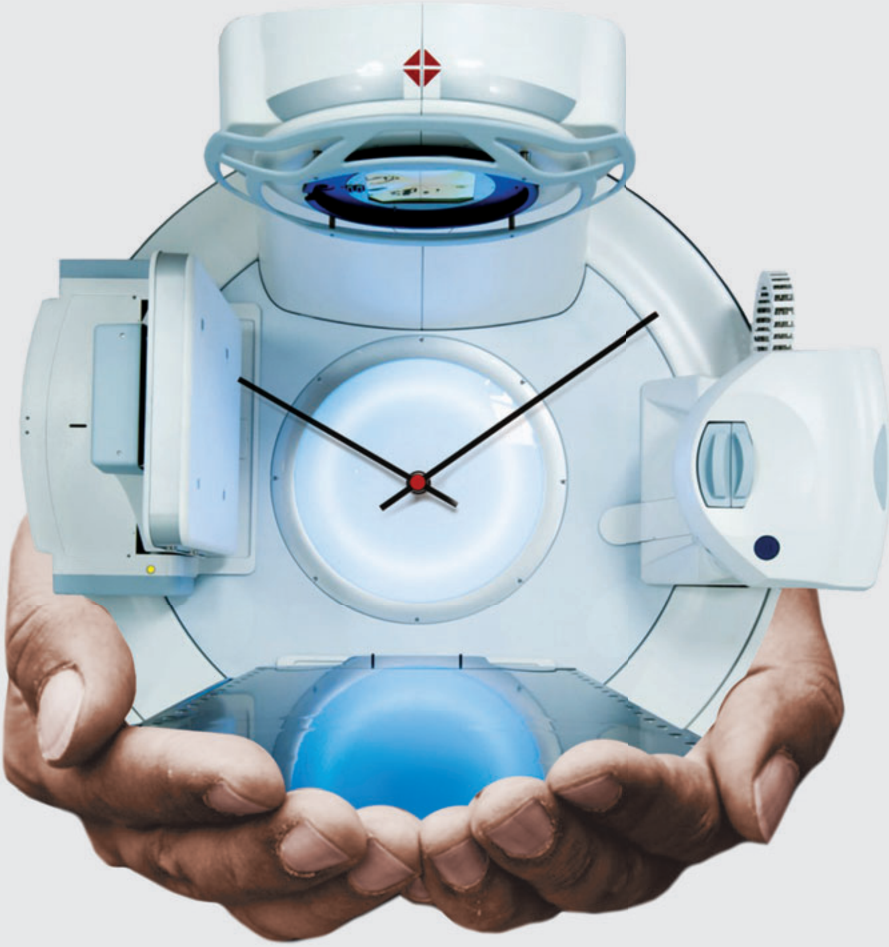
Logical Optimization Of Radiotherapy Treatments



Bruno Vieira

LOGISTICAL OPTIMIZATION OF RADIOTHERAPY TREATMENTS

Bruno Vieira

Graduation committee

Chairman / secretary:	prof. dr. T.A.J. Toonen *University of Twente, Enschede, the Netherlands*
Supervisors:	prof. dr. W.H. van Harten *University of Twente, Enschede, the Netherlands*
	prof. dr. ir. E.W. Hans *University of Twente, Enschede, the Netherlands*
Co-supervisor:	dr. J.B. van de Kamer *Netherlands Cancer Institute, Amsterdam, the Netherlands*
Committee members:	prof. dr. R.J. Boucherie *University of Twente, Enschede, the Netherlands*
	prof. dr. N.J.J. Verdonschot *University of Twente, Enschede, the Netherlands*
	dr. A. Viana *Polytechnic Insitute of Porto, Porto, Portugal*
	prof. dr. C.R.N. Rasch *Leiden University, Leiden, the Netherlands*
	prof. dr. C.A.M. Marijnen *Leiden University, Leiden, the Netherlands*

Ph.D. thesis, University of Twente, Enschede, the Netherlands
Department of Health Technology and Services Research
Center for Healthcare Operations Improvement and Research

This thesis is part of the Health Science Series, HSS 20-32, department Health Technology and Services Research, University of Twente, Enschede, the Netherlands. ISSN: 1878-4968.

Printed by Ridderprint BV, www.ridderprint.nl

ISBN 978-90-365-4997-4
DOI 10.3990/1.9789036549974

LOGISTICAL OPTIMIZATION OF RADIOTHERAPY TREATMENTS

DISSERTATION

to obtain
the degree of doctor at the University of Twente,
on the authority of the rector magnificus,
Prof. dr. T.T.M. Palstra,
on account of the decision of the Doctorate Board
to be publicly defended
on Friday 26th June 2020 at 14h45

by

Bruno Miguel Soares Vieira

born on the 18th November 1987
in Porto, Portugal

This dissertation has been approved by:

Supervisors:
prof. dr. W.H. van Harten
prof. dr. ir. E.W. Hans

Co-supervisor:
dr. J.B. van de Kamer

Contents

CHAPTER 1

Introduction

1.1 Research motivation and scope

Radiotherapy (RT) is a treatment modality in cancer care that uses ionizing radiation to kill tumor cells. In external-beam RT, a machine called "linear accelerator" (linac) is used to deliver high-energy radiation beams onto the tumor area while minimizing the exposure of surrounding healthy tissue. To achieve this, a number of preparation steps (e.g. computer tomography (CT) scan, tumor contouring, treatment planning) need to be performed as a “pre-treatment" phase before the treatment is delivered in a series of (usually daily) irradiation sessions. Several health professionals are involved in the delivery of RT treatments. Imaging scans, treatment planning, and irradiation sessions are conducted by radiation therapy technologists (RTTs), while first consultation, tumor contouring, and follow-up appointments are performed by radiation oncologists (doctors).

With the increase in cancer rates [14] and given that around half of all cancer patients receive radiotherapy as part of their treatment [28], demand for RT services has been continuously growing [94]. This makes the efficient planning and control of resources and patients’ care pathways especially important in order to ensure timely and quality treatments. In RT, delays in the start of treatment can negatively affect the patient’s outcome and quality of life [64]. Not only have delays been associated with an increased risk of local recurrence and tumor progression [20], it has also been shown that patients experience higher levels of psychological distress and prolonged symptoms when they are subject to longer waiting times [63]. Therefore, achieving timely RT is essential in oncologic care. In the Netherlands, timeliness standards are defined by the Dutch Society for Radiation Oncology (NVRO), which recommends a maximum waiting time between referral and start of treatment of 10 calendar days for subacute (semi-urgent) patients, and 28 days for regular patients [68]. Acute (urgent) patients should receive treatment on the referral day. However, several complexities make an efficient organization of resources in RT especially difficult. Due to factors such as fluctuations in patient inflow, highly specialized treatment pathways, variability in the processing times required by certain activities, and partial availability of highly specialized personnel, bringing supply into line with demand has become increasingly difficult for RT centers. With the increase in the dimension and complexity of RT logistical problems, research questions arise: Would an optimal allocation of radiation therapy technologists to the several tasks they perform increase the number of patients starting their treatment within the desired targets? What is the impact of scheduling the start of treatment right after the first consultation instead of at the end of treatment planning on waiting times? Can all patient preferences regarding appointment times be satisfied in the linacs' weekly schedules? These are types of questions that we intend to answer using the most advanced analytical methods for the benefit of cancer patients and health professionals.

Operations Research (OR) is a discipline that focuses on the efficient design, control, and optimization of processes using techniques such as mathematical programming, computer simulation, queuing theory and (meta)heuristics [84]. Integer linear programming methods have proven effective for optimally solving problems of combinatorial nature with several resource constraints and multiple objectives. Metaheuristics are suitable for complex processes and/or large instances where exact (optimal) solutions are hard to achieve in reasonable computation time. Computer simulation and queuing theory are useful for modeling stochastic processes where the uncertainty inherent to certain inputs (e.g. patient arrivals and care content) is high. Several OR-based tools have been successfully developed for decision-making support in the context of healthcare logistics [83]. Resource capacity planning in healthcare processes is especially complex due to the high number of constraints that usually need to be satisfied, the inherent stochasticity in patient inflow, care trajectories, and activities' processing times. In RT, several additional complexities such as time constraints between the pre-treatment and treatment stages, or the combination of RT and other treatment modalities (chemotherapy and/or surgery) arise. This makes the application of OR methods for planning and control purposes even more important and challenging.

In this thesis, OR methods are used to research how RT operations, personnel and equipment can be efficiently coordinated considering the variability in the amount and type of demanded care. The main goal is to prevent delays in the start of treatment while optimizing other KPIs related to timeliness, patient-centeredness, and quality of labor. We demarcate our scope on the tactical and operational levels (see [40]) of the RT organizational structure, and propose innovative OR-based methods for optimized decision-making in problems such as staff allocation, patient scheduling, capacity dimensioning and other resource planning problems commonly faced by RT centers. Moreover, by working together with managers and clinicians of collaborating RT centers in the Netherlands, we have designed and validated models and corresponding solutions in order to facilitate the translation of the obtained knowledge into clinical practice.

This research work focuses on external-beam photon RT using linacs as the treatment device, and thus any kind of internal RT (e.g. brachytherapy) or other types of external-beam RT (e.g. orthovolt) treatments are not dealt with. Also, we only consider care pathways of patients that have already been referred to receive radiotherapy, i.e. any diagnosis/treatment activities held before the start of the radiotherapy chain of activities are excluded from our scope. Furthermore, while there are interesting medical problems in RT (such as the optimization of the radiation angles and intensity for treatment planning) that can highly benefit from OR knowledge, we focus on the logistical aspects of the RT process only.

1.2 Operations research and radiotherapy logistics

In the last two decades, several OR methods have been proposed to solve logistics problems in RT. In Chapter 2 we present an in-depth overview of the state of the art in the development and application of OR models for RT logistics with a thorough categorization in terms of, e.g., hierarchical level and extent of implementation. Among OR methods, literature shows that computer simulation is the most popular method for solving problems at the strategic and tactical levels where patient flow analysis and capacity allocation problems are the subject of research. Werker *et al*. [110], Crop *et al*. [26] and Joustra *et al*. [52] used discrete-event simulation (DES) to model the pre-treatment phase of the RT process and test "what-if" scenarios that could potentially reduce patients' waiting times. While Werker *et al*. [110] showed that shorter oncologist-related delays decrease the overall planning time by more than a day, Joustra *et al*. [52] found that the outpatient department was the main bottleneck hampering RT patient flows in the Amsterdam Medical Center, the Netherlands. They then developed a queuing model to reduce fluctuations in the capacity of the outpatient department to increase the number of patients complying with the waiting time targets from 39% to 92%. Crop *et al*. [26] were able to increase the control over the work-in-progress by proposing a method that only allows a new patient to enter the system after another patient leaves. Their strategy is evaluated using DES, which showed that a 32% increase in the total number of treatments was attainable while relieving workload-related stress for personnel. Monte Carlo simulation modeling was used by Thomas *et al*. [97] to estimate the number of linacs needed to cover the expected demand in RT centers and calculate the percentage of spare capacity required to keep waiting times low. Amongst other insights, they found that about a 10% of spare capacity is needed to ensure that 86% of patients start radiotherapy up to one week after the treatment planning is finished.

Mathematical programming techniques are the most used techniques in the literature to address logistical problems at the operational level, with an emphasis on treatment scheduling problems. Conforti *et al*. [22–24], Leite-Castro *et al*. [18] and Burke *et al*. [15] developed mixed-integer linear programming (MILP) models to find optimal weekly schedules for irradiation sessions amongst the available linacs for a known pool of patients. Their models can find (near-

)optimal solutions in reasonable computation time for real-world size instances considering several medical and technological constraints. Legrain *et al*. [59] proposed a two-step mathematical model to optimize the scheduling of RT treatments on a patient-by-patient basis. By considering scenarios of future patient arrivals to predict upcoming workload, their method showed an average decrease of patients breaching the waiting time targets by 50% for acute patients, and 81% for subacute patients when comparing to clinical practice of a Canadian RT center. A combination of mathematical programming and computer simulation is proposed by Bikker *et al*. [5] to optimally align the doctors' agendas with demand regarding contouring and consultation activities. They used DES to evaluate the solution obtained by the MILP model and found that waiting times before the start of treatment could potentially be reduced by 15% for regular patients.

Metaheuristics and constructive heuristics are mainly proposed in the literature for optimizing the pre-treatment phase, and for larger instances of the RT treatment scheduling problem where MILP models become intractable in terms of computation time. Genetic algorithms are used by Petrovic *et al*. [73, 74, 77] to optimize pre-treatment patient flows by efficiently scheduling treatment planning and physics-related operations. Their methods showed a potential decrease of the average waiting times for radical (from 35.0 to 21.5 days) and palliative (from 15.0 to 13.1 days) patients. They have also found that enabling doctors to approve treatment plans right after they have been finished has a significant impact on the average waiting times in all patient categories (35% reduction). Constructive heuristics are high-level procedures that allow to iteratively build solutions based on the empirical knowledge of the system being optimized. Petrovic *et al*. [75, 76, 78] proposed four different constructive approaches for scheduling treatment sessions at the Nottingham University Hospitals Trust (UK). A decrease in the percentage of late patients of up to 40% for palliative patients and 4% for radical patients was achieved by using a just-in-time algorithm that assigns the latest feasible start date of the first treatment to each patient from a prioritized list.

Most of the studies presented in the literature focus on the scheduling of RT treatment sessions. Fewer studies optimize the pre-treatment workflow, despite the potential benefits that can be achieved in reducing waiting times before treatment. Although staff members often perform several activities during a workday, only one study has been found to optimize the allocation of staff (doctors) to the several operations they perform. No studies have been found for the allocation of RTTs to multiple operations. As for the scheduling of treatment sessions on linacs, there is a vast literature of many well-performing models that have been developed considering a variety of operational constraints and objectives. Nevertheless, no studies showed to e.g. integrate patient preferences into the scheduling routines or minimize empty time periods between sessions in the linacs' schedules. We address these problems in this thesis using OR-based methods for the efficient management of RT resources and patients. Since the rate of implementation of OR methods in healthcare settings is rather low when compared with other industries [11, 100], we work in close collaboration with managers and clinicians of real-world RT departments. To avoid that optimization tools are developed for a specific case, we have set-up a collaboration with six Dutch RT centers and performed a survey on their logistics in order to bring the validation and implementation of the developed tools one step closer to clinical practice, as we describe next.

1.3 A survey on the logistics of 6 radiotherapy centers

Several OR models have been proposed for the efficient management of RT processes such that personnel and equipment can be efficiently coordinated with the demanded care to ensure patient-centeredness and timely treatments. However, the generalization of these models is not always straightforward, as RT centers may have different strategic positions which impact the way they organize their processes to optimize logistics according to their goals. Therefore, studies comparing the logistics environment amongst diversified RT centers are crucial for the development of innovative planning tools for sustained, analytical-based decision-making that can be representative of current clinical practice. To this end, we have conducted a survey on several logistical aspects of the RT process amongst six collaborating RT centers in the Netherlands. Information has been gathered by means of a series of meetings and correspondence exchange held with managers and clinicians of the collaborating centers over three years (2015-2017). The survey covers subjects such as the size and resources of each center, patient flow, workflow analysis, and the study of the most important key performance indices (KPIs) to be optimized in clinical practice.

1.3.1 The collaborating RT centers

The following six RT centers have collaborated with our survey:

- The Netherlands Cancer Institute (NKI)
- Bernard Verbeeten Instituut (BVI)
- Amsterdam Universitair Medische Centra (AUMC)
- Radiotherapeutisch Intituut Friesland (RIF)
- Radiotherapiegroep - Arnhem (ARTI)
- Radiotherapiegroep - Deventer (RISO)

Table 1.1 presents the size and number of existing resources in each of the RT centers involved in our study. The RT department of the NKI is the largest

Table 1.1 Comparison of the number of new treatments, personnel and equipment between the different RT centers by 2017.

Description	NKI	BVI	AUMC	RIF	ARTI	RISO
Annual no. new treatments	~5100	~4300	~2500	~2500	~2850	~2200
No. satellite locations	1	2	1	0	1	0
No. CT scanners	2	2	2	1	1	1
No. MRI scanners	1	0	1	0	0	0
No. PET-CT scanners	1	1	1	0	0	0
No. linear accelerators	9	8	6	4	6	4
No. radiation oncologists	26	11	12	10	11	11
(FTE)	(21)	(n.a.)	(12)	(8)	(10)	(10)
No. radiation therapy technologists	115	65	70	65	62	44
(FTE)	(100)	(n.a.)	(50)	(43)	(50)	(36)

center, treating over 5000 new patients per year using nine linacs. The NKI provides high-quality treatments with patient-specific care pathways. The RT department of the AUMC hospital treats around 2500 patients using six linacs. By having easy access to other resources (e.g. MRI, PET-CT) and treatment modalities (e.g. surgery) available within the hospital, they have been able to provide good-quality, personalized treatments while keeping waiting times within the required targets, similarly to the NKI. The BVI, RIF, ARTI and RISO are independent groups dedicated to providing RT services in the provinces of North Brabant, Friesland, Arnhem, and Deventer, respectively. With mainly standardized treatment pathways, they aim for timely, patient-friendly treatments while ensuring quality of labor by, e.g., avoiding overtime work. As we can see in Table 1.1, all centers have at least one CT scanner, while only the NKI and the AUMC are provided with an MRI scanner. When non-existent resources (e.g. MRI and PET-CT) are needed by some centers, the required capacity is contracted through agreement with nearby hospitals. Moreover, four RT centers have at least one satellite location, i.e. a secondary physical location where part of the linacs (and corresponding staff) operate.

The involvement of RT centers with varied sizes and value propositions have allowed us to design, develop, and validate tools such that the obtained solutions are robust enough to accommodate logistical differences between centers. We describe these in more detail in the next section.

1.3.2 The radiotherapy process

The RT treatment chain is characterized by a sequence of operations, which depends on the characteristics of the tumor (such as tumor site, level of advancement, etc.), urgency level, amongst other factors. Figure 1.1 depicts a possible deployment flowchart of the RT process and specific operations involved which captures the variations encountered amongst the workflows of the surveyed RT centers.

Patients are normally referred to receive radiotherapy upon indication by

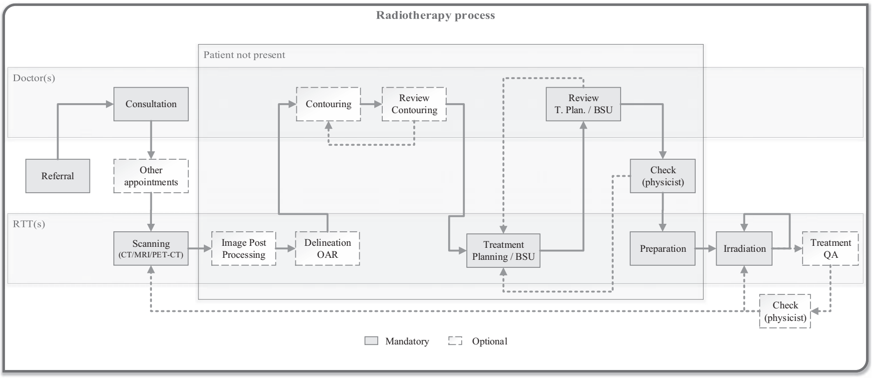


Figure 1.1 Overview of the RT process considering the logistical differences between centers.

a group of specialists during multi-disciplinary meetings undertaken together with referring hospitals. After referral, patients are scheduled for a consultation with a radiation oncologist. Based on the available diagnostic and patient information, the doctor defines a preliminary treatment protocol that includes the dose and the fractionation scheme, and decides upon the urgency level (acute, subacute, or regular) and specific care pathway intended for the patient. This includes, for instance, the necessary pre-treatment imaging scans and the treatment planning modality. The scanning phase always includes a CT scan, but it may also include an MRI scan and a PET-CT scan. Before the CT scan, several additional appointments may be needed. These may include molding, blood analysis, dentist, dietician, etc. In case of multiple imaging modalities, image post processing (IPP) is necessary for matching and/or optimizing the scanned images. Afterwards, delineation of the tumor and the organs-at-risk (OAR) takes place. While in some centers (AUMC, BVI, ARTI, RISO) the delineation of OAR, undertaken by an RTT, takes place before contouring, in other centers (NKI and RIF) it is done in conjunction with treatment planning, after the doctor has contoured the tumor volume. The contouring is then either peer-reviewed or discussed with a group of physicians, which may indicate the need for changes for improvement. Once contouring is approved, treatment planning follows. In this step, the angles and intensity of the irradiation beams are optimized to deliver the desired dose in the target area while minimizing the exposure of surrounding healthy tissue. Each treatment plan needs to be revised and approved by the doctor(s) and a medical physicist. After approval, the plan is uploaded to the linear accelerator scheduled for the patient before the first irradiation session, as a preparation stage. A pre-defined number of irradiation sessions follows, among which a cone-beam CT may be required for position verification purposes. If there are significant anatomical changes or dosimetry deviations, a medical physics expert in consultation with a radiation oncologist may indicate that the process needs to be re-started from the scanning stage with a new CT. After the last irradiation session has been delivered, the treatment is finished and a follow-up period starts.

As for the scheduling moments, it has been verified that acute and subacute patients are scheduled their start of treatment date right at referral/consultation in all centers, while the (first) treatment sessions of regular patients may be either booked right after consultation (AUMC, ARTI, RIF) or only when treatment planning has been completed (NKI, BVI, RISO). It is common practice to inform the patients about the scheduled irradiation sessions on a weekly basis.

1.3.3 Key Performance Indicators (KPIs)

Performance indicators make it possible to evaluate the logistical performance of a center with respect to its specific goals. To ensure that the objectives to be optimized by the analytic models proposed in this thesis are in line with the objectives of most RT centers, we have surveyed the key performance indicators (KPIs) that are consistently monitored and optimized by RT managers and planners. To this end, we used a restricted short list of KPIs selected from an international benchmarking of cancer centers involved in research and training undertaken by van Lent *et al*. [99]. Such KPIs include waiting times, utilization levels, or patient preference satisfaction levels. In our study, we have assessed which of the selected KPIs are considered by the managers and planners of the collaborating centers (Table 1.2). As we can observe, timeliness is a major concern, with all RT centers attempting to minimize waiting times before treatment. The fulfillment of the national waiting time standards is constantly monitored by all centers. Moreover, machines' idle times and overtime labor have been identified as the main objectives to be minimized in at least 3 centers. From a patients' perspective, two centers have indicated that patients preferences regarding appointment times are concurrently considered during scheduling (NKI and RIF), and two centers (BVI and RIF) mentioned having the goal of keeping the time patients wait in the waiting room as short as possible.

Table 1.2 Main Key Performance Indicators (KPIs) monitored and optimized by the collaborating RT centers by 2017.

KPI	Description	NKI	BVI	AUMC	RIF	ARTI	RISO
Waiting times	Time between referral (or consultation) and first fraction	x	x	x	x	x	x
Access times	Time between specific operations (e.g. consultation – CT scan)	x	x		x	x	x
Wait-in-room times	Amount of time a patient waits in the waiting room		x		x		
Patient preferences	Satisfaction of patients' preference requests regarding appointment times	x			x		
Utilization levels	Ratio between the total time a machine is utilized and the total available time	x		x		x	x
Overtime	Amount of time that doctors and RTTs work overtime		x	x			x

1.3.4 Conclusions

Working in collaboration with managers and planners of the six RT centers has allowed to understand that, although some logistical differences amongst RT centers exist (e.g. the moment of delineating OAR, the peer-review process of contouring / treatment planning), the sequence and resources involved in the main steps of the RT process follows a similar structure. Consultation, imaging, treatment planning, and treatment execution are standard sequential steps in all surveyed centers. Radiation oncologists are responsible for the first consultation, contouring, and reviewing of the treatment planning activities, while RTTs guide the patients through imaging scans and treatment sessions, and perform basically all treatment planning activities. As for the objectives to be optimized by the collaborating centers, we found that timeliness is the main objective. All centers work towards timely delivery of treatments according to national waiting time standards. Patient-centeredness (satisfy patient preferences and minimize wait-in-room times) and other indicators of efficiency (minimize overtime and machine idle times) are also relevant KPIs to be considered.

The efficient coordination between patients, machines and staff is crucial to be able to provide quality RT treatments that meet the necessary medical and technical requirements. In this regard, information about the way RT centers plan and organize the delivery of RT treatments is necessary such that the proposed solutions can be generally applicable in RT, thus increasing the chances for implementation. This survey allowed us to assess the most relevant logistics problems in RT, which originated research questions for Chapters 3, 4, and 5. The involvement of RT centers allowed for the adjustment and fine tuning of solutions to their specific needs or desires. Thus, the OR models developed within this thesis aim to represent and optimize RT processes considering the logistical characteristics and goals found within the surveyed centers. However, the logistical problems and case studies in this thesis are largely inspired by the NKI.

1.4 Thesis outline

Chapter 2 presents a literature review on the development and application of OR methods for logistics optimization in RT, at various managerial levels. By means of a literature search performed in six databases covering several disciplines, we categorize studies in terms of the subject of research, the OR methods used, the extent of implementation according to a six-stage model and the (potential) impact of the results in practice.

Chapter 3 proposes a stochastic MILP model that optimizes the allocation of RTTs to multiple operations in RT over a set of scenarios of patient inflow. Multiple scenarios are generated from historical patient data, and the final RTT allocation covers the workload associated with all scenarios. The goal is to maximize the (expected) number of patients completing pre-treatment within their maximum waiting time targets according to national standards.

Chapter 4 uses discrete-event simulation to model the patient flow of the radiotherapy department of the NKI. A staff survey, interviews with managers, and historical data are used to generate model inputs, in which fluctuations in patient inflow and resource availability are considered. The objective of the study is to assess the impact of using pull and push strategies on workflow control and explore alternative interventions for improving timeliness in radiotherapy.

Chapter 5 proposes a MILP model for scheduling and sequencing treatment sessions on the available linacs. The objective of the model is to maximize the number of sessions scheduled within the time window preferences given by patients for a one-week planning horizon. To use the model for larger (three or more linacs) centers in acceptable computational time, we propose a heuristic method that pre-assigns patients to linacs to decompose the problem in subproblems (clusters of linacs) before using the MILP model to solve the problem in a sequential manner.

Chapter 6 we use the MILP model proposed in **Chapter 5** to generate schedules for the RT treatment scheduling in two Dutch RT centers in view of the practical implementation of the model. In this study, the theoretical model is iteratively adjusted to fulfill the specific technical and medical constraints of each center until a valid model was attained. Real patient data was collected for the planning horizon of one week, and the feasibility of the obtained (final) schedules was verified for applicability by the staff of each center. The optimized solutions are compared with the ones manually developed in practice.

CHAPTER 2

Operations research for resource planning and -use in radiotherapy: a literature review

2.1 Background

Due to the growing numbers of cancer patients, demand for RT has been continuously increasing [94]. According to Delaney *et al*. [27, 28], the optimal rate for the use of RT in some part of the treatment in cancer care should be around 50%, although this figure has not yet been achieved in practice [91]. In addition, RT has proven to be at least as cost-effective as both chemotherapy and surgery when all costs across the life cycle of patients are considered [91], making it more likely that demand for RT will keep growing over the coming years. In RT, timeliness is crucial and literature shows that delays in the start of treatment increase the risk of local recurrence and tumor progression [20]. In both breast cancer [47] and radical cervix cancer [29], longer radiotherapy waiting times were found to be associated with diminished survival outcomes, and previous research has shown that delay in initiation of radiotherapy may be associated with a clinically important deterioration in local control rates [63]. Besides, unavailability of medical staff was pointed out as one of the main causes for patient dissatisfaction regarding pain management [79]. In RT resources are expensive and limited in capacity, and treatments are prepared and delivered by a multidisciplinary group of specialists with multiple functions and restricted time availability [111]. In addition to variable patient inflows, medical and technological progress makes treatments more and more specialized. Therefore, resource planning and control in RT are complex and time-consuming activities. In this context, advanced analytical models from fields such as systems engineering or applied mathematics have been proposed to help managers of RT centers make better decisions. A recent report published by the Institute of Medicine claims that using systems engineering, timeliness and patient-centeredness in healthcare delivery can be significantly increased [57]. This paper reviews the extent to which operations research techniques have been used to support decision-making in RT, evaluates their (potential) added value and draws lines for future research.

