal solutions in both cases (deterministic and robust counterpart), the robust optimization approach significantly increases the size of the models, but the deterministic has a simulated occupancy rate of more than 100%, which is unrealistic as completely disruptive in a real scenario.

For the patients it is extremely important that the surgery is performed in a timely manner without compromising their health. Persson and Persson (2018) developed an optimization model for scheduling patients in which the objective is to perform the maximum number of surgeries by minimizing the cost of not scheduling patients. This cost is based on patients' prioritization, waiting time and public costs that it can generate. Thus, this model allows the decrease in the patients' waiting time. On the other hand, Cardoen *et al*. (2009) present an exact model and a heuristic method based on integer programming and on a branch-and-bound approach in which they are concerned with patients' preferences. There are many aims to be achieved in this work: surgeries with children, surgeries that have been canceled before and surgeries for very urgent patients should be scheduled as early as possible in the day; patients that travel a long distance to get to the hospital should be scheduled after a certain reference period since they need more time to plan the trip and get to the hospital; minimization of the number of periods in which recovery care has to be given after the surgical day-care center closes to avoid hospitalizations; minimization of peak number of bed occupancy in the recovery units. All these objectives can improve patients' satisfaction and level the workload for the medical staff which also contributes to staff satisfaction.

To facilitate the choice of which patients to treat in the first place, a scoring system can be defined. The creation of a good system involves a clarified definition of the main parameters to be considered when ordering the patient list. It is known that the scheduling priority of a patient can be defined by the patient's position on a waiting list, by the severity of its medical condition or by a combination of factors. The main pitfalls of the current prioritization systems are: insufficient

robustness of the results to face the typical uncertainty of this environment; underestimation of health risks associated with patient waiting time; the decision continues to be mostly taken by the surgeon; static prioritization procedures that do not consider evolution over time, while waiting lists are dynamic; the fact that there is no comprehensive framework for the prioritization of surgical patients on the waiting list (Rahimi *et al*., 2016). Dexter *et al*. (1999) propose several methods of prioritization according to the objectives that are intended to be achieved. One goal is to minimize the average time a patient waits until surgery. In this case, surgeries must be ordered in ascending order of the expected length of surgery. The second objective is to sequence the cases in the same order in which they are submitted in the waiting list and in this case a First-Come-First-Served approach is used. Finally, the last objective is to order patients according to their medical priority considering not only the duration of surgery but also the time difference between the occurrence and the maximum time before surgery. On the other hand, in Hans *et al*. (2008) the objective is to plan surgeries optimizing the utilization of the operating rooms and minimizing overtime. As a rule of priority, First Fit is used: the surgery at the top of the waiting list is assigned to the first operating room in which it is feasible. The list of surgeries is sorted according to Longest Processing Time which means that patients are sorted in descending order of the expected duration of the surgery. Durán *et al*. (2017) use a prioritization formula with a measure called Need-Adjusted-Waiting-Days (NAWD) that is a function of both the biomedical category of the patient (which depends on the severity of his condition) and the number of waiting days. In this situation, the patient list is sorted in descending order of NAWD, with the first patient being the one that needs a faster treatment. Besides assigning a penalty for time extensions and rewarding the early scheduling of urgent patients, this work has the objective of maximizing the compliance with patient priority. In this case, four approaches are tested and compared. The first approach relies on an integer linear programming model that uses a priority formula that grows exponentially as the patient index on the list ordered by the NAWD values increases too. The second approach is a variant of the first integer linear programming model in which a new priority formula that includes the surgery duration is considered to avoid the exponential growth of the priority value. The third approach is based on an integer programming feasibility model algorithm, for which the problem is divided in two: an assignment and a timetable problem. The assignment problem is solved in the same order as the computed priority and determines the list of patients that can be feasibly assigned. Then for the timetable problem the first integer programming model is applied without the priority term. Finally, the fourth approach is a constructive algorithm that comprises four steps: preprocess, construction of feasible sub-assignments, dominance and choosing the solution. Analyzing which patients are scheduled according to each approach and comparing their priorities, it is showed that the results obtained with the first integer linear programming model (first approach), the feasibility model (third approach) and the constructive algorithm (fourth approach) are quite good and similar in regard of the patient priority compliance, while the variant of the integer linear programming model (second approach) presented worst results. However, the algorithm execution time for the feasibility model is significantly higher than for the remaining approaches and it depends on the number of patients for both of the integer linear programming models.

In fact, the prioritization formulas currently used may be difficult to interpret and use, and therefore it is difficult to use them as a factor in the scheduling of surgeries. Nonetheless, this can be a very powerful tool in attending patients prioritized needs. Research on prioritization issues for elective patients is still very scarce. Currently, there is a static management of the waiting list, where the health status of the patient is assumed to be the same over the waiting time, and methods for surgery scheduling rely often on the application of strategies such as First-Come-First-Served (FCFS), First Fit or even Longest Processing Time (LPT), matching the goals of surgeons and hospitals. However, these methods can fail to protect patients' interest and needs and do not consider patients' situation evolution. As a matter of fact, most times, the order in which patients are treated can be biased my surgeons, hospital secretaries or even patients. The aim of this study is to incorporate in the operating room scheduling a relative score assigned to each elective patient that considers not only their priority but also personal and social issues that evolve over time. In this way, patients are assessed in an equitable way and according not only the severity of their medical condition but also regarding the strategic value of the surgery and the impact that it might have in the patients' life. Thus, depending on the evaluation of all these factors and the constraints of the hospital under study, patients' surgeries can be scheduled with the main purpose of attending and meeting their prioritized needs.

2.4. Chapter considerations

There are many approaches that can be followed to model an operating room scheduling problem. Those approaches differ mainly on the type of decisions to be made, the patients considered, the conditions of the hospital under study and the objectives to be achieved.

A FCFS approach is usually considered in the literature. Concerning the waiting list management, in the best-case scenario, only the patient's clinical situation at the time of registration in the waiting list and their seniority on the list are considered to select patients to be scheduled. In addition, prioritization is generally a process disintegrated from operating room scheduling. Therefore, this dissertation fills this research gap by integrating in the process of optimization of the operating room scheduling a new dynamic priority system to obtain more equitable and effective schedules for patients on the waiting list that answers to real as updated needs.

3. Case Study

This dissertation is integrated in a project motivated by a collaboration between HEC Montréal and a urology department from a university hospital in Canada.

In Canada, access to medical services is guaranteed to the entire population and is financed through taxes paid to provincial, territorial, and even federal governments. In this way, access to care is not limited to the most deprived population. In fact, under Canadian law there is the Canada Health Act (CHA), which lays down laws for publicly funded health care insurance, ensuring that all taxpayers have reasonable access to medical services without having to pay on the spot.

However, Canada was considered the country with longer waiting times in access to care among other 11 countries. This can be tackled by paying monetary incentives to each hospital based on the number of performed surgeries, incorporating virtual care or remove patients from the waiting list that are not in dire need of surgeries. Indeed, in British Columbia, one of the provinces of Canada, 32% of the people waiting for cataract surgery had near-perfect vision (Urbach, 2018).

In Canada, the currently used prioritization system involves the assignment of some level to the patient according to his/her clinical condition and based on that level a maximum waiting time before the surgery is imposed. An example of the priority levels and correspondent maximum time before surgery, used in the urology department of the hospital under study, is shown in Table 1.

Priority level	Maximum time before surgery
P1	2 weeks
P2	4 weeks
P3	12 weeks
P4	1 year

Table 1: Maximum time before surgery according to the assigned priority level, in Canada.

Figure 4 shows the median value of the number of weeks that patients wait between the specialist appointment and the surgery, according to the Canadian province in which they are performed (Barua, 2017). This figure also shows total values to the overall country.

Between 1993 and 2017, waiting times for elective surgeries increased in most provinces. Considering the entire country, the number of waiting weeks has increased from 5.6 to 10.9, which represents a 95% growth. The province of Quebec, where the hospital under study is located, more than duplicated the number of weeks waiting for surgery. This hospital is part of a general health center that is one of the largest in the country. This great health center comprises 5 hospitals in which there are 13500 employees distributed among 25 medical specialties.

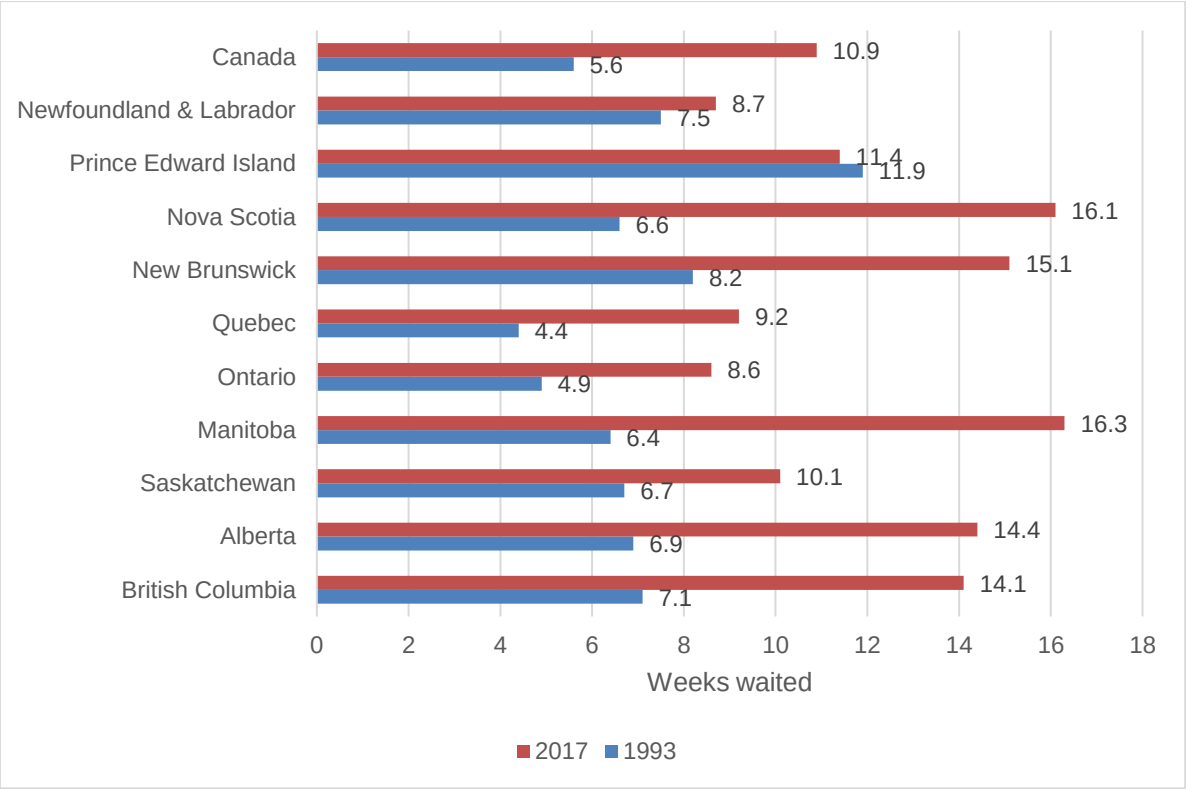

Figure 4: Median wait between appointment with specialist and treatment, by province, 1993 and 2017.

Anesthesiology, cancer, oral and maxillofacial surgery, general surgery, orthopedic surgery, plastic surgery, gastroenterology, gynecology, ophthalmology, otorhinolaryngology, neurosurgery and urology are some of the specialty surgeries performed in this hospital. It serves approximately 2 million people and 67088 surgeries were performed in the financial year of 2016-2017. In each day, approximately 216 new requests are made for elective surgeries. Moreover 223 requests were in the waiting lists for longer than one year.

In this hospital, after the specialist doctor recommends a surgery and the patient agrees, patient's health is assessed, and a level of priority is assigned accordingly. The information of the patient is given to a hospital secretary, who deals with the case loading. The secretary assigns manually some session, operating room and doctor according to their knowledge and best belief, based on limited patients' information.

For scheduling purposes, this hospital's department does not distinguish sessions within a day, so one session corresponds to one day. Surgeries are scheduled from Monday to Friday between 8h00 and 16h00. The scheduling is usually done with a planning horizon of 4 weeks. Moreover, currently, no doctors' overtime work is allowed but there is some discussion within the hospital on the benefits of allowing the extension of some surgery time after the regular time on the waiting list and on the patients.

In the urology department, usually five or six surgeries are performed in each session-operating room block. Every day of the week there are some operating rooms assigned to urology specialty as

defined by the master surgery schedule. On Mondays, Tuesdays and Fridays there are two available operating rooms to be used and on Wednesdays and Thursdays there are three available operating rooms.