2.1.1 Operations research and healthcare

Operations research (OR)[1] is a discipline that combines knowledge from fields such as applied mathematics, computer science, and systems engineering. It encompasses a wide range of techniques for improved decision-making, commonly for real-world problems [84]. Originally, OR emerged as a way to improve military material production during the second world war but methods have continuously grown to model and solve problems in business and industry since then. During the last decades, a wide range of problems have been addressed to support strategic decision making, facilitate day-to-day hospital management, and solve medical problems related to the healthcare practice [83]. Among the existing OR applications for hospital management and logistics optimization, well-known problems include appointment scheduling [39], staff rostering [34] and operating room planning and scheduling [16]. Given the growing acceptance of OR models to solve problems in healthcare, research on modeling emerging problems receives increased attention, and both a taxonomy for resource capacity planning and control decisions in healthcare and algorithms to solve the most relevant ones have been proposed [44].

2.1.2 The radiotherapy treatment chain of operations

The RT treatment chain is characterized by a sequence of operations, which depends on the characteristics of the tumor (such as location, level of advancement, etc.). Figure 2.1 depicts a deployment flowchart of the operations involved in external-beam RT. After referral, patients have a consultation with a radiation oncologist, who prescribes one or more diagnostic examinations, such as a computer tomography (CT) scan, a magnetic resonance imaging (MRI) exam, or a positron emission tomography-computer tomography (PET-CT) scan. Thereafter, in most cases the target area is contoured, and the delineation of organs-at-risk takes place in a digital planning system. Once the treatment plan is completed and approved, it is transferred to a linear accelerator (linac) before the first irradiation session. In some other cases, a "beam set-up" is done instead. Here, a skilled RTT defines the angles and intensities of the beams to be irradiated in a certain location, similarly to treatment planning. After a specified number of irradiation sessions, a follow-up period takes place. Although in most types of external-beam RT irradiation sessions can be delivered by a single machine working independently, in other types, such as proton therapy, delivery rooms have a more complicated logistics structure that is not captured by the deployment flowchart of Figure 2.1.

The flow of both patients and information is usually influenced by medical and technological constraints. Medical constraints arise when RT is dependent on other forms of treatment such as chemotherapy and/or surgery. In such cases, a time constraint that encompasses a planned delay in the start of treat-

[1]Sometimes referred to as Management Science.

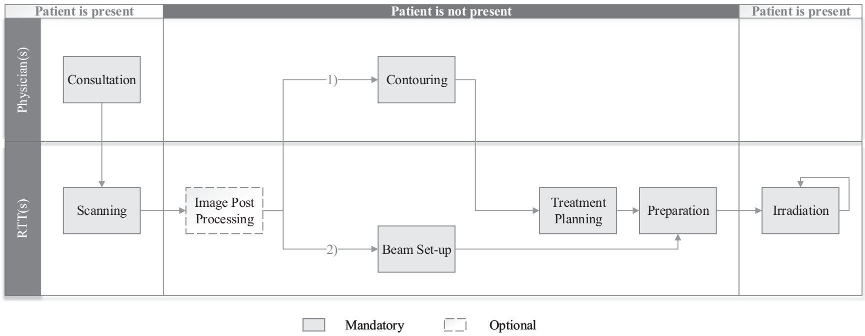


Figure 2.1 Deployment flowchart of the RT process

2

ment emerges. An example is when a patient has surgery before RT and radiation can only be delivered when the wound has healed. Or when a patient receives chemotherapy and a time window for radiation must be followed to ensure the effectiveness of the combined treatment. Technological constraints might occur when only some radiation therapy technologists (RTTs) are trained to carry out a novel treatment or when only a subset of the available linear accelerators (LINACs) is technically capable of delivering RT to a particular cancer type. Moreover, as shown in Figure 2.1, staff members (radiation oncologists, RTTs, etc.) are responsible for performing several operations throughout the RT chain, raising the question of how much of their available time should be allocated to each of these operations. In addition, other appointments (e.g. dentist, dietitian) that depend on the availability of the corresponding professionals and can only be undertaken during certain time slots may be needed before the scanning stage, implying increased waiting times for some patients' throughput. Besides, RT is subject to a considerable number of uncertainties. Daily inflow of new patients, duration of treatment planning activities, and a large number of variables affecting individual care pathways throughout the RT chain appear to be the most significant. Due to this complex logistic environment, the relation between supply and demand in different steps of the chain is not straightforward, and factors limiting the performance of the system - "bottlenecks" - may not always be easy to find. All these factors make the delivery of RT a process with particular characteristics, which brings the need for the development of 'ad hoc' approaches to support recurrent decision-making. Nevertheless, knowledge from the OR community can provide the starting point to optimizing RT logistics through the development of innovative, but yet effective decision support systems [38].

2.1.3 Research aims

There is a wide range of OR applications to solve problems related to medical physics in radiation oncology. A popular example is the design of fluence maps in intensity modulated radiotherapy, i.e. find a fluence pattern over a collection of angles that minimizes the deviation from the desired dose. These applications are discussed by Ehrgott and Holder in [32], but in their review as few as 3 papers covering the logistics aspect of RT treatments are cited. Kapamara *et al*. [56] showed that patient scheduling in RT can be seen as a special case of job-shop scheduling. However, their paper focuses on methods for solving job-shop problems rather than reviewing the application of OR to the RT delivery process. Although OR methods have been extensively applied to solve problems in RT, literature reviews focusing on resource planning problems are scarce, despite the practical relevance of these problems. To fill this gap, in this paper we identify, study and classify OR models that aim to support managerial decision-making in RT. To that end, the research aims of this study are defined as follows:

1. Identify research papers that cover managerial problems in RT using OR methodologies with at least some empirical material.
2. Position the literature by classifying the studies based on several factors such as the subject of research, the hierarchical nature of decision making and the OR technique(s) employed.
3. Examine the maturity level of implementation of the models and the (potential) impact they have created in practice.
4. Identify the shortcomings in the current literature and provide guidelines for future research.

2.2 Methods

2.2.1 Scope

Radiotherapy encompasses a wide range of problem types that can benefit from the OR knowledge. According to the framework proposed by Hans *et al*. [40], managerial decisions can be divided in four areas: medical planning, resource capacity planning, materials planning and financial planning. In this work, we focus on resource capacity planning problems. Our goal is to investigate how resources, staff and patients can be efficiently coordinated to optimize objectives such as the minimization of waiting times, or the maximization of capacity use. Therefore, medical or financial problems are excluded from the scope of this study. On the other hand, we focus on OR methods that quantitatively model those problems with measurable performance indicators. While the spectrum of OR methods is wide and not always consistent amongst researchers [96, 112], we classify the methods in six categories: computer simulation, constructive heuristics, metaheuristics, queuing theory, mathematical programming and Markov decision processes. A list of abbreviations and a short description of these methods can be found in Table 2.1.

Table 2.1 Description of the OR methods.

OR method (abbreviation)	Description
Computer simulation (CS)	Process of building an abstract model that mimics the behavior of a real-world or theoretical system, executing the model on a computer and analyzing the output [58].
Constructive heuristics (CH)	Heuristic methods to create and/or improve a candidate solution, step by step, according to a set of rules defined beforehand, which are built based on the specific characteristics of the problem to be solved [93].
Metaheuristics (MH)	General-purpose heuristic algorithms that iteratively improve a candidate solution, designed to solve a wide range of hard optimization problems without having to deeply adapt to the problem at hand [6]. Contrary to CH, MH are problem-independent techniques that can be used as 'black boxes'. CH and MH are approximation methods, i.e. they do not guarantee that an optimal solution is found. They are used when exact approaches take too much computational time, or when feasibility (or speed) are more important than optimality.
Markov decision processes (MDP)	Mathematical methods to model complex multi-stage decision problems in situations where outcomes are partly random and partly under the control of a decision maker [82].
Mathematical programming (MP)	Optimization methods that aim to mathematically represent a decision problem by defining a set of constraints that bound the values of a set of decision variables, and an objective function to be either minimized or maximized until an optimal solution is found [10].
Queuing theory (QT)	Mathematical methods to model the arrival and departure processes of waiting lines (queues), in order to analyze the congestion and decide the amount of resources required to provide a certain service [113].

2

2.2.2 Data sources and search strategy

We performed searches in 6 databases, divided in three categories: medical, technical and multidisciplinary. To find papers within the medical field, we searched EMBASE and PubMed. To look for literature more geared towards engineering approaches, we searched EBSCO Business Search Elite (BSE). In addition, we carried out searches in two multi-disciplinary databases: Web of Science and Scopus. Besides, a search was performed in ORchestra [43], a database created and maintained by the Center for Healthcare Operations Improvement and Research (CHOIR) containing references from the fields of OR and healthcare categorized by medical and mathematical subject. The full strategy and search terms are provided in Additional file 1. As a means to achieve relevant publications not covered by the chosen databases we also checked the references list of the selected papers for snowballing.

Table 2.2 Inclusion and exclusion criteria.

Inclusion criteria	Exclusion criteria
Journal paper, conference paper or book chapter	Paper published before 2000
Paper uses an OR method or technique	Paper written in other languages than English
Paper addresses one or more logistics problem in RT	Paper tackles a medical problem
	Paper focuses on macro-planning
	Abstract not available online

2.2.3 Inclusion/exclusion criteria and paper selection method

Inclusion and exclusion criteria are presented in Table 2.2. In the aforementioned database search we restricted the search to journal/conference papers and book chapters, and limited the results to papers written in the English language. Besides, due to the fast evolution of both information technologies and algorithms for decision support, we consider that literature studies published before the year 2000 are not likely to be relevant for the purpose of this work. The literature search resulted in 163 different abstracts, from a total of 301 results. Two authors participated in the selection of papers according to the remaining inclusion/exclusion criteria presented in Table 2.2. We decided to neglect papers focusing on macro-planning, i.e., papers proposing analytical models that support decision making for large scale planning, e.g. involving several RT centers at a regional or national level. Instead, this review focuses on models that aim at solving managerial problems of a single RT center.

The first author read the title and abstract of all the 163 papers and selected 30 relevant papers. Thereafter, the fifth author read the title and abstract of a random sample of 25% of the 163 papers (41). The matching rate between the authors was 98% (40 in 41), thus the selection procedure undertaken by the first author was considered valid. We were able to obtain, online, the full text of all papers but 3. These 3 papers were submitted to conference proceedings that we were not able to track. The cross reference checks of the remaining 27 papers resulted in 6 additional papers. Therefore, a total of 33 papers were included in this review. Figure 2.2 depicts an overview of the selection process.

2.2.4 Data extraction

For each paper included in the review we extracted the following information: 1) Subject of research; 2) Hierarchical level; 3) OR method(s); 4) Extent of implementation and 5) (Potential) impact on performance. The subject of research states the type of intervention expected to be taken in practice by the proposed study. It may refer to the problem(s) verified in practice that may have caused the need for a research study, for example. The hierarchical (or organizational) structure was defined in four levels [44]: strategic, tactical, operational offline and operational online. To evaluate the extent of implementation of the models proposed in the literature, we further apply a six stage maturity model as seen in Figure 2.3. The maturity model includes the stages through which OR models typically undergo from the end of the development phase to the observation of

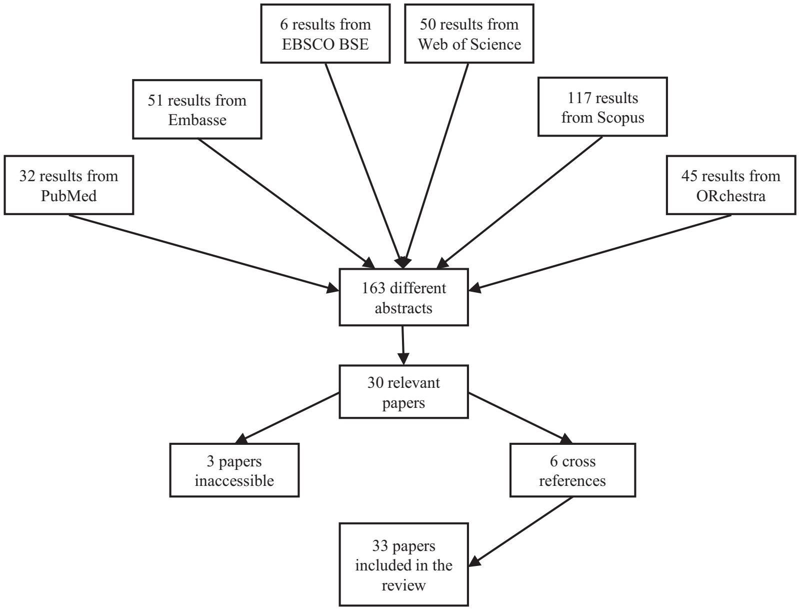


Figure 2.2 Overview of the selection process.

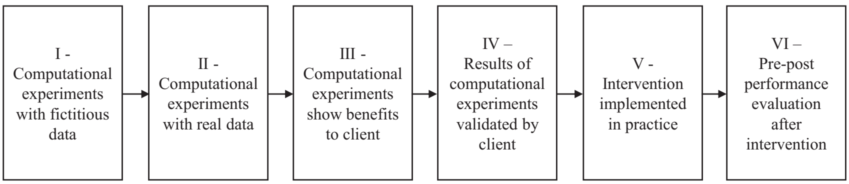


Figure 2.3 Phases for assessing the extent of implementation.

practical operations improvement.

2.2.5 Categorization of results

Managerial decisions for planning and control in RT may vary in purpose, scope or objectives, and may be oriented to the long-term, mid-term or short-term operation. We grouped our findings in four sections according to the structure of the decision problems being tackled: 1) Strategic managerial decision making; 2) Resource capacity planning; 3) Patient prioritization; 4) Scheduling.

Strategic managerial decision making refers to finding the best policies that enhance the long-term operation of an RT center. These decisions are commonly linked to the organization's mission and strategic direction, involving problems such as capacity dimensioning, or the definition of the healthcare delivery process. Strategic decisions usually involve capital investment and are therefore made by on top-level positions of the center's administration. Because there is a high degree of uncertainty at this level, decisions have a long term planning horizon based on highly aggregated (forecasted) information.

Models for resource capacity planning aim to find the best policies to manage the available capacity of existing machines and staff. These usually cover made for a mid-term planning horizon, and involve the combination of forecasted and known information. Decisions on capacity planning guide and restrict the decisions made at lower levels of the center's hierarchy. This can be achieved, for instance, by efficiently assigning the available time slots of machines to certain patient groups in order to guide the appointment office when booking appointments for patients, or optimizing the throughput time of a specific process (e.g. the time slot duration for a CT scan). At this level there is a limited flexibility for capacity expansion.

Patient prioritization models attempt to maximize the tumor control probability (TCP) by making decisions on the urgency levels assigned to patients undergoing treatment; certain patients require shorter access times than others. This stratification is related to the characteristics of the tumor and risk of metastasis. Thus, a proper patient prioritization results in a maximized level of satisfaction for the overall population of patients in a waiting list, even if some patients have their waiting time extended in detriment of others.

Scheduling models aim to generate scheduling decisions for patients throughout the RT chain. The goal is to make an efficient planning of the machines' available capacity by organizing patients in such a way that overall access and waiting times are minimized, delays are avoided, and utilization rates of machines are maximized. Contrary to the previous sections, scheduling decisions typically have a short-term planning horizon, aiming to support the execution of the healthcare delivery process. Although there is a low flexibility on the supply side, at this level the amount of information available is high. The end goal is to balance the workload in such a way that it can be covered by the available capacity. Studies within this section may be oriented towards a specific operation, or integrate scheduling decisions for a part of the chain of operations, such as the pre-treatment stage, i.e. from referral to the first fraction.

2.3 Results

2.3.1 Strategic managerial decision making

Table 2.3 shows the 8 papers that fall within the category of strategic managerial decision making. The subject of research varies among the different scientific publications, with throughput optimization problems being studied the most (50%). Because the majority of the papers address problems at the strategic level (7 in 8), computer simulation is the predominant methodology. Potential improvements were reported, such as the combination of computer simulation and queuing theory performed by Joustra *et al*. [52], which has proven to be capable of increasing the percentage of patients complying with waiting time targets from 39% to 92%. With a similar subject of research, Werker *et al*. [110] presented results that could potentially reduce patients' waiting times by 20%. Results of both studies were accepted by the corresponding clients, implementation was not reported upon.

Table 2.3 Results for strategic managerial decision making.

Reference	Subject of research	Hierarchical level	OR method(s)	Extent of implementation	(Potential) Impact on performance
Thomas [97]	LINACs'capacity dimensioning	Strategic	CS	II	86% patients begin treatment within 10 days for a spare capacity >= 10%
Proctor *et al*. [81]	Patient flow analysis	Strategic	CS	III	82% of patients begin treatment within 14 days
Kapamara *et al*. [54]	Patient flow analysis	Strategic	CS	II	2% reduction in patients' waiting times
Werker *et al*. [110]	Throughput optimization in RT (pre-treatment stage)	Strategic	CS	IV	20% reduction in patients' waiting times
Joustra *et al*. [52]	Throughput optimization in RT	Strategic	CS + QT	IV	Percentage of patients treated within 21 days increase from 39% to 92%
Aitkenhead *et al*. [1]	Throughput optimization in a proton therapy facility	Tactical	CS	III	Deliver over 100 fractions per working day with 4 delivery rooms
Shtiliyanov *et al*. [90]	Evaluation of radiotherapy centers	Strategic	MP + CS	III	Not mentioned
Price and Wasil [80]	Throughput optimization in a proton therapy facility	Strategic	CS	II	Average increase of 2.1 patients treated per hour

2

2.3.2 Resource capacity planning

Five papers tackling resource capacity planning problems were found (see Table 2.4). Results show that queuing theory and mathematical programming techniques may be very useful to find appropriate solutions within a reasonable time. By efficiently planning of the capacity of treatment machines using these techniques, Li *et al*. [61] were able to reduce the number of weekly time slots needed by 12%. At the tactical level, Bikker *et al*. [5] developed a mixed-integer programming model to allocate the doctors' capacity for consultation and contouring tasks, as a function of the workload predicted for a mid-term planning horizon. The authors showed a potential access times' reduction of 15% for regular patients and 16% for subacute patients. These results have been validated by a University Medical Center, and the model is under consideration for implementation. No other implementation reports were found.

Table 2.4 Results for resource capacity planning.

Reference	Subject of research	Hierarchical level	OR method(s)	Extent of implementation	(Potential) Impact on performance
Ogulata *et al*. [69]	Capacity planning of a cobalt device	Operational offline	CE + CS	III	No delays in the start of treatment if slack capacity >= 4 patients per day
Joustra *et al*. [53]	Waiting lists management	Tactical	QT + CS	III	Separate queues require 50% less capacity to achieve targets
Li *et al*. [62]	LINACs' capacity planning	Tactical	QT + MP	I	Not mentioned
Li *et al*. [61]	LINACs' capacity allocation	Operational Offline	MP + QT	I	Reduction of number of required weekly time slots from 125 to 110
Bikker *et al*. [5]	Doctors' capacity allocation	Tactical	MP + CS	IV	Access times reduction of 15% for regular patients and 16% for subacute patients

2.3.3 Patient prioritization

Two papers for patient prioritization were found (see Table 2.5). Ebert *et al*. [30] presented a non-linear programming model that applies a utilitarian prioritization for patients being treated with curative intent. Their results demonstrated large gains in TCP for some groups of patients at the expense of small reductions in TCP for other groups. However, the simulations revealed to be computationally unrealistic for direct application in a clinical setting. To tackle this drawback, Ebert *et al*. [31] developed an analytical solution that quickly prioritizes patients on a waiting list under the same circumstances as in [30], but using a Lagrangean Multiplier method [85] that leads to the same solution in a much faster way. However, this research is still in a very early stage.

Table 2.5 Results for patient prioritization.

Reference	Subject of research	Hierarchical level	OR method(s)	Extent of implementation	(Potential) Impact on performance
Ebert *et al*. [30]	Patient prioritization	Operational offline	MP	I	55% patients with TCP increase
Ebert *et al*. [31]	Patient prioritization	Operational offline	MP	I	Computational time reduction from 1 hour to 1 min

2.3.4 Scheduling

The literature search returned 18 papers addressing scheduling problems (see Table 2.6). Because both the degree of flexibility and the level of uncertainty are low, these models fall within the operational level of a center's hierarchy. Most authors apply mathematical programming techniques (9 in 18), thus achieving (near) optimal solutions. However, (meta)heuristic methods appear as a viable supplement or alternative (8 in 18). Optimizing the overall RT chain using both constructive heuristics and metaheuristics, Petrovic *et al*. [72] achieved considerable reductions in waiting times for palliative (34%) and radical patients (41%).

Focusing on the pre-treatment stage, Petrovic *et al*. [73] explored similarities between radiotherapy and job-shop scheduling problems commonly encountered in industrial processes, using genetic algorithms to minimize both the average waiting times and the average delays in the start of treatment. Results showed that these indicators were reduced by 35% and 20%, respectively. From the 18 papers found, 12 (67%) propose models for scheduling patients on LINACs. Sauré *et al*. [88] formulated the problem as a discounted infinite-horizon Markov decision process to identify policies that can better allocate the LINACs' capacity to reduce waiting times. The percentage of treatments initiated within 10 days, for a clinical data-set provided the British Columbia Cancer Agency increased, on average, from 73% to 96%. In contrast, Legrain *et al*. [59], in collaboration with the Centré Integré de Cancérologie de Laval (CICL), proposed a two-step stochastic algorithm for optimal scheduling in an online fashion. Results of computational experiments undertaken using real data instances provided by the CICL showed an average decrease in the number of patients breaching the standards of 50% for acute patients and 81% for subacute patients. As in the previous sections, none of the papers reported a full implementation of the results, with 56% of the studies performing computational experiments only, either with fictitious or real data.

2.4 Discussion

We observed that there is a growing trend towards applying OR methods for improved decision making in RT over the last 15 years: one paper was published between 2000 and 2005, 13 papers in 2006-2010 and 19 papers in 2011-2015. A total of 33 papers met the inclusion criteria, covering a wide range of problems at various organizational levels with promising results. As for strategic managerial decision making a total of 8 papers were found. At this level, machines' capacity dimensioning and throughput optimization are the most studied problems with computer simulation as the preferred technique. The 5 papers on resource capacity planning show that suggestions for potential improvements mainly refer to increasing the flexibility by, e.g. implementing a dynamic way of reserving time slots for different patient types, allowing breaks between fractions, or letting treatments start in any weekday. For this type of problems, finding a good balance between demand and supply is of special importance to ensure timely treatments.

We found that scheduling problems are the most studied, with 18 out of the 33 papers (55%). Mathematical programming and (meta)heuristics are the preferred OR methods for patient booking throughout the whole RT chain of operations. We presume that decision makers prefer to get approximate (not optimal) solutions in less computational time, as solutions need to be implemented in a daily/weekly basis. However, the problem structure is usually too complex for applying mathematical programming techniques, which require a high computational effort. From the 18 papers focusing on scheduling problems, 12 (36% of

Table 2.6 Results for scheduling.

Reference	Subject of research	Hierarchical level	OR method(s)	Extent of implementation	(Potential) Impact on performance
D. Petrovic *et al*. [73]	Pre-treatment scheduling	Operational offline	MH	III	Reduction of average waiting times and tardiness by 35% and 20%, respectively
Kapamara and D. Petrovic [55]	Radiotherapy scheduling	Operational offline	CH + MH	II	Average waiting times of 1.6, 19.1 and 19.4 days for emergency, palliative and radical patients, respectively
S. Petrovic and Castro [77]	Pre-treatment scheduling	Operational offline	MH	II	Not mentioned
Castro and Petrovic [18]	Pre-treatment scheduling	Operational offline	MP + CH	II	11% of all patients exceed the waiting time targets, in average
D. Petrovic *et al*. [74]	Pre-treatment scheduling	Operational offline	MH	II	Reduction of average waiting times for radical (35 to 21.48 days) and palliative (15 to 13.10) patients
D. Petrovic *et al*. [72]	Radiotherapy scheduling	Operational offline	CH + MH + CS	III	Average waiting times of palliative and radical patients reduced by 34% and 41%, respectively
S. Petrovic *et al*. [76]	Treatment scheduling	Operational offline	CH	II	Decrease in the percentage of late patients of up to 40% for palliative patients and 4% for radical patients
S. Petrovic and Leite-Rocha [75, 78]	Treatment scheduling	Operational offline	CE + MH	I	Average weighted tardiness of 0.935 days
Conforti *et al*. [23, 24]	Treatment scheduling	Operational online	MP	III	Increase of 47% in the number of booked treatment sessions
Conforti [22]	Treatment scheduling	Operational offline	MP	I	LINACs' utilization rates of 95%, in average
Jacquemin al. [49]	Treatment scheduling	Operational offline	MP	I	Admission rate of 25.4 patients per week in a fictitious center with 2 LINACs
Burke *et al*. [15]	Treatment scheduling	Operational offline	MP	V	27% of patients breaching the norms
Jacquemin *et al*. [50]	Treatment scheduling	Operational offline	MP	I	4% increase on the percentage of patients treated
Sauré *et al*. [88]	Treatment scheduling	Operational offline	MDP + MP	III	Increase the average percentage of new patients treated within 10 days, from 73% to 96%
Cares *et al*. [17]	Treatment scheduling	Operational offline	MH	I	Not mentioned
Legrain *et al*. [59]	Treatment scheduling	Operational offline	MP	IV	Decrease on the average number of patients breaching the standards by 50% for acute patients and 82% for subacute patients

the total papers) address the problem of scheduling patients on treatment machines. An elegant example of finding a proper balance between the processes' workload and smooth patient flows is a model that focuses on the scheduling of patients throughout the entire RT chain. To demonstrate that, Petrovic *et al*. [72] achieved impressive reductions in waiting times for palliative (34%) and radical (41%) patients using heuristic algorithms and computer simulation together. We found only two papers integrating scheduling decisions for the overall RT chain. The enormous complexity of the optimization models bringing all these scheduling decisions together might explain the low rate of development of scientific studies within this context.

Table 2.7 summarizes the extent of implementation of the papers included in the literature review. No paper reported a full implementation and performance evaluation of recommendations or software tools, with only one paper referring to a practical implementation being undertaken at the time of publication. Moreover, only four studies had their results validated by the client. Earlier research also reported low levels of actual implementation [100] but publication bias can also play a role. Although we recognize that the extent of implementation of the (scientific) interventions reviewed in this paper may be higher than those reported in the articles, it is also clear that there are many reasons that hamper the translation of theoretical models into practice. First, there are still major issues in getting OR models accepted by clinicians, even when (potential) benefits of innovations are evident [11]. Another factor concerns the development of software tools to be used in the clinic. We found promising models resulting from "in silico" or desk research and/or modelling whereas the translation of the models into a reliable, user-friendly, and bug-free software tool is not straightforward as this part usually falls outside the OR experts' background. A joint teamwork between software developers and operations researchers is needed to overcome this issue. Data availability may be another reason for the low implementation rates; 9 papers were tested using fictitious data rather than real data. Thus, both the verification and validation of the results become an issue that hampers the acceptance of the model by managers. Further clinicians and OR researchers have different publishing routes and priorities; the former aim at improving effectiveness and efficiency directly in practice, whereas publishing new theoretical findings or innovative algorithms is often sufficient for the latter. A last very practical reason for limited findings of implementation may be that generating evidence on operations improvement is not common practice in healthcare and many incremental improvements are implemented in rapid improvement cycles or by trial and error.

Although not within the focus of our study, we verified the topic of facility planning on macro level in an additional search. Decisions on long term capacity need and size of RT centers can be of great influence on cost effective allocation of funds. We could only find one study by Shukla et.al. [91], as referred to in the background section, so it is clear that further research on the application of OR methods in RT macro-planning is very relevant.

Table 2.7 Results for the extent of implementation.

Extent of implementation	Number of papers
I - Computational experiments with fictitious data	9
II - Computational experiments with real data	8
III - Computational experiments show benefits to client	9
IV - Results of computational experiments validated by client	4
V - Intervention implemented in practice	1
VI - Performance evaluation after intervention	0

2.4.1 Research limitations

We may have missed relevant papers, possibly due to the fact that it concerns an interdisciplinary field. The fact that we found six papers by snowballing demonstrates this.

Although we recognize that more papers within the defined scope might be publicly available, we decided to exclude non-peer reviewed articles in this review. Firstly because a search strategy for these papers may be hard to design, and secondly because these may lack scientific rigor. Yet, we made no distinction between papers based on other factors such as the journal's impact factor or the quality of the design and data management in the paper.

Implementation stages were scored according to the reported stages in the papers, and no follow-up investigation has been done in this review. This is a laborious exercise and has shown to reveal limited response [100]. It is thus not possible to report on the most actual extent of implementation, but we have no indications that implementation in practice is very different from what we found.

Further, there is no deterministic way to define exactly what constitutes an OR methodology, or what the main results of a complex and detailed research work are. Therefore, the data extraction process may have a bias towards the authors' perspectives.

Still, we believe that this review provides a good overview of the application of scientific knowledge from OR, applied mathematics and systems engineering to operations improvement in RT.

2.4.2 Future research

Although the range of OR applications in RT is broad and promising results have been reported and some achieved, there is room for future improvement in many directions. Due to new scientific findings related to cancer treatment and technological progress, treatments are getting more specialized and the number of possible care pathways is constantly increasing. This issue creates the need for research in clustering care plans based on the similarities encountered on the corresponding care pathways. Moreover, new devices for improved imaging (such as positron emission tomography–computed tomography) or enhanced radiation delivering (such as the magnetic resonance-LINAC) have been developed. These machines have their own features and limitations, raising the need for new capacity allocation models, as well as the adaption of current models to these new devices. Besides, optimization models should be tested for several real-world data instances in order to strengthen the evidence found by the scientific approaches and ensure the generalization of the models to many different RT centers.

This research produced only one paper proposing a model for scheduling patients in an operational online manner. An investigation area could be the development of innovative models to book patients' sessions on-site immediately after referral or during consultation. These approaches usually involve the use of stochastic programming methods to find good solutions in the presence of the patient, integrating his/her preferences [36].

Another line for further research is the development of more thorough maturity models to assess the extent of implementation, and identify the main causes of the low implementation rates of OR studies in the healthcare field. Due to the assumptions and simplifications of reality usually done in scientific approaches, it would be interesting to see how the implemented solutions perform in comparison with the theoretical findings. The real extent of implementation could be surveyed by approaching the original authors; earlier experience showed however that this requires creativity and perseverance as organizations and staff positions change frequently, and research is published years after the actual projects took place.

2.5 Conclusions

We show that the literature on OR applications in RT covers a wide range of problems, and considerable benefits can be achieved in terms of both waiting times and resource utilization. But there are still major lines for further research, such as the improved coordination of imaging tests, or the development of online models that enable on-site scheduling of patients immediately upon arrival. With respect to the daily flow of patients, results indicate that scientists and managers tend to believe that bottlenecks are most likely to occur on treatment machines. However, research studies have shown that large gains in waiting times reduction can be achieved if the pre-treatment stage is optimized jointly.

Despite the potential benefits of applying OR methods in RT, implementation rates are still low. We provide suggestions for further development of methods as well as for research priorities.

CHAPTER 3

A mathematical programming model for optimizing the staff allocation in radiotherapy under uncertain demand

3

3.1 Introduction

Radiation therapy (RT) is a treatment method for cancer care that involves ionizing radiation to kill cancerous cells. External-beam RT uses a machine called "linear accelerator" (linac) to deliver high-energy radiation beams onto the target area in a series of (usually daily) treatment sessions. As a result of the growing number of cancer patients, demand for RT has been increasing over the years [94], and the global use of radiotherapy in cancer care is estimated to be around 50% [28].

In RT, delays in the start of treatment increase the risk of local recurrence and tumor progression [20], and patients experience greater psychological distress and prolonged symptoms when subject to longer waiting times [63]. For these reasons, standards for maximum waiting times [1] have been set by the Dutch Society for Radiotherapy and Oncology (NVRO) [68]. Besides, resources are highly expensive and RT centers are encouraged to contain their overhead costs by limiting the acquisition of extra capacity. As a result, RT centers aim to organize their resources in the most efficient manner, by promoting policies that smooth patient flow and provide quality of labor while maintaining waiting times low. However, the relation between demand and supply is not straightforward. The RT process is subject to a considerable number of uncertainties, such as the inflow of patients and care pathways, and the time spent on treatment planning activities. On the supply side, staff members, and in particular radiation therapy technologists (RTTs), are multi-skilled highly specialized technicians who combine clinical duties with research and administrative activities. The variety of competences, together with their rotation needs, partial availability and unforeseen no-shows make the planning of RTTs a complex task, potentially leading to situations of overstaffing, understaffing, and jeopardizing the fulfillment of the waiting time standards. In fact, unavailability of staff is found to be one

[1] time between referral and first irradiation session.

of the main causes for patient dissatisfaction regarding pain management in RT [79]. All these factors bring the need for the development of advanced analytical approaches to support recurrent decision making of staff planning in RT.

Operations research (OR) methods have been proposed to optimize resource planning and control in RT, as shown by Vieira *et al*. [107]. In their review, it is shown that although most of the published studies focus on the scheduling of treatment sessions (such as [22], [88], and [59]), greater benefits in terms of waiting time reduction and resource utilization can be achieved if the pre-treatment stage is optimized jointly. Nevertheless, the problem of planning RTTs throughout the RT chain of operations has not been addressed so far.

In this paper, we focus on the tactical planning of RTTs throughout the pre-treatment process of the RT treatment chain considering stochastic patient arrivals. We propose a stochastic mixed-integer linear programming (MILP) model that is able to integrate actual and future demand and maximizes the number of patients finishing the radiotherapy pre-treatment stage within the maximum waiting time target. Our method allows RT centers to optimally allocate RTTs on a mid-term planning horizon, and get insights in the staffing levels that should be allocated to each of these operations, while satisfying the capacity constraints and avoiding the need for RTTs to work overtime.

3.2 Problem statement and background

3.2.1 Radiotherapy process

The RT treatment chain can be divided into two stages: the pre-treatment and the treatment stage. In the former, the patient is scanned, the specific area to be treated is localized and a treatment plan is generated. As for the latter, ionizing radiation is delivered in a pre-defined number of fractions. The delivery of the first fraction determines the start of treatment. The care pathway of a given patient throughout the pre-treatment stage is characterized by a sequence of operations that depends on the characteristics of the tumor (e.g. tumor site, level of advancement, etc.) and urgency level of the patient. Generally, there are three urgency levels: acute (emergency), subacute (urgent), and regular (routine) patients, which determine the recommended maximum waiting time between referral and first fraction. Table 3.1 shows the maximum waiting times in the Netherlands, as recommended by the NVRO.

Table 3.1 Waiting time targets defined by the NVRO, in calendar days.

Category	Target standards	Maximum waiting time
Acute	-	1 day
Subacute	80 % of patients treated within 7 days	10 days
Regular	80 % of patients treated within 21 days	28 days

Figure 3.1 depicts a typical deployment flowchart of the main operations involved in external-beam RT. At referral, patients are scheduled for a consulta-

tion with a doctor (radiation oncologist). During the first consultation, the doctor fills a medical form (hereby denoted as "PlanRT") which includes, amongst others, the urgency level and the care content of the patient. For acute and subacute patients, the first fraction of the patient is determined directly after consultation and all the pre-treatment operations have to be done before the scheduled day. For the imaging of the tumor, a patient may be prescribed one or more examination scans, such as computer tomography (CT), magnetic resonance imaging (MRI), or positron emission tomography-computer tomography (PET-CT). Depending on the care plan, other appointments (e.g. moulding, dentist, dietitian ...) may be needed before the scanning process. In case of multiple imaging scans, an image post-processing (IPP) is needed to match and process the images. Thereafter, the target area is delineated (contoured) by the doctor and subsequently a treatment plan is generated in a digital planning system. Depending on the tumor site and urgency level of the patient, a beam set-up (BSU) may be done instead or in addition to treatment planning (TP). In BSU a skilled RTT defines the angles and determines the intensities of the irradiation beams. BSU can be seen as a straightforward TP step in which simpler techniques such as the two-field technique called "anterior-posterior-posterior-anterior" (AP/PA) are applied. The more elaborate planning techniques such as VMAT and IMRT are applied in TP. Thus, RTTs who can perform TP can also perform BSU, but not all RTTs who are able to undertake BSU can perform TP. Once the final plan is completed, the medical physics unit checks the treatment plan, which is then uploaded to the linac before the treatment can start. After a specified number of irradiation sessions, the treatment finishes and a follow-up period takes place.

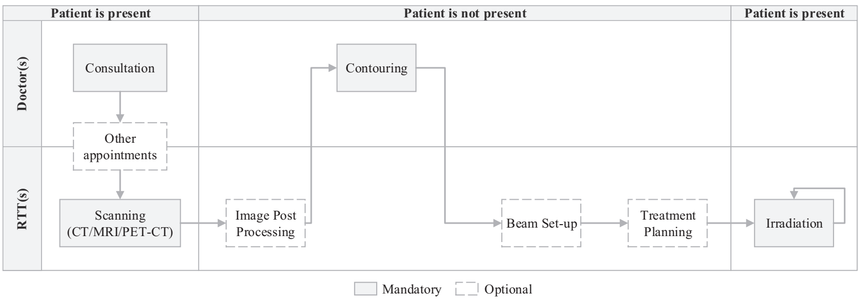


Figure 3.1 Deployment flowchart of the RT process.

3.2.2 The RTT planning problem

RTTs are health professionals who prepare, plan and administer RT treatments, thereby playing a key role throughout the RT chain of operations (see Figure

3.1) and are responsible for the imaging stage, by operating the scanners and performing post-processing of multiple imaging scans. As for the planning part, they are responsible for the execution of BSU and treatment planning, and provide support to the patient while operating the linacs during the irradiation fractions. From a managerial viewpoint, making a schedule for RTTs is not straightforward as several operational restrictions need to be satisfied. For instance, some RTTs with multiple skills have certain rotation requirements to ensure that they remain capable of performing these activities. Also, there is a minimum number of RTTs needed to operate machines (scanners and linacs) to ensure that minimum standards are kept. Moreover, some RTTs combine clinical duties with research projects, patient preparation sessions and other administrative tasks and are only partially available to the clinic.

In practice, RT managers produce a mid-term master schedule with a planning horizon varying between one and six months, in which, apart from having enough RTTs to cover the workload in each operation, all the resource-related constraints are met. However, this is usually done manually, and estimations regarding current and future workload are not fully taken into account. This may lead to suboptimal solutions (overstaffing and understaffing), particularly when fluctuations in workload occur, especially for treatment planning operations. The RT managers often perform last-minute adjustments to the schedule. This may increase the stress and dissatisfaction levels of RTTs and jeopardizes the fulfillment of the waiting time targets. With the growing demand for RT and the resulting increase in the number of RTTs employed, the problem is becoming increasingly relevant to the RT community. Thus, the development of advanced analytical models producing efficient and stable solutions is needed to support the planning of RTTs.

In this paper we propose a stochastic MILP model that aims to optimize the number of patients fulfilling the waiting time standards by allocating RTTs, at the tactical level, to cope with the workload generated from several scenarios of patient arrivals. We start with an overview of the related research studies found in the literature and highlight the contributions of our study. In Section 3.3 we present our innovative solution methodology designed to solve the RTT planning problem. Section 3.4 presents the computational experiments designed to test our methodology using real data. In Section 3.5 we present the results obtained from the case study and elaborate a discussion of the results, and in Section 3.6 we draw the conclusions of this study and elaborate on future research lines.

3.2.3 Related literature

The problem of assigning RTTs to multiple operations in RT has not been studied in the literature. However, other problems such as resource allocation and scheduling of the (pre-)treatment activities of the RT process have been addressed by the OR community [107]. Bikker *et al.* [5] proposed a MILP model to optimize the doctors' agenda regarding consultation and contouring operations, and evaluated the output solutions in a stochastic environment using discrete-event simulation (DES). Although presenting promising results, which include a potential waiting time reduction of 15%, the stochasticity inherent to patient arrivals is not taken into account in the model, which reduces the robustness of the final solutions. Mathematical programming and heuristics have been used by Castro and Petrovic [18] to solve the pre-treatment scheduling problem. Their method is able to minimize the number of patients exceeding the waiting time targets while minimizing the lateness of these patients under resource (doctors and machines) capacity constraints. However, the model is deterministic as the patients are assumed to be known in advance. Legrain *et al*. [60] combined a genetic algorithm with constraint programming to schedule patients for dosimetry (i.e. treatment planning) and linacs, respectively. They generate several scenarios of patient arrivals to estimate future workload and minimize the expected number of patients being delayed for the first fraction. Despite their positive results, many key operations in the pre-treatment part are not modeled and RTT allocation is not optimized simultaneously. Petrovic *et al*. [74] propose a multi-objective genetic algorithm approach for daily scheduling of categorized patients. Innovative coding and decoding procedures and appropriate operators are proposed, however additional metrics and performance evaluations are needed to prove the effectiveness of the proposed method. At the strategic level, Werker *et al*. [110] used DES to mimic the workflow throughout the entire pre-treatment process of the RT department of the British Columbia Cancer Agency (BCCA), and provided sensitivity analysis to identify potential improvements. They found that reducing the variability and the length of the oncologist-related delays reduces the average access time by 25%. Joustra *et al*. [52] hybridized computer simulation and queuing theory to reduce the fluctuations in the outpatient department capacity and increased the number of patients complying with the waiting time targets from 39% to 92%.

Resource allocation and staff rostering problems with a comparable structure have been addressed in other healthcare settings [83]. Hulshof *et al*. [42] proposed a MILP model for allocating hospital resources which copes with multiple patient groups with various uncertain treatment paths. Their results lead to more balanced distribution of resources without jeopardizing patient access times and the number of patients served. Nevertheless, their model is designed to optimize the allocation of resources healthcare settings in which waiting lists arise between care processes. Jerié and Figueira [51] proposed a method for scheduling medical treatments with allocation of physicians' capacity. The problem is formulated as a multi-objective MILP model, and three types of metaheuristics are proposed to solve it: a variable neighborhood search method, scatter search-based method and a non-dominated sorting genetic algorithm. However, their approach is deterministic as the number of treatments and corresponding procedures are assumed to be known in advance. Monte Carlo simulation was used by van Lent *et al*. [101] to reduce the access time between consultation and CT in a radiology department by changing the allocation of CT capacity to different patient groups. By simulating a series of interventions, at the strategic level, they were able to increase the number of patients completing the diagnostic track by 52%.

We conclude that models specifically designed to solve the RTT planning do not exist in the literature, despite the complexity and practical relevance of the problem. Existing decision support models for resource allocation in RT mainly optimize the doctors' agenda, and appear to be deterministic. On the other hand, scheduling models show potential improvements in patient waiting times but do not optimize staffing decisions simultaneously. To fill in this gap, we develop a mixed integer linear program to solve the problem of allocating RTTs to multiple operations in RT that achieves a (near) optimal solution in acceptable computation time. We integrate the stochasticity inherent to the patient inflow in cancer care by optimizing the allocation decisions over a set of scenarios of patient arrivals, and test the effectiveness of our method by applying it to a real-world comprehensive cancer center.

3.3 A mathematical programming model for the RTT allocation problem

We propose a stochastic MILP model that takes into account the patients that are already undergoing the RT pre-treatment process (actual patients), and scenarios of future patient arrivals and corresponding care content. We assume that each patient is aimed to start treatment within the maximum waiting time, which is in accordance with the waiting time standards for his/her urgency level (Table 3.1). The solutions found by our method comply with the practical constraints verified in the pre-treatment process, both for patient throughput and resource (RTTs) planning, and are able to accommodate the fluctuations inherent to the inflow of patients and care content. The goal is to maximize the (expected) number of patients finishing the pre-treatment stage within a predefined waiting time target. RTT allocation decisions are optimized to cover the expected demand in the imaging and planning processes, while the number of RTTs needed to operate linacs is given as input. Thus, we model the scheduling of patients to six different operations: CT, MRI, PET-CT, IPP, BSU, and TP. Operations are divided in "continuous" and "non-continuous" operations. Continuous operations are the ones in which the patient is present, and therefore cannot be interrupted and resumed at a later period (CT, MRI and PET-CT). Non-continuous operations can be interrupted and completed in later time periods (IPP, BSU, and TP). In addition, we include the occurrence of a randomly-generated delay between the last imaging operation and the planning process, different per patient group, to account for the time needed by the doctors of the corresponding specialty to perform and discuss the delineation. Similarly, a safety time margin to account for the time that might be needed for other appointments before the imaging process is included. The planning horizon is discretized in time periods of one day length. In this section, we present the assumptions, formal description, and the mathematical formulation of the problem.

3.3.1 Assumptions

In our model, we make the following assumptions regarding the RT delivery process:

1. Patients joining the system have already had a consultation with the doctor and are assigned a maximum waiting time target, a patient group (tumor site) and the operations included in his/her care pathway.
2. Each patient is assigned a date from which they can start the imaging process. A delay may be needed for other appointments (e.g. dentist, blood analysis, etc.).
3. Each patient has a preparation target time, which is known at the moment of the PlanRT. This corresponds to the date by which the patient is aimed to have finished the RT preparation stage for a timely start of the RT treatment.
4. A care pathway includes at least one imaging operation (CT, MRI, PET-CT), and at least one planning operation (BSU, TP).
5. Operations are delivered in sequence, in the order presented in Figure 2.1.
6. Continuous operations are processed by one machine (number of RTTs per machine is known) and non-continuous operations are processed by one RTT.
7. Processing times of operations vary per patient group.
8. Imaging machines (CTs/MRIs/PET-CTs) are always functional (no breakdowns).
9. Imaging rooms of the same type are interchangeable, i.e. there are no technical differences between the same type of machines.
10. The number of RTTs needed to operate linacs is known in advance.

3.3.2 Notation

The sets, subsets and indices used in the MILP model can be found in Tables 3.2 and 3.3. The list of input parameters is presented in Table 3.4.

Table 3.2 Sets and indices of the MILP model.

Set	Description
$\mathcal{P}$	set of all (current and future) patients ($p \in \mathcal{P}$)
$\mathcal{O}$	set of operations ($i \in \mathcal{O}$)
$\mathcal{R}$	set of RTTs ($r \in \mathcal{R}$)
$\mathcal{M}$	set of machines ($m \in \mathcal{M}$)
$\mathcal{T}$	set of time periods (workdays) ($t \in \mathcal{T}$)
$\mathcal{G}$	set of patient groups ($g \in \mathcal{G}$)
Ω	set of scenarios ($\omega \in \Omega$)

Table 3.3 Subsets of the MILP model.

Subset	Description
$\mathcal{P}^A$	set of actual patients
$\mathcal{P}^\omega$	set of future patients in scenario ω
$\mathcal{P}^g$	set of patients of patient group g
$\mathcal{O}^p$	set of operations to be performed for patient p
$\mathcal{O}^C$	set of continuous operations ($\mathcal{O}^C \subset \mathcal{O}$)
$\mathcal{O}^N$	set of non-continuous operations ($\mathcal{O}^N \subset \mathcal{O}$)
$\mathcal{O}^i$	set of operations preceding operation i
$\mathcal{M}^i$	set of machines available for performing operation i

3.3.3 Decision variables

Decisions regarding the allocation of RTTs are modeled by variables X_{rit} and L_{rt}. We define X_{rit} to be 1 if RTT r is allocated to work in operation i in time period t and 0 otherwise, and L_{rt} to be 1 if RTT r works on linacs in time period t and 0 otherwise. To model patient scheduling decisions, we use a combination

Table 3.4 Input parameters of the MILP model.

Parameter	Description
Integer	
b_p	time period from which patient p is ready to start the RT process, $\forall p \in \mathcal{P}$
l_t	length, in minutes, of each time period t, $\forall t \in \mathcal{T}$
d_p	target date for finishing the RT pre-treatment process of patient p, $\forall p \in \mathcal{P}$
k_i	number of RTTs needed to operate one machine of continuous operation i, $\forall i \in \mathcal{O}^C$
f_t	number of RTTs needed to operate linear accelerators in time period t, $\forall t \in \mathcal{T}$
tra_{ri}	minimum number of time periods that RTT r has to work on operation i, $\forall r \in \mathcal{R}, \forall i \in \mathcal{O}$
lin_r	minimum number of time periods that RTT r has to work on linear accelerators, $\forall r \in \mathcal{R}$
c_p	waiting time, in time periods, between imaging and planning for contouring of patient p, $\forall p \in \mathcal{P}$
p_{gi}	processing time, in minutes, of operation i of patient group g, $\forall g \in \mathcal{G}, \forall i \in \mathcal{O}$
h_i	minimum amount of time, in minutes, spent in a non-continuous operation i per period $\forall i \in \mathcal{O}^N$
dl_p	delay, in time periods, between planning and the first fraction of patient p, $\forall p \in \mathcal{P}$
Binary	
rav_{rt}	1 if RTT r is available during time period t, 0 otherwise, $\forall r \in \mathcal{R}, \forall t \in \mathcal{T}$
cap_{ri}	1 if RTT r is capable of performing operation i, 0 otherwise, $\forall r \in \mathcal{R}, \forall i \in \mathcal{O}$
q_r	1 if RTT r is capable of operating linear accelerators, 0 otherwise, $\forall r \in \mathcal{R}$
Continuous	
mav_{imt}	time, in minutes, that machine m of operation i works during period t, $\forall i \in \mathcal{O}^c, \forall m \in \mathcal{M}^i, \forall t \in \mathcal{T}$

of binary variables and continuous variables. We let Y_{pit} be 1 if operation i of patient p is performed during time period t, and 0 otherwise, and define the continuous variables T_{pit} as the time, in minutes, spent on operation i of patient p during period t. In addition, we introduce the auxiliary variables W_{pt} to model the completion of the pre-treatment stage for each patient. Binary variables W_{pt} take the value 1 if patient p has finished the RT pre-treatment stage before or during t, 0 otherwise. A formal description of the decision variables used in our model can be found in Table 3.5.

Table 3.5 Decision variables of the MILP model.

Variable	Description
Binary	
X_{rit}	1 if RRT r is allocated to work in operation i in time period t, 0 otherwise, $\forall r \in \mathcal{R}, i \in \mathcal{O}, \forall t \in \mathcal{T}$
L_{rt}	1 if RTT r is allocated to work on linacs in time period t, 0 otherwise, $\forall r \in \mathcal{R}, \forall t \in \mathcal{T}$
Y_{pit}	1 if operation i of patient p is performed in time period t, 0 otherwise, $\forall p \in \mathcal{P}, \forall i \in \mathcal{O}, \forall t \in \mathcal{T}$
W_{pt}	1 if patient p has completed pre-treatment before or during t, 0 otherwise, $\forall p \in \mathcal{P}, \forall t \in \mathcal{T}$
Continuous	
T_{pit}	time, in minutes, spent on operation i of patient p during period t, $\forall p \in \mathcal{P}, \forall i \in \mathcal{O}, \forall t \in \mathcal{T}$

3

3.3.4 Constraints

Constraints for patient scheduling

The scheduling component assigns the demand that is expected at each time period to operations of the RT process. The problem structure has similarities with the hybrid flow-shop scheduling problem [87]. All constraints have to be met for all patients already in the system and for future patients included in each scenario $\omega \in \Omega$. Constraints (3.1) ensure that patient operations can only be scheduled once the patient is ready for RT. Constraints (3.2) ensure that each continuous operation is performed in at most one time period. Precedence relations for continuous operations and non-continuous operations are modeled by Constraints (3.3) and (3.4), respectively. Inequalities (3.5) ensure that the time needed for processing operations for a given patient during the same time period do not exceed the period's length l_t. Precedence relations for operations performed within the same time period are not guaranteed by the model, but remain feasible due to Constraints (3.5). Constraints (3.6) force a delay c_p between the imaging planning processes for contouring of the tumor by the doctor(s). Finally, Constraints (3.7) and (3.8) bound the values of the continuous variables T_{pit}.

$$Y_{pit} = 0, \forall p \in \mathcal{P}, \forall i \in \mathcal{O}^p, \forall t = 1..b_p - 1 \tag{3.1}$$

$$\sum_{t \in \mathcal{T}} Y_{pit} \leq 1, \forall p \in \mathcal{P}, \forall i \in \{\mathcal{O}^p \cap \mathcal{O}^C\} \tag{3.2}$$

$$\sum_{n=1}^{t} Y_{pi'n} \geq Y_{pit},$$
$$\forall p \in \mathcal{P}, \forall i \in \mathcal{O}^p, \forall i' \in \{\mathcal{O}^i \cap \mathcal{O}^p \cap \mathcal{O}^C\}, \forall t \in \mathcal{T} \tag{3.3}$$

$$\sum_{n=1}^{t} T_{pi'n} \geq Y_{pit}.p_{gi'},$$
$$\forall g \in G, \forall p \in \mathcal{P}^g, \forall i \in \mathcal{O}^p, \forall i' \in \{\mathcal{O}^i \cap \mathcal{O}^p \cap \mathcal{O}^N\}, \forall t \in \mathcal{T} \tag{3.4}$$

$$\sum_{i \in \{\mathcal{O}^p \cap \mathcal{O}^N\}} T_{pit} + \sum_{i \in \{\mathcal{O}^p \cap \mathcal{O}^C\}} Y_{pit}.p_{gi} \leq l_t, \forall g \in G, \forall p \in \mathcal{P}^g, \forall t \in \mathcal{T} \tag{3.5}$$

$$(1 - Y_{pi't}).|\mathcal{T}| \geq \sum_{n=t}^{\min\{|\mathcal{T}|, t+c_p\}} Y_{pin}, \forall p \in \mathcal{P}, \forall t \in \mathcal{T},$$
$$i' = \max\{a \in \mathcal{O}^p \cap \{1..4\}\}, i = \min\{b \in \mathcal{O}^p \cap \{5..6\}\} \tag{3.6}$$

$$0 \leq \sum_{t \in \mathcal{T}} T_{pit} \leq p_{gi}, \forall g \in G, \forall p \in \mathcal{P}^g, \forall i \in \{\mathcal{O}^p \cap \mathcal{O}^N\} \tag{3.7}$$

$$Y_{pit}.h_i \leq T_{pit} \leq Y_{pit}.l_t, \forall p \in \mathcal{P}, \forall i \in \{\mathcal{O}^p \cap \mathcal{O}^N\}, \forall t \in \mathcal{T} \tag{3.8}$$

Constraints for RTT allocation

We aim to satisfy the restrictions encountered by RT managers when producing RTT master mid-term schedules. Constraints (3.9) assign each RTT to at most one operation, if he/she is available in the corresponding time period. Inequalities (3.10) and (3.11) ensure that an RTT is not assigned to an operation that he/she cannot perform. Relations (3.12) match demand and supply restrictions for non-continuous operations in terms of RTT availability, and Constraints (3.13) serve the same purpose but for continuous operations as a function of machines' availability. Constraints (3.12) and (3.13) ensure that there is enough supply to cover the estimated workload, per operation and time period, in all scenarios. Restrictions (3.14) and (3.15) assign the necessary number of RTTs to operate imaging machines and linacs, respectively.