Besides assigning a session and an operating room to each surgery, it is also necessary to assign a doctor to perform it. The urology surgeons can be divided in 5 groups according to the types of surgeries that they can execute. Each type of surgery can only be performed by doctors with the right qualifications. Many times, when a doctor follows the case before surgery s/he should be the one doing it; so, when this patients' information is given to the secretary, sometimes there is already an assigned doctor to the surgery. However, normally, a doctor does not work in the hospital every day of the week. It can happen that the doctor is not available to perform a surgery on a patient that is being followed by them, and therefore another doctor with the skills to perform it must be assigned. Normally, the procedure at this hospital is that the doctor who should be performing it is the one who recommends some colleague to do it.

The fact that this process is done in an empirical way can make the hospital resources not be used to the best advantage. The current management of the operating room complicates access to elective surgeries. The fact that this procedure is still carried out by the doctors and secretaries in this hospital serves as a motivation for this project, and in particular to create a mathematical model that could process all these constraints while generating schedules that are equitable regarding the real needs of the patients on the waiting list. The application of the prioritization method that uses utility measures, and their integration into a systematic approach to the scheduling of the operating room contributes to the patients' needs being met, without the rules currently implemented in the hospital being disregarded.

4. Mathematical Model

In this chapter a model is proposed to formulate this operating room scheduling problem. The notation used to describe the objective function and constraints of the integer linear programming model is also presented (Section 4.1.). The model presented in Section 4.2. is intended to be the most general as possible, so that it can be applied to other contexts and other situations. This chapter ends with some conclusions (Section 4.3.).

4.1. Notation

This section introduces the indices, sets, parameters and decision variables used to formulate the model. The indices and sets are summarized on Table 2 while parameters are presented in Table 3. Decision variables are described in Table 4.

Indices and sets	Description
$\boldsymbol{s} \in \boldsymbol{S}$	Set of sessions in the planning horizon
$\boldsymbol{p} \in \boldsymbol{P}$	Set of patients in the waiting list
$\boldsymbol{d} \in \boldsymbol{D}$	Set of doctors working in the hospital
$\boldsymbol{o} \in \boldsymbol{O}$	Set of operating rooms in the hospital
$\boldsymbol{subspecialty_p} \in \boldsymbol{SP}$	Set of surgeries' sub-specialties
$\boldsymbol{D_s}$	Doctors available to perform surgeries in session s

Table 2: Summary of indices and sets.

A patients' set P are on a waiting list to have some surgery: each patient $p \in P$ has an associated utility value $utility_p \in [0,100]$ and a duration value $duration_p$ that represents the number of minutes of the expected duration of the surgery to be performed on patient p. Each one of the surgeries on the waiting list is connected to some sub-specialty $subspecialty_p \in SP$. A sub-specialty is some type of surgeries within the specialty under study. The waiting time of the patient is measured by a number of sessions $waitingsess_p$ which represents the number of sessions in which patient p has been on the waiting list without being scheduled, computed in the beginning of each planning period.

Parameters	Description
$\boldsymbol{utility_p} \in \mathbb{N}$	Utility value of patient p
$\boldsymbol{duration_p} \in \mathbb{N}$	Average surgery duration of patient p
$\boldsymbol{waitingsess_p} \in \mathbb{N}$	Number of waiting sessions of patient p on the waiting list on the beginning of the planning horizon
$\boldsymbol{skillsbin_{dp}} \in \{\mathbf{0}, \mathbf{1}\}$	Equals 1 if doctor d has the required skills to perform patient p's surgery; 0 otherwise
$\boldsymbol{skills_{dp}} \in \mathbb{R}^+$	Scale factor for the duration of a surgery related to the experience of doctor d on the surgery required by patient p
$\boldsymbol{predoc_{pd}} \in \{\mathbf{0}, \mathbf{1}\}$	Equals 1 if doctor d is pre-assigned to patient p's surgery; 0 otherwise
$\boldsymbol{ot} \in \mathbb{N}$	Maximum overtime capacity in minutes
$\boldsymbol{rt} \in \mathbb{N}$	Maximum regular time capacity in minutes
$\boldsymbol{mns} \in \mathbb{N}$	Maximum number of surgeries to be performed by each doctor on a session
$\boldsymbol{mwd} \in \mathbb{N}$	Minimum number of operating rooms to be assigned to each doctor on the planning horizon
$\boldsymbol{noa_s} \in \mathbb{N}$	Number of operating rooms available in session s

Table 3: Summary of parameters.

In addition, assuming a doctors' set D performing surgeries on some hospital: each doctor $d \in D$ has an associated skill represented by a binary value $skillsbin_{dp} \in \{0,1\}$ which equals one if doctor d is able to perform a surgery from sub-specialty $subspecialty_p$.

$$skillsbin_{dp} = \begin{cases} 1, if\ doctor\ d \in D\ has\ the\ required\ skill\ to\ perform\ the\ surgery\ associated\ to\ patient\ p \in P \\ 0, otherwhise \end{cases}$$

Furthermore, in case doctor $d \in D$ is able to perform surgeries from sub-specialty $subspecialty_p$ ($skillsbin_{dp} = 1$) another value $skills_{dp} \in [0.8,1.2]$ is assigned, which represents how fast doctor d can perform surgeries that belong to that $subspecialty_p$. The need to distinguish the skill level of various physicians was felt because more experienced doctors generally have more standardized procedures

that allow them to perform tasks quickly and react more easily to some unforeseen event that may occur.

Each patient $p \in P$ has a pre-assigned doctor $d \in D$ to perform the surgery; usually the specialist that identified the need for surgery and registered the patient for surgery is the one that performs it. Thus, a parameter $predoc_{pd} \in \{0,1\}$ is considered meaning that some patient $p \in P$ might have pre-assigned doctor $d \in D$ to perform the surgery.

$$predoc_{pd} = \begin{cases} 1, if\ doctor\ d \in D\ is\ pre-assigned\ to\ the\ surgery\ associated\ to\ patient\ p \in P \\ 0, otherwise \end{cases}$$

However, to improve schedules or because of availability reasons, it is possible that this patient's surgery is not performed by the doctor that is pre-assigned to it, but by some colleague with the right skills to perform the surgery.

In fact, because doctors work in multiple facilities in parallel and are not always free to perform surgeries, it is necessary to model the availability of physicians. For that reason, a sub-set of doctors D_s is created to represent the doctors available on session s.

Besides, three decision variables are defined and are summarized in Table 4.

Decision variables	Description
$\boldsymbol{x_{spdo}} \in \{\mathbf{0}, \mathbf{1}\}$	Equals 1 if patient p is scheduled for session s, operating room o and doctor d; 0 otherwise
$\boldsymbol{y_{sdo}} \in \{\mathbf{0}, \mathbf{1}\}$	Equals 1 if a doctor d is assigned to session s and operating room o; 0 otherwise
$\boldsymbol{overtimebin_{so}} \in \{\mathbf{0}, \mathbf{1}\}$	Equals 1 if overtime is used in session s and operating room o; 0 otherwise

Table 4: Summary of decision variables.

The first set of decision variables, $x_{spdo} \in \{0,1\}$, delineates if patient $p \in P$ is assigned to some session $s \in S$, on operating room $o \in O$ and with doctor $d \in D$.

$$x_{spdo} = \begin{cases} 1, \text{if patient } p \in P \text{ is scheduled to session } s \in S, \text{on operating room } o \in O \text{ and doctor } d \in D \\ 0, otherwise \end{cases}$$

Variables y_{sdo} specify if doctor $d \in D$ is assigned to the operating room $o \in O$, on session $s \in S$.

$$y_{sdo} = \begin{cases} 1, \text{if doctor } d \in D \text{ is assigned to session } s \in S \text{ and operating room } o \in O \\ 0, otherwise \end{cases}$$

Ideally, each session-operating room block is used only on regular time. But sometimes it is necessary to perform some work in overtime and therefore a binary variable $overtimebin_{so} \in \{0,1\}$ is defined.

$$overtimebin_{so} = \begin{cases} 1, if\ overtime\ is\ used\ on\ session\ s \in s\ and\ operating\ room\ o \in O \\ 0, otherwise \end{cases}$$

4.2. Model formulation

In this section, the constraints and objective function of the integer linear programming model are explained, using the sets, indices, parameters and decision variables previously described in this chapter.

The problem constraints are formulated with Expressions (1) to (10).

It is necessary to establish a maximum value to the time capacity of each session and each operating room. Thus, not only regular time but also overtime need to be considered. This is done through Constraints (1): the sum of the expected durations of scheduled surgeries (that depends on the average duration $duration_p$ and on doctors' skills $skills_{dp}$) cannot be higher than the regular time in the slot and operating room (rt) plus the overtime available (ot) in case it is used (in this case, $overtimebin_{so}$ = 1).

$$\sum_{p \in P} \sum_{d \in D} duration_p\ skills_{dp}\ x_{spdo} \leq ot \times overtimebin_{so} + rt\ , \forall\ s \in S, o \in O \tag{1}$$

It is necessary to define that overtime is only used when necessary. Indeed, it is only needed if the real total duration of scheduled surgeries is higher than rt minutes, as shown in Expression (2).

$$\sum_{p \in P} \sum_{d \in D} duration_p\ skills_{dp}\ x_{spdo} \geq rt\ overtimebin_{so}\ , \forall\ s \in S, o \in O \tag{2}$$

Moreover, each surgery on the waiting list can only be scheduled once in the planning horizon. In case a patient needs more than one surgery, there are more than one registration in the waiting list, one to each required surgery.

$$\sum_{s \in S} \sum_{d \in D} \sum_{o \in O} x_{spdo} \leq 1 \, , \forall \, p \in P \tag{3}$$

Constraints (4) guarantee that the doctors only participate on surgeries if they have the right skills.

$$x_{spdo} \leq skillsbin_{dp} \;\; , \forall \, s \in S, p \in P, d \in D_s, o \in O \tag{4}$$

A maximum number of surgeries mns to be performed by each doctor in each session is also established in Expression (5).

$$\sum_{p \in P} \sum_{o \in O} x_{spdo} \leq mns \, , \forall \, s \in S, d \in D_s \tag{5}$$

Constraints (6) avoid doctors overlapping among the various operating rooms in a session, stating that each doctor, on a certain session, can only be assigned to one of the available operating rooms.

$$\sum_{o \in O} y_{sdo} \leq 1 \;\; , \forall \, s \in S, d \in D_s, o \in O \tag{6}$$

To balance the doctors' work schedules and to contribute to a fairness perception, Constraints (7) guarantee that each doctor is assigned to a minimum number of sessions and operating room blocks within the planning horizon, mwd.

$$\sum_{s \in S} \sum_{o \in O} y_{sdo} \geq \; mwd \, , \forall \, d \in D_s \tag{7}$$

Moreover, each session $s \in S$ has some number of available operating rooms, noa_s, to be assigned to surgeries from the specialty under study, as defined in Expression (8).

$$\sum_{d \in D} \sum_{o \in O} y_{sdo} \leq noa_s \, , \forall \, s \in S \tag{8}$$

Two sets of linking constraints are required to connect variables x_{spdo} and y_{sdo} - Expressions (9) and (10). These constraints mainly allow a surgery to be performed by a doctor that is working on the corresponding slot and operating room. One can see in the opposite direction: if a patient p is

assigned to doctor d on session s and operating room o, then doctor d must be working in session s and operating room.

$$x_{spdo} \leq y_{sdo}\ , \forall\ s \in S, p \in P, d \in D_s, o \in O \tag{9}$$

$$\sum_{p \in P} x_{spdo} \geq y_{sdo}\ , \forall\ s \in S, d \in D_s, o \in O \tag{10}$$

There are many objectives that can be considered when planning the operating rooms. In this case study, four objectives are considered (see Expressions (11) to (14)). Expressions (11) to (13) model individual objectives while Expression (14) represent a sum of the previous three objectives.

One of the main objectives is presented in Expression (11) and is related to the compliance with patients' prioritized needs. For that reason, the sum of the scheduled patients' utilities must be maximized in the planning horizon. Thus, it is guaranteed that the patients with higher values of utilities and that are waiting for longer times are scheduled. Patients should be scheduled in a way that patients with higher utilities, representing the ones in most need of a surgery, are assigned as early as possible in the planning horizon. To comply with the patients' needs, Objective Function (11) maximizes the utility value of scheduled patients, while minimizing the number of the assigned session.

$$Maximize \sum_{s \in S} \sum_{p \in P} \sum_{d \in D} \sum_{o \in O} \frac{utility_p\ x_{spdo}}{s+1} \tag{11}$$

Additionally, pre-assigned doctors must be respected as much as possible, if they do not disrupt the scheduling process, as shown by Expression (12). In case they are not respected, some other doctor with the skills to perform the surgery can be assigned.

$$Maximize \quad \sum_{s \in S} \sum_{p \in P} \sum_{d \in D} \sum_{o \in O} predoc_{pd}\ x_{spdo} \tag{12}$$

One commonly used objective in the literature is also used in this case: overtime minimization. With Expression (13), it is intended that the number of session-operating rooms blocks in which overtime is used is minimal.

$$Maximize \quad -\sum_{s \in S} \sum_{o \in O} \quad overtimebin_{so} \tag{13}$$

A sum of Objectives (11) to (13) is formulated in Expression (14). These expression models the maximization of the utility compliance term and of the pre-assigned doctors' compliance term and the minimization of the number of sessions in which overtime is used term.

$$Maximize \sum_{s\in S}\sum_{p\in P}\sum_{d\in D}\sum_{o\in O}\frac{utility_p\, x_{spdo}}{s+1} + \sum_{s\in S}\sum_{p\in P}\sum_{d\in D}\sum_{o\in O} predoc_{pd}\, x_{spdo} - \sum_{s\in S}\sum_{o\in O} overtimebin_{so} \tag{14}$$

When the three objectives come together (Expression (14)), there is a disregard for the order of magnitude that each term takes on. For this reason, it is necessary to normalize the function.