$$\sum_{i\in\mathcal{O}} X_{rit} + L_{rt} \leq \text{rav}_{rt}, \forall r \in \mathcal{R}, \forall t \in \mathcal{T} \tag{3.9}$$

$$X_{rit} \leq \text{cap}_{ri}, \forall r \in \mathcal{R}, \forall i \in \mathcal{O}, \forall t \in \mathcal{T} \tag{3.10}$$

$$L_{rt} \leq q_r, \forall r \in \mathcal{R}, \forall t \in \mathcal{T} \tag{3.11}$$

$$\sum_{p\in\{\mathcal{P}^A\cup\mathcal{P}^\omega\}} T_{pit} \leq \sum_{r\in R} X_{rit}.l_t, \forall i \in \mathcal{O}^N, \forall t \in \mathcal{T}, \forall \omega \in \Omega \tag{3.12}$$

$$\sum_{g\in G}\Big(\sum_{p\in\{\mathcal{P}^g\cap(\mathcal{P}^A\cup\mathcal{P}^\omega)\}} Y_{pit}.p_{gi}\Big) \leq \sum_{m\in\mathcal{M}^i} \text{mav}_{imt},$$
$$\forall i \in \mathcal{O}^C, \forall t \in \mathcal{T}, \forall \omega \in \Omega \tag{3.13}$$

$$\sum_{r\in\mathcal{R}} X_{rit} = \sum_{m\in\mathcal{M}^i} k_i. \min\{\text{mav}_{imt}, 1\}, \forall i \in \mathcal{O}^C, \forall t \in \mathcal{T} \tag{3.14}$$

$$\sum_{r\in\mathcal{R}} L_{rt} = f_t, \forall t \in \mathcal{T} \tag{3.15}$$

3

Constrains (3.16) set the values of the variables W_{pt} to be 0 if a patient p has not completed the planning process (BSU or TP) before or during period t, and constraints (3.17) ensure that variables W_{pt} are equal to 1 if a patient p has completed the planning process (BSU or TP) before or during period t.

$$p_{gi} - \sum_{n=1}^{t} T_{pin} \leq (1 - W_{pt}).p_{gi},$$
$$\forall g \in G, \forall p \in \mathcal{P}^g, i = \max\{i' : i' \in \mathcal{O}^p\}, \forall t \in \mathcal{T} \tag{3.16}$$

$$\sum_{n=1}^{t} T_{pin} - p_{gi} \leq W_{pt} - 1,$$
$$\forall g \in G, \forall p \in \mathcal{P}^g, i = \max\{i' : i' \in \mathcal{O}^p\}, \forall t \in \mathcal{T} \tag{3.17}$$

3.3.5 Objective function

The objective function (3.18) aims to maximize the number of actual patients ($\mathcal{P}^A$) and the expected number of future patients ($\mathcal{P}^\omega$) amongst scenarios $\omega \in \Omega$ finishing the pre-treatment stage within the waiting time target d_p.

$$\max \sum_{p\in\mathcal{P}^A} W_{p,d_p} + \frac{1}{|\Omega|} \sum_{\omega\in\Omega} \sum_{p\in\mathcal{P}^\omega} W_{p,d_p} \tag{3.18}$$

3.3.6 Model extension - RTT rotation needs

RTTs who are capable of performing multiple activities have certain rotation requirements to ensure that they keep having enough practical experience to keep performing them. However, rotation requirements can in practice be relaxed or satisfied in a long (strategic) planning horizon. Nevertheless, the integration of constraints ensuring enough rotation between operations for RTTs might be relevant for RT managers producing their schedules. Thus, we introduce as model extensions the constraints (3.19) and (3.20), which model the rotation needs for pre-treatment operations and linacs, respectively.

$$\sum_{t \in \mathcal{T}} X_{rit} \geq \text{tra}_{ri}, \forall r \in \mathcal{R}, \forall i \in \mathcal{O} \tag{3.19}$$

$$\sum_{t \in \mathcal{T}} L_{rt} \geq \text{lin}_{r}, \forall r \in \mathcal{R} \tag{3.20}$$

3.4 NKI case study

3.4.1 Experiment data

We tested our model using data from the radiotherapy department of the Netherlands Cancer Institute (NKI), a comprehensive cancer center located in Amsterdam, the Netherlands. Since our contribution is at the tactical level, i.e. the model is to be run in the beginning of a mid-term planning horizon, we ran our experiments for the month of November 2016. We chose this month since it was the most recent one without public holidays. Thus, the whole data-set used in our experiments is from November 2016. The scenarios for future patient arrivals and care content were generated based on historical data measured in practice during the year 2016. In addition to the 22 workdays of November 2016, we included a *cool-down* period of 12 workdays to ensure that operations of patients whose maximum waiting time target date falls after November were also performed. Thus, the length of the planning horizon is 34 time periods of one day. The clinic is open on workdays from 08h30 to 17h30, with a 1-hour break time. Thus, each time period has a length of 8 hours, i.e. 480 minutes. At the beginning of the planning horizon there are 132 patients who did not complete the pre-treatment process. The care plan, urgency level, specific care pathway and the date of the first fraction of those 132 patients are given as input. Patients are categorized in 53 different patient groups, which are related to the tumor type. The precedence relations between the six modeled operations are as follows: CT, MRI and PET-CT have no precedence relation, and are precedent to IPP, BSU, and TP. IPP is precedent to both BSU and TP.

The studied department has 2 CT scanners, 1 MRI scanner, and 1 PET-CT scanner. The PET-CT is shared with the radiology department. The total time, in minutes, that each machine is available per weekday for the department is shown in Table 3.6. Each working CT requires at least 2 RTTs, while the MRI and the PET-CT require one RTT each from the RT department to operate. There are 9 linacs functioning on a daily basis, with 4 RTTs each. Continuous operations (CT, MRI and PET-CT) have a time slot length that is independent of the patient group, as follows: 25 minutes for a CT, 45 minutes for an MRI, and 45 minutes for a PET-CT. According to the managers, an IPP operation is also expected to take around 45 minutes, independent of the patient group. Standard processing times for BSU and TP vary considerably per patient group, ranging from 60 (e.g. bone metastasis) to 120 (e.g. breast) minutes for a BSU, and from 150 (e.g. prostate) to 960 (e.g. head-and-neck) minutes for creating a treatment plan. If an RTT works in a non-continuous operation on a given day, then he/she should do it for at least 15 minutes, i.e. $h_i = 15$.

Table 3.6 Daily availability, in minutes, of the imaging machines of the RT department at the NKI.

Weekday	CT1	CT2	MRI	PET-CT
Monday	480	0	180	135
Tuesday	480	240	300	0
Wednesday	480	240	420	0
Thursday	480	240	420	180
Friday	480	240	300	135

The department has a total of 109 full-time and part-time RTTs, corresponding to 75 in full-time equivalence (FTE). The availability of each RTT in each time period is subject to a no-show probability of 4%, according to historical data. From the whole group of 109 RTTs, the break down of RTTs' capabilities per operation is as follows: 22 are trained for CT, 8 for MRI, 5 for PET-CT, 7 for IPP, 27 for BSU, 24 for TP, and 92 for linacs. Table 3.7 shows the skills for a subset of the RTTs of the NKI.

Table 3.7 RTT skils in the RT department of the NKI for a subset of RTTs.

	CT	MRI	PET-CT	IPP	BSU	TP	linacs
RTT 1	x		x				x
RTT 2					x	x	
RTT 3	x	x		x			x
...	...	...	...	...	...	...	...
RTT109							x

3.4.2 Scenario generation

Scenarios of patient arrivals (PlanRTs) and corresponding care content during the whole planning horizon are generated according to probability distributions and corresponding parameters calculated using historical data gathered from MOSAIQ [33], a patient information management system for radiation oncology.

We used data from the year 2016 to estimate probability distributions and parameters to generate patient data for future time periods. We studied the historical records of PlanRTs filled in by the doctors after consultation over the year, excluding weekends and public holidays, in a total of 5531 records. Since doctors have their own agenda and work routines, we started by performing an analysis of variance (ANOVA) with post-hoc Bonferroni test to study whether there were statistically significant differences on the number of new patients between workdays. Results (see Table 3.8) showed that Tuesday, Wednesday and Thursday are not significantly different from each other. However, Monday has significantly fewer new patients compared to Tuesday and Wednesday, and Friday has significantly fewer patients compared to Tuesday, Wednesday and Thursday. Monday and Friday are comparable. Due to the identified differences, we analyzed patient arrivals per weekday independently.

Table 3.8 ANOVA of the patient arrivals in the NKI during the year 2016.

(I) Weekday	(J) Weekday	Mean Diff. (I-J)[a]	Sig.	95% Conf. Interval	
				Lower B.	Upper B.
Monday	Tuesday	-5.4*	0.0	-8.7	-2.1
	Wednesday	-4.4*	0.0	-7.7	-1.0
	Thursday	-31.0	0.1	-6.4	0.2
	Friday	12.3	1.0	-2.1	4.5
Tuesday	Wednesday	10.4	1.0	-2.2	4.3
	Thursday	23.0	0.5	-1.0	5.6
	Friday	6.6*	0.0	3.4	9.9
Wednesday	Thursday	12.6	1.0	-2.0	4.5
	Friday	5.6*	0.0	2.3	8.9
Thursday	Friday	4.3*	0.0	1.1	7.6

[a] The mean difference is statistically significant below the 0.05 level (*).

A statistical analysis using a Kolmogorov-Smirnov test with a cut-off value of 0.05 showed that patient arrivals can be represented by a Poisson distribution on each workday. Table 3.9 shows the mean, standard deviation and corresponding p-values of the test. As can be seen, Monday and Friday have considerably fewer incoming patients in comparison to the other three weekdays. All p-values are above 0.05, which indicates that the hypothesis of the patterns of patient arrivals following a Poisson distribution cannot be rejected. Therefore, we use a Poisson distribution to generate a random number of new patients, per time period, with a mean equal to the corresponding weekday.

Table 3.9 Statistical analysis of the patient arrivals of the NKI, per weekday, during the year 2016.

Weekday	Sample size	Prob. Dist.	Mean	Std. Dev.	p-value
Monday	49	Poisson	19.4	4.9	0.42
Tuesday	52	Poisson	24.8	6.4	0.25
Wednesday	51	Poisson	23.7	5.7	0.19
Thursday	51	Poisson	22.5	6.0	0.09
Friday	52	Poisson	18.1	6.1	0.66

The care content of each patient includes i) the patient group, ii) the urgency level, iii) the specific operations to be performed, iv) the delay to start the RT process (due to medical reasons), v) the delay needed for contouring, and vi) the delay needed for the start of treatment due to the scheduling on the linacs. These data are generated according to empirical distributions built from measurements performed in the clinic during the year 2016. The patient group of each patient is generated independently of the remaining parameters, while the remaining data are generated proportionally to the empirical measurements from the sample that corresponds to his/her patient group. Table 3.10 shows the distribution of patients amongst the five largest patient groups and the corresponding distribution of urgency levels. Similarly, for each patient we generate the (possible) delay for the start of the pre-treatment (0,5,..,40 workdays for regular patients only), the delay for contouring (0,..,3 workdays for subacute and 0,..,5 for regular patients, 0 for acute patients), and the delay needed for the first fraction due to the scheduling on the linacs (0,..,3 workdays for subacute and 0,..,5 for regular patients, 0 for acute patients) according to the measurements for the sample of his/her patient group and urgency level.

Table 3.10 Distribution of patients for the five largest patient groups in 2016 as a percentage of the total number of patients, and corresponding distribution in terms of urgency level per patient group.

Patient group	% Total	% Acute	% Subacute	% Regular
Bone metastasis	24.9%	5.7%	87.9%	6.5%
Mamma	15.2%	0.0%	0.6%	99.4%
Lung	5.6%	0.0%	3.6%	96.4%
Head-and-neck	4.2%	0.4%	2.2%	97.4%
Prostate	3.7%	0.0%	2.0%	98.0%

3.4.3 Solution generation approach

The RT department of the NKI aims to fulfill the quality standards defined by the NVRO, as presented in Table 3.1. There are two distinct recommended standards for timely delivery of the RT in the Netherlands: i) a maximum waiting time for 100% of the patients of each urgency level, hereby referred to as “100%-target”, and ii) a maximum waiting time for (at least) 80% of the patients of each urgency level, hereby called “80%-target”. After a number of interviews with the department managers, it was agreed that both targets should be considered in the optimization, as both are crucial for the quality assessment of RT centers in the Netherlands. Also, it was agreed that it is more important to ensure that patients start treatment within the maximum waiting time standard, i.e. objective i) has priority over objective ii). Besides, since it is not possible to know in advance how many patients can be actually processed given the existing resources, and given that not all patients will be able to finish pre-treatment within the planning horizon (e.g. those arriving in the last time periods), a preliminary stage is needed to maximize the number of patients finishing pre-treatment. Following this, we apply the model presented in Section 3.3 in three sequential stages, as a hierarchical optimization approach (see Figure 3.2). In the first stage, we aim to maximize the number of patients completing the pre-treatment stage within the planning horizon. Therefore, in the first stage we solve the following MILP program:

$$\begin{aligned}&\max z_1 = \sum_{p \in \mathcal{P}^A} W_{p,|\mathcal{T}|} + \frac{1}{|\Omega|} \sum_{\omega \in \Omega} \sum_{p \in \mathcal{P}^\omega} W_{p,|\mathcal{T}|} \\ &\text{subject to: } (3.1) - (3.16)\end{aligned} \tag{3.21}$$

In the second stage, we add the objective value z_1^* obtained by solving problem (3.21) as a new constraint, and solve the MILP model setting as target date (d_p^1) of each patient p the maximum waiting time corresponding to his/her urgency level (see Table 3.1) minus the access time needed due to the scheduling on the linacs (dl_p). Thus, we solve the following problem:

$$\begin{aligned}&\max z_2 = \sum_{p \in \mathcal{P}^A} W_{p,d_p^1} + \frac{1}{|\Omega|} \sum_{\omega \in \Omega} \sum_{p \in \mathcal{P}^\omega} W_{p,d_p^1} \\ &\text{subject to: } (3.1) - (3.16), z_1 \geq z_1^*.\end{aligned} \tag{3.22}$$

In the third stage, the objective value z_2^* obtained by solving the program (3.22) is passed as an additional constraint to another MILP problem, which has as target date (d_p^2) of each patient p the 80%-target corresponding to his/her urgency level minus the delay needed to access the linacs, as follows:

$$\begin{aligned}&\max z_3 = \sum_{p \in \mathcal{P}^A} W_{p,d_p^2} + \frac{1}{|\Omega|} \sum_{\omega \in \Omega} \sum_{p \in \mathcal{P}^\omega} W_{p,d_p^2} \\ &\text{subject to: } (3.1) - (3.16), z_1 \geq z_1^*, z_2 \geq z_2^*.\end{aligned} \tag{3.23}$$

When the model is to be used on a certain frequency basis (e.g. monthly) in practice, new scenarios must be generated in the beginning of the planning horizon (first day of the month) and used in conjunction with the available system data (e.g. patients undergoing pre-treatment) to feed the model and apply the solution methodology, obtaining a new RTT allocation for the whole month. The process is then repeated every month. The output allocation solution given by the model will thus be optimized to cover the workload associated with the system patients and the future patients of all generated scenarios.

Scenarios and input-output validation

The waiting times of patients throughout the RT chain are heavily influenced by the highly variable patient inflow. Therefore, considering only one scenario of patient arrivals per day may not give a sufficiently realistic representation of

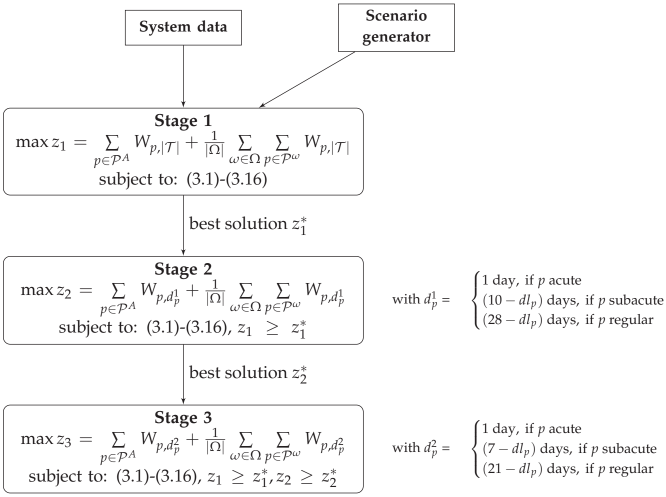


Figure 3.2 Overview of the solution approach designed for the NKI case study.

the possible variation of patient inflow between consecutive days. To define the minimum number of scenarios needed in our experiments, we analyzed how much variability increases with the increase of the number of scenarios throughout the one-month period. To this end, we ran the scenario generator for 1, 2,...,100 scenarios (100 runs per scenario number), and calculated the maximum variation in patient arrivals between two consecutive periods within a scenario. The results showed that the steepest variation occurs between the use of 1 scenario and 6 scenarios (from 15.4 to 21.6), whilst the increase of the maximum variation in the number of patient arrivals between consecutive periods from 6 scenarios to 100 scenarios is not as significant (from 21.6 to 27.0). In addition, we ran a set of computational experiments by solving the model using the NKI data-set, starting with 2 scenarios and increasing the number of scenarios one by one until the solver reached a CPU time limit of 4 hours. We observed that the time limit would be exceeded when using more than 5 scenarios. Therefore, a total of 5 training scenarios were used for generating solutions as described in Section 3.4.3. As a validation step, we performed an accuracy check of the scenario generator by comparing the generated patient data with both the input parameters used to generate them and the data recorded in the real system (Tables 3.9 and 3.10). Figure 3.3 depicts the number of patient arrivals per scenario over the 22 time periods (solid lines), and the daily number of new patients measured in practice during the month of November 2016 (dashed line). As we can see, there are considerable variations in the patient inflow between consecutive days (from 18 to 31, and from 31 to 17 patients between periods 15 and 17), which are hard to predict. Nevertheless, we can see that, overall, the diversity introduced by the scenarios cover the fluctuations associated with the patient inflow during a one-month period.

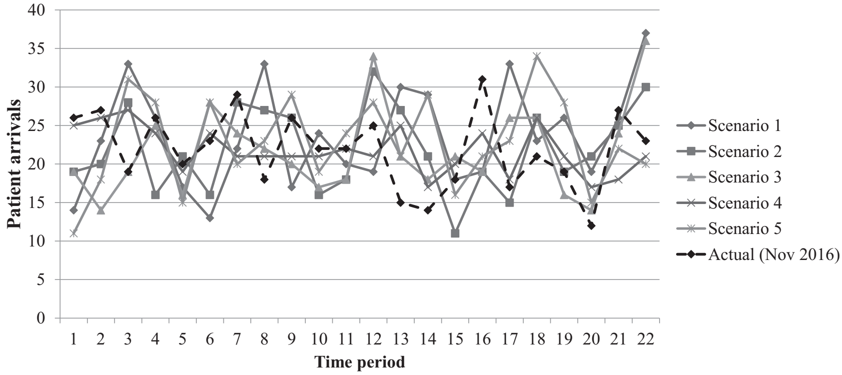


Figure 3.3 Patient arrivals, per time period, for the scenarios created by the scenario generator (solid lines) for generating the RTT allocation solution, and patient arrivals measured in practice during November 2016 (dashed line).

The number of patient arrivals and distribution of patients per urgency level can be seen in Table 3.11. The average number of arrivals (492.2) is above the monthly average throughout 2016 (460.9), but close to the patient arrivals during November 2016 (480). This is because the evaluation month of November 2016 has 22 workdays, while the number of monthly labor days in 2016 averaged 20.6. Analyzing Table 3.11 per scenario, we confirm the validity of the generated data. The scenario generator introduces variability on the number of patients between scenarios in each urgency category, with the average over all scenarios approaching the records measured in practice. For instance, the proportion of subacute patients varies between 29.17% (scenario 1) and 35.55% (scenario 2), but the average over scenarios is only 0.7% away from the corresponding proportion verified in 2016. Similarly, Table 3.12 shows the distribution of patients per patient group, per scenario, from which comparable conclusions may be drawn. Additional checks for the remaining input data were performed, reaching similar conclusions.

The scenario generator and the MILP model were programmed in C++ using Visual Studio 2015 and the Concert Technology of CPLEX v12.7.0, which was used as a solver. Experiments were run in a desktop computer with a processor Intel i7 3.6 GHz and 16 GB of RAM using up to 8 threads, running on a 64-bit version of Windows 10. To speed up the search process, the solution obtained on each stage was passed to the next stage as an initial solution for a warm MILP start. In addition, we set an optimality gap tolerance of 5% and a CPU time limit of 14400 seconds per stage. Table 3.13 presents the performance of the proposed methodology in achieving the final solution within the defined tolerances in each stage. The optimality gap for stages two and three are be-

Table 3.11 Patient arrivals and distribution of patients per urgency level per scenario, and the average measured in practice in 2016.

Sample	Arrivals	Acute	Subacute	Regular
Scenario 1	521	1.7%	29.2%	69.1%
Scenario 2	481	1.5%	35.6%	63.0%
Scenario 3	477	0.8%	34.8%	64.4%
Scenario 4	479	0.6%	34.7%	64.7%
Scenario 5	503	1.0%	34.2%	64.8%
Total (avg.)	492.2	1.1%	33.7%	65.2%
Practice - Year 2016 (avg.)	460.9	1.7%	33.0%	65.3%

Table 3.12 Distribution of patients amongst the five largest patient groups for the scenarios created by the scenario generator, corresponding average, and the average measured in practice throughout 2016.

Sample	Bone met.	Head&Neck	Lung>44Gy	Mamma	Prostate
Scenario 1	23.8%	4.6%	5.2%	15.9%	4.0%
Scenario 2	26.4%	3.3%	5.8%	16.0%	2.7%
Scenario 3	25.2%	3.6%	6.9%	15.5%	3.6%
Scenario 4	26.3%	4.4%	5.8%	15.0%	2.7%
Scenario 5	23.7%	5.0%	6.4%	14.5%	4.8%
Total (avg.)	25.1%	4.2%	6.0%	15.4%	3.6%
Practice - 2016 (avg.)	24.9%	4.2%	5.6%	15.2%	3.7%

low 2%, which means that output solution is very close to optimal. In terms of computational time, we can see that there is some variability between the optimization problems associated with the different stages. Stage 3 took around one minute to be completed, which may be a result of receiving the final solution from stage 2 as a warm MILP start. The computational experiments for generating the RTT allocation solution using 5 scenarios took a total wall-clock time of approximately 33 minutes. As the proposed model is designed to be used on a tactical (mid-term) level and the time to achieve a final solution can in practice go up to several hours, this computational time is considered acceptable. The RTT allocation solution output by running the solution generation approach with the described data-set can be found in Table 3.14.

Table 3.13 Solver information for the solution generation.

Stage	Wall-clock time (min)	CPU total time (min)	Objective value	Opt. gap
1	9.5	24.4	787.2	3.41%
2	22.4	43.8	772.4	1.27%
3	1.1	1.5	760.2	1.89%

Table 3.14 Results of the allocation of RTTs for the NKI case study.

Period	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
Weekday	Tu	We	Th	Fr	Mo	Tu	We	Th	Fr	Mo	Tu	We	Th	Fr	Mo	Tu	We	Th	Fr	Mo	Tu	We
CT	4	4	4	4	2	4	4	4	4	2	4	4	4	4	2	4	4	4	4	2	4	4
MRI	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
PET-CT	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0
IPP	1	1	1	2	1	1	1	1	2	1	1	1	1	1	2	1	1	1	1	1	1	1
BSU	1	1	1	2	2	2	2	1	2	1	1	2	2	2	1	1	1	2	1	2	2	1
TP	12	10	7	8	6	11	7	12	15	10	15	15	12	10	10	13	11	13	13	11	13	10
LINACs	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36
Total	55	53	51	54	49	55	51	56	61	52	58	59	57	55	53	56	54	58	57	54	57	53
Available	68	65	68	71	68	69	64	68	68	71	71	65	71	71	71	69	65	71	68	65	67	66
Difference	13	12	17	17	19	14	13	12	7	19	13	6	14	16	18	13	11	13	11	11	10	13

3.4.4 Solution evaluation

To evaluate the generated RTT solution, we re-run the solution framework depicted in Figure 3.2 for a new set of test scenarios, but restricting the values of decision variables X_{rit} and L_{rt} to the RTT allocation solution found in the solution generation run. By reducing the number decision variables associated with the MILP model, we were able to run the method for a total of 10 test scenarios without exceeding the CPU limit time of 14400 seconds per stage with a reduced maximum optimality gap of 1%. The test scenarios used for the evaluation of the RTT allocation solution are presented in Figure 3.4. The number of total patient arrivals varies from 461 (scen. 8) to 527 (scen. 3), with an average of 490.6. Validation steps were performed for the remaining patient data as in Section 3.4.3, reaching similar conclusions. The remaining parameters were kept the same as in the solution generation experiments.

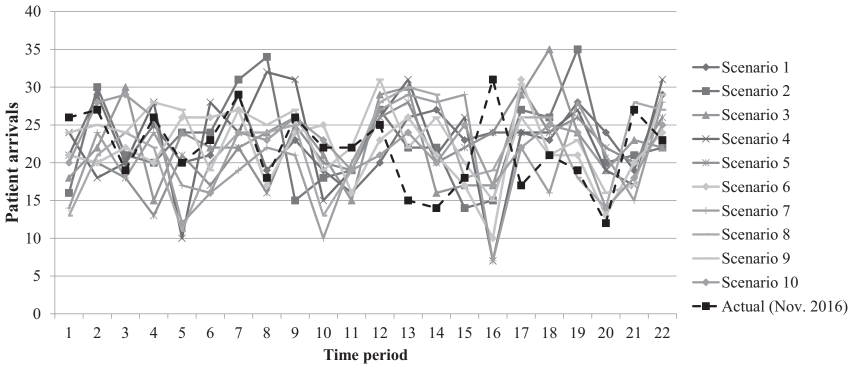


Figure 3.4 Patient arrivals, per time period, for the scenarios created by the scenario generator (solid lines) for evaluating the RTT allocation solution, and patient arrivals measured in practice during November 2016 (dashed line).

3.5 Results and discussion

Table 3.15 shows the results, regarding patient throughput, obtained by running the solution approach with the (fixed) RTT allocation solution for the 10 new generated scenarios (Figure 3.4), and the performance measured in practice both as the average throughout 2016 and the month of November 2016. The experiment results refer to the performance during the month of November, i.e. the first 22 time periods of the planning horizon. In the evaluation run, the optimality gap was found at 0.64%, 0.76%, and 0.96% for stages 1, 2, and 3, respectively. Analyzing the results of Table 3.15, we can see that with our model significant improvements in terms of timeliness can be achieved. The average number of patients completing the pre-treatment stage (458.0) is higher than both the records from November of 2016 and the average throughout 2016. Yet, the average percentage of patients completing the pre-treatment stage within the maximum waiting time target (100%-target) increases from 90.0% to 99.1%. Comparing with the performance in the clinic over the same one-month period, we can see that similar improvements could be achieved, with the percentage of patients breaching the main target going down from 9.1% to approximately 0.9%. In fact, results show that, with our solution, at most 11 patients would breach the 100%-target (scenario 7) for the used data set, while in scenario 8 all patients would be able to start the treatment within the recommended time standards. Regarding the 80%-target, potential improvements are similar. The NVRO recommends that at least 80% of subacute and regular patients start treatment within 7 and 21 days, respectively. Although the clinic was already fulfilling these standards, performance regarding these targets can potentially increase from 82.4% to 98.9%, on average. Figure 3.5 shows the distribution of waiting times amongst all patients. The circles mark the average waiting times, the dark bars indicate the standard deviation, and the grey bars represent the minimum/maximum waiting times. Even though the average waiting times from our models are similar to the ones measured in practice, the maximum waiting times are significantly lower using our approach. In fact, the maximum waiting time can decrease from 25 workdays (Nov. 2016) to, at most, 17 workdays (scenarios 6 and 7) by using our model. The standard deviation in each scenario is also slightly smaller when comparing to practice.

Figures 3.6, 3.7, and 3.8 show the results regarding the fulfillment of the waiting time standards for acute, subacute and regular patients, respectively. As shown in Figure 3.6, acute patients would be able to receive treatment within a day in at least 66.7% of the times (scenario 7), with the average over all scenarios (91.9%) outperforming the average of around 70.8% verified in practice in 2016 and the percentage of about 50.0% measured during November 2016. Similarly, subacute patients (Figure 3.7) can benefit significantly from our model according to the NKI case study results. In 2016, the department had approximately 91.3% of patients within this category finishing pre-treatment within the maximum waiting time target, while the proposed solution approach performed at an av-

Table 3.15 Comparison of the theoretical results per scenario and the corresponding average, with the performance measured in practice (November 2016 and the average throughout 2016).