Therefore, , the maximum value that each objective term can take when optimized individually: $norm_1$ for the term related to utility, $norm_2$ for the term related to the number of pre-assigned doctors and $norm_3$ for the term related to the use of overtime. It is then possible to transform the objective function into a normalized function where the sum of the percentages of each objective is considered.

The normalized multi-objective function, with all the objectives with the same order of magnitude, shown by Expression (15) can, then, assume a value between 0 and 2, being 0 when every term is null and reaching 2 when the two maximization terms assume their maximum value and no overtime is used in the planning horizon.

$$Maximize \frac{1}{norm_1}\sum_{s\in S}\sum_{p\in P}\sum_{d\in D}\sum_{o\in O}\frac{utility_p x_{spdo}}{s+1} + \frac{1}{norm_2}\sum_{s\in S}\sum_{p\in P}\sum_{d\in D}\sum_{o\in O} predoc_{pd}\, x_{spdo} - \frac{1}{norm_3}\sum_{s\in S}\sum_{o\in O} overtimebin_{so} \tag{15}$$

4.3. Chapter considerations

An integer linear programming model is proposed in this chapter for a generalized version of the operating room schedule problem based on the characteristics of the hospital under study. The model links the new prioritization system with the optimization of the operating room scheduling to answer to this dissertation's objective.

Multiple objectives are considered in the model: surgeries' utility values compliance, pre-assigned doctor compliance and minimization of the overtime use. The utility value assigned to each patient is incorporated in the objective function, trying to maximize the sum of the utility values of the patients that are scheduled in the planning horizon. This intends to better select the patients of the waiting list to be scheduled to attend their real needs and to avoid the consequences on the physical, psychological and social conditions caused by long waiting times. The multiple objectives are

normalized in order to obtain the same order of magnitude and summed up into a single objective function.

The instances used as input in the computational experiments are presented in Chapter 5 and results are discussed in Chapter 6.

5. Instances

The model formulation described in Chapter 4 demand a specific type of input. This chapter is organized in two sections: Section 5.1. explains the main characteristics of the input; Section 5.2. describes how the patients', surgeries' and doctors' characteristics used to test and validate the model are created.

5.1. Instances description

Instances are required for the computational experiments, including model validation and sensitivity analysis. Although it was not possible to obtain real data on waiting lists, some insights of the hospital dimensions, regarding the number of operating rooms, doctors and operations rules were provided by the staff. This information is presented in this section.

The characteristics shown in Table 5 are based in the hospital data. For problem complexity simplification, instead of generating schedules for 4 weeks at once, the planning horizon is cut in half: patients are scheduled in two weeks that corresponds to 10 sessions. Based on the hospital information, a maximum of 3 operating rooms are assigned to the urology department of the hospital. Besides, 8 doctors are distributed in 5 sub-specialties surgeries. A waiting list with a larger number of patients than the scheduling capacity of surgeries in the planning horizon is created: 200 patients were generated, in order to be more evident the consequences of the method in waiting times and patients' needs.

Number of sessions	Number of operating rooms	Number of doctors	Number of patients	Number of sub-specialties
10	3	8	200	5

Table 5: Main characteristics of the instances.

Based on the problem definition, the model's parameters must also be set according to the rules of the hospital under study. In fact, it is necessary to define the number of minutes in which a doctor can work within the regular time and the maximum extension in minutes that some session work can have. Moreover, the maximum number of surgeries that can be performed by some doctor on each session needs to be stated. Finally, there is a limited number of operating rooms available in each session to be used by the department under study that is necessary to establish. These parameters are summarized in Table 6.

Parameter	Description	Value
$\boldsymbol{ot} \in \mathbb{N}$	Maximum overtime in minutes in which surgeries can be performed	30
$\boldsymbol{rt} \in \mathbb{N}$	Maximum regular time in minutes in which surgeries can be performed	480
$\boldsymbol{mns} \in \mathbb{N}$	Maximum number of surgeries to be performed in each session by each doctor	6
$\boldsymbol{mwd} \in \mathbb{N}$	Number of session-operating room blocks in which a doctor is assigned within the planning horizon	2
$\boldsymbol{noa_s} \in \mathbb{N}, \forall \boldsymbol{s} \in \{\mathbf{0}, \mathbf{1}, \mathbf{4}, \mathbf{5}, \mathbf{6}, \mathbf{9}\}$	Number of available operating rooms in which surgeries can be planned, on Mondays, Tuesdays and Fridays	2
$\boldsymbol{noa_s} \in \mathbb{N}, \forall \boldsymbol{s} \in \{\mathbf{2}, \mathbf{3}, \mathbf{7}, \mathbf{8}\}$	Number of available operating rooms in which surgeries can be planned, on Wednesdays and Thursdays	3

Table 6: Mathematical model's parameters definition.

Considering that surgeries are performed between 8h00 and 16h00, it must be established that a maximum of 8 hours, so 480 minutes, of surgeries can be scheduled in each session and operating room. Accordingly, it is defined that $rt = 480$. Currently, besides the established regular time, overtime is not allowed in the hospital, but it is aimed to analyze the impact of permitting it. For that reason, a maximum of 30 minutes of overtime is granted in each session-operating room block. Thus, $ot = 30$ to analyze the benefits that could be extracted by allowing the surgeries to last for half hour after the end of the usual operating room schedules.

Additionally to the time constraints, the number of authorized surgeries performed within a day are also limited. Based on the hospital information, a maximum of 6 surgeries per session should be assigned to each doctor in order to control the workload within a day and among surgeons. For that reason, $mns = 7$. In fact, besides limiting the amount of work that each doctor has in a day, it is also necessary, because of economical and personal factors, to force that each doctor is assigned to some surgeries within the planning horizon. In the specific case of the hospital under study and considering the number of doctors and the length of the planning horizon it was established that each doctor should be assigned to, at least, two session per every two weeks. Due to that, $mwd = 2$. It was stated by the staff that working less than this minimum amount could have repercussions in this surgery skills of a doctor.

Lastly, it is necessary to take in account the master surgery schedule of the hospital. As explained in the definition of the problem, the number of available operating rooms in each session-operating

room blocks for the urology department of the hospital depend on the day of the week. So, for Mondays, Tuesdays and Fridays, two operating rooms are available: $noa_s = 2,\ s \in \{0, 1, 4, 5, 6, 9\}$. On the other hand, on Wednesdays and Thursdays, an additional operating room is available: $noa_s = 3, \forall s \in \{2,3,7,8\}$.

5.2. Instances generation

To validate and test the quality of the solutions, it is necessary to have patients' and doctors' characteristics input data. The model is structured and parameterized according to the data provided by the hospital under study. However, it was not possible to obtain real data from the patients on the waiting list for elective surgeries. For this reason, a programming model was coded in Java to generate random instances that reasonably represent the waiting list of patients and the characteristics of the surgeons.

After defining that there are 200 patients on the waiting list to be scheduled in 10 sessions and 3 operating rooms and that the number of existent doctors in the urology department of the hospital is 8 and the number of existing sub-specialties is 5, the data correspondent to each one of them needs to be predicted.

A summary of the instances' generator is presented in Table 7.

Parameter	Generator procedure
$\boldsymbol{utility_p}$	$utility_p \sim \mathrm{N}(65,15) \in]0,100]$
$\boldsymbol{duration_p}$	$duration_p \sim \mathrm{Lognormal}(\mu,\sigma,\gamma)$
$\boldsymbol{subspecialty_p}$	$subspecialty_p \sim \mathrm{U}(1,5)$
$\boldsymbol{waitingsess_p}$	$waitingsess_p \sim \mathrm{U}(0,20)$
$\boldsymbol{predoc_{pd}}$	$predoc_{pd} \sim \mathrm{U}(0,8)$
$\boldsymbol{skillsbin_{dp}}$	$skillsbin_{dp} \sim \mathrm{Bern}(0.7)$
$\boldsymbol{skills_{dp}}$	$skills_{dp} \sim \mathrm{U}(0.8,1.2)$
$\boldsymbol{doctoravailability_{ds}}$	$doctoravailability_{ds} \sim \mathrm{Bern}(0.9)$

Table 7: Instance generator's procedure summary.

The following information was generated to each patient in the waiting list: utility value, surgery average durations and sub-specialty, the number of sessions waiting and if there was any doctor pre-assigned to the surgery to be performed. The utility value of patient p, $utility_p$ is achieved based on a normal distribution with mean value of 65 and a standard deviation of 15 and between 0 and 100. On the other hand, surgery average duration of patient p, $duration_p$, is generated by a 3-parameter lognormal distribution using data from a urology department database provided by the CHOIR group of the University of Twente[1] . The patient p's surgery sub-specialty, $subspecialty_p$, the waiting time sessions, $waitingsess_p$, and the pre-assigned doctors, $predoc_{pd}$, are obtained according to a uniform distribution between 1 and the number of sub-specialties, between 0 and 20 sessions and between 0 and the number of doctors, respectively.

Regarding doctors, it is necessary to generate their skills in each one of the 5 sub-specialties and their availability in each session. For doctors' skills, first, it is necessary to determine if they are able to perform surgeries of each sub-specialty, $skillsbin_{dp}$. Thus, a 70% chance of being capable and a 30% chance of not being capable is established. In the case of being capable, a number between 0.8 and 1.2 is assigned to $skills_{dp}$, to represent how fast a doctor can perform that type of surgery, meaning that, if a doctor is assigned with 1, the surgery performed by them will last for the average duration, but if the number is higher or lower than 1, then the doctor will perform it faster or slower, respectively. Moreover, for doctors' availability, $doctoravailability_{ds}$, there is a 90% chance of a doctor being assigned as available on each session and a 10% chance of not being available.

This procedure is applied to generate the 10 instances used in the computational experiments. These experiments are discussed in the next chapter.

[1] available at https://www.utwente.nl/en/choir/research/BenchmarkORScheduling/. Consulted on 21-06-2018

6. Computational experiments

In this chapter, the model is validated through ten instances and shows good results within a short running time (Section 6.1.). In addition, a randomly picked instance is used to analyze the dimensions of the problem and sensitivity analyses on changes in parameters and on the objective function are performed with the instance with best results (Section 6.2.). Finally, a rolling horizon approach is proposed and the impact of the dynamic utility is analyzed (Section 6.3.). The chapter ends in Section 6.4. with some conclusions.

6.1. Model validation

The tests are performed using a portable computer with Intel ® Core™ i7-3632QM CPU @ 2.20GHz 2.20 GHz – RAM 6.00GB and the Windows 10 operating system. The model was implemented in Java with Eclipse Java Mars.2, using the callable library of ILOG CPLEX 12.8.0. CPLEX is a powerful tool to model and solve optimization problems. It uses a branch-and-cut method to solve the integer programming models ("IBM Knowledge Center," n.d.). Tests are performed in the 10 instances described in the previous chapter.

For each test, the process followed is presented on Figure 5.

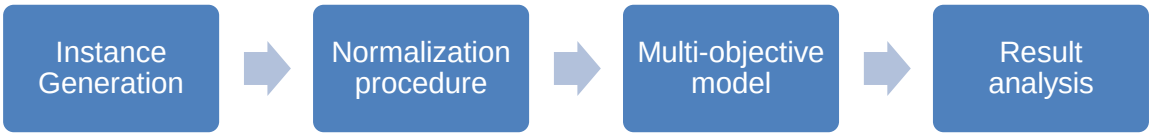

Figure 5: Process used to perform tests on the model.

For the normalization procedure it is necessary to calculate $norm_1$, $norm_2$ and $norm_3$. These three parameters represent the maximum values that each term of the objective function can take, when considered as the only objective. $norm_1$ corresponds to the utility compliance term; $norm_2$ corresponds to the pre-assigned doctors' compliance term; and $norm_3$ corresponds to the overtime use term. These three values are computed running the model with the instance to be tested, with a 10-minute running time limit. The results obtained on the objective function in each one of the three tests are assigned to the parameters. After computing those three values, the model runs with the normalized multi-objective function, for which the time limit is extended to 60 minutes. To refine the results, a coefficient of 1000 was multiplied to each term of the objective function.

The number of constraints and the number of variables generated with the model and the relative gap for each one of the 10 instances in 1-hour tests are described in Table 8. The relative gap is

calculated as the percentage that the difference between the best bound and the best integer solution represents when compared to the best integer solution value[2].

Instance ID	Number of variables	Number of constraints	Relative gap (%)
1	21093	36448	0.57
2	21699	38858	1.72
3	25920	47293	0.92
4	25314	44883	0.13
5	22302	40063	2.48
6	22905	41268	1.10
7	25317	46088	0.94
8	22905	41268	0.62
9	23508	42473	1.07
10	25314	44883	2.51

Table 8: Results for the number of variables, number of constraints and relative gap (%) for ten instances with 1-hour time limit.

Considering 10 instances with the refered dimensions, meaning, with 200 patients, 3 operating rooms, 8 doctors and 10 sessions generate problems to be solved with 36448 to 47293 constraints and 21093 to 25920 variables. The best result is obtained for Instance 4, with a relative gap of 0.13% and the worst result is obtained for Instance 10, with a relative gap of 2.51%. For these problem sizes, the model can generate good results in a reasonable time.

[2] Relative gap = $\frac{Best\ Bound - Best\ Integer}{Best\ Integer}$

6.2. Sensitivity Analysis

In this section, the sensitivity of the model to changes in parameters and instances and results from a practical point of view are analyzed. The impact of changes in the quality of the solutions obtained is also evaluated. Instance 3 is randomnly picked to be used to test the sensitivity of the model to the dimensions of the problem and its complexity and the practical results of interest to the hospital are analyzed.

The following changes are tested: considering only the first half of the patients on the waiting list (fewer patients) and considering only a one-week planning horizon (shorter planning horizon). The minimum workload Constraints (7) were adjusted for the shorter planning horizon tests. So, it is defined that each doctor must be assigned, at least, once in the planning horizon (mwd =1) for this case. Moreover, changes are applied and tested on the model with Instance 3 regarding time capacity, surgery duration, doctors' skills and availability and on the constraints. All these changes impact the symmetry of the problem and on the results. One-hour computation time limit is also used in these tests.

Table 9 shows, for instance 3, the results on the number of variables, number of constraints and relative gap when only the first 100 patients on the waiting list are considered and when the planning horizon is reduced to one week and compares it with the results with no changes in the number of patients and the number of sessions (200 patients and 10 sessions).

Instance 3	Number of variables	Number of constraints	Relative gap (%)
No changes	25920	47293	0.92
100 patients	13020	23793	0.05
5 sessions	14463	24348	~0.00

Table 9: Results on sensitivity analysis for Instance 3 regarding the number of variables, number of constraints and the relative gap (%).

Considering only the first 100 patients of the waiting list and considering a planning horizon of only one week and at least one slot assigned to each surgeon, generate problems with a reduced number of variables and constraints, which lead to a better relative gap value. Indeed, when considering only half of the length of the planning horizon, there is a higher impact in relative gap within the computing time limit: the results are retrieved after 7 seconds, because the relative gap value dropped below the limit of 0.01%. However, the relative gap does not depend only on the number of variables and on the number of constraints.

In fact, the problem's complexity can depend on many factors, namely the way in which the problem is formulated, their parameters or even the instance's characteristics. Table 10 presents the results obtained from several changed scenarios tests with computing time limit of 1 hour and compares it with the results retrieved by the model on Table 8, regarding Instance 3.

Scenario	Relative gap (%)
Instance 1 - base	0.92
CAP30010	4.11
SDur25^{+}	3.57
SDur25^{-}	~0.00
MaxNbSur5	0.09
EqSkills1	0.92
AllSkills1	3.55
DAvail	2.80
MinW1	1.56

Table 10: Results regarding the relative gap (%) for Instance 3 and changed models.

Instance 1 - base represents the scenario in which no changes are applied to the model parameters or to the instance – base scenario. As mentioned previously, Instance 3 is a randomly picked instance from Table 8.

The first tests, CAP30010, SDur25^{+} and SDur25^{-} analyse how changes regarding time can impact the results. This model deals with time in the capacity of the operating rooms and the duration of surgeries.

CAP30010 represents the scenario in which the maximum regular time capacity (ot) is set to 300 minutes and the maximum overtime capacity (rt) is 10 minutes. In this case, a considerably greater relative gap was obtained. While limiting the time capacity, the number of scheduled patients' is reduced and a more detailed selection of patients must be done to fit on fewer available slots,

considering the surgery's duration and the contribution that each surgery has to the objective function, which may increase the complexity of the problem.

SDur25$^+$ represents the scenario in which the average surgery duration $duration_p$ of each patient is increased by a 25% factor. This change lead to an increase in the relative gap but kept retrieving good results in a reasonable time. This case can be justified by the same reason as in the scenario CAP30010. In fact, having longer surgeries hamper the schedules to have the same number of scheduled surgeries and will complicate the time management process. With this surgery duration, it is hard to keep the same utilization rates, because there will not be short surgeries to fill the breaches that may arise in the schedules.

SDur25$^-$, on the other hand, represents the scenario in which the average surgery duration $duration_p$ of each patient is decreased by a 25% factor. Contrarily to SDur25$^+$, SDur25$^-$ allows to reduce complexity as the relative gap gets lower than 0.01% (gap tolerance) after 12 seconds running time. In this case, a greater number of patients can be scheduled, and it is easier to fit a surgery in the schedule, which can decrease the problems' complexity.

Besides, the capacity can also be limited to each surgeon when establishing a maximum number of surgeries that can be performed by each one of them in a session. MaxNbSur5 represents the scenario in which a maximum of 5 surgeries, instead of 6, can be performed by the same doctor on a session (parameter mns assumes value 5). By decreasing the maximum number of surgeries to be performed by each doctor in a session, a reduced relative gap is obtained, because, although less surgeries can be scheduled on that session, there is a smaller number of surgery combinations that can be done.

The assignment of a doctor to surgeries, in addition to being limited by the number of surgeries, can also be affected by their skills and availability during the planning horizon. In fact, a doctor can only perform a surgery in some session if that doctor has the right skills and is available on the assigned session.

As noted in Chapter 3, doctors are divided into 5 groups referring to types of sub-specialties of surgeries that they can perform. In addition, if the doctor performs a certain type of surgeries, a number between 0.8 and 1.2 is also assigned ($skills_{dp}$), that reflects the experience and speed with which the doctor performs this type of surgeries. EqSkills1 represents the scenario in which doctors that are able to perform surgeries from some sub-specialty do it in the same rate, meaning that every skilled doctor performs the surgery in their average expected value ($skills_{dp} = 1, \ \forall s \in S, d: skillsbin_{pd} = 1$). This test showed a similar relative gap value to the base scenario, which may be due to the fact that $skills_{dp}$ can only take a value between 0.8 and 1.2 and so changing it to 1 does not have a significant impact on the results.

Besides, another test was performed regarding doctors' skills. AllSkills1 represents the scenario in which all doctors are equally skilled, i.e. every doctor is able to perform every type of surgery and at the same speed ($skillsbin_{dp} = 1$ and $skills_{dp} = 1, \ , \forall \ d \in D, p \in P$). This scenario generates a higher

relative gap value, 3.55%, because adds great symmetry which impacts on the complexity of the problem. Indeed, having the possibility to assign any doctor to any surgery with no difference in the real duration of the surgery, makes the problem harder to solve.

In addition to doctors' skills, their availability must also be considered. DAvail represents the scenario in which every doctor is completely available for the operating room, i.e., every surgeon is available to perform surgeries in every session ($doctoravailability_{ds} = 1, \forall\, d \in D, s \in S$). Again, allowing every doctor to be scheduled in any session also increases the symmetry of the problem and thus the relative gap also increases to 2.80%. However, the increase is not as high as in the previous case since in this case the doctor can be assigned to any of the 10 sessions, while in the previous case the doctor could be assigned to any of the 200 surgeries on the waiting list.

Additionally to the characteristics of doctors, in order to ensure that there is some equity in the surgeons' access to the operating room, it is established that each doctor must be assigned, at least, to two operating rooms in the planning horizon. MinW1 represents the scenario in which this constraint is relaxed, by defining that a doctor should be assigned, at least, to only 1 session during the planning horizon (mwd=1). Although there is a small increase in the relative gap value, this modification does not have significant impact the relative gap. In fact, by analyzing the results obtained, all physicians were assigned to more than one session.

The results on Table 10 also validate the model as showing that, despite having some higher and some lower relative gap values, this model can obtain good results (relative gap value inferior to 5%) in a reasonable time (1 hour), independently of the parameters and of the changes that may occur to the parameters of the instances. This also validates the use of this model in other contexts or case studies.

Considering the base scenario (initial characteristics and parameters), it is also possible to analyze the potential practical results of the model. Using as an example, again, Instance 3, Table 11 shows the results for the value of each term of the objective function, when running the normalized multi-objective function for 1 hour. It also shows the percentage that the term values represent comparing to their maximum values ($norm_1$, $norm_2$, $norm_3$) computed as single objectives.

Utility value term	Pre-assigned doctors' compliance term	Overtime use term
2752.62	10.0	0.00
100.00 %	16.39 %	0.00 %

Table 11:Results on each objective function term for Instance 3.

It is shown that 10 out of the 61 pre-assigned doctors (16.39% of the pre-assigned doctors) that were possible to schedule, in the best-case scenario, were assigned to some surgery on the planning horizon. Additionally, no overtime was used in any session and operating room during the two weeks of the planning horizon. Regarding the utility and waiting time value term, it is difficult to analyse the results based on their contribution to the objective function.

For that reason, it is shown, in detail, on Table 12 and Table 13 the average, minimum value, maximum value and standard deviation on the utility values and on the waiting time of scheduled patients, respectively.

Average utility value of scheduled patients	Minimum utility value of scheduled patients	Maximum utility value of scheduled patients	Standard deviation of utility value of scheduled patients
69.56	31.00	100.00	14.03

Table 12: Analysis on the utility values of scheduled patients of Instance 3.

The scheduled patients have an average utility value of 59.56, with a standard deviation of 14.03. This average value represents a relatively high number. Besides, it is known that a patient with a utility value of 31.0 is scheduled. In fact, patients with lower utility values are not and should not be neglected in the scheduling process. Even though patients with highest utility values should have some kind of priority in the scheduling process, there must have some balance and patients with low utility values cannot be waiting for too long, or else, they will also suffer the consequences on their conditions. Moreover, all patients on the waiting list with utility value 100 (total of 2) are scheduled.

Average waiting time of scheduled patients (number of sessions)	Minimum waiting time of scheduled patients (number of sessions)	Maximum waiting time of scheduled patients (number of sessions)	Standard deviation of waiting time of scheduled patients (number of sessions)
9.94	0.00	19.00	5.74

Table 13: Analysis on the waiting time of scheduled patients of Instance 3.

The average waiting time of scheduled patients was 9.94 sessions with a standard deviation of 5.74. Considering only one static planning horizon, there can be patients waiting for approximately one month, 20 sessions, using the instance generator described before. For that reason, the maximum waiting time of scheduled patients is 19 sessions.

Minimum workload per doctor (minutes)	Maximum workload per doctor (minutes)	Number of scheduled surgeries
924.33	2364.70	125

Table 14: Results on minimum and maximum workload per doctor (in minutes) and number of scheduled surgeries for Instance 3.

Moreover, Table 14 shows the number of scheduled minutes, considering the doctors' skills, that were assigned to the doctor with least work and to the doctor with most work. It also shows the number of scheduled surgeries on the 10 sessions.

As stated on Expression (7), each doctor must be working at least on two sessions during the planning horizon. The minimum workload of 924.33 minutes, that corresponds to approximately the time capacity of two sessions (960 minutes), can be justified by that constraint. The doctor with most amount of work is then assigned to five sessions (maximum regular time capacity of 2400 minutes). There is no big imbalance to be registered on doctors' workload. However, for a different number of doctors or for a different planning horizon length, it is necessary to re-consider this parameter that defines the minimum number of sessions in which a doctor must be assigned.

As a matter of fact, this model implements some equity on the patients' scheduling considering their needs:the average values of utility of scheduled patients is higher than those average values when considering the whole waiting list. As shown in the literature review, patients' surgery utility is not a usual objective in the operating room schedules. Normally, the patient scheduling is done in a FCFS basis or even considering succinct levels of prioritization and the patients' real needs are not considered.

To better answer the research question, some modifications are done in the objective function term that considers the new priority measures and the waiting times for surgeries. Tests are performed to analyze what is the impact of not considering the prioritization term as an objective.