Sample	Patients completed (z_1)	Within 100-target (z_2)		Within 80-target (z_3)	
Scenario 1	459	455	99.1%	455	99.1%
Scenario 2	460	459	99.8%	459	99.8%
Scenario 3	465	457	98.3%	456	98.1%
Scenario 4	460	459	99.8%	458	99.6%
Scenario 5	459	452	98.5%	450	98.0%
Scenario 6	465	459	98.7%	458	98.5%
Scenario 7	478	467	97.7%	465	97.3%
Scenario 8	442	442	100.0%	441	99.8%
Scenario 9	436	434	99.5%	434	99.5%
Scenario 10	456	454	99.6%	454	99.6%
Total (avg.)	458.0	453.8	99.1%	453.0	98.9%
Practice - Nov. 2016	449	408	90.9%	360	80.2%
Practice - Year 2016 (avg.)	443.1	398.9	90.0%	365.2	82.4%

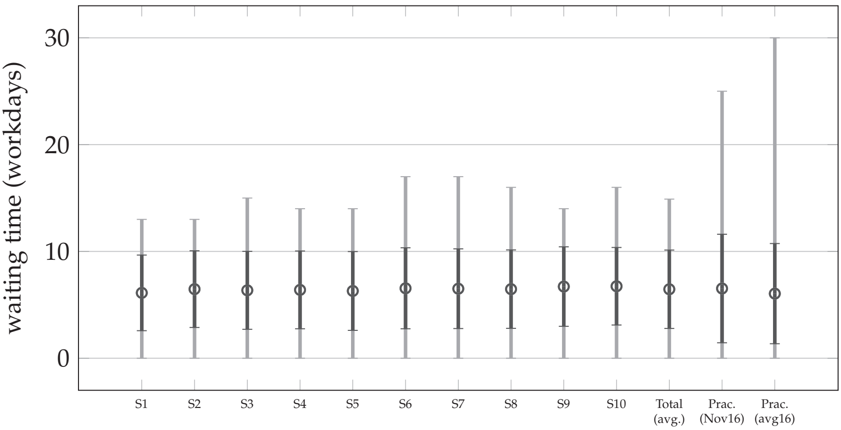


Figure 3.5 Distribution of average waiting times (in workdays) and corresponding standard deviation and maximum/minimum values of the computational results and the samples measured in practice (November 2016 and the average throughout 2016).

erage of 97.9% amongst scenarios, with a worst-case scenario of 95.0% (scenario 7) fulfillment of the 10 calendar days target. For the 80%-target, the performance can potentially grow from 77.9% to 97.6%. As for regular patients (Figure 3.8), although the results show that improvements can be achieved, they are not as impacting as for subacute patients. This is because the clinic is performing best for this urgency category, with 96.3% of the patients starting treatment within the 28 days target, and 91.0% of them starting treatment within 21 days, on average, in 2016. Still, the number of patients who are able to start treatment within the recommended maximum waiting time can increase to 100.0% for regular patients, while the percentage within the 80%-target can still rise up to 99.9%.

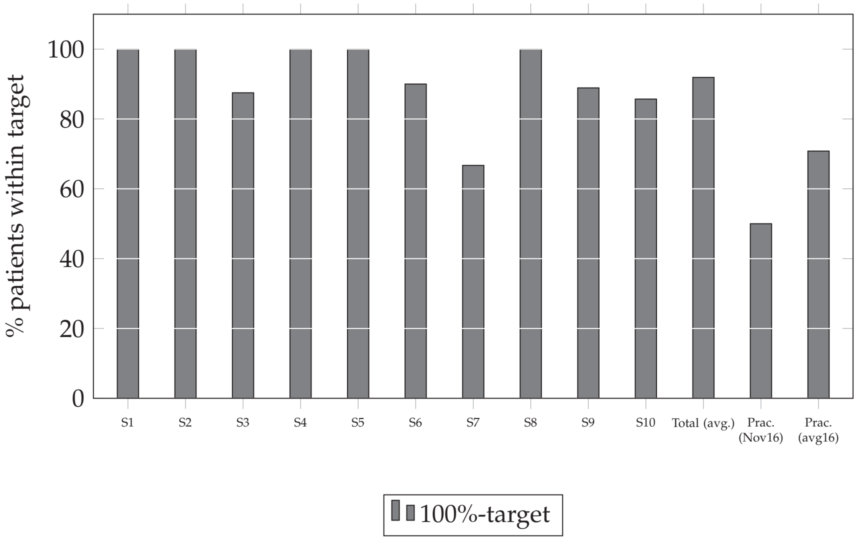


Figure 3.6 Comparison of the theoretical results per scenario and the corresponding average, with the performance measured in practice (November 2016 and the average throughout 2016) in terms of target satisfaction for acute patients.

The number of RTTs allocated to each operation in the final solution and the total RTT availability in the clinic are presented in Table 3.14. We can observe that the number of staff members allocated to imaging machines and linacs on each day is as low as the minimum number of people required to operate them. However, the number of people assigned to non-continuous operations is not as straightforward. For instance, there are considerable variations in the number of people needed for TP amongst workdays, ranging from 6 (period 5) to 15 (periods 9, 11, and 12). In practice, the number of people assigned to work on TP is often more stable, which leads to situations of understaffing and overstaffing when peaks and valleys in workload occur. A similar situation may be encountered, to a lesser extent, for IPP and BSU (range between 1 and 2), which is in accordance with the number of RTTs working daily on those operations at the NKI. We believe that such a dynamic scheme would be beneficial to RT centers. In addition, we observe that, according to the output solution given by the model, the number of RTTs assigned to the modelled operations could be reduced by 6 and there would still be enough staff members to cover the demand in all time periods. This represents a possible reduction of 5.5% in RTT capacity.

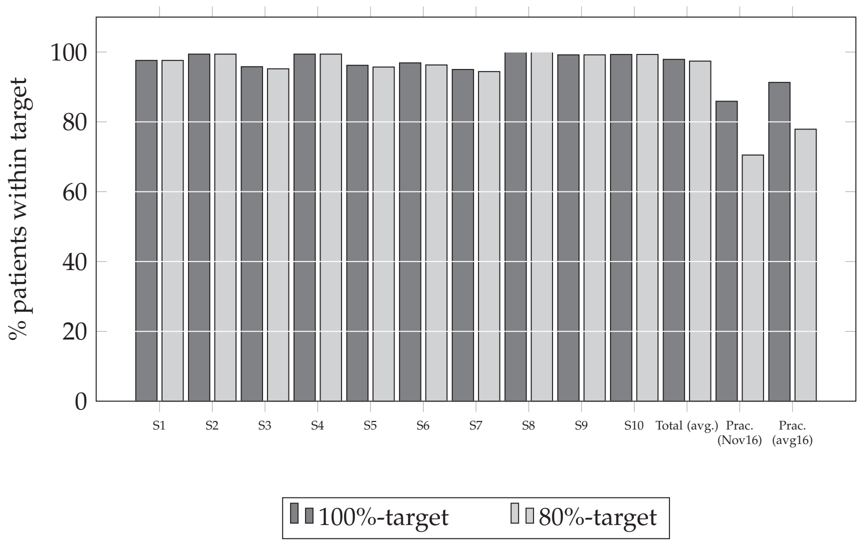


Figure 3.7 Comparison of the theoretical results per scenario and the corresponding average, with the performance measured in practice (November 2016 and the average throughout 2016) in terms of target satisfaction for subacute patients.

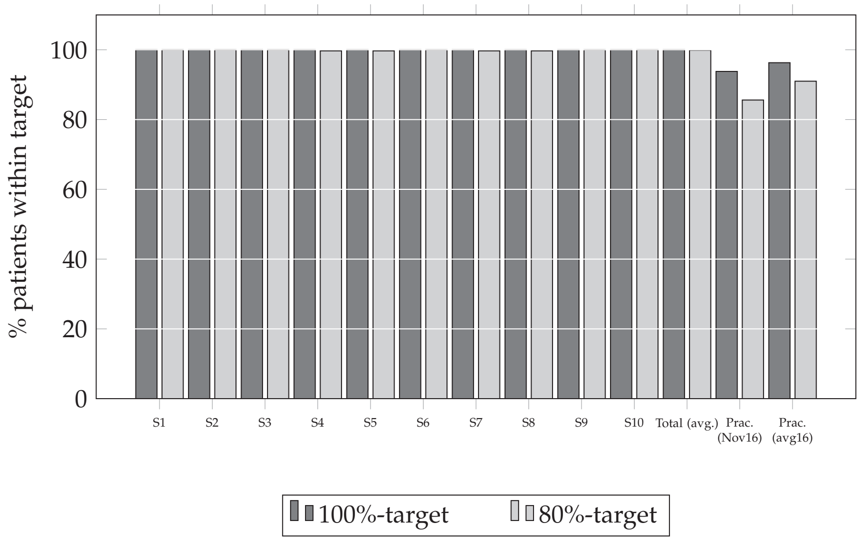


Figure 3.8 Comparison of the theoretical results per scenario and the corresponding average, with the performance measured in practice (November 2016 and the average throughout 2016) in terms of target satisfaction for regular patients.

3.5.1 Rotation needs

Rotation requirements for RTTs usually need to be satisfied on a yearly basis. However, if the proposed methodology is to be used in a systematic way with shorter planning horizons (e.g. one month), then the constraints regarding RTT rotation needs ((18)-(19)) have to be considered. For instance, the resource allocation provided by our framework resulted in RTTs 53 and 80 not being assigned to CT scanning in the whole period, although they are capable of operating CT scanners. Similarly, RTTs 23 and 30 are not allocated to work on MRI scanning in any day of the month even though they are eligible to do so. However, according to the managers' input, each RTT has to work, on average, one day per month in each operation he/she can perform. To test the effect of including the constraints regarding training needs of RTTs in the model, we ran the same methodology using exactly the same input data, training scenarios (5), and tolerances of the experiments described in Sections 3.4.2 and 3.4.3, but integrating constraints (18)-(19) into the MILP models of stages 1, 2, and 3 of the solution approach. With the inclusion of constraints (18)-(19), the total CPU time increased slightly from 69.7 to 72.0 minutes, and the optimality gap values were 1.47%, 0.19%, and 2.59% for stages 1, 2, and 3, respectively. We then fixed the (new) RTT allocation and re-ran the experiments using the same 10 test scenarios as presented in Section 3.4.4. Results are presented in Table 3.16. Comparing these with the results for the base case (Table 3.15) we observe that, although the number of patients finishing pre-treatment and the waiting times have slightly worsen as a whole, the inclusion of constraints (18)-(19) does not decrease performance regarding the achievement of the waiting time targets. The fulfillment of the 100-target has actually increased from 99.1% to 99.2%, while the percentage of patients able to receive the first fraction within the 80%-target has increased from 98.9% to 99.9%, on average. Similar results were obtained per urgency category. For the evaluation run, the optimality gap was 0.69%, 0.79%, and 0.54% for stages 1, 2, and 3, respectively.

3.5.2 Research limitations

Despite the promising results hereby presented, there are some research limitations of this study. First, the presence of the doctor may be needed during some operations, for example, to oversee a CT scan with intravenous contrast in case complications arise, or during TP to assist RTTs performing complicated treatment plans. However, the doctor may not be available when needed, delaying the execution of these operations. Second, additional imaging scans may be required before or during treatment, and treatment planning may be required to be re-done after the verification performed by the doctors and the medical physics unit. Nevertheless, both of these occurrences are not common in practice (less than 1%). Third, the conducted experiments lead to a rather strict roster that, by not considering overtime or the possibility of working only half days, may not fully represent the actual practice of the clinic. In addition, the model

Table 3.16 Results for the NKI case study including RTT training needs constraints.

Sample	Patients completed (z_1)	Within 100-target (z_2)		Within 80-target (z_3)	
Scenario 1	449	446	99.3%	446	99.3%
Scenario 2	441	440	99.8%	440	99.8%
Scenario 3	447	440	98.4%	439	98.2%
Scenario 4	453	453	100.0%	453	100.0%
Scenario 5	461	456	98.9%	454	98.5%
Scenario 6	461	456	98.9%	454	98.5%
Scenario 7	463	453	97.8%	452	97.6%
Scenario 8	419	417	99.5%	415	99.0%
Scenario 9	425	422	99.3%	422	99.3%
Scenario 10	440	438	99.5%	438	99.5%
Total (avg.)	445.9	442.1	99.2%	441.3	99.0%
Practice - Nov. 2016	449	408	90.9%	360	80.2%
Practice - Year 2016 (avg.)	443.1	398.9	90.0%	365.2	82.4%

does not take into account RTTs that may start or finish a long-term sick leave during the planning horizon. Another limitation relates to the uncertainty inherent to the time needed for treatment planning. Although these processing times per technique are usually standardized, for some complicated tumors the treatment optimization may take up to several days to be completed. Finally, unavailability of imaging machines due to breakdowns or maintenance actions are also not considered, although downtime rates verified in practice are also low (around 1%).

3.6 Conclusions and further research

This research proposes a stochastic MILP-based approach for RTT allocation to improve timeliness in the delivery of the RT treatment while taking into account the stochasticity inherent to the RT process. With our model, allocation decisions cover the demand represented by a sufficient set of scenarios of patient inflow and care content. The proposed approach optimizes the number of patients finishing the pre-treatment stage before the maximum waiting time target for two different quality standards in a hierarchical manner.

Results for a case study using real data from the NKI, a large cancer center in the Netherlands, over an evaluation period of one month show that an adequate allocation of RTTs to multiple operations in RT can considerably reduce waiting times. The number of patients finishing the pre-treatment stage within the primary target (100%-target) increases from 90.0% to 99.1%, on average, while the fulfillment of the secondary target (80%-target) rises from 82.4% up to 98.9%. Per urgency level, the average percentage of acute patients able to receive the first fraction in a timely manner increases from 70.8% to 91.9%, while for subacute patients this percentage can increase from 91.3% to 97.9%. The proportion of regular patients starting treatment within their recommended maximum waiting time target can grow from 96.3% to 100.0%. In addition, the RTT capacity needed to cover the expected demand in the modeled operations can be reduced by 5.5%.

The obtained results suggest that the proposed model can be useful for RT managers in practice. Besides giving decision support for the allocation of RTTs, our model can also provide insights in finding bottlenecks, slack/surplus of RTT capacity, and lack of specific RTT skills. Moreover, the model can be used to provide an initial monthly schedule upon which demand-driven daily adjustments can be performed. By achieving a (near) optimal solution in just over an hour of CPU time, the MILP program solves in reasonable computational time. The RTT allocation solution has been validated by the managers of the RT department of the NKI and the results were considered very positive. The possible implementation of the model is under consideration in the department. Furthermore, as the modeled processes and resources are standard among RT centers and the case mix at the NKI is representative of the case mix in RT, the proposed methodology can be applied to other RT centers.

CHAPTER 4

Improving workflow control in radiotherapy using discrete-event simulation

4.1 Background

Radiotherapy (RT) is a therapy modality for cancer treatment that requires several preparation steps consisting of imaging and treatment planning. RT resources are expensive and limited in capacity, and treatments are prepared and delivered by a multidisciplinary group of specialists with multiple activities and limited time availability [111]. As demand for RT continues to grow [92] and cancer treatments become more personalized [19], ensuring a timely delivery of RT for each patient trajectory without jeopardizing the timeliness of the other patients is not straightforward. Earlier research has shown that the dynamic nature of treatment scheduling in RT, in which scheduled and non-scheduled patients have to be queued up for undergoing pre-treatment, can considerably impact access times for RT [26, 35]. Long waiting times have been associated with negative clinical outcomes such as higher risk of local recurrence [20], increased tumor progression [64] and prolonged psychological distress in patients [63]. In fact, the unavailability of medical staff was pointed out as one of the main causes for this [79]. Related to this, Hutton *et al.* found that RT professionals in the UK are prone to the effects of compassion fatigue and burnout and that special attention must be paid to workload and its impact on practitioners' job satisfaction [45].

The RT treatment process starts with referral, followed by a consultation with a radiation oncologist, who prescribes the necessary steps needed (referred to as "pre-treatment workflow") before the treatment starts. The pre-treatment workflow includes imaging (CT, MRI, PET-CT), contouring of the tumor and organs-at-risk, and treatment planning, and is commonly driven by the scheduling of the first irradiation session, which is usually set immediately after consultation. This demands pre-treatment workflow to be programmed *a priori* before the scheduled starting date for treatment. We refer to this strategy as the "pull" strategy [26], a term derived from logistics and supply chain management where manufacturing is driven by customer demand and resources are expected to be available at each operation when needed for just-in-time production. In RT, a pull strategy foresees that a date for the start of treatment is set right after consultation, and the scheduling of pre-treatment workflow is performed in a "backwards" fashion, ensuring that the necessary rooms and staff will be available when needed to meet timeliness targets. However, for some patient types, the first irradiation is scheduled after (some) the pre-treatment steps have been completed, typically at the start or at the end of treatment planning This is referred to as "push" strategy, which in logistics terms refers to a continuous flow of products throughout the system, with no specific due date, typically leading to store inventory. By applying a push strategy in radiotherapy flexibility to perform pre-treatment activities and consequently a low number of first linac appointment rebooks can be expected. However, setting a treatment start date right after consultation (pull strategy) may lead to increased patient and staff (doctors) satisfaction, particularly when time slots for doctors' activities (e.g. contouring of the tumor) are pre-allocated in coordination with treatment scheduling decisions. It may also increase control over the work in progress, leading to a reduced number of patients breaching the waiting time targets. Therefore, appropriate workflow management systems (e.g. scheduling routines) and the design of efficient resource planning schemes are crucial to meet the intended waiting time targets [68] while ensuring patient centeredness and quality of labor.

Operations research (OR) methods have been successfully used to support decision-making in healthcare in general [83], and increasingly in radiotherapy [107]. Among OR methods, discrete-event simulation (DES) stands out as a powerful tool to find logistical interventions for performance improvement by modeling the behavior of complex systems as a series of discrete events occurring over time [58]. DES has been proven useful in testing operational changes in several healthcare settings [48], such as analyzing optimal discharge rates in acute care [25], capacity management and patient scheduling in outpatient clinics [2], and decreasing throughput times for CT scanning in radiology departments [101, 102]. In the field of radiotherapy, a few DES studies have been conducted for process improvement and resource planning. Kapamara *et al.* [54] performed a patient flow simulation analysis to find bottlenecks in the Arden Cancer Center, UK, to reduce waiting times and maximize patient throughput. The authors were able to model three treatment modalities (conventional external-beam, brachytherapy, and unsealed sources therapy), and found that an extension of clinical shift hours reduces patients' waiting times by 2%. Proctor *et al.* [81] modeled patient care pathways from arrival to discharge to estimate the impact of increased levels of demand in the performance of the department of RT of the Walsgrave hospital, UK. They reported that reducing the percentage of patients seeing their own doctor on the simulator from 71% to 35% and extending the linacs' operating hours by 38% would provide the best performance, with 82% of the patients starting treatment within the desired target. Werker *et al.* [110] used DES as an attempt to improve the RT planning process of the RT center of the British Columbian Cancer Agency in Canada, finding that reducing delays associated with the oncologists' tasks would reduce the planning times by 20%. Babashov *et al*. [3] included the treatment stage of the RT trajectory, thus modeling the process from patient arrival to treatment completion. They found that adding one more full-time oncologist would reduce the waiting times by 6.55%, leading to around 85% of the patients starting treatment within 14 calendar days. Crop *et al*. [26] studied an alternative workflow control system for robotic stereotactic RT by testing a constant work-in-progress system that only allows new patients to start pre-treatment when a patient leaves the system, in an attempt to keep workload constant. Results showed that a hybrid constant work-in-progress workflow could potentially increase the number of irradiation sessions per day by 32%, while the time between CT and start of treatment remained stable at an average of 9 days.

Computer simulation studies of RT are available but mainly focus on finding operational improvements by re-dimensioning workforce, expanding machine capacity/availability, or extending clinical opening times, whilst the impact of implementing alternative scheduling routines and different workflow control systems are rarely found. In this work, we model the RT pre-treatment workflow using DES to quantify the operational impact of using pull and push strategies in RT scheduling. As a secondary goal, we try to find interventions (e.g. increase treatment planning capacity) that maximize the number of patients starting treatment within the intended targets and allow for minimal waiting times.

4.2 Methods

We used DES modeling to construct a model on the flow of patients receiving external-beam RT in the Netherlands Cancer Institute (NKI) from consultation to the start of treatment (first fraction). The model was built using Tecnomatix Siemens Plant Simulation 13.2 by Siemens PLM Software [4]. After the model was validated, we studied the impact of increasing the number of pull patients starting from the baseline case representing the current practice (40% pull / 60% push), as well as other possible interventions for performance improvement.

4.2.1 The RT treatment workflow in the NKI

Figure 4.1 depicts the RT workflow in the NKI. Upon referral, patients are scheduled for a consultation (Moment 1) with a radiation oncologist, who becomes responsible for monitoring the patient's care trajectory. At consultation, the doctor meets the patient and assesses all the information needed to plan an RT treatment. After consultation, the doctor fills in a form (PlanRT) with the medical information and sets up a preliminary treatment plan outlining the care pathway intended for the patient. The pre-treatment workflow starts after consultation, when patients are scheduled for a CT scan, but a delay before pre-treatment starts, due to other appointments (e.g. IV-contrast, blood analysis, manufacturing of patient-specific aids such as masks etc.) may be needed, as well as additional imaging examinations (MRI and PET-CT). In case a 4DCT has been taken, imaging motion compensation is needed (warping). If multiple imaging scans are involved, then the registration of the different datasets is also necessary (image registration). Thereafter, the doctor delineates the target area (contouring), right before treatment planning. At this step, beam set-up (simplified treatment planning such as the two-field technique "anterior-posterior-posterior-anterior") may be done instead or in conjunction with regular treatment planning. Once treatment planning is finished, the generated plan is uploaded to the corresponding linac and the treatment can start. The modeled pre-treatment workflow, indicated by the black bounding box in Figure 4.1, starts right after consultation (PlanRT) and ends at the start of treatment. The time needed to complete the pre-treatment phase is referred to as "waiting time" in this study.

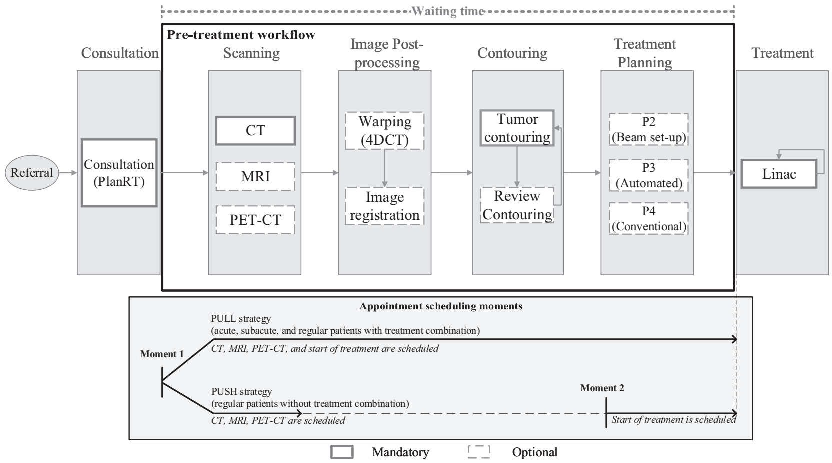


Figure 4.1 Flowchart of the complete RT treatment workflow in the NKI.

Regarding the appointment scheduling process, Figure 4.1 shows that upon submission of the PlanRT sheet after consultation, an appointment officer schedules all the necessary imaging scans for all patients. This moment in time is represented by "Moment 1" in Figure 4.1. At Moment 1, acute patients, subacute patients, and regular (i.e. non-urgent) patients who have a combination of RT with other treatment modality (surgery or chemotherapy) are also scheduled for all irradiation sessions right after consultation. We refer to these as "pull" patients. Acute and subacute patients are scheduled in a pull manner as a timely start of treatment needs to be ensured due to the urgency of their treatment. Regular patients with a treatment combination between RT and other treatment modality (e.g. chemotherapy or surgery) also need to be scheduled right at consultation. For these patients, a proper time coordination between irradiation sessions and the other treatment modality is necessary to maximize the effectiveness of the combined treatment. For pull patients, pre-treatment activities need to be given enough time to be completed before the pre-scheduled starting date to avoid linac sessions' rebooks. Alternatively, regular patients without a combination of treatment modalities, indicated as "push" patients in this study, are scheduled for the start of treatment only once contouring has been done and treatment planning has started, as indicated by Moment 2 in Figure 4.1.

4.2.2 Model inputs

In DES, a number of inputs are needed to generate events (e.g. patient arrivals, processing times, resource availability) that represent the behavior of the real system. In our model, we used historical data from the whole year 2017 (January 01 to December 31) as model inputs to (randomly) generate those events. To obtain data that was not available in the internal databases, we conducted several interviews with radiation oncologists, radiation therapy technologists (RTTs), managers and appointment schedulers to estimate the most realistic values for each input parameter. Table 4.1 presents an overview of all input parameters of our DES model.

4.2.3 Model development

The modeled steps, scheduling routines and their relationship with the input parameters are depicted in Figure 4.2. The specific workflow and data contained in each component are explained in more detail throughout this section. Patient arrivals are generated using records of PlanRT form creation dates (after consultation), followed by the creation of patient care content according to the probability distributions mentioned in Table 4.1. At this point, push patients will be scheduled the necessary imaging scans, and will proceed to the pre-treatment workflow CT/MRI/PET-CT/IPP, contouring and treatment planning. Pull patients will also be scheduled the start of treatment before following the same route. The start of treatment of push patients is then scheduled at treatment planning. "Resource availability" and "processing times" contain the logistics data used in the scanning, contouring, image post-processing, and treatment planning steps.

Patient arrivals

We used historical data from the year 2017 to determine probability distributions for the arrival processes in the NKI, which are used in the DES model to generate patient arrivals. We considered the historical records of all PlanRT forms filled in by the doctors after consultation as patient arrivals, excluding weekends and public holidays. In total, we included 4973 patient care pathways recorded in 2017 for external-beam RT treatments. Earlier research has shown that there were statistically significant differences in the patient arrivals

Table 4.1 Input parameters of the DES model.

Name	Description	Probability Distribution	Dependencies
Patient arrivals	Patient arrival rates per weekday, per tumor site (8 independent generators)	Poisson	-
Care plan	Proportion of patients in each of the 62 possible care trajectories depending on tumor site (generators)	Empirical	Tumor site
Urgency level	Proportion of acute, subacute, and regular per care plan	Empirical	Care plan
Steps needed	Proportion of patients with CT, MRI, PET-CT, warping, image registration, contouring, and treatment planning type, per care plan, per urgency level	Empirical	Care plan, Urgency level
CT/MRI/PET-CT processing times	CT = 25 min. MRI = 45 min, and PET-CT = 45 min, regardless of other parameters.	-	-
Image Post-processing times	Mean and standard deviation of the duration for processing warping and image registration.	Lognormal	-
Contouring time	60 min for tumor contouring, and 60 min for peer-review review.	-	-
Treatment planning times	Processing times of P2, P3, and P4, depending on the care plan.	-	Care plan
Scheduling of first fraction	Proportion of patients for each possible duration of the time-to-treatment (0…21 days) per urgency level, per weekday	Empirical	Weekday, Urgency level
Planned delay	Proportion of patients with a planned delay before pre-treatment, and the length of the delay (1…8 weeks), per care plan	Empirical	Care plan
Machine availability	Time of the day each CT, MRI, and PET-CT is available to be operated	-	-
Doctors' agenda	Start time and end time for each day of the simulation period, and parts of the day the doctor is unavailable due to other scheduled activities (meetings, research, etc.)	-	-
RTTs' agenda	Start time and end time for each day of the simulation period	-	-
Public holidays and days-off	Days of the simulation period in which the clinic is not operating, and days each RTT and doctor is unavailable (days-off)	-	-

between workdays, and that patient arrival patterns follow a Poisson distribution in each workday [105]. An updated ANOVA analysis with the 2017 data using the probability-distribution fitting software EasyFit [66] resulted in the same conclusions (Table 4.2), i.e. patient arrivals were found to follow a Poisson

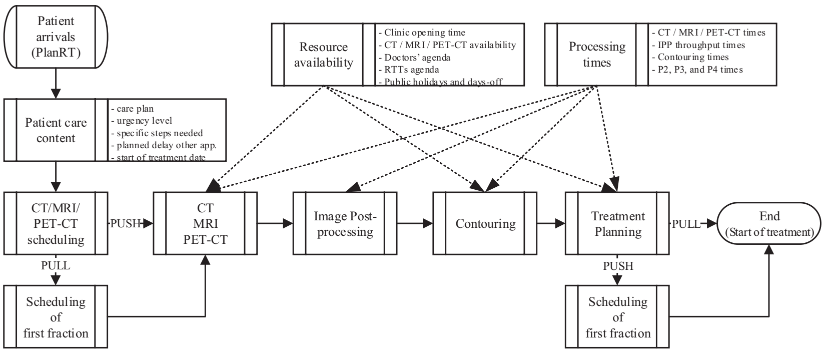


Figure 4.2 Flowchart of the complete RT treatment workflow in the NKI.

Table 4.2 Patient arrival statistical analysis for the 2017 data.

Weekday	Sample Size	Prob. Dist.	Mean (SD)	p-value
Monday	859	Poisson	17.5 (4.7)	0.72
Tuesday	1067	Poisson	20.9 (5.7)	0.24
Wednesday	1208	Poisson	23.2 (6.7)	0.61
Thursday	1063	Poisson	21.7 (5.9)	0.51
Friday	776	Poisson	15.5 (5.4)	0.25

distribution for every weekday.

In the NKI, patients are assigned one of eight possible tumor sites upon referral: Bone metastasis, Breast, Lung, Brain, Prostate, Head-and-neck, Chest wall, or Others, as depicted in Figure 4.3. Each tumor site has a different consultation pattern over the week. For instance, consultations for (regular) lung patients are mostly held on Wednesday mornings. Therefore, we generated patient arrivals in the model by using the mean arrival rate per tumor site, per weekday, according to a Poisson distribution (Table 4.2) and using the proportions presented in Figure 4.3.

Patient care content

The attributes of each patient (care plan, urgency level, specific steps needed, planned delay before pre-treatment, and start of treatment date) were randomly assigned based on the historical breakdown measured in 2017. After consultation, the doctor selects one of 62 possible trajectories for the patient, which depends on the tumor site for that patient (see Appendix 4.6). For instance, a lung patient may be assigned the palliative trajectory, or the regular trajectory, which would yield a different care pathway. The care trajectory defines whether a patient would require MRI (18.5% of the population), PET-CT (3.9%), Warping (12.4%), Image registration (29.7%) or Beam set-up (34.7%). All patients require

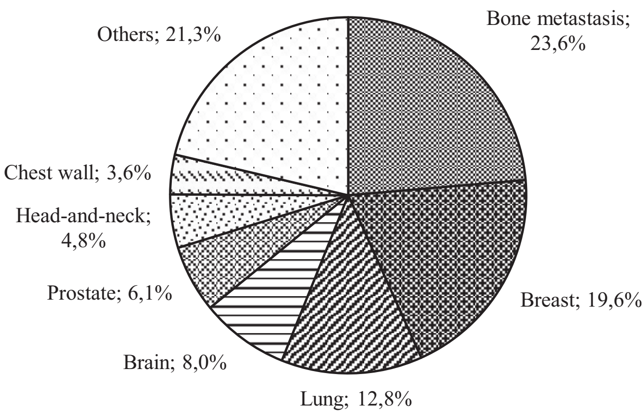


Figure 4.3 Distribution of patients by tumor site in 2017.

a CT, contouring, and treatment planning. The urgency level indicating whether a patient is acute (1.3% of the patient population), subacute (30.8%), or regular (67.9%) was generated based on the historical proportions verified for the corresponding trajectory. Moreover, measured data shows that 650 out of the 4973 patients (13%) have a planned delay before starting pre-treatment (CT) due to medical reasons (e.g. RT after surgery, dentist) or patient preferences (e.g. holidays), the delay ranging between 1 and 8 weeks. In the 2017 data, we found that 40.8% of the patients were scheduled in a pull fashion (SD = 5.8%), while the remainder 59.2% were scheduled using on a push fashion. Empirical distributions using the above-mentioned proportions were used to create patient care content in each replication of each computational experiment.

CT/MRI/PET-CT scheduling

Scheduling of scanning appointments in imaging rooms are assigned on a first-come-first-planned basis, except for some appointments in CT scanners, where a pre-allocation of specific time slots exists. For instance, the first two time slots in the morning cannot be assigned to patients who need IV-contrast before the CT, as the corresponding doctor must be present in the department but may not have started his/her shift before 08h30. Similarly, there is one time slot exclusively available for acute patients per day.

Contouring

Doctors are grouped in teams based on their specialty: Breast, Lung, Urology, Head-and-neck, Gynecology, Gastrointestinal tract, and Central nervous system. Table 4.3 presents the total number of doctors per specialty. Depending on the specific tumor site, a doctor belonging to the corresponding specialty is

Table 4.3 Doctor teams and corresponding number of elements in the NKI during 2017.

Specialty	No. doctors
Lung	7
Head-and-neck	9
Breast	9
Central nervous system	3
Gynecology	4
Gastrointestinal tract	5
Urology	7
Palliative	All (44)

assigned to the patient using empirical distributions from the 2017 data. Contouring of palliative patients (acute and bone metastasis), accounting for 815 of the 4973 patients, can be undertaken by any available doctor right after scanning. Pending contouring activities waiting in queues are sorted on an Earliest Due Date (EDD) basis, giving priority to the patients with the earliest date for start of treatment. For push patients, who have not been scheduled at this point, we considered the target date for start of treatment according to the national targets.

Treatment planning

Treatment planning is divided in three types: P2, P3, and P4. There used to be a P1 type that does not currently exist in the NKI. P2, also referred to as beam set-up, is a simpler form of planning mostly undertaken for bone metastasis and some breast cancer patients. P3 is a form of automated planning in which a computer software performs the planning autonomously. P4 is the conventional treatment planning modality, in which beam angles and intensities are iteratively optimized with the help of a computer software. P3 is immediately assigned to all breast, rectum, and prostate patients, as the planning of these tumor sites was automated in 2017. P4 will be assigned to all patients belonging to the other patient groups who have not been assigned P2 or P3. The assignment of P2 is modeled by means of empirical distributions that vary per care plan, i.e. the probability of a patient being assigned P2 varies depending on the care plan of that patient (see Appendix 4.6). For instance, 93% of all bone metastasis patients will have a P2 type of planning, while a head-and-neck patient will never be assigned P2, which means that he/she will always be assigned P4. Out of the 24 planning RTTs available, 3 hold a P2 level, 7 are skilled at level P3, and the remaining 10 are considered at level P4. P4 planners are also able to perform P3 and P2, and P3 planners can also perform P2. Moreover, P3 and P4 level planning RTTs can process 2 plans simultaneously. As with the previous step, treatment planning of acute patients and bone metastasis patients can be performed by any available planner right after scanning, and queued tasks are prioritized on an EDD basis.

Scheduling of first fraction

A statistical analysis showed that the time between arrival and the start of treatment do not follow any specific probability distribution with sufficient statistical significance (p-value > 0.05). Therefore, we used empirical distributions to randomly assign a date for start of treatment for both pull and push patients. For pull patients (40% of the total population), a treatment start date is generated based on the historical records upon first consultation. Since certain care plans have starting date requirements (e.g. head-and-neck patients must start on a Monday), we generated this time to treatment depending on the weekday of the request. This means that, for instance, a regular head-and-neck patient having the first consultation on a Tuesday will most likely be assigned a time to treatment of 6 or 13 days. According to the measured data, time to treatment of pull patients ranges between 0 and 1 day for acute patients, between 1 and 8 days for bone metastasis and subacute patients, and between 3 and 21 days for regular patients. Push patients (60%) are assigned a time between treatment planning and start of treatment that can range between 1 and 7 days, also generated on a weekday basis.

Resource availability

The RT department of the NKI operates from 07h30 to 17h30 on every weekday except public holidays. Staff members work 8 or 9-hour shifts (with breaks) while rooms and machines are available during the 10-hour period. The department has 2 CT scanners, 1 MRI scanner, and 1 PET-CT scanner. The PET-CT scanner is shared with the diagnostics department. In total, there are 26 time slots of 25 minutes available per day for CT scanning, 37 weekly slots of 45 minutes for MRI, and 5 weekly time slots of 45 minutes for PET-CT. As for staff members, the department hosts a total 113 RTTs (75 FTE), of which 24 can do treatment planning. In addition, there are 44 practitioners (26 FTE) in the department, which include radiation oncologists, residents and physician assistants. Their main duties include patient consultations, regular meetings (such as multi-disciplinary, RT treatment discussions, and research) and other administrative tasks. In the NKI, a doctor is available to perform contouring whenever he/she is not scheduled to do any of the pre-allocated tasks. Except for scheduled activities, the doctor gives priority to perform contouring over the other non-scheduled duties. The weekly schedule and absent days (incl. holidays, sick leave, conferences, training, etc.) of each staff member throughout 2017 have been used for the staff availability of our model.

Processing times

A CT scan has a time slot duration of 25 minutes, while an MRI and a PET-CT usually take approximately 45 minutes each. We included two possible tasks (warping and image registration) for IPP based on the historical records, which

Table 4.4 Statistical analysis of IPP tasks: processing times for both CT-Warping and Scanning-Image registration follow a lognormal distribution (p-value >0.05).

Time	Sample Size	Prob. Dist.	Mean (SD)	p-value
CT – Warping	608	Lognormal	0.4 (0.6)	0.35
Scanning-Image registration	1306	Lognormal	0.1 (1.0)	0.60

were found to follow a lognormal distribution with the mean and standard deviation presented in Table 4.4. If warping is needed for a patient, a delay corresponding to the time between CT and warping (CT-Warping) is generated. In case a patient needs multiple scans and thus has the need for image registration, we forced a delay respective to the time between the last scan (warping included) and image registration (Scanning-Image registration).

In the NKI, a contouring typically takes up to 30 minutes for acute and subacute patients, and one hour for regular patients to be completed according to the interviewed doctors. Moreover, each contouring needs to be peer-reviewed and approved by another doctor before the process moves on to treatment planning. In the NKI this step is done right after contouring, with the doctor in charge asking a colleague to double check the contouring on site. This extra step takes at most 60 minutes. Therefore, we have added 60 minutes to the processing time of each contouring to account for the peer-review task. Standard processing times for beam set-up and treatment planning vary considerably per care trajectory, ranging from 60 (e.g. bone metastasis) to 120 (e.g. breast) minutes for a beam set-up, and from 150 (e.g. prostate) to 960 (e.g. head-and-neck) minutes for treatment planning.

Model verification

The model was built iteratively in constant interaction with managers and clinicians from the RT department of the NKI. Components of the model as described in “model inputs”, such as patient arrivals generators, staff management tools, and processing units were added step by step after conducting interviews with the staff members of the NKI responsible for that step. The scheduling routines and simplifications introduced in each process were carefully discussed and approved by the manager in charge of the corresponding process.

Performance metrics

The most important Key Performance Indicators (KPIs) to evaluate the performance of our model are related to timeliness: the waiting times (in calendar days) and the percentage of patients breaching the waiting time targets. Maximum waiting time targets defined by the Dutch Society for Radiation Oncology (11) state that acute patients should be treated within one day, subacute patients should start treatment within 10 calendar days, and regular patients should start treatment within 28 days. In addition, we also look at the percentage of first fraction rebooks, i.e. the percentage of (pull) patients that have their start of treatment postponed as the pre-treatment phase cannot be completed in due time.

Warm-up period and number of replications

Since the model starts in an empty state with no queues and idle resources, we introduced a warm-up period by running the model for one-year data to assess the time needed for the resources to be occupied and the queues filled up. By measuring the evolution of patients' waiting times over time, the warm-up analysis showed that a steady state is achieved at around 130 days (see Figure 4.4). Therefore, during the 130 first simulation days of our computational experiments, output measurements are not included in the results. The 130-day warm-up period runs before the simulation run length of 365 days, which corresponds to the year 2017.

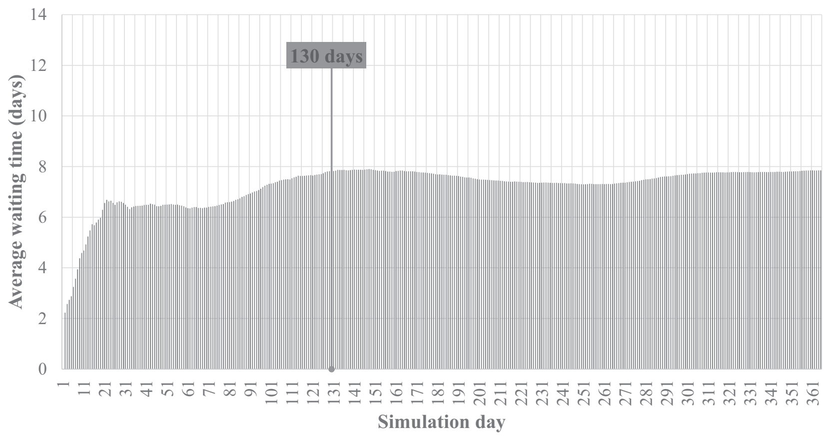


Figure 4.4 Distribution of patients by tumor site in 2017.

In order to find the proper number of replications, we performed several computational experiments with a different number of replications ($n = 2, 3, 4, ...$) until the relative error of the halfwidth of the confidence interval of the average waiting times ($\bar{x}$) measured across n was sufficiently small ($\gamma' < 0.05$), according to equation (4.1). Since the sample size (number of replications) is small and thus the real variance is unknown, we use a student's t-distribution to estimate the confidence interval of $\bar{x}$ for the corresponding number of replications n being tested. The halfwidth of the confidence interval is therefore obtained by $t_{n-1,1-\frac{\alpha}{2}}.\frac{s}{\sqrt{n}}$, with s being the variance of the waiting times for n replications, and $t_{n-1,1-\frac{\alpha}{2}}$ being the percentile of the Student-t distribution for $n-1$ degrees of freedom at $t_{1-\frac{\alpha}{2}}$ for a confidence level $(1-\alpha)$. In our experiments, since we consider a 95% confidence level, thus we set $\alpha = 0.05$.

$$\frac{t_{n-1,1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}}{x} < \gamma \tag{4.1}$$

By measuring the relative error according to the left-hand side of equation (4.1) for each replication number ($n = 2, 3, 4, ...$), we found that the relative error was smaller than $\gamma' = 0.05$ for $n = 15$ replications, with a relative error of 0.048. Therefore, we decided to run 15 replications of each computational experiment in our case study.

Workflow control analysis

To test the impact of increasing the number of patients being scheduled with a pull strategy starting from the baseline case, we gradually added subpopulations of patients based on tumor sites to the current pool of patients being scheduled with a pull strategy. The more complex the pre-treatment process of a patient is, the higher the uncertainty regarding the time needed to complete pre-treatment. Therefore, we started adding patients from the simplest to the most complex tumor types in terms of treatment preparation.

Scenario analysis

In conjunction with the workflow control analysis, we have investigated the impact of additional interventions that may lead to performance improvements in the NKI. The following scenarios were tested on the baseline case (i.e. with only 40% pull patients):

1. *Spreading consultation slots throughout the week*: We tested the impact of spreading the consultation time slots over the week by setting the same patient arrival mean on every weekday per care trajectory. The overall mean arrival rate, per care trajectory, remains constant.
2. *No pre-allocated time slots for CT*: We tested the impact of removing the pre-allocated slots from the CT tactical plan, by allowing full flexibility to schedule any patient in any available slot as they arrive.
3. *Balancing doctor availability for contouring*: We re-arranged the doctors' agenda such that each doctor is available for contouring for (at least) 2 hours a day, while working the same number of hours per week.
4. *P3 planners can process lung and chest wall patients*: We studied the influence of having P3 planners capable of performing treatment planning of lung and chest wall patients (16.4% increase), in addition to the current tumor sites (rectum, prostate, and breast).

Table 4.5 Comparison between the clinical performance and the DES model for validation purposes.

Performance metric	Actual system	DES model (95% conf. interval)
Waiting time (total)	7.9	7.8 (7.5, 8.1)
Waiting time (pull)	5.9	5.6 (5.4, 5.9)
Waiting time (push)	9.7	9.7 (9.4, 10.0)
No. patients breaching WT target	92	87.7 (68.1, 107.4)

5. *One more full-time P4 planner*: we evaluated the possible gain in waiting times by having one more planning RTT of level P4 (thus capable of performing P4, P3, and P2).

4.3 Results

For model validation, we have compared several outputs of the model for the baseline case with the clinical performance regarding the main KPIs that could be measured in practice for the year 2017 (Table 4.5). We verify that the total average waiting time (WT) output by the DES model (7.8 days) is very close to the one measured in the actual system, i.e. in the NKI practice (7.9), with the actual system value falling within the 95% confidence interval of the DES model. A similar behavior is observed for the pull and push patient trajectories, with pull patients having lower overall waiting times than average, as in current practice most of these patients are subacute. Regarding the timeliness target fulfillment, the model outputs an average of 85.13 patients breaching their targets, below the value observed in practice (92). Moreover, generated input data, including patient arrival histograms, care content, urgency level and process times, have been compared and found to be consistent with the historical data. The outcomes measured in the actual system and the output values obtained by the model were considered close enough to regard the DES model as a close representation of the actual system behavior, and therefore validated. The final DES model and corresponding outcomes therefore served as the baseline case for running the computational experiments previously described.

Figure 4.5 shows the effect of increasing the number of pull patients on the overall waiting times. The grey boxes indicate the 95% confidence interval of the average, while the whiskers represent the minimum and maximum values found over the 15 replications. Results show that with the increase of pull patients, the waiting times tend to slowly increase, ranging from 7.8 on the baseline case to an 8.9 maximum, when all patients are scheduled on a pull way. Nevertheless, the addition of some tumor sites like lung or prostate, to a pull strategy, do not impact waiting times considerably. Figure 4.6 shows the evolution of the number of patients breaching the national waiting time targets: 1 day for acute patients, 10 days for subacute, and 28 days for regular. Overall, the number of breaching patients tends to decrease with the use of a pull strategy. The average number of patients starting treatment after their due date goes down from 87.7

to 51.9, with the maximum topping at 118 patients over all replications when all patients are scheduled on a pull fashion. Figure 4.7 shows how a pull strategy affects number of first fraction rebooks, i.e. when the pre-treatment workflow cannot be completed before the pre-scheduled date. The more pull patients, the more rebooks occur, with an increase from 69.5 (baseline) to 132.7 (all) in the average number of occurrences.

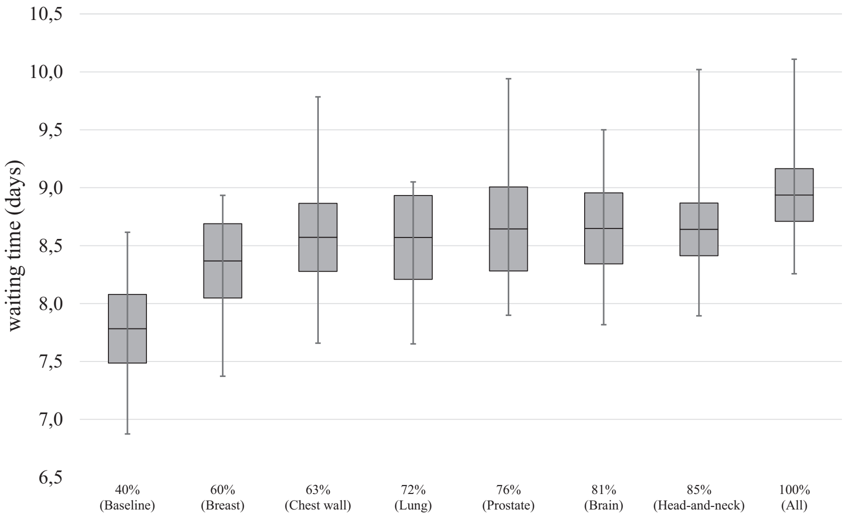


Figure 4.5 Box plot of the average waiting time (days) for different percentages of patients being scheduled in a pull manner for the workflow control analysis.

Table 4.6 shows the results of the scenario analysis. Balancing the consultation slots had the greatest impact on the performance, by decreasing waiting times from 7.8 to 6.2 days (20.8%) while providing a reduction in the number of patients breaching their waiting time targets from 88 to 23 (74%). Similarly, by not having a pre-allocation of time slots in the CT scanners results show that lower waiting times (17.3%) and fewer patients breaching their targets (57.8%) could be achieved. As for treatment planning, results indicate that performance would modestly improve by either having P3 planners doing lung and chest wall patients (1.6%) or hiring an extra P4 full-time planner (1.4%). Balancing the doctors' time available for contouring throughout the week has shown not to improve performance, providing the same average waiting time as the baseline case.

4.4 Discussion

We have developed a discrete-event simulation model to assess the optimal balance between two different strategies for patient scheduling in RT: pull (sched-

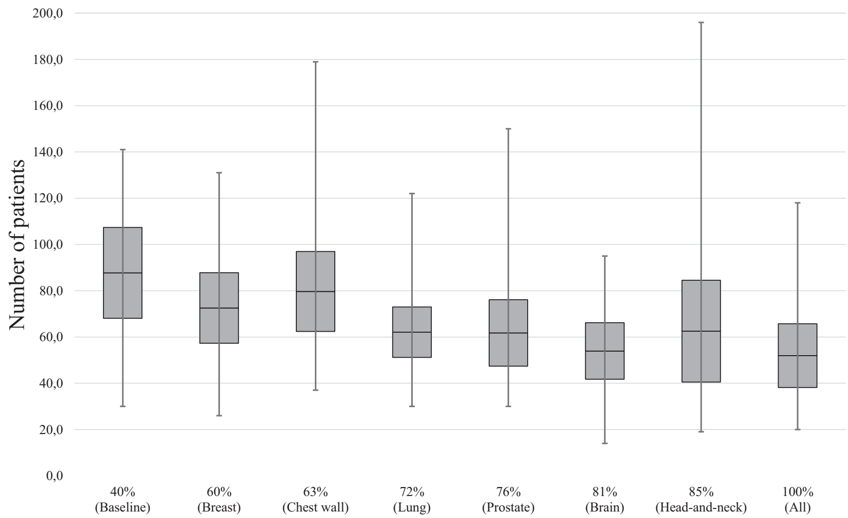


Figure 4.6 Box plot of the average number of patients starting treatment after the desired waiting time for different percentages of patients being scheduled in a pull manner for the workflow control analysis.

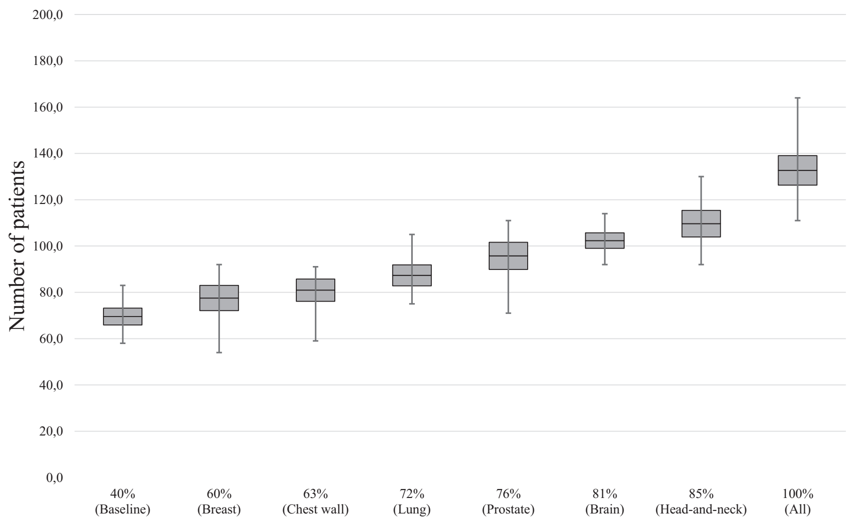


Figure 4.7 Box plot of the average number of start of treatment rebooks for different percentages of patients being scheduled in a pull manner for the workflow control analysis.

Table 4.6 Results of the scenario analysis for the baseline case (i.e. 40% pull patients).

Scenario	Average WT days (95% CI)	# patients breaching WT target (95% CI)	# first fraction rebooks (95% CI)
Baseline (DES model)	7.8 (7.5, 8.1)	87.7 (68.1, 107.4)	69.5 (65.9, 73.2)
Spread consultation slots over the week	6.2 (6.1, 6.3)	22.5 (19.0, 26.0)	60.7 (56.4, 65.1)
No pre-allocation for CT	6.4 (6.4, 6.5)	37.1 (31.8, 42.4)	65.6 (62.4, 68.8)
Balance doctor availability for contouring	7.8 (7.5, 8.0)	80.9 (66.1, 95.6)	76.9 (73.4, 80.5)
Increase automated planning by 16.4 %	7.7 (7.4, 7.9)	74.2 (61.0, 87.4)	67.5 (62.9, 72.2)
One more full-time P4 planner	7.7 (7.4, 7.9)	77.3 (62.3, 92.4)	64.3 (60.3, 68.2)

ule at first consultation) and push (schedule after treatment planning), based on the actual system data of the NKI. Results showed that increasing the pull strategy from 40% to 100% reduces the number of patients starting treatment after the WT target date from 87.7 to 51.9 (Figure 4.6), on average. By setting a start of treatment right at the beginning of the process, the control over the work-in-progress obviously increases and there is a lower risk of having delayed patients. This can be achieved at a cost of a maximum of one day increase in the average waiting times (Figure 4.5). A push strategy, by allowing work to flow continuously throughout the RT chain, provides up to 1.1 days reduction in the average waiting times. However, in moments of high workload and/or reduced staff availability while using a push system, some patients may have to wait longer than desired and consequently breach their WT target date, which can be mitigated by a pull strategy. As expected, the percentage of first appointment rebooks gradually increases with a pull strategy, due to non-completion of the pre-treatment phase on time to a maximum of 2.7% (Figure 4.7). Moreover, we have found that applying a pull strategy for certain tumor sites has greater impact on performance than for others. For instance, by adding prostate, brain and head-and-neck patients to the pull group, we verified that waiting times remained constant while the number of breaching patients slightly decreased. This may indicate that there is enough capacity in the department to accommodate these patients working on a pull strategy without increasing waiting times. In fact, the process of increasing the number of patients working on a pull fashion can be gradual. For instance, by scheduling all breast patients in addition to the baseline case, thus increasing the total number of pull patients from 40% to 60%, may allow achieving a 17.3% decrease on patients breaching the waiting time targets, with an increase on the average waiting time (6.4%) and the number of first appointment rebooks (11.5%).

A scenario analysis of possible interventions performed on the baseline case (40% pull patients) has shown that distributing consultation time slots evenly throughout the week has the highest impact on the measured performance. As shown in Table 4.6, by spreading consultations slots evenly over the week and thus keeping workload less variable throughout the chain, average waiting times can potentially decrease from 7.8 to 6.2 days. Although we understand that this may not be straightforward to implement due to the complex doctor schemes and busy agendas, it is an insight that may encourage decision makers to strive for consultation slots spread throughout the week as much as possible for each specialty. In addition, by not having pre-allocated time slots for CT scheduling the average waiting times and number of patients breaching the targets can potentially decrease by 12.7% and 57.8%, respectively (see Table 4.6). However, since most of the allocated time slots are dedicated to acute and bone metastasis patients, the impact on delays of these patient types would need to be further explored before an actual implementation. Our findings also showed that spreading the availability of doctors to perform contouring over the week does not increase performance, suggesting that the current doctors' agenda is well synchronized with the patient throughput for contouring. Moreover, our study showed that the increasing the number of planning RTTs does not improve performance significantly when compared to other scenarios, as the addition of an extra full-time RTT with the highest skill level of planning provided a marginal decrease of 1.4% in waiting times and 11.9% in the number of patients breaching the WT targets. Similarly, we found that upgrading the skill level of P3 planners to perform lung and chest wall patients did not impact results considerably from a logistics point of view.