For Instance 3, the model was modified to consider as only objective the scheduling of patients already waiting for the longest time on the earliest possible session, in the form of Expression (16). With this test, it is intended to observe the impact in the utility values measures of scheduled patients of not considering it in the objective.

$$Maximize \sum_{s \in S} \sum_{p \in P} \sum_{d \in D} \sum_{o \in O} \frac{waitingsess_p \, x_{spdo}}{s+1} \tag{16}$$

Using once again a 1-hour computing time, the results regarding Objective (16), on the utility values of scheduled patients, the waiting times of schedules patients and regarding the workload and number of scheduled surgeries, are presented in Table 15, Table 16, Table 17 and Table 18 respectively, and compared with the original model. A relative gap of 1.46% was retrieved.

Objective function	Utility value term	Pre-assigned doctors' compliance term	Overtime use term
Expression (16)	2469.00	8.00	22.00
Expression (15)	2752.62	10.00	0.00

Table 15: Results on each objective function term for Instance 3, while using Expression (16) and Expression (15) as the objective function.

As expected there is less compliance with pre-assigned doctors and overtime is used in almost every session-operating room block, because the maximization of pre-assigned doctors' compliance and the minimization of overtime use are not forced. The utility value term has a higher value when using Expression (15) than the value obtained when Expression (16) is used, and to better understand this result, Table 16 and Table 17 give some insights on the utility values and waiting times of scheduled patients.

Objective function	Average utility value of scheduled patients	Minimum utility value of scheduled patients	Maximum utility value of scheduled patients	Standard deviation of utility value of scheduled patients
Expression (16)	64.43	23.00	100.00	14.89
Expression (15)	69.56	31.00	100.00	14.03

Table 16: Analysis on the utility values of scheduled patients of Instance 3, while using Expression (16) and Expression (15) as the objective function.

The average utility value of the scheduled patients is 64.43, a smaller number than in the multi-objective function, which makes sense since the objective does not include the maximization of the sum of the utility values of scheduled patients. In fact, in this case, the average utility value of scheduled patients is inferior to the average of utility values of all patients on the waiting list, that takes value 64.82.. Thus, using this objective, that mimics more or less the FCFS process, the patients' needs are not considered.

Objective function	Average waiting time of scheduled patients (number of sessions)	Minimum waiting time of scheduled patients (number of sessions)	Maximum waiting time of scheduled patients (number of sessions)	Standard deviation of waiting time of scheduled patients (number of sessions)
Expression (16)	12.94	1.00	19.00	4.03
Expression (15)	9.94	0.00	19.00	5.74

Table 17: Analysis on the waiting times of scheduled patients of Instance 3, while using Expression (16) and Expression (15) as the objective function.

On the other hand, regarding waiting times, there are improved results. The average waiting time of scheduled patients increased from 9.94 sessions to almost 13 sessions. Besides, there is a smaller standard deviation of the waiting time in this test. In fact, no patients with no waiting time are scheduled, independently on their characteristics and condition. Considering the maximization of waiting times of scheduled patients, although making a better selection of patients that are waiting for longer times, ignores patients in more need of a surgery.

Objective Function	Minimum workload per doctor (minutes)	Maximum workload per doctor (minutes)	Number of scheduled surgeries
Expression (16)	1000.50	2075.90	135
Expression (15)	924.33	2364.70	125

Table 18: Results on minimum and maximum workload per doctor (in minutes) and number of scheduled surgeries for Instance 3, while using Expression (16) and Expression (15) as the objective function.

With Expression (16), ten more surgeries are planned and, although we have increased values for the minimum workload and reduced values for the maximum workload, there is no significant disparity in the difference between those two values, since they correspond to the time of two sessions and five sessions, respectively, but now considering the overtime use.

Thus, considering only the waiting time of scheduled patients in the objective function does not impact greatly the workload of surgeons or the number of scheduled patients. However, since only patients waiting for longer times, the patients' needs are not taken in account, and so, patients with more urgent needs of surgeries are not scheduled if they did not wait for long times.

Alternatively, a test was done considering utility compliance as the only objective of the model, meaning that the patients with higher utility values should be scheduled as early as possible in the

planning horizon, independently for how long the patients is on the waiting list. This test aims to analyse what would be the impact of not considering the other two objectives of pre-assigned doctors and overtime compliance. The objective function that was used is presented on Expression (17). The results are retrieved with a 0.82% relative gap.

$$Maximize \sum_{s \in S} \sum_{p \in P} \sum_{d \in D} \sum_{o \in O} \frac{utility_p \, x_{spdo}}{s+1} \tag{17}$$

The results on having the maximization of the utility values of scheduled patients can be seen and compared with the model that considers Expression (15) as the objective function on Table 19, Table 20, Table 21 and Table 22, regarding, respectively, the values that utility, pre-assigned doctors and overtime use would get with the obtained results, the analysis of the utility values of scheduled patients, the analysis of the waiting times of scheduled patients and the doctors' workload and number of scheduled surgeries.

Objective function	Utility value term	Pre-assigned doctors' compliance term	Overtime use term
Expression (17)	3012.41	8.00	23.00
Expression (15)	2752.62	10.00	0.00

Table 19: Results on each objective function term for Instance 3, while using Expression (17) and Expression (15) as the objective function.

Again, as happened for Expression (16) that only considered the waiting times of scheduled patients, there is less compliance with pre-assigned doctors and no significative compliance with the minimization on overtime use when considering only the utility values of scheduled patients. However, a increased value for the utility term is obtained: this result is analysed in more depth with Table 20 and Table 21.

Objective function	Average utility value of scheduled patients	Minimum utility value of scheduled patients	Maximum utility value of scheduled patients	Standard deviation of utility value of scheduled patients
Expression (17)	71.30	43.00	100.00	11.22
Expression (15)	69.56	31.00	100.00	14.03

Table 20: Analysis on the utility value of scheduled patients of Instance 3, while using Expression (17) and Expression (15) as the objective function.

Analysing the results regarding the utility values of scheduled patients, when considering only the maximization of these values, it can be seen the only patients with utility value above 43 are scheduled and the standard deviation is smaller. In fact, the average value of utility increases from 69.56 to 71.30.

Objective function	Average waiting time of scheduled patients (number of sessions)	Minimum waiting time of scheduled patients (number of sessions)	Maximum waiting time of scheduled patients (number of sessions)	Standard deviation of waiting time of scheduled patients (number of sessions)
Expression (17)	9.49	00.00	19.00	5.84
Expression (15)	9.94	00.00	19.00	5.74

Table 21: Analysis on the waiting times of scheduled patients of Instance 3, while using Expression (17) and Expression (15) as the objective function.

Although the minimum and the maximum waiting time are the same, when using Expression (15) and Expression (17) as the objective function, there is a smaller average waiting time for the case being analysed.

Objective function	Minimum workload per doctor (minutes)	Maximum workload per doctor (minutes)	Number of scheduled surgeries
Expression (17)	1025.20	2110.4	138
Expression (15)	924.33	2364.70	125

Table 22: Results on minimum and maximum workload per doctor (in minutes) and number of scheduled surgeries for Instance 3, while using Expression (17) and Expression (15) as the objective function.

Regarding doctors' workload and the number of scheduled surgeries, the conclusions are the same as when using Expression (16) as the objective function: although the minimum workload per doctor, the maximum workload per doctor and the number of scheduled surgeries increase when using Objective (17), there is no significative difference between the number of work sessions for the doctor with least and for the doctor with most work .

Thus, considering only the utility values compliance as an objective to the model does not impact greatly the workload balance between surgeons nor the number of scheduled surgeries. Although

considering Objective (17) allows a better selection of patients regarding their needs and avoid situations as scheduling patients that are not in dire need of a surgery (as the example given on Chapter 3), it does not consider the other objectives marked as important by the hospital staff.

Besides, improving equity in access to elective surgeries, the hospital under study also intends to analyse the impact of allowing overtime in operating rooms. As seen in Table 11, for the multi-objective function, allowing a 30-minutes overtime period did not have any impact on the results, because no overtime is used to improve schedules. To examine if extending the maximum extraordinary work time could be beneficial to the scheduled, the test is repeated considering a maximum overtime of 1 hour, instead of only 30 minutes ($ot = 60$ minutes). Although in the normalization process the results obtained are different, using the normalized multi-objective function, the results are the same as in Table 11 and still no overtime is used. In consequence, it can be concluded that, in this context and with this model, the overtime allowance inferior or equal to 1 hour will not impact the results, which supports the hospital decision in not using only operating room regular time.

However, objectives as overtime minimization are very common in the literature. Besides cost or waiting time minimization or even maximization of number of performed surgeries are some other of the most common ones. In fact, maximizing the number of scheduled surgeries can improve operating room efficiency. For this reason, a test was also done to investigate what is the cost of not consider this objective in the proposed model. Thus, the objective function was replaced by Expression (18). The results and comparison with the model that uses Objective (15) are presented in Table 23, Table 24, Table 25 and Table 26, for which an optimal solution was found in 18 seconds.

$$Maximize \sum_{s \in S} \sum_{p \in P} \sum_{d \in D} \sum_{o \in O} x_{spdo} \tag{18}$$

Objective function	Utility value term	Pre-assigned doctors' compliance term	Overtime use term
Expression (18)	2402.09	9.00	17.00
Expression (15)	2752.62	10.00	0.00

Table 23: Results on each objective function term for Instance 3, when considering Expression (18) and Expression (15) as the objective function.

As happened for Objective (16) and Objective (17), there was a reduced compliance of pre-assigned doctors and minimization of overtime use, since those objectives are not considered. Regarding the utility and waiting time term, a much lower number was obtained when using Expression (18) as the objective. Table 24 and Table 25 show the average, minimum, maximum and standard deviation of utility value and waiting times of scheduled patients.

Objective function	Average utility value of scheduled patients	Minimum utility value of scheduled patients	Maximum utility value of scheduled patients	Standard deviation of utility value of scheduled patients
Expression (18)	64.25	19.00	100.00	14.88
Expression (15)	69.56	31.00	100.00	14.03

Table 24: Analysis on the utility value of scheduled patients of Instance 3, while using Expression (18) and Expression (15) as the objective function.

Regarding the utility values of scheduled patients, no significative differences to the model using Expression (15) is found concerning the maximum and the standard deviation values. However, there is a significative reduction in the average utility value of scheduled patients.

Objective function	Average waiting time of scheduled patients (number of sessions)	Minimum waiting time of scheduled patients (number of sessions)	Maximum waiting time of scheduled patients (number of sessions)	Standard deviation of waiting time of scheduled patients (number of sessions)
Expression (18)	10.20	0.00	19.00	5.78
Expression (15)	9.94	0.00	19.00	5.74

Table 25: Analysis on the waiting times of scheduled patients of Instance 3, while using Expression (18) and Expression (15) as the objective function.

The results regarding the waiting times also show no significative differences for the average, minimum, maximum and standard deviation values.

Objective function	Minimum workload per doctor (minutes)	Maximum workload per doctor (minutes)	Number of scheduled surgeries
Expression (18)	969.39	2621.80	138
Expression (15)	924.33	2364.70	125

Table 26: Results on minimum and maximum workload per doctor (in minutes) and number of scheduled surgeries for Instance 3, while using Expression (18) and Expression (15) as the objective function.

With respect to the workload, despite having almost the same minimum workload per doctor, the workload difference between the doctor with most work and the doctor with least work increases by 212.04 minutes. Besides, the number of scheduled surgeries, supported by the objective function, also increases.

Analysing the results, it can be seen that the utility term and the pre-assigned doctors' compliance term values decrease and the overtime term value increases, meaning that this model does not comply with any of the original objectives. Moreover, although minimum, maximum and standard deviation values regarding utility values and waiting times of scheduled patients are similar to the results of the described model, the average utility value is reduced when comparing both cases. The workload of the doctor with most work also increases, while the workload of the doctor with least work suffer just a slightly increase, which causes a greater imbalance (although not significant) of work time between surgeons. All these changes are caused by an increase of only 10 surgeries on the number of scheduled surgeries in the two weeks of the planning horizon. Accordingly, the model that considers the patients' utility values, pre-assigned doctors' compliance and overtime use represents a good trade-off between equity, needs and efficiency.

The results on this section validate the model regarding utility values and efficiency and equity on schedules. In fact, considering this new prioritization method in the optimization system for surgery scheduling can generate more equitable schedules, since patients can be selected considering their need of surgery, instead of considering only the waiting time. Although overtime is used when single objectives were optimized, as maximizing the utility values of scheduled patients' compliance or the maximization of waiting time of scheduled patients, when the multi-objective function is considered, overtime of 30 minutes or 60 minutes is not used. This allows to conclude that overtime does not improve the surgical schedules and thus it is a good decision for the hospital under study to not allow overtime. The model also shows good levels of efficiency, because the number of scheduled surgeries is not much far to the ideal number.

6.3. Rolling horizon

After validating the model in one planning horizon period of two weeks, it is important to predict the longer-term impact that this model formulation can have in the patients' conditions and their stay in the waiting lists.

Normally, waiting lists are considered as static lists of patients to facilitate research purposes. However, it is known that this is not true. In fact, there are always patients entering and leaving the surgeries' waiting lists. Moreover, when managing operating room scheduling, the evolution on patients' condition is often not taken in account because it is very difficult to measure and predict. This could change the selection of patients since patients' evolution rates differ from each other. A rolling horizon approach is proposed in this section to deal with the fact that waiting lists and patients' conditions are dynamic.

Thus, after generating the schedule for the first two weeks, it is necessary to update the waiting list. At the end of each planning horizon, the characteristics of patients that were not scheduled must be updated, the patients that were scheduled have to be deleted and some new patients need to be entered there. The proposed approach to update the waiting list after the generation of a schedule is shown in Figure 6.

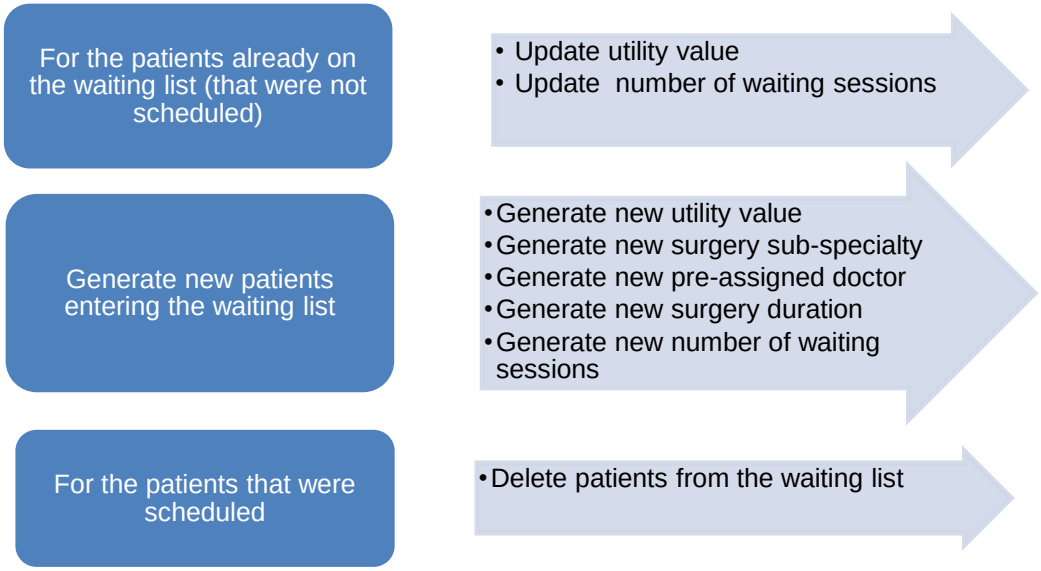

Figure 6: Overall scheme on the waiting list update process after each planning horizon.

The scheduled patients should be deleted from the waiting list, since there is already a session, operating room and doctor assigned to them in the previous considered planning horizon. Hence, each one of the deleted patients is replaced by new patients generated by the same process as the first group of patients, described on Chapter 5. The utility value of the new patients is assigned using a normal distribution with mean value 65 and standard deviation 15; the new surgery duration is computed based on a lognormal distribution; the sub-specialty and the pre-assigned doctor are defined based on a uniform distribution between 1 and 5 and between 0 and 8 respectively. However, the definition of the waiting time on the list needs to be done in a different way: in that case, a uniform distribution is also used but a number between 0 and 10 is generated. That change is justified by the fact that the new patient entered the waiting list during the two weeks period of the previous planning

horizon in which they were not considered, and so the maximum time that may have passed since they entered the list is of 10 sessions.

Besides, for the patients that were not scheduled, the waiting time and the utility value must be updated. For the waiting time, 10 sessions are added to represent the planning horizon that already have passed in which they were not scheduled. The utility value, on the other hand, can be replaced considering many factors.

A formula to update utility of patient p to be considered in the planning horizon t, $utility_p^t$, is presented by Expression (19) , using new parameters summarized on Table 27. This formula aims to represent the evolution of the patient's condition caused by the time spent on the waiting list without having their surgery performed.

Notation	Description
$\boldsymbol{utility_p^t} \in \mathbb{N}\backslash\{\mathbf{0}\}$	Utility value of patient $p \in P$ on the planning horizon $t \in \mathbb{N}$.
$\boldsymbol{evl_p} \in \mathbb{R}^+$	Reflects the prediction of how fast the condition of patient $p \in P$ will evolve.

Table 27: Summary on parameters' notation used for the proposed rolling horizon approaches.

$$\boldsymbol{utility_p^t = utility_p^{t-1} \times evl_p}, \forall\, \boldsymbol{p} \in \boldsymbol{P}, \boldsymbol{t} \in \mathbb{N} \tag{19}$$

Expression (19) updates utility making use of a new parameter evl_p, which is a factor that is assigned to each patient on the waiting list when the surgery is recommended that predicts how fast the condition can evolve considering not only physical but also social characteristics of the patient. This parameter should be defined by some expert that evaluates the patient when the need of a surgery is identified, but since it was not possible to test and create a systematic method, a random number between 1 and 1.3 is generated to represent the evolution of each patient's condition. evl_p is multiplied to the utility value of the patient in the previous planning horizon, to compute the value of utility to be considered in the planning horizon that is being analysed.

Computational experiments are performed to test the impact of assuming a dynamic waiting list, when comparing to a static waiting list. Using again Instance 3 and considering as results obtained for the first planning horizon the ones used to fill Table 11, the utility values of patients on the waiting list are updated according to Expression (19) for the dynamic version and it is not updated for the static version. Besides, the results are also compared to the model that considers an objective function that aims to comply with waiting times instead of utility values.

Considering the next 5 planning horizons, and assuming that, during each planning horizon, 150 new patients are added to the waiting list, the results regarding the selection of patients and waiting times for the dynamic model being proposed are presented on Figure 7 and Figure 8, respectively.

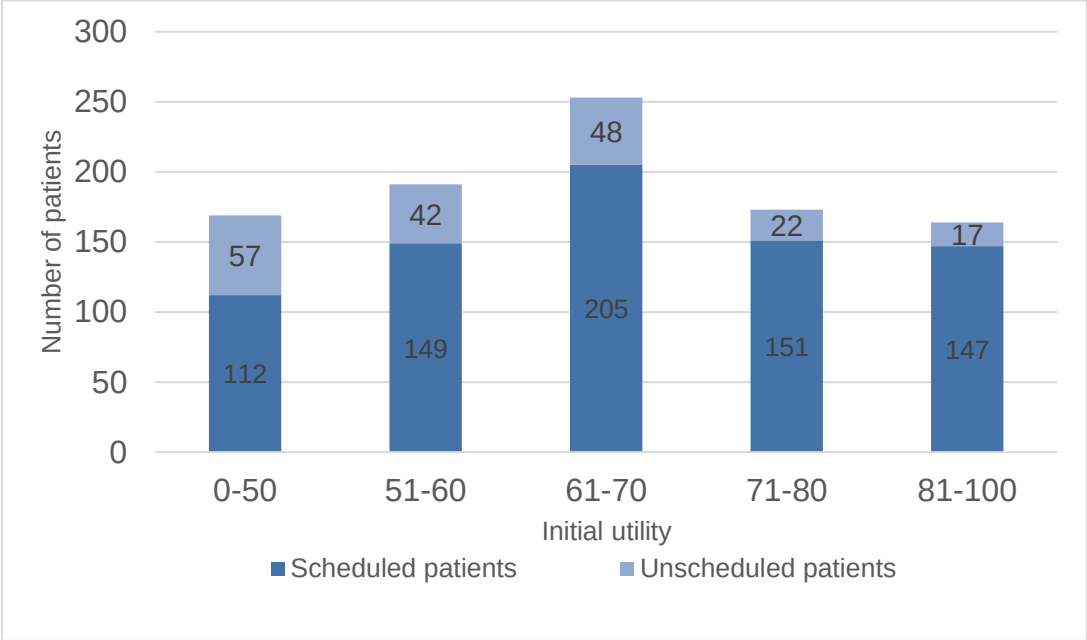

Figure 7: Proportion of scheduled and unscheduled patients for each class of patients (defined by their initial value of utility).

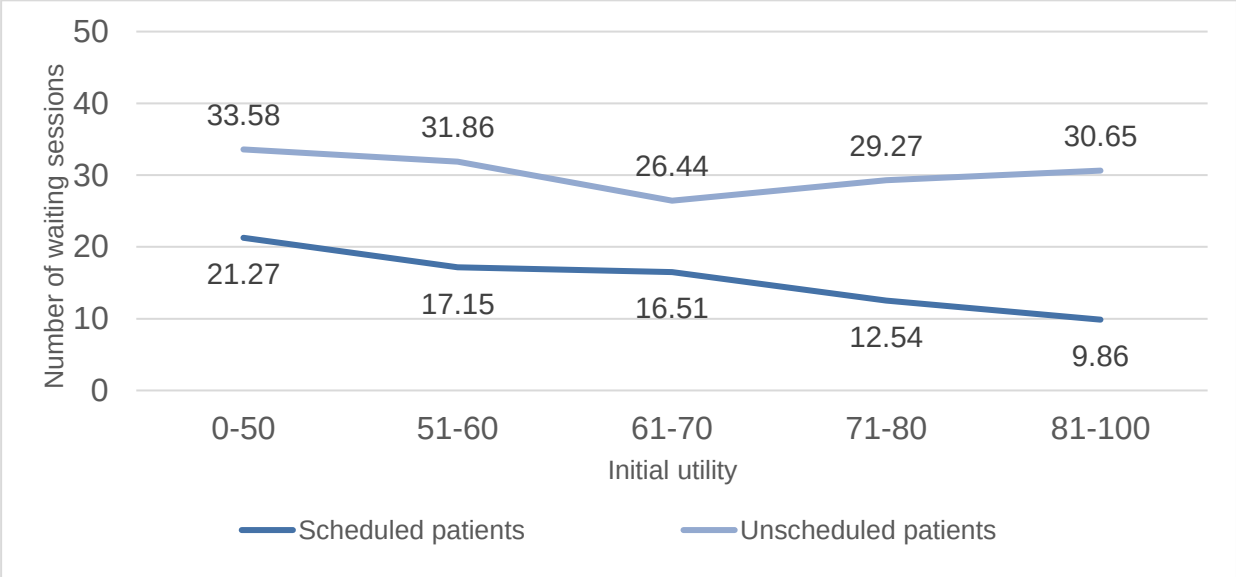

Figure 8: Average number of waiting sessions for scheduled and unscheduled patients of each class of patients defined by their initial values of utility.

For comparison reasons, the results, regarding selection of patients and waiting times, obtained for the same model, but without considering the update of utility values and the results obtained when the utility term is replaced by the maximization of waiting times (to mimic the FCFS approach) are also presented on

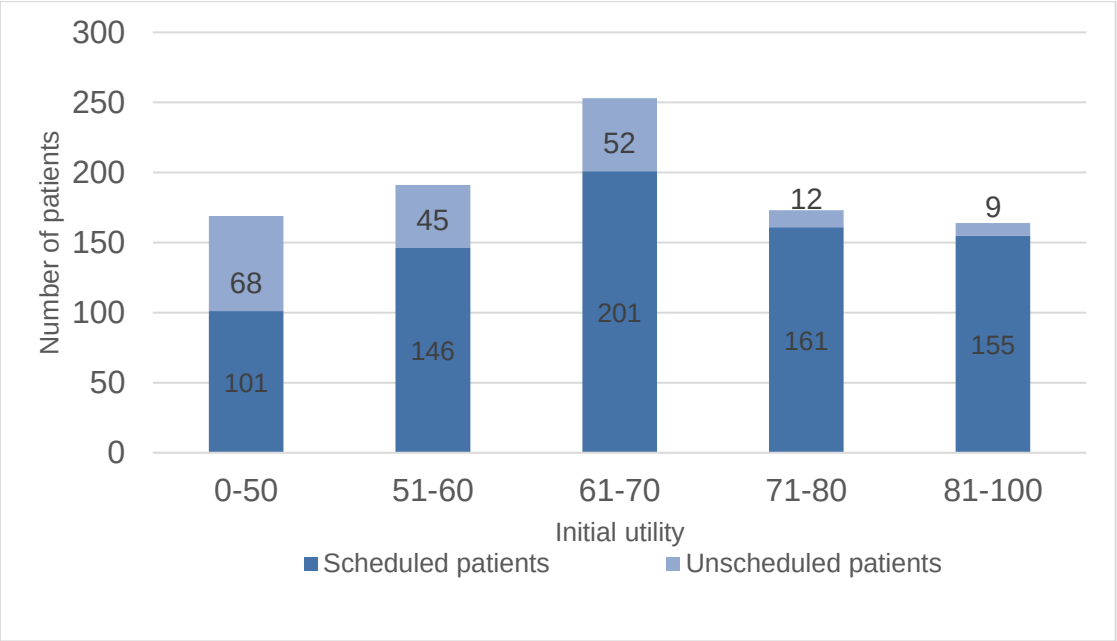

Figure 9: Proportion of scheduled and unscheduled patients for each class of patients (defined by their initial value of utility), for static values of utility on the proposed model.