Despite all the insights obtained with the DES model, there are a few limitations to our simulation study. The model is not able to fully capture the behavior of clinicians, as they may for example stay at work longer than expected to finalize certain tasks and avoid delaying the process of more urgent cases or skip certain meetings to do contouring when their clinical workload is high. Given the lack of clinical data regarding these situations, we overlook this possibility in the model. Moreover, each treatment plan needs to be checked and approved by a medical physicist before the first fraction is delivered. However, in the NKI a medical physicist is called by the planning RTT right after completion of the treatment plan. Therefore, there is no delay due to this step. In addition, the treatment plan may need to be improved or modified as a result of the medical physics check, thus requiring extra time to complete the treatment planning phase. We have overlooked these situations in our model as they account for less than 1% of the cases.

4.5 Conclusions

A 100% pull strategy, in which patients are scheduled a start of treatment right after consultation, provides increased predictability on the fulfillment of waiting time targets in detriment of a small increase in the average waiting times when compared to a push strategy. These findings are useful to support policy making in RT regarding their workflow control strategies and help RT centers achieve a desired service level within their resource constraints. Some centers may accept having slightly longer waiting times if that means having their patients informed about the start date for treatment date right at consultation, thus decreasing the discomfort and psychological distress associated with waiting for a date to start treatment. Moreover, DES has proved to be a powerful tool that provides an overview of the actual system and can help RT managers find bottlenecks and opportunities for performance improvement with recourse to visualization tools. Managerial interventions can be tested with little effort after a valid and robust model has been constructed, and the consequences of alternative input parameters can be quickly estimated.

As a follow up of this study, we want to implement and test extending the number of patients being scheduled in a pull way in the RT department of the NKI (e.g. all breast patients) and perform a pre-post performance evaluation to verify whether our theoretical results hold in practice. Furthermore, as the modeled processes and the patient mix are standard among RT centers, the proposed model can also be applied to other centers with a similar workflow and resource schemes.

4.6 Appendix

Table 4.7 List of care plans in the NKI (2017) and corresponding data used for model inputs.

Care plan	Tumor site	# patients	# acute	# subacute	# regular	# patients with planned delay	Prob. P2	P2 time (min)	P3 time (min)	P4 time (min)
Anus +/- inguinal lymph node	Others	22	0	0	22	1	0.05	60	-	480
Adrenal Stereotaxic	Others	7	0	0	7	0	0.00	-	-	420
Bladder	Others	43	0	2	41	2	0.02	60	-	240
Bladder (partial)	Others	22	0	0	22	5	0.00	-	-	240
Chest wall	Chest wall	45	0	0	45	20	0.02	60	-	120
Chest wall+Axilla	Chest wall	114	0	0	114	57	0.85	90	-	240
Chest wall+Axilla+Parasternal	Chest wall	15	1	1	13	7	0.60	120	-	240
Chest wall+Parasternal	Chest wall	3	0	0	3	2	0.33	60	-	240
Chest wall (bsu)	Chest wall	2	0	1	1	1	1.00	60	-	-
Bone metastasis	Bone met.	1119	53	1006	60	32	0.93	60	-	300
Bone metastasis Stereotaxic	Bone met.	56	0	1	55	1	0.14	60	-	420
Brachy Bladder	Others	15	0	0	15	0	0.00	-	-	-
Brachy Cilinder	Others	21	0	0	21	0	0.05	60	-	-
Brachy Intrauterine	Others	27	0	0	27	0	0.00	-	-	-
Brachy IOBT	Others	1	0	0	1	0	0.00	-	-	-
Brachy Oesophagus	Others	1	0	0	1	0	0.00	-	-	-
Brachy Prostate	Others	6	0	0	6	0	0.00	-	-	-
Uterus	Others	43	0	2	41	4	0.02	60	-	360
Endometrium	Others	17	0	0	17	1	0.06	60	-	360
Neck (bsu)	Others	9	0	8	1	1	1.00	60	-	-
Brain 1 fraction	Brain	140	0	10	130	2	0.14	60	-	360
Brain several fractions	Brain	94	0	5	89	3	0.09	60	-	360
Brain (whole)	Brain	166	2	134	30	8	1.00	60	-	-
Brain Electrons	Others	6	0	0	6	1	0.17	60	-	360
Lymph node Stereotaxic	Others	22	0	0	22	3	0.00	-	-	300
Head-and-neck	Head&neck	237	0	2	235	21	0.00	60	-	480
Head-and-neck (palliative)	Head&neck	4	0	0	4	1	0.25	60	-	480
Larynx 2vs	Others	2	0	0	2	0	0.00	-	-	480
Liver	Others	34	0	1	33	1	0.03	60	-	600
Lung (palliative)	Lung	65	1	51	13	5	0.42	60	-	690
Lung	Lung	283	1	8	274	15	0.08	60	-	690
Lung (bsu)	Lung	50	0	46	4	2	0.90	60	-	690
Lung Stereotaxic	Lung	239	0	2	237	11	0.01	60	-	420
Lymphoma	Others	43	0	2	41	6	0.07	60	-	300
Lymphoma (bsu)	Others	18	0	3	15	2	0.78	60	-	300
Stomach	Breast	14	1	1	12	2	0.21	60	-	360
Breast	Breast	777	0	7	770	276	0.03	60	120	120
Breast+Axilla	Breast	180	0	0	180	69	0.63	90	240	240
Breast+Axilla+Parasternal	Breast	19	0	0	19	9	0.53	120	240	240
Breast+Parasternal	Others	1	0	0	1	0	1.00	60	120	120
Esophagus	Others	78	1	5	72	9	0.09	60	-	360
Esophagus (palliative)	Others	22	1	18	3	1	0.55	60	-	360
Axilla (virtual)	Others	5	0	4	1	0	0.60	60	-	300
Orbit (eye socket)	Others	1	0	1	0	0	0.00	-	-	480
Ovaries	Others	5	0	2	3	1	0.00	-	-	360
Others	Others	186	0	138	48	18	0.07	60	-	300
Others (bsu)	Others	62	1	56	5	2	0.87	60	-	300
PAO (+/- iliac single-sided)	Others	1	0	0	1	0	0.00	-	-	240
Penis	Others	11	0	1	10	0	0.00	-	-	480
Prostate	Prostate	243	0	1	242	12	0.01	60	150	150
Prostate+Pelvic lymph nodes	Prostate	61	0	0	61	8	0.00	-	420	420
Prostatic bed	Prostate	36	0	0	36	2	0.00	-	300	300
Prostatic bed+Pelvic lymph nodes	Prostate	25	0	0	25	5	0.00	-	420	420
Rectum / Sigmoid	Others	88	0	4	84	6	0.02	60	300	300
Rectum 13 x 3 Gy	Others	6	0	5	1	0	0.00	-	300	300
Rectum 5 x 5 Gy	Others	57	1	0	56	2	0.00	-	300	300
Sarcoma abdominal/thoracic wall	Others	13	0	0	13	4	0.00	-	-	240
Sarcoma extremity	Others	31	0	4	27	2	0.03	60	-	480
Sarcoma retroperitoneal	Others	9	0	0	9	0	0.00	-	-	240
Spinal cord	Others	37	0	1	36	4	0.16	60	-	480
Vagina	Others	5	0	0	5	1	0.00	-	-	360
Vulva +/- inguinal lymph nodes	Others	9	0	0	9	2	0.00	-	-	420
Total		4973	63	1533	3377	650				
% of patient population			1.3%	30.8%	67.9%	13.1%				

CHAPTER 5

Radiotherapy treatment scheduling considering time window preferences

5.1 Introduction

With the increasing demand for radiotherapy (RT) services [92], which is expected to grow by an average of 16% until 2025 [8], the complexity related to the administration of existing RT resources (machines and staff) has become increasingly relevant [105, 106]. Radiotherapy treatments, usually given in a set of (daily) irradiation sessions, are administered by a machine called linear accelerator (linac), which is able to kill cancer cells by delivering high-energy radiation directed to the tumor. The growing number of treatment sessions to be booked amongst the available machines makes the scheduling process especially complex for RT centers aiming at delivering timely and patient-friendly treatments. Not only has it been shown that delays in the start of treatment may induce greater psychological distress in patients subject to longer waiting times [63], but also that 80% of the patients prefer a short interval (two weeks or less) between referral and first oncology consultation [70]. The problem of scheduling RT treatment sessions for large varieties of treatment care pathways and technical constraints has been tackled by several studies in the current literature [107]. Models exist for assigning patients' irradiation sessions to linacs and days [22, 88], with some studies addressing not only the scheduling component but also the sequencing of patients throughout the day [60]. While an overview on RT capacity in European countries [86] has shown that RT centers in most Western European countries are provided with enough capacity to treat all patients in due time, a survey amongst six Dutch RT centers within this project has shown the need to include patient preferences in the scheduling process. For these centers, asking patient preferences and integrating them into the schedule production process was common practice as they wanted to provide a better treatment experience to patients who want to maintain their routines and daily schedules during treatment. They showed that the quality of care, from a patients' perspective, increased when patients feel involved into the scheduling process and experience the provider trying to satisfy their personal preferences for the (several) number of visits they must pay to the hospital. Moreover, literature shows that patients have different preferences regarding the time of their appointments, emphasizing the importance of fulfilling those for increased patient-centeredness [70]. The goal of these RT centers is to schedule irradiation sessions such that all patients start treatment in due time, medical and technological constraints are satisfied, and the fulfilment of patient preferences regarding the starting time of their sessions is maximized. According to these RT centers, patients have shown the desire for specific appointment times for a variety of reasons, such as avoid traffic peak times, manage to keep their normal work schedule, or coordinate the RT treatment with their daily routines and hobbies. However, manual endeavours to produce such a schedule by (several) staff members are usually time consuming, prone to errors, and keen to find sub-optimal solutions regarding the fulfillment of patient preference requests.

Previous studies have approached different variants of the RT treatment scheduling problem and several methods have been proposed to solve it [107]. Sauré *et al*. [88] formulated the problem as a discounted infinite-horizon Markov decision process, showing that the percentage of treatments initiating treatment within 10 days can potentially increase from 73% to 96%. Legrain *et al*. [59] proposed a two-step stochastic algorithm for online scheduling of RT sessions, with results showing an average decrease in the number of patients breaching the standards of 50% for acute patients and 81% for subacute patients. Conforti *et al*. [22] developed an integer linear optimization program modeled in a non-block scheduling strategy, ensuring a linacs' utilization rate of 95% while minimizing the mean waiting times. Petrovic *et al*. [74] propose three genetic algorithms (GAs) for minimizing waiting time target breaches when scheduling emergency, palliative and radical patients. Results showed a potential reduction of average waiting times for radical (35 to 21.48 days) and palliative (15 to 13.10 days) patients. Although efficient methods for scheduling RT sessions have been proposed, the literature in relation to optimizing the sequencing of patients throughout the day is rather scarce [107]. However, as discussed above, besides complying with timeliness requirements and technical constraints, RT centers are often faced with the problem of finding a schedule that maximizes patient preferences regarding the starting time of irradiation sessions. In more recent years, two models have been proposed [65, 109] to optimize intra-day linac schedules in a way that starting time of irradiation sessions do not deviate from a pre-defined target time by more than a certain threshold (30 minutes in both [65] and [109]). Vogl *et al*. [109] modeled the problem for an ion beam facility (in which a single particle beam serves multiple treatment rooms). They included time window constraints whose violations are penalized in the objective function, which minimizes the idle time of the particle beam unit. Using real-world inspired data, they found that a combination of two stand-alone metaheuristic approaches leads to the best results when compared to a genetic algorithm and iterated local search. Maschler and Raidl [65], on the other hand, proposed an enhanced iterated greedy (EIG) metaheuristic to solve the patient scheduling problem with limited starting time variation between sessions in particle therapy. Computational experiments using fictitious data showed that the EIG method outperforms two other metaheuristics in 26 out of 30 instances. However, these two studies focus on particle therapy (PT), and thus cannot be applied directly to conventional external-beam RT since the technical and medical constraints vary considerably. For instance, in particle therapy a single beam source is used by multiple treatment rooms, but only one room can use the beam source at a time. Moreover, in both studies, patient preferences are not taken into account, and only approximation methods are able to solve the problem in acceptable time due to the complexity of the mathematical formulations and high number of constraints involved. On a hospital-wide setting, Gartner *et al*. [37] present exact and heuristic methods for the scheduling and routing of physical therapists where scheduled treatment sessions are bounded to pre-defined time windows. However, they optimize their models for the minimization of waiting times only and do not allow for sessions to be scheduled outside the required time window.

Overall, most models presented in the current literature focus on deciding on the specific day and linac of each irradiation session, with the sequencing of patients in each linac and each day being either neglected or determined in a secondary stage. Studies addressing the sequencing problem considering time windows are developed in the context of PT, thus they are not directly applicable to conventional RT. No studies have been found where optimization models integrate patient preference structures when deciding on the appointment times of irradiation sessions in conventional external-beam RT. In this paper, we propose a mixed-integer linear programming (MILP)-based approach for scheduling and sequencing RT treatment sessions. Our model takes all the medical and technical constraints into account, and maximizes the satisfaction of time window preferences given by patients for the starting time of their appointments. To solve the problem more efficiently for larger instances, we propose a heuristic procedure that pre-assigns patients to linacs before using the MILP model to solve each of the subproblems (subset of patients and linacs) independently. We compare the performance of the MILP model alone and the combined approach regarding solution quality and CPU time for different instance sizes. The feasibility of our algorithm is tested using real-world data from the Netherlands Cancer Institute (NKI), a large RT center located in Amsterdam, the Netherlands, with approximately 5000 new treatments per year and eight linacs operating on a daily basis. Although patient preferences are currently not recorded in the NKI and thus data regarding patient preferences is not available, we have performed a sensitivity analysis on both the preference structure breakdown and the size of the possible time windows being made available for patients to choose from.

This paper is organized as follows: Section 5.2 contains the formal problem description. The methodology including our MILP model and the algorithm to pre-assign patients to linacs are presented in Section 5.3. Section 5.4 presents the computational experiments performed using real-world data from a large RT department. The analysis and discussion of the results are described in Section 5.5, and Section 5.6 outlines the major conclusions of this study.

5.2 Problem description

In the RT scheduling problem, the aim is to schedule a set of treatment sessions for a set of cancer patients $\mathcal{P}$ over a given planning horizon $\mathcal{T}$, discretized in time periods $t = 1, ..., |\mathcal{T}|$. Each patient has a certain due date d_i, which defines the maximum date a patient should start treatment before the maximum waiting time target is achieved. Treatment sessions are delivered by a set of linear accelerators $\mathcal{K}$. The capacity of each linac is given by the number of available time slots $|\mathcal{S}|$ of duration l. Each session of each patient $i \in \mathcal{P}$ has an estimated processing duration quantified as a pre-defined number of time slots p_i. Most sessions are delivered on a daily basis, however some patients (e.g. hypofractionation schemes) may require (at least) one day off between two consecutive treatment sessions. Typically, every linac is capable of treating patients from all tumor types. However, some RT centers such that of the NKI may have a master schedule (pre-allocation) indicating that some patient groups must be assigned to a restricted set of linacs. For instance, brain patients may only be allowed to be scheduled on the technologically most advanced linacs as patients with this tumor site benefit the most from higher precision levels of the linac's delivery. Factors such as the accuracy level and other technologies (such as cone-beam CT) of the linacs, departments may want to pre-allocate certain patient groups to certain linacs. In case a pre-allocation exists, each patient must receive treatment in one of the linacs pre-allocated to his/her patient group ($\mathcal{K}^i$). The duration of treatment sessions typically vary per patient group, but sessions of each individual patient usually have the same duration throughout the whole treatment. Besides, we assume that RT centers aim at delivering all RT sessions on the same linac for each given patient. Although there are no technical or medical constraints that enforce this as a necessary condition, from a patient perspective it is highly desirable that patients receive their daily sessions on the same linac such that they always see the same facilities and personnel throughout most of the treatment. In addition, due to the combination of RT with other treatment modalities, such as chemotherapy, some patients may need to start treatment on a Monday to guarantee a proper coordination between the different treatment modalities. Moreover, because the first irradiation fraction of each patient may take longer than expected due to the need of explaining the whole process to the patient, RT centers commonly set a threshold T limiting the number of new patients who are scheduled to start treatment on the same linac and the same day in order to avoid congestion. Besides, for some patients there may be the need of guaranteeing that specialized staff (e.g. doctors) are in the department during the delivery of irradiation sessions to certain patients ($\mathcal{P}^f$) in case unexpected complications occur. In these cases, a time frame $[\underline{f}^t, \overline{f}^t]$ must be set to bound the starting time of all irradiation sessions of those patients.

Apart from the fulfillment of all the medical and technological constraints, in this problem we consider that RT centers are interested in finding a (weekly) schedule that minimizes the number of appointments scheduled outside the preferential time window requested by patients. This means that RT centers can run the model during the last workday of the previous week (i.e. Friday). Thus, data regarding (regular) patients to be scheduled is known by the beginning of the planning horizon, allowing to build a deterministic model to be used at an offline operational level (for a definition of the different hierarchical planning levels, see Hulshof *et al*. [44]).". Although other sources of uncertainty (session duration, no-shows, sessions' cancellations, linac breakdowns) exist, we verified that the percentage of occurrences of deviations between the planned and the realized values was lower than 1% for each case. Thus, we assumed these input parameters as deterministic, focusing on the performance of the "planned" solution regardless of the modifications that may be required at an online operational level. In our model, the goal is that the starting time of the scheduled sessions falls within the patients' desired time window $[t_i^{\text{min}}, t_i^{\text{max}}]$ consistently. Figure 5.1 depicts a possible weekly schedule of a linear accelerator in external-beam RT with time window preferences. Note that appointments times are, for most patients, consistent throughout the week. Let us assume that, in this schedule, Patient 1 had requested their sessions to be booked between slots 1-3 inclusive. Then, two (Tuesday and Thursday) out of five sessions would fall outside the desired window. Considering that all appointments of all other patients in Figure 5.1 are set within the requested time window, then the performance value of the solution for this linac would be equal to 23/25, i.e. 92% of the appointments are booked within the requested time window. In this paper, we propose a method that aims at maximizing the percentage of sessions falling within the requested time window for patients and linacs of real-world RT centers.

<table>
<tr><th></th><th>Monday</th><th>Tuesday</th><th>Wednesday</th><th>Thursday</th><th>Friday</th></tr>
<tr><td>slot 1</td><td>Patient 1</td><td>Patient 6</td><td>Patient 1</td><td>Patient 6</td><td>Patient 1</td></tr>
<tr><td>slot 2</td><td rowspan="2">Patient 2</td><td rowspan="2">Patient 2</td><td rowspan="2">Patient 2</td><td rowspan="2">Patient 2</td><td rowspan="2">Patient 2</td></tr>
<tr><td>slot 3</td></tr>
<tr><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td></tr>
<tr><td>slot s</td><td rowspan="2">Patient 3</td><td>Patient 1</td><td rowspan="2">Patient 3</td><td></td><td rowspan="2">Patient 3</td></tr>
<tr><td>slot $s+1$</td><td></td><td>Patient 1</td></tr>
<tr><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td></tr>
<tr><td>slot $|\mathcal{S}|-2$</td><td rowspan="2">Patient 4</td><td rowspan="2">Patient 4</td><td rowspan="2">Patient 4</td><td>Patient 5</td><td>Patient 5</td></tr>
<tr><td>slot $|\mathcal{S}|-1$</td><td rowspan="2">Patient 4</td><td rowspan="2">Patient 4</td></tr>
<tr><td>slot $|\mathcal{S}|$</td><td>Patient 5</td><td>Patient 5</td><td>Patient 5</td></tr>
</table>

Figure 5.1 Example of a weekly schedule of a linear accelerator in RT.

5.3 Methodology

In this section we present the methodology developed to solve the RT scheduling problem with time window preferences given by patients for the starting time of their sessions. We use the notation presented in Table 5.1 to formulate the MILP model and the heuristic procedure designed to pre-allocate patients to linacs presented in Algorithm 1.

Table 5.1 Notation of the MILP model.

Parameter	Description
$\mathcal{P}$	set of patients to be scheduled ($i, j \in \mathcal{P}$)
$\mathcal{K}$	set of linear accelerators ($k \in \mathcal{K}$)
$\mathcal{S}$	set of time slots per linac ($s \in \mathcal{S}$)
$\mathcal{T}$	set of time periods (days) in the planning horizon ($t \in \mathcal{T}$)
$\mathcal{P}^n$	set of patients who have not started treatment ($\mathcal{P}^n \subseteq \mathcal{P}$)
$\mathcal{P}^m$	set of patients who must start treatment on monday ($\mathcal{P}^m \subseteq \mathcal{P}$)
$\mathcal{P}^f$	set of patients with restricted time frame for treatment sessions ($\mathcal{P}^f \subseteq \mathcal{P}$)
$\mathcal{K}^i$	set of feasible linacs for treating patient i ($\mathcal{K}^i \subseteq \mathcal{K}$)
T	maximum number of patients starting treatment in the same linac and same day
l	time slot duration, in minutes, in each linac, each day
a_{kst}	1 if slot s of linac k is available on time period t, 0 otherwise
$\underline{f}^t, \overline{f}^t$	lower and upper bound of the restricted time frame set for time period t
I_i	number of total remaining sessions to be delivered to patient i
d_i	due date: time period by which patient i must start treatment
p_i	duration, in number of time slots, of each session of patient i
b_i	number of time periods needed between sessions of patient i (1 for consecutive daily sessions)
$t_i^{\min}, t_i^{\max}$	lower and upper bound of the time window preference for patient i
c_i	linac in which patient $i \notin \mathcal{P}^n$ is currently undergoing treatment
C_k	weekly capacity, in minutes, of linac k
$vol(i)$	expected weekly session time, in minutes, for patient i
$WL(k)$	workload, in minutes of session time, assigned to linac k
Variable	**Description**
x_{iks}^t	1 if patient i is scheduled a session starting on time slot s of linac k in day t, 0 otherwise
y_{ik}^t	1 if new patient i starts treatment in period t and linac k, 0 otherwise
$\Delta_{it}^-, \Delta_{it}^+$	lower and upper deviation, in minutes, from preference time window of patient i in time period t

5.3.1 Mathematical programming model

The problem is formulated such that the capacity of each linac is divided in time slots $s = 1, ..., |\mathcal{S}|$ of duration l. When scheduled, patients' sessions are assigned a certain starting time slot on a certain linac and day. To this end, we introduce binary variables x_{iks}^t, which take the value 1 if patient i is scheduled for a session starting on time slot s of linac k in day t, and 0 otherwise. If a certain starting slot is assigned to a patient, we prevent the following slots needed to achieve the corresponding patient's session duration on that same linac and day from being assigned to other patients.

Objective function

The objective (5.1) is to minimize the overall deviation between the bounds of the preferred time window $[t_i^{\min}, t_i^{\max}]$ given by patients and the starting time of their appointments. Real variables Δ_{it}^{-} and Δ_{it}^{+} are used to represent this deviation for each patient in each day. Binary variables y_{ik}^{t} are auxiliary variables, which will be equal to 1 if a new patient starts his/her treatment in period t and linac k, and 0 otherwise.

$$\min \sum_{i \in \mathcal{P} \setminus \{1\}} \sum_{t \in \mathcal{T}} (\Delta_{it}^{-} + \Delta_{it}^{+}) \tag{5.1}$$

The technical and medical constraints described in Section 5.2 are modeled as follows:

Sessions' assignment constraints

Inequalities (5.2)-(5.4) ensure that patients receive their sessions on the same linac and with the required frequency b_i until the number of sessions or the end of planning horizon is reached. Constraints (5.2) and (5.3) force the necessary sessions to be booked, at least every b_i days, as soon as a first session is scheduled. Constraints (5.4) avoid unnecessary sessions from being scheduled in days occurring between the days of the sessions booked by constraints (5.2)-(5.3) when $b_i > 1$.

$$\begin{aligned}
&\sum_{s \in \mathcal{S}} x_{iks}^{t} - \sum_{s \in \mathcal{S}} \sum_{t'=1}^{t-1} x_{iks}^{t'} \leq \sum_{s \in \mathcal{S}} x_{iks}^{n}, \\
&\forall i \in \mathcal{P}, \forall k \in \mathcal{K}, \forall t = 2, ..., \mathcal{T}, \\
&\forall n = t + b_i, t + 2b_i, ..., \min\{|\mathcal{T}|, t + b_i(I_i - 1)\} && (5.2) \\
&\sum_{s \in \mathcal{S}} x_{iks}^{1} \leq \sum_{s \in \mathcal{S}} x_{iks}^{n}, \forall i \in \mathcal{P}, \forall k \in \mathcal{K}, \\
&\forall n = b_i + 1, 2b_i + 1, ..., \min\{|\mathcal{T}|, b_i(I_i - 1) + 1\} && (5.3) \\
&1 - \sum_{s \in \mathcal{S}} x_{iks}^{t} \geq \sum_{s \in \mathcal{S}} x_{iks}^{n}, \forall i \in \mathcal{P}, \forall k \in \mathcal{K}, \\
&\forall t = 1, ..., |\mathcal{T}| - b_i, \forall n = t + 1, ..., t + b_i - 1, b_i \geq 2 && (5.4)
\end{aligned}$$

Limitations on the number of sessions

Inequalities (5.5) limit the number of sessions that each patient can receive to a maximum of one per day. Constraints (5.6) restrict the number of sessions delivered during the planning horizon to the number of remaining sessions for that patient.

$$\sum_{k \in \mathcal{K}} \sum_{s \in \mathcal{S}} x_{iks}^{t} \leq 1, \forall i \in \mathcal{P}, \forall t \in \mathcal{T} \tag{5.5}$$

$$\sum_{k \in \mathcal{K}} \sum_{s \in \mathcal{S}} \sum_{t \in \mathcal{T}} x_{iks}^{t} \leq I_i, \forall i \in \mathcal{P} \tag{5.6}$$

Timeliness constraints

Constraints (5.7) impose that every patient starts treatment before their due date d_i. Note that for patients who need to start treatment on a Monday one can set $d_i = 1$.

$$\sum_{k\in\mathcal{K}}\sum_{s\in S}\sum_{t=1}^{d_i} x_{iks}^{t} \geq 1, \forall i \in \mathcal{P} \tag{5.7}$$

Linacs' capacity constraints

Constraints (5.8) ensure that each (available) slot of each linac is scheduled at most one session per day, and restrictions (5.9) ensure that each patient is assigned to a feasible linac by preventing sessions of being assigned to slots of linacs that do not belong to $\mathcal{K}^i$.

$$\sum_{i\in\mathcal{P}} x_{iks}^{t} \leq a_{kst}, \forall k \in \mathcal{K}, \forall s \in \mathcal{S}, \forall t \in \mathcal{T} \tag{5.8}$$

$$\sum_{s\in\mathcal{S}}\sum_{t\in\mathcal{T}} x_{iks}^{t} \leq 0, \forall i \in \mathcal{P}, \forall k \in \mathcal{K} \setminus \{\mathcal{K}^i\} \tag{5.9}$$

Maximum number of patients starting treatment per linac per day

Constraints (5.10)-(5.11) force variables y_{ik}^{t} to take the value 1 if a new patient i starts treatment on linac k and day t, while equations (5.12) use these auxiliary variables to limit the number of patients starting treatment on the same linac and same day to the pre-defined threshold C.

$$y_{ik}^{t} \geq \sum_{s\in\mathcal{S}} x_{iks}^{t} - \sum_{s\in\mathcal{S}} x_{iks}^{t'}, \forall i \in \mathcal{P}^n, \forall k \in \mathcal{K}, \forall t = 2, ..., \mathcal{T},$$

$$t' = \max\{1, t - b_i\} \tag{5.10}$$

$$y_{ik}^{1} \geq \sum_{s\in\mathcal{S}} x_{iks}^{1}, \forall i \in \mathcal{P}^n, \forall k \in \mathcal{K} \tag{5.11}$$

$$\sum_{i\in\mathcal{P}^n} y_{ik}^{t} \leq C, \forall k \in \mathcal{K}, \forall t \in \mathcal{T} \tag{5.12}$$

Session duration constraints

Restrictions (5.13) prevent the remainder of the time slots needed to achieve the session duration p_i after the chosen starting slot (x_{iks}^{t}) from being assigned to other patients on the same linac and day. Inequalities (5.14) ensure that the starting slot of sessions with a duration of two or more slots are not assign to the last slot(s) of the day.

$$x^t_{iks} \leq 1 - \sum_{i' \in \mathcal{P}} x^t_{i',k,s'}, \forall i \in \mathcal{P}, \forall k \in \mathcal{K},$$
$$\forall s = 1, ..., |\mathcal{S}| - p_i + 1, \forall t \in \mathcal{T},$$
$$\forall s' = s + 1, ..., s + p_i - 1, p_i \geq 2 \tag{5.13}$$
$$x^t_{iks} = 0, \forall i \in \mathcal{P}, \forall k \in \mathcal{K}, \forall s = |\mathcal{S}| - p_i + 2, ..., S,$$
$$\forall t \in \mathcal{T}, p_i \geq 2 \tag{5.14}$$

Time window constraints

Constraints (5.15) force treatment sessions of each patient to fall within the restricted time frame set by the department due to the need of ensuring that specialized staff are present during the sessions of the applicable patients ($\mathcal{P}^f$). Equations (5.16) set variables Δ^-_{it} and Δ^+_{it} to take a non-zero value if a session's starting time deviates from the desired lower and upper bounds, respectively, and constraints (5.17) are the non-negativity constraints associated with the real variables.

$$x^t_{iks} \leq 0, \forall i \in \mathcal{P}^f, \forall k \in \mathcal{K}, \forall s \in \mathcal{S}, \forall t \in \mathcal{T},$$
$$s < \underline{f}^t, s > \overline{f}^t \tag{5.15}$$
$$t_i^{\min} x^t_{iks} - \Delta^-_{it} \leq l(s-1) x^t_{iks} \leq t_i^{\max} x^t_{iks} + \Delta^+_{it},$$
$$\forall i \in \mathcal{P}, \forall k \in \mathcal{K}, \forall s \in \mathcal{S}, \forall t \in \mathcal{T} \tag{5.16}$$

Non-negativity and integrality constraints

$$\Delta^-_{it} \geq 0, \Delta^+_{it} \geq 0, \forall i \in \mathcal{P}, \forall t \in \mathcal{T} \tag{5.17}$$
$$x^t_{iks}, y^t_{ik} \in \mathbb{B}, \forall i \in \mathcal{P}, \forall k \in \mathcal{K}, \forall s \in \mathcal{S}, \forall t \in \mathcal{T} \tag{5.18}$$

5.3.2 Patient-to-linac assignment

As we demonstrate in Section 5.4, the proposed MILP model alone is not capable of solving the problem for larger RT centers (three linacs or more) in acceptable computation time. In these cases, we apply a heuristic procedure (Algorithm 1) to pre-assign patients to linacs, and use the MILP model to solve the sequencing problem for each subset of linacs, hereby referred to as "clusters". In Algorithm 1, C_k represents the weekly capacity of the linacs, in minutes, $vol(i)$ represents the total session time expected during the whole planning horizon for patient i, while $WL(k)$ contains the workload, measured in total minutes of session time, in each linac k.