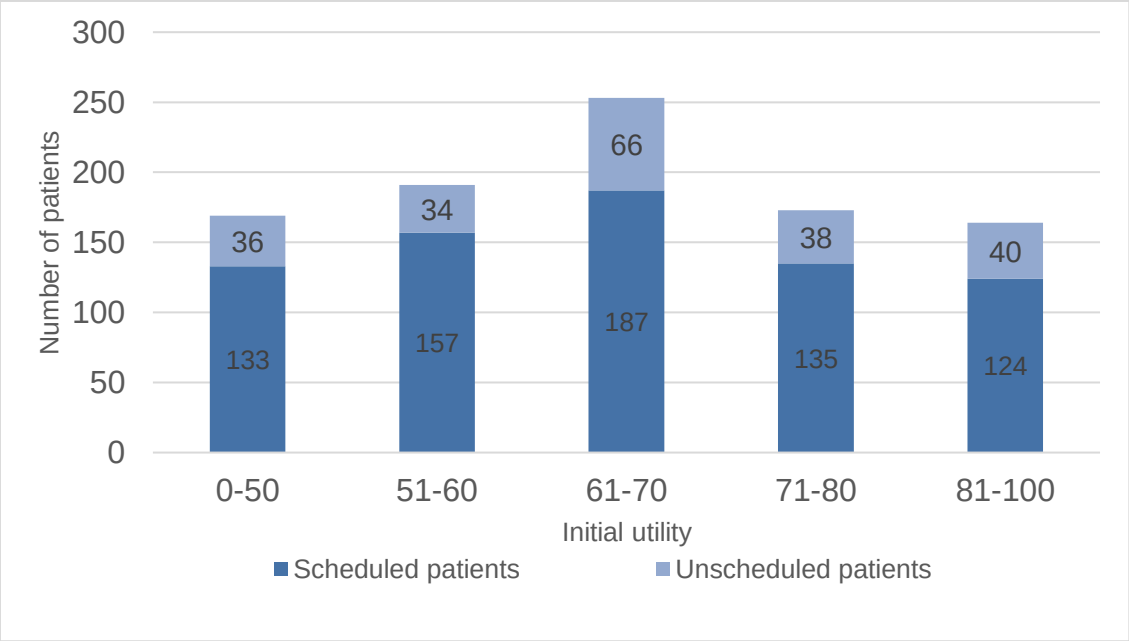

Figure 10: Proportion of scheduled and unscheduled patients for each class of patients (defined by their initial value of utility), with static values of utility, on a FCFS approach.

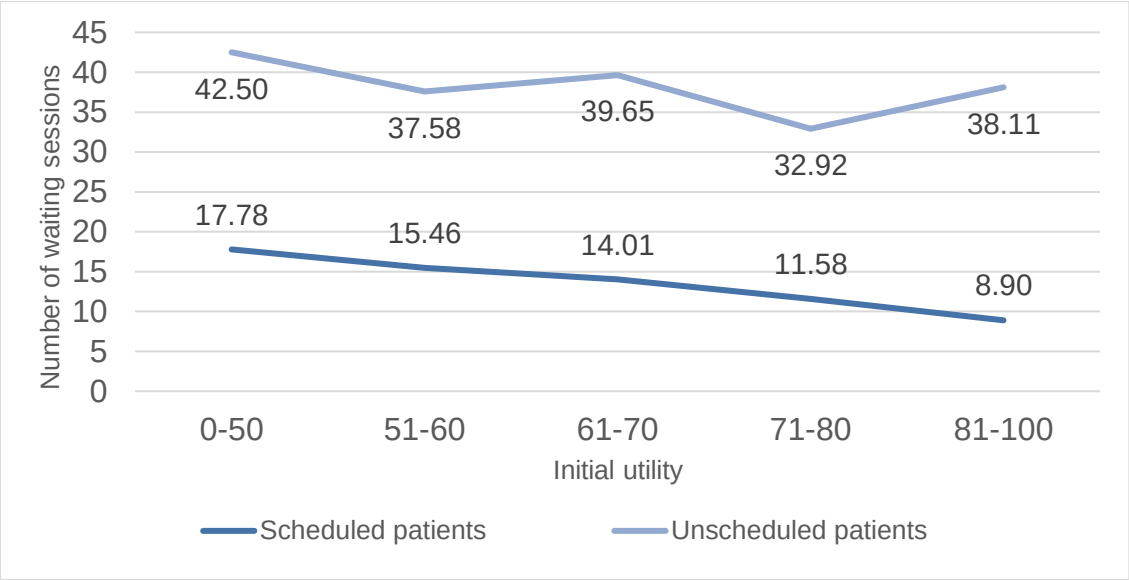

Figure 11: Average number of waiting sessions for scheduled and unscheduled patients of each class of patients defined by their initial values of utility, on the proposed model but with static values of utility.

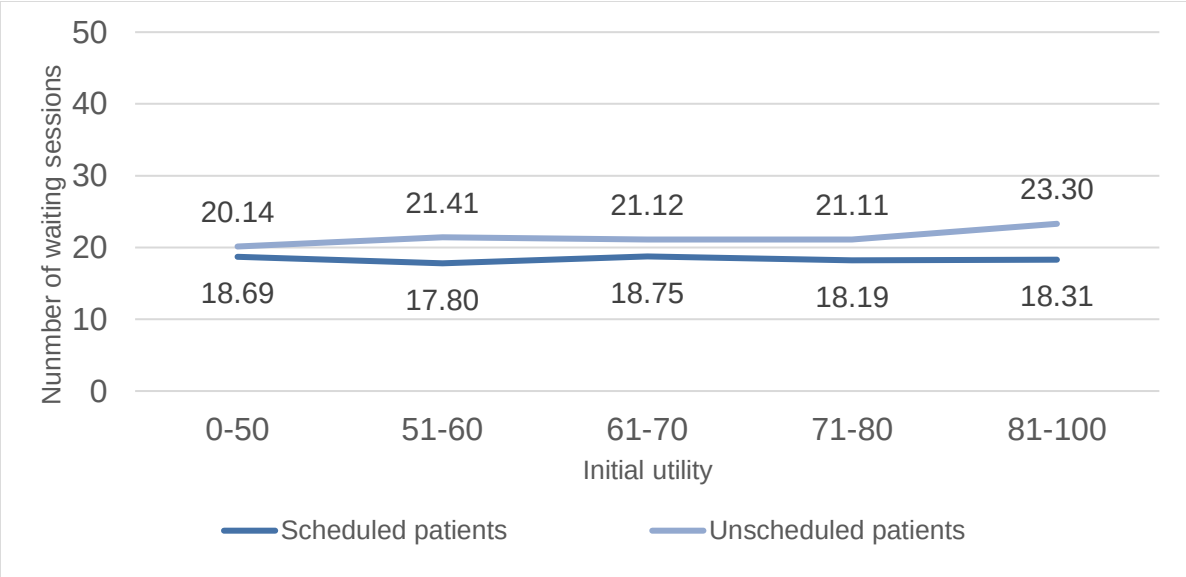

Figure 12: Average number of waiting sessions for scheduled and unscheduled patients of each class of patients defined by their initial values of utility, on a model based on FCFS approach.

With the static model of utility, there has been a marked increase in the proportion of patients scheduled as the utility values are increasing. For the static model of waiting time, the proportion of patients scheduled for the different classes is more or less the same. For the dynamic utility model, the same behaviour is observed as in the static model, in which, for patients with higher utility values, the proportion of patients scheduled is higher; however, there is a greater proportion of patients with less utility scheduled, because this model should also select patients with lower values of utility but higher accumulated waiting times.

Regarding the waiting times, the model that considers the waiting times, presents little difference between the waiting times of the scheduled and not scheduled patients. Furthermore, the values are constant regardless of the utility values, that is, regardless of the actual needs of the patients. For the

model that consider utility, it is observed that waiting times for patients with higher utility values are lower than waiting times for patients with lower utility values. However, the dynamic model selects patients with longer waiting times than the static utility model. In fact, the dynamic model proposed in this thesis presents a behaviour between the one of the model that considers waiting time and the one that considers only utility. Therefore, it can be stated that it constitutes a balanced approach between patients' real needs and waiting times.

6.4. Chapter considerations

In this chapter, computational experiments are performed with the model to validate it and test the impact in changing instances, parameters and objectives. The results obtained can be used to answer the research question. The utility value is intended to represent the real needs of patients considering factors beyond their clinical situation. Integrating this value into the scheduling assigns priority to patients with an effective need for surgery. Thus, a better selection of the patients to be scheduled is made. The waiting time of the patient is also considered through a dynamic approach, which makes the access to the treatments more equitable and prevent patients, who do not have such high utility values but whom are waiting the longest, to be neglected in the process. Since waiting time is also a scheduling factor and patient conditions are updated with rolling horizon approaches, consequences on the health status and well-being of the patients are prevented.

Initially, ten different instances, created based on the generation approach and the data provided by the hospital previously described on Chapter 5, are used to validate the model. Good results, with relative gaps below 3%, are obtained within 1 hour computation time.

Sensitivity analyses are performed to one of the instances, which show that the change in the size of the problem (number of variables and number of restrictions) can impact the gap obtained in these tests, modified through the number of sessions and patients on the waiting list. However, it is shown that complexity of the problem has a greater impact on the results. Increasing or decreasing symmetry of the problem by changing parameters and characteristics of skills and availability of doctors or the duration of waiting list surgeries, induce a considerably greater impact in the relative gap value.

Besides, the results show that the model makes a good selection of patients according to their needs and has some compliance with pre-assigned doctors and overtime. Using only utility or only waiting time in the objective function has a strong impact on the indicator that is not considered. The results regarding the use of overtime support the hospital's decision in not allowing it. It is also analysed the impact of not using the common goal of maximizing the number of scheduled surgeries, in which it is revealed that the use of the utility model does not seriously affect the number of scheduled surgeries.

Finally, an approach is proposed for considering a rolling horizon, which shows that, considering a static approach causes patients with a condition that evolves during waiting time to be neglected.

The utility value update formula should be designed and chosen based on the goals of the scheduling process.

The overall conclusions show that the model can retrieve good results in a reasonable time, but the waiting list should be dynamic in order to attend patients' real needs.

7. Conclusions

This chapter concludes this dissertation. The main results and the limitations on the work are emphasized, and directions for future research on this topic are recommended.

7.1. Concluding remarks

This work is motivated by a project that aims to develop a new and more generalized prioritization formula that includes the patient's medical, social and psychological condition in the surgical scheduling process. The inclusion of a new utility value is intended to better respond to the patient's real needs and the strategic value of his surgeries. This project is designed to answer to needs identified by a general hospital in Canada, where access to elective surgery is increasingly difficult. Therefore, the approach chosen to deal with this problem is to make a more equitable selection of the patients to be scheduled in terms of real surgical needs.

Generally, elective patients are treated according to a FCFS approach on the waiting lists, without considering the evolution of the patients' health status. Besides, research on prioritization procedures is very scarce and is done separately from the scheduling procedures.

This dissertation focuses on the patient selection method integrated in the operating room scheduling. An integer linear programming model is proposed to tackle this problem. To include patients' needs compliance as on objective, the sum of the utility values of scheduled patients is maximized. Since it was not possible to collect real data on waiting lists, an instance generator is created based on the data provided by the hospital and information available in the literature based on real case scenarios. The model is tested and validated with these instances.

The proposed model selects the patients with the highest utility values while considering the whole waiting list, the limitations on the capacity and the rules of the hospital, instead of only considering the waiting time of the patients. This allows that patients in most need for surgery are scheduled first. The model also generates schedules in which a balance between the workload of surgeons is guaranteed and a considerable high number of surgeries should be scheduled in each planning horizon.

The results obtained with the computational experiments allows answering the research question: "What is the impact of considering a utility function representing the real needs of the patients and its evolution in an optimization system for surgery scheduling, on the selection of patients and waiting times?". The computational experiments show that the model returns good results (gaps lower than 3%) in a short period of time (1-hour computation time limit). After validating the model, a rolling horizon approach is developed to directly answer the research question. Thus, regarding the evolution of the patients' condition, a rolling horizon approach is also proposed to mimic the dynamic waiting lists. A formula is suggested to update utility value in order to predict the evolution on the condition of

the patient while waiting and not only the situation of the patient when entered the waiting list. This approach schedules surgeries of patients for whom the condition worsen while waiting, contrarily to the static version of the model in which the patients are scheduled according to the initial assigned utility value. Selected patients have high utility values. Indeed, the results show that this model generates schedules that prioritizes patients with higher utility values but does not neglect patients with lower utility values, if they are waiting for long times. A good balance and compromise are established between the real needs of patients and their waiting times. The results demonstrate that if a static condition of the patient is considered, patients for whom the condition can evolve during the waiting time are prejudiced and may not be considered to schedule, which can have serious consequences on their health and treatment. It is also shown that the choice of the formula to be used to update utility values highly influences the results and so, it should be designed according to the objectives of the scheduling procedure. The computational experiments, in which the utility value is incorporated, allow selecting patients for the scheduling process not only by their waiting time but also for their needs. By considering a dynamic utility function, representing the real (updated) needs of the patients, patients with worsening conditions on the waiting list are selected, thus decreasing their waiting period and preventing the patient from suffering health consequences caused solely by the long waiting times for surgery.

In general, and despite the assigned parameters' values or the instances' characteristics to be used as an input, this model can generate more equitable schedules from the elective patients' point-of-view and considering real and updates needs for surgery.

7.2. Limitations and future research

To conclude this dissertation, some limitations of work are enumerated, which may be used to improve the proposed approach and explore future issues.

The computational experiments show that the implemented mathematical model returns reasonable results in a short period of time (1-hour computation limit is used). Nevertheless, although the model is reliable, it does not consider some aspects of real scenarios in hospitals such as the uncertainty of the duration of the surgeries or cancellations of surgeries. In addition, although generated instances are built according to real scenarios, the model still needs to be validated with real data from the hospital under study. Namely it is of interest to understand the impact on the performance of the model regarding waiting list and dimension of the problem and on the results. Furthermore, the efficiency of model needs to be improved to enable the extension of the planning horizon to 4 weeks, instead of the 2 weeks used in the computational tests, because currently 4 weeks schedules are generated empirically on the hospital. Besides, a more in-depth investigation of the rolling horizon approach is needed to evaluate how changing the update formula can impact the patients' selection. Finally, a user-friendly interface should be created to able the use of this approach by hospital staff.

This approach can be extended to other queuing systems. Indeed, the programming model can be applied in several contexts, according to the parameters set. Appointments, meetings or even queues in customer service for any areas can adapt this model to be used in the scheduling process.

In general, the results of the model are good, and further research on this topic is recommended in order to attend as much as possible to people's real and updated needs and avoid consequences on the condition of the patient during the long waiting times, while scheduling elective surgeries.

References

Abbasgholizadeh Rahimi, S., Jamshidi, A., Ruiz, A., & Ait-kadi, D. (2016). A new dynamic integrated framework for surgical patients' prioritization considering risks and uncertainties. *Decision Support Systems*, *88*, 112–120. https://doi.org/10.1016/j.dss.2016.06.003

Adan, I., Bekkers, J., Dellaert, N., Jeunet, J., & Vissers, J. (2011). Improving operational effectiveness of tactical master plans for emergency and elective patients under stochastic demand and capacitated resources. *European Journal of Operational Research*, *213*(1), 290–308. https://doi.org/10.1016/j.ejor.2011.02.025

Barua, B. (2017). *Waiting Your Turn: Wait Times for Health Care in Canada, 2017*. Retrieved from http://www.deslibris.ca/ID/10094207

Beliën, J., & Demeulemeester, E. (2008). A branch-and-price approach for integrating nurse and surgery scheduling. *European Journal of Operational Research*, *189*(3), 652–668. https://doi.org/10.1016/j.ejor.2006.10.060

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 and Analgesia*, *94*(1), 143–148, table of contents.

Canada's health care system. (2016). Retrieved April 3, 2018, from https://www.canada.ca/en/health-canada/services/canada-health-care-system.html

Cardoen, B., Demeulemeester, E., & Beliën, J. (2009). Optimizing a multiple objective surgical case sequencing problem. *International Journal of Production Economics*, *119*(2), 354–366. https://doi.org/10.1016/j.ijpe.2009.03.009

Cardoen, B., Demeulemeester, E., & Beliën, J. (2010a). Operating room planning and scheduling: A literature review. *European Journal of Operational Research*, *201*(3), 921–932. https://doi.org/10.1016/j.ejor.2009.04.011

Cardoen, B., Demeulemeester, E., & Belien, J. (2010b). Operating room planning and scheduling problems: a classification scheme, 14.

Day, R., Garfinkel, R., & Thompson, S. (2012). Integrated Block Sharing: A Win–Win Strategy for Hospitals and Surgeons. *Manufacturing & Service Operations Management*, *14*(4), 567–583. https://doi.org/10.1287/msom.1110.0372

Denton, B., Viapiano, J., & Vogl, A. (2007). Optimization of surgery sequencing and scheduling decisions under uncertainty. *Health Care Management Science*, *10*(1), 13–24. https://doi.org/10.1007/s10729-006-9005-4

Dexter, F., Macario, A., & Traub, R. D. (1999). Optimal sequencing of urgent surgical cases. Scheduling cases using operating room information systems. *Journal of Clinical Monitoring and Computing*, *15*(3–4), 153–162.

Durán, G., Rey, P. A., & Wolff, P. (2017). Solving the operating room scheduling problem with prioritized lists of patients. *Annals of Operations Research*, *258*(2), 395–414. https://doi.org/10.1007/s10479-016-2172-x

Erdem, E., Qu, X., & Shi, J. (2012). Rescheduling of elective patients upon the arrival of emergency patients. *Decision Support Systems*, *54*(1), 551–563. https://doi.org/10.1016/j.dss.2012.08.002

Fei, H., Meskens, N., & Chu, C. (2006). An operating theatre planning and scheduling problem in the case of a “block scheduling” strategy (pp. 422–428). IEEE. https://doi.org/10.1109/ICSSSM.2006.320500

Ferrand, Y., Magazine, M., & Rao, U. (2010). Comparing two operating-room-allocation policies for elective and emergency surgeries (pp. 2364–2374). IEEE. https://doi.org/10.1109/WSC.2010.5678933

Glover, F., & Marti, R. (2006). Tabu Search. In E. Alba & R. Martí (Eds.), *Metaheuristic Procedures for Training Neutral Networks* (Vol. 36, pp. 53–69). Springer US. https://doi.org/10.1007/0-387-33416-5_3

Gomes, C., Almada-Lobo, B., Borges, J., & Soares, C. (2012). Integrating Data Mining and Optimization Techniques on Surgery Scheduling. In S. Zhou, S. Zhang, & G. Karypis (Eds.), *Advanced Data Mining and Applications* (Vol. 7713, pp. 589–602). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-35527-1_49

Guerriero, F., & Guido, R. (2011). Operational research in the management of the operating theatre: a survey. *Health Care Management Science*, *14*(1), 89–114. https://doi.org/10.1007/s10729-010-9143-6

Gul, S., Denton, B. T., Fowler, J. W., & Huschka, T. (2011). Bi-Criteria Scheduling of Surgical Services for an Outpatient Procedure Center: Bi-Criteria Scheduling of Surgical Services. *Production and Operations Management*, *20*(3), 406–417. https://doi.org/10.1111/j.1937-5956.2011.01232.x

Hadorn, D. C. (2000). Setting priorities for waiting lists: defining our terms. Steering Committee of the Western Canada Waiting List Project. *CMAJ: Canadian Medical Association Journal = Journal de l'Association Medicale Canadienne*, *163*(7), 857–860.

Hall, R. (Ed.). (2013). *Patient Flow* (Vol. 206). Boston, MA: Springer US. https://doi.org/10.1007/978-1-4614-9512-3

Hans, E. W., van Houdenhoven, M., & Hulshof, P. J. H. (2012). A Framework for Healthcare Planning and Control. In R. Hall (Ed.), *Handbook of Healthcare System Scheduling* (Vol. 168, pp. 303–320). Boston, MA: Springer US. https://doi.org/10.1007/978-1-4614-1734-7_12

Hans, E., Wullink, G., van Houdenhoven, M., & Kazemier, G. (2008). Robust surgery loading. *European Journal of Operational Research*, *185*(3), 1038–1050. https://doi.org/10.1016/j.ejor.2006.08.022

Healthcare Financial Management Association. (2003). Achieving operating room efficiency through process integration. *Healthcare Financial Management: Journal of the Healthcare Financial Management Association*, *57*(3), suppl 1-7 following 112.

Heng, M., & Wright, J. (2013). Dedicated operating room for emergency surgery improves access and efficiency. *Canadian Journal of Surgery*, *56*(3), 167–174. https://doi.org/10.1503/cjs.019711

IBM Knowledge Center. (n.d.). Retrieved May 22, 2018, from https://www.ibm.com/support/knowledgecenter/SSSA5P_12.6.2/ilog.odms.cplex.help/CPLEX/UsrMan/topics/discr_optim/eg_col_gen/02_col_gen_defn.html

Kharraja, S., Albert, P., & Chaabane, S. (2006). Block Scheduling: Toward a Master Surgical Schedule (pp. 429–435). IEEE. https://doi.org/10.1109/ICSSSM.2006.320501

Magerlein, J. M., & Martin, J. B. (1978). Surgical demand scheduling: a review. *Health Services Research*, *13*(4), 418–433.

Mallawaarachchi, V. (2017). Introduction to Genetic Algorithms—Including Example Code. Retrieved May 22, 2018, from https://towardsdatascience.com/introduction-to-genetic-algorithms-including-example-code-e396e98d8bf3

Marques, I., & Captivo, M. E. (2017). Different stakeholders' perspectives for a surgical case assignment problem: Deterministic and robust approaches. *European Journal of Operational Research*, *261*(1), 260–278. https://doi.org/10.1016/j.ejor.2017.01.036

Marques, I., Captivo, M. E., & Vaz Pato, M. (2012). An integer programming approach to elective surgery scheduling: Analysis and comparison based on a real case. *OR Spectrum*, *34*(2), 407–427. https://doi.org/10.1007/s00291-011-0279-7

Marques, I., Captivo, M. E., & Vaz Pato, M. (2015). A bicriteria heuristic for an elective surgery scheduling problem. *Health Care Management Science*, *18*(3), 251–266. https://doi.org/10.1007/s10729-014-9305-z

Min, D., & Yih, Y. (2014). Managing a patient waiting list with time-dependent priority and adverse events. *RAIRO - Operations Research*, *48*(1), 53–74. https://doi.org/10.1051/ro/2013047

Persson, M., & Persson, J. A. (n.d.). Optimization modelling of hospital operating room planning: analyzing strategies and problem settings, 11.

Samudra, M., Van Riet, C., Demeulemeester, E., Cardoen, B., Vansteenkiste, N., & Rademakers, F. E. (2016). Scheduling operating rooms: achievements, challenges and pitfalls. *Journal of Scheduling*, *19*(5), 493–525. https://doi.org/10.1007/s10951-016-0489-6

Stuart, K., & Kozan, E. (2012). Reactive scheduling model for the operating theatre. *Flexible Services and Manufacturing Journal*, *24*(4), 400–421. https://doi.org/10.1007/s10696-011-9111-6

The Federal Government Role in Reducing Health Care Wait Times. (2014). Retrieved June 2, 2018, from http://www.waittimealliance.ca/wta-reports/federal-role-health-care-wait-times-in-canada

Urbach, D. (2018). How to shorten hospital wait times in Canada. Retrieved from https://www.thestar.com/opinion/contributors/2018/05/22/how-to-shorten-hospital-wait-times-in-canada.html

van Essen, J. T., Bosch, J. M., Hans, E. W., van Houdenhoven, M., & Hurink, J. L. (2013). Reducing the number of required beds by rearranging the OR-schedule. *OR Spectrum*. https://doi.org/10.1007/s00291-013-0323-x

Van Houdenhoven, M., Hans, E. W., Klein, J., Wullink, G., & Kazemier, G. (2007). A Norm Utilisation for Scarce Hospital Resources: Evidence from Operating Rooms in a Dutch University Hospital. *Journal of Medical Systems*, *31*(4), 231–236. https://doi.org/10.1007/s10916-007-9060-5

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*(1), 91–104. https://doi.org/10.1007/s10479-009-0619-z

Van Riet, C., & Demeulemeester, E. (2015). Trade-offs in operating room planning for electives and emergencies: A review. *Operations Research for Health Care*, *7*, 52–69. https://doi.org/10.1016/j.orhc.2015.05.005

Vanberkel, P. T., Boucherie, R. J., Hans, E. W., Hurink, J. L., van Lent, W. A. M., & van Harten, W. H. (2011). An exact approach for relating recovering surgical patient workload to the master surgical schedule. *Journal of the Operational Research Society*, *62*(10), 1851–1860. https://doi.org/10.1057/jors.2010.141

Wullink, G., Van Houdenhoven, M., Hans, E. W., van Oostrum, J. M., van der Lans, M., & Kazemier, G. (2007). Closing Emergency Operating Rooms Improves Efficiency. *Journal of Medical Systems*, *31*(6), 543–546. https://doi.org/10.1007/s10916-007-9096-6