Algorithm 1: Pseudo-code of the procedure that pre-allocates patients to linacs.

initialization;
compute $C_k = \sum_{s \in \mathcal{S}} \sum_{t \in \mathcal{T}} a_{kst}.l, \forall k \in K$;
compute $vol(i) = \max\{I_i, \left\lceil \frac{|\mathcal{T}|}{b_i} \right\rceil\}.p_i, \forall i \in \mathcal{P}$;
compute $WL(k) = \sum_{\substack{i \in \mathcal{P} \setminus \{\mathcal{P}^n\}, \\ c^i = k}} vol(i), \forall k \in K$;
for $i \in \mathcal{P}^n$ **do**
    sort $\mathcal{K}^i$ in decreasing order of $WL(k)$;
    for $k \in \mathcal{K}^i$ **do**
        if $WL(k) + vol(i) \leq C_k$ **then**
            assign patient i to linac k ;
            $WL(k) \leftarrow WL(k) + vol(i)$;
            remove patient i from $\mathcal{P}^n$;
            break last cycle;
        end
    end
end

The algorithm initiates by computing the initial values of each parameter. Note that patients undergoing treatment ($i \in \mathcal{P} \setminus \{\mathcal{P}^n\}$) are already assigned to a linac, and thus the initial values $WL(k)$ will be given by the total patient volume undergoing treatment on linac k by the beginning of the planning horizon. The variable $WL(k)$ associated with linac k keeps track of the total volume (number of sessions times the duration) assigned to that linac. Next, and for each new patient $i \in \mathcal{P}^n$, the algorithm sorts the set feasible linacs $\mathcal{K}^i$ in increasing order of $WL(k)$. This way, the algorithm first searches the less busy linacs to foster a balanced workload. Starting from the top of the list of feasible linacs for patient i, the algorithm checks whether the current patient's volume $vol(i)$ fits the current available capacity C_k and, if so, assigns the patient to that linac, updating the value of $WL(k)$ by the total volume $vol(i)$ of the patient being assigned. If the patient has been assigned to k, patient i is removed from the list and the algorithm proceeds to the next patient. If not, the procedure continues to search for the next linac on the list until a feasible linac is chosen. Algorithm 1 therefore assumes that there is enough linac capacity to treat the whole patient population $\mathcal{P}$ being scheduled in order to find a feasible pre-assignment solution.

5.4 Computational experiments

This section presents the results of the computational experiments we have performed with our model. Section 5.4.1 describes the instance generator using historical patient data from the RT department of the NKI. Section 5.4.2 shows the results for several instance sizes using the NKI historical records to generate patient data for a reduced number of linacs. In Section 5.4.3 we solve the problem for the NKI size (eight linacs) using a method that combines our MILP model and the Algorithm 1, and Section 5.4.4 describes a sensitivity analysis performed to analyze the impact of the variation of patient requests and the competition for the same time window.

The MILP model and Algorithm 1 were coded in C++ using Visual Studio 2017 and the Concert Technology of CPLEX v12.8.0, which was used as a solver. All experiments were conducted on a desktop computer with a processor Intel i7 3.6 GHz and 16 GB of RAM using up to 8 threads, running on a 64-bit version of Windows 10. In our case study, the goal is to find a weekly schedule (Monday to Friday), which means that RT centers can hypothetically run the model during the last workday of the previous week (i.e. Friday) so that the maximum amount of patient data is known. The maximum allowed CPU time was set to 28800 seconds (8 hours) per run.

5.4.1 Historical patient data used for generating test instances

Patient characteristics are generated according to empirical distributions generated using historical data collected throughout 2017 (number of new treatment courses = 4720). In our instance generator, we start by randomly attributing a care plan (i.e. care trajectory) to each patient. There are 56 care plans in total, with the largest being "Bone metastasis" (23.3%) and “Breast” (16.5%). Thereafter, we generate the number of sessions I_i of each patient, which can vary between 1 and 35 sessions depending, to a large extent, on the care plan. For instance, nearly half of all prostate patients will undergo 35 sessions, while 65% of all bone metastasis patients are prescribed 3 sessions or less. Similarly, the urgency level of each patient, which can be either urgent (34%) or regular (66%), is randomly assigned according to historical data associated with his/her care plan. The due date ($d_i = 1, ..., 5$) of new patients (35% of the patient population) is generated as follows: we calculate, for each patient, the difference between the maximum waiting time according to the standards defined by the Dutch Society for Radiation Oncology (NVRO)[68] (21 days for urgent patients, and 28 days for regular patients) and the number of days elapsed from referral to treatment planning. For instance, if for a regular patient i, who needs to start treatment within 28 days of referral, the pre-treatment phase (referral to treatment planning) took 25 days to complete, this value would be equal to 3. We use the values verified in practice and corresponding proportions in the whole 2017, per care plan, to build empirical distributions from which due dates d_i

are randomly generated. The due date will be equal to 1 in case a patient is already undergoing treatment, and in case of new patients who need to start their treatment on a Monday ($\mathcal{P}^m$). The target due date already takes into account eventual delays due to medical or personal reasons, which are usually known by the beginning of the pre-treatment phase (consultation) and will be reflected in the waiting time target date set by the department. Furthermore, the duration of each session p_i is also assigned empirically on a care plan basis, ranging from 10 to 30 minutes, in multiples of 5 minutes. Data shows that the majority of patients will be scheduled 15-min sessions (60.5%), with 19.9% patients having sessions of 20 minutes or longer.

The daily available time for delivering irradiation sessions in the clinic ranges from 07h30 to 17h30, thus $|\mathcal{S}| = 120$ by considering $l = 5$ min. We solve the problem for a planning horizon of one labour week, discretized in time periods of one day ($|\mathcal{T}| = 5$). The restricted time frame ($\underline{w}^t, \overline{w}^t$) due to the need of guaranteeing that doctors are always present when irradiating ranges from 08h30 to 17h00, which means that $\underline{w}^t = 60$, $\overline{w}^t = 570$, $\forall t \in \mathcal{T}$. A total of 12 care plans (21% of the patient population) will require patients to be scheduled between 08h30 and 17h00. In the NKI, patients belonging to a total of 4 care plans (namely stereotactic and hypofractionation schemes) need at least 2 days (48h) break between two sessions, corresponding to a total of 6% of the patient population. Moreover, in the NKI a maximum of six new patients are allowed to start treatment on the same linac and same day, thus $C = 6$.

In the NKI, patient preferences, i.e. the preferences given by patients for the desired starting time of their irradiation sessions, are taken into consideration when scheduling irradiation sessions. However, data regarding patient preferences are not currently recorded in the clinic, and thus real data for patient preferences cannot be used. Interviews with the appointment planners showed, according to their empirical knowledge, that around 1 in 4 patients will have a preference for the early morning ($<$ 09h00), and 1 in 4 patients will want to receive their sessions in the end of the day ($>$ 16h00). From there, we assumed that all other patients, including those who do not have a preference, will be included in the group of the remaining 50% (09h00 - 16h00). Therefore, we randomly generated time window preferences $[t_i^{\min}, t_i^{\max}]$ for both urgent and regular patients as follows: 25% of patients have a preference for the morning window ($w_1 = [0, 90]$), 25% have a preference for receiving their RT sessions later in the day ($w_3 = [510, 600]$), while the remainder 50% of patients have a preference set for the time window ($w_2 = [90, 510]$). A sensitivity analysis on the possible variations on these proportions is presented in Section 5.4.4.

5.4.2 Results for several instance sizes

To test our model, we generated a set of test instances with various sizes regarding the number of patients ($|\mathcal{P}|$) and available linacs ($|\mathcal{K}|$) for a planning horizon of one labour week ($|\mathcal{T}| = 5$). We generated patient data by using his-

Table 5.2 Results of the MILP model for several instance sizes using NKI patient data.

# patients	# linacs	# sessions scheduled		# sessions outside window		average deviation (min)		CPU time (s)	
		milp	heur+milp	milp	heur+milp	milp	heur+milp	milp	heur+milp
33	1	123	123	4	4 (3.25%)	5	5.0	99.9	99.9
66	2	254	247	1	5 (2.0%)	15	8.0	499.6	17.2
99	3	-	379	-	23 (6.1%)	*	16.5	*	218.7
132	4	-	361	-	22 (5.2%)	*	28.2	*	1780.5
165	5	-	580	-	33 (5.7%)	*	24.8	*	1175.8
198	6	-	700	-	40 (5.7%)	*	24.0	*	3318.3
231	7	-	810	-	50 (6.2%)	*	25.3	*	2705.2
260	8	-	925	-	31 (3.4%)	*	13.2	*	393.5

torical data from the RT department of the NKI, a large cancer center operating in the Netherlands provided with eight full-time working linacs and scheduling an average of 260 patients per week.

We started by running experiments using the MILP model alone. In each test instance, the patient-to-linac ratio of the NKI, i.e. 33 patients per linac per week, is maintained. Given that we are scaling down the NKI problem for a subset of the linacs, the pre-allocation of linacs to patient groups in the appendix of Section 5.7 cannot be applied. Thus, we consider that all patients can be treated by all linacs ($\mathcal{K}^i = \mathcal{K}, \forall i \in \mathcal{P}$) by relaxing constraints (5.9).

As we can observe in Table 5.2, the proposed MILP model is able to find an optimal solution that schedules all sessions within the desired time window for all instance sizes up to 66 patients and 2 linacs within the CPU time limit. For the instance with 2 linacs, the proposed formulation proved effective in finding the optimal weekly schedule in less than 10 minutes of CPU time. However, the CPU time limit of 8 hours was achieved for instances with 99 patients and higher, possibly due to complexity introduced by the exponentially higher number of variables and constraints. Since the model was able to solve the problem for one linac in just 100 seconds, we used Algorithm 1 to pre-allocate patients to linacs as a pre-processing step (considering that all linacs are empty, i.e. $WL(k) = 0, \forall k \in \mathcal{K}$). We then apply the MILP model to solve the problem for each linac independently. We denote this combined approach as “heur+milp” in Table 5.2. We found that the combined approach is able to find a near-optimal solution in a total CPU time of 55 minutes or less for all instance sizes. For the NKI size (260 patients and 8 linacs), as few as 31 in 925 sessions were scheduled outside the desired window in just 7 minutes of CPU time. The solutions obtained by the combined approach schedule at most 6.2% of the sessions outside the preferred time window, with the percentage lowering down to 2% for the instance where the optimal solution is known (2 linacs). This indicates that the solutions found by partitioning the problem are probably even closer to optimality than the obtained percentages. Moreover, sessions scheduled outside the time window were, on average, under 30 minutes away from the corresponding windows. This shows that, with our approach, even sessions that are scheduled outside the preferred time window are still close to the target window bounds.

5.4.3 Results for the NKI size

In this section, we combine Algorithm 1 with the MILP model to solve the problem for the NKI size but now including the pre-allocation constraints (5.9). This means that patients undergoing treatment are pre-assigned to the linac they have been receiving their treatment, which is assigned randomly based on historical data. Therefore, the initial workload values $WL(k)$ of each linac k are pre-processed before running the cycle of Algorithm 1 to allocate new patients to linacs. Table 5.3 shows the results obtained after running Algorithm 1 for the 260 patients generated for the NKI size instance. As we can see, the pre-assignment solution provides a balanced workload amongst linacs, with an average of 2051.3 minutes and a maximum workload difference of 80 minutes between any pair of machines. Moreover, the number of patients assigned to each linac is considered stable, with a standard deviation of 3.3 patients amongst the 8 linacs at an (expected) average of 33 patients per linac. Utilization rates representing the percentage of the daily capacity (3000 minutes) assigned to each linac show that linacs have an utilization rate that ranges between 68.0% and 69.5%. This confirms the availability of slack capacity to apply the proposed heuristic effectively. The available capacity can be used to accommodate urgent patients arriving and having to start treatment during the planning horizon.

Table 5.3 Pre-assignment results after running Algorithm 1 for the NKI test instance.

Linac	no. patients assigned	workload (WL)	utilization rate (%)
L1	35	2055	68.5%
L2	27	2050	68.3%
L3	33	2055	68.5%
L4	39	2085	69.5%
L5	32	2040	68.0%
L6	32	2005	66.8%
L7	32	2080	69.3%
L8	30	2040	68.0%
avg	32.5	2051.3	68.4%
st. dev.	3.3	23.4	0.8%

Following the pre-assignment of patients to linacs, we cluster the linacs (L1,...,L8) in different groups based on the characteristics of the available machines at the NKI. Thus, new patients can still be assigned any linac belonging to the cluster his/her linac belongs to. For instance, if a cluster contains L1 and L2, a new patient pre-assigned to L1 can still be assigned to L2 by the MILP model (if that patient type can be treated in L2). Moreover, from the 8 NKI linacs running on a daily basis, two (L7 and L8) are located in a satellite location. Decisions on whether patients will be receiving treatment in the satellite or the main location are made right after referral. In those cases, the pre-assignment cannot be changed between the main and satellite locations by the moment our model is intended to be used, which is at the beginning of the planning horizon. This means that patients pre-assigned to L7 can still be allocated to L8 and vice-versa, however the linac assignment cannot be changed by any of the remaining 6 linacs running on the main location. Furthermore, we have clustered the linacs by level of similarity in terms of the total patient volume that they are able to treat in common according to the table of Section 5.7, while ensuring that L7 and L8 are not clustered together with other linacs. Each experiment runs the MILP model sequentially for a certain number of times, which corresponds to the number of clusters. Table 5.4 outlines the experimental setup and its considered clusters, as well as the corresponding results in terms of solution quality and computational (CPU) time. In the first experiment, each linac is a cluster in and of itself and therefore the MILP model is run 8 times in a row. For this test instance, a cumulative deviation of 1085 minutes over a total of 41 sessions scheduled outside the desired window (out of 918) was achieved, in a combined CPU time of around 5 hours. This means that, in around 5 hours of running time, our method was able to find a feasible solution with only 4.5% of the sessions being scheduled outside the preference time window. When 4 clusters are used, results are improved further. The extra flexibility provided by having 2 linacs per cluster decreased the overall objective value to 545, with as few as 26 sessions out of 936 (2.8%) being scheduled outside the intended time window. The combined CPU time was also reduced to around 3.5 hours. The CPU time limit of 8 hours per run was achieved when applying the same methodology for 2 clusters (main and satellite locations). Nevertheless, the solver was able to find a feasible solution, which did not improve the solution found with 4 clusters. A higher number of sessions (60) were found breaching the time window preference, in a total of 928 sessions scheduled. Note that, with the introduction of constraints (5.9), the MILP model was able to find a feasible solution the problem for an instance size of 6 linacs (L1,...,L6 in the experiment with 2 clusters) within the CPU time limit, although the CPU time limit was achieved in one of the runs. As with the experiments conducted in Section 5.4.3, the 8 hours of CPU time limit was achieved when attempting to solve the problem for the 8 linacs combined in a single cluster without any integer solution being found. We also note that, overall, the performance of the combined approach decreases in terms of CPU time (from 7 minutes to 5 hours) after considering the pre-allocation constraints (5.9) as part of the pre-assignment process. This shows that the existence of such pre-allocation, such as the one in the NKI, may decrease the performance of the proposed methodology by providing less flexibility when pre-assigning patients to linacs.

5.4.4 Sensitivity analysis

In this part, we investigated the impact of varying two input parameters: the probability breakdown for the possible time windows being chosen by patients, and the size (minutes) of the time windows being made available to patients. In our sensitivity analysis, we vary the size of both time windows w_1 and w_3

Table 5.4 Results for pre-assignment heuristic and MILP model for the NKI size.

No. clusters	Linacs belonging to each cluster	cumulative deviation (obj. value)	# sessions scheduled	# sessions outside window	average deviation (min)	total CPU time (s)
8	[L1] [L2] [L3] [L4] [L5] [L6] [L7] [L8]	1085	918	41 (4.5%)	26.5	18143
4	[L1,L3] [L2,L6] [L4,L5] [L7,L8]	545	936	26 (2.8%)	21.0	11763
2	[L1,L2,L3,L4,L5,L6] [L7,L8]	7695*	928*	60 (6.5%)*	128.3*	31181*
1	[L1,L2,L3,L4,L5,L6,L7,L8]	-	-	-	-	-

* CPU time limit achieved in at least one cluster

Table 5.5 Total cumulative deviation, in minutes.

Total deviation (Obj. value)			
Probabilities	Window size (min)		
(w_1 / w_2 / w_3)	*150*	*120*	*90*
0.25 / 0.5 / 0.25	0	65	650
0.5 / 0.25 / 0.25	320	4659*	25614*
0.125 / 0.75 / 0.125	0	0	55

* CPU time limit achieved

Table 5.6 Percentage of sessions breaching the time window preference.

% sessions outside window			
Probabilities	Window size (min)		
(w_1 / w_2 / w_3)	*150*	*120*	*90*
0.25 / 0.50 / 0.25	0.0%	0.5%	3.4%
0.5 / 0.25 / 0.25	1.5%	10.1%*	28.8%*
0.125 / 0.75 / 0.125	0.0%	0.0%	0.5%

* CPU time limit achieved

Table 5.7 Average deviation, in minutes, of the sessions outside window.

Average deviation (min)			
Probabilities	Window size (min)		
(w_1 / w_2 / w_3)	*150*	*120*	*90*
0.25 / 0.50 / 0.25	0.0	13.0	20.3
0.5 / 0.25 / 0.25	22.9	50.1*	97.4*
0.125 / 0.75 / 0.125	0.0	0.0	11.0

* CPU time limit achieved

Table 5.8 CPU time, in seconds, of the 4 runs combined.

Total CPU time (s)			
Probabilities	Window size (min)		
(w_1 / w_2 / w_3)	*150*	*120*	*90*
0.25 / 0.50 / 0.25	43.3	108	11299
0.5 / 0.25 / 0.25	5341.5	18981*	28800*
0.125 / 0.75 / 0.125	60.8	42	39

* CPU time limit achieved

from 90 minutes to 120 and 150 minutes, and test the probability breakdown $[w_1, w_2, w_3]$ at [50%,25%,25%] and [12.5%,75%,12.5%] in addition to the original [25%, 50%, 25%]. All combinations between these scenarios are tested using the instance with four clusters, since it has shown to provide the best performance regarding both the solution quality and overall CPU time needed (see Table 5.4). The maximum allowed CPU time in these experiments is set to 7200 seconds (2 hours) per run, for a combined CPU time limit of 28800 seconds (8 hours). Results (Tables 5.5-5.8) show that the patient satisfaction levels may significantly increase by enlarging the window size to 120 minutes, with the percentage of sessions being scheduled outside the preferential window decreasing to 0.5% (Table 5.6). Besides, the total CPU time needed to solve the four subproblems associated with the original problem decreases from 3.5 hours (11299 seconds) to only 108 seconds. On the other hand, extending the windows w_1 and w_3 to $[0, 150]$ and $[450, 600]$, respectively, allowed our methodology to schedule all sessions within the desired time window in just 43.3 seconds when the baseline probabilities are maintained.

By changing the probability of patients choosing w_1 from 25% to 50%, we observe that the complexity of the problem increases substantially. In fact, with the increased competition level for a window size of 90 min, the CPU time limit of 7200 seconds was reached in all four clusters. For a window size of 120, 2 out of the 4 runs achieved the time limit before optimality could be proved. Still, a feasible solution has been found in both cases, with only 10% of the sessions being booked outside the desired window for the 120-min case, and approximately 29% time preference breaching rate when the window size is equal to 90 minutes. An opposite phenomenon is observed when the probabilities are set to [12.5%, 75%, 12.5%], with an ideal solution (all sessions within the desired time window) being found for the 150 and 120-min time window sizes, and only 0.5% of sessions being scheduled outside the time window for the 90-min case. All the instances for this last probability breakdown were solved in under a minute of total CPU time.

5.5 Discussion

The proposed MILP model has proven to be efficient in achieving an optimal solution for small instances of up to 66 patients and two linacs. For larger cancer centers (three linacs or more), the combination of the pre-assignment heuristic and the MILP model was able to provide near-optimal solutions (maximum of 6.2% optimality gap) quickly (less than 1 hour), thus ensuring the scalability of our combined approach. Given the elevated fulfillment rates provided by these solutions, RT centers may opt for the combined approach in order to ensure low computation times without loosing significant levels of solution quality. By running the combined approach for the NKI size considering pre-allocation constraints and patients with ongoing treatments, the combination between the MILP model and the heuristic procedure allowed to find a solution in which as few as 2.8% of the sessions are scheduled outside the preference time window in around 3.5 hours of CPU time. The positive performance of our methodology when partitioning the problem suggests that large RT centers should divide the main problem in subproblems (subsets of linacs) and use the MILP model to solve each subproblem separately. Centers may split the fleet of linacs based on their location, technological specifications (e.g. cone-beam CT embedded or not) or based on staff planning. Although the obtained solution for the original problem may not be proven optimal, we believe that our methodology can be effective in generating a (near-)optimal schedule in due time for most real-world RT centers. According to the International Atomic Energy Agency (IAEA), the NKI ranks amongst the largest RT centers in the Netherlands with a total of 10 radiation machines (orthovoltage machines included) registered by 2017 [46], just below the Erasmus University Medical Center (12) and the University Medical Centre Utrecht (11) but above all the other 21 Dutch centers. Comparing with the US, the NKI size matches those of the largest cancer centers, paired with e.g. the Stanford Hospital (11) and the New York Memorial Sloan-Kettering Cancer (10).

A sensitivity analysis revealed that the larger the preferential time window, the easier it is for our approach to fit the irradiation sessions within the corresponding window preferences. Although enlarging the window can be seen as an advantage from a model viewpoint, we lack evidence on whether the patient would still be satisfied with such window size. Moreover, it has been verified that when the competition for the same time window increases (from 25% to 50% of the patients), the size of the time window to be chosen must be enlarged (from 90 to 150 minutes) in order to keep the computation times low. RT centers using the proposed approach should then monitor the percentage of patients asking for their sessions to be scheduled in the same time period, and re-dimension the time window put available accordingly.

While we believe our model captures the operational constraints encountered in the vast majority of RT centers, additional features may need to be considered before a practical implementation. For instance, it is known that most patients need to have a weekly consultation with the radiation oncologist in charge of the follow-up during the course of their treatment. A possible extension of the model could be to include the availability of doctors' agenda to ensure a proper coordination between the weekly consultation and (one of the) treatment sessions. Moreover, linacs need to undergo maintenance on a certain frequency basis. Maintenance operations are usually undertaken during office hours, with the linac under maintenance being replaced by a "back-up" linac. In the NKI, this linac is not able to treat certain care plans, since it does not have an embedded cone-beam CT scanner. Therefore, an operational offline re-scheduling of patients with a proper allocation to other machines may be needed when implementing the solution output by our model.

In case some restrictions (e.g. constraints (5.12) or (5.15)) are not part of the planning process of the RT center interested in using our methodology, the MILP model can be easily manipulated to account for those differences by relaxing (excluding) or adding constraints before solving the problem. Furthermore, the solution found by our model can be merely used as a basis where adaptations that fit specific needs of individual patients can be integrated to build a more robust, personalized solution. On the other hand, collection of real data and information regarding patient preferences would further increase the robustness of our solutions. Although planners and clinicians have provided insights on the usual requests asked by patients based on their experience, real data on patient perspectives regarding the desired time window sizes and actual time preferences for appointment times would allow for more concrete and realistic conclusions. For instance, some patients may be more interested in consistency amongst the appointment times of their treatment sessions rather than specific time preferences for their appointments. Other patients may not have a preference at all, being more interested in starting treatment as soon as possible. In these cases, one may use the extra flexibility associated with those patients to further improve the fulfilment of requests of patients who actually have a preference. Moreover, in case some RT centers do not consider patient preferences when scheduling irradiation sessions, they may still use our approach to ensure consistency between appointment times. For instance, a "fictitious" time window of a pre-defined size may be chosen for each patient before running the MILP model, which then outputs a solution with the desired degree of consistency amongst appointment times for each patient. A possible extension of our approach could be to apply a formulation inspired in the Multi-Mode Resource Constrained Project Scheduling Problem (MMRCPSP) [21]. For example, the linacs could be modelled as a renewable resource with a fixed capacity, and variables X decide upon a "mode", defined as a combination of linacs and time slots. Time window violations could, for instance, be calculated using the finish-start precedence relations between activities, computed as time increments.

Efforts to assess the practical feasibility of the obtained solutions need to be performed by RT managers and/or planners before using the model in practice. This and other implementation steps are currently being made together with department managers and clinicians of the NKI in order to perform an implementation of the proposed methodology in the clinic.

5.6 Conclusions

Earlier research on the problem of scheduling RT sessions considering time windows is scarce, with existing models being developed in the context of particle therapy, which makes them directly non-applicable to conventional RT. In this study, we propose a MILP model that is able to solve the RT scheduling problem to optimality in reasonable computation time for RT centers with up to two linacs. For RT centers with three linacs or more, we propose a heuristic procedure that is capable of pre-assigning patients to linacs while maintaining a balanced workload between linacs. Combining the pre-assignment heuristic with the MILP model allowed to solve the problem in less than 3.5 hours of CPU time with 97.2% of the sessions scheduled within the desired time window for a large cancer center operating with eight linacs.

Besides providing automated decision making for scheduling RT treatments which allows managers and planners of RT centers to save time and effort during the scheduling process, our algorithm is capable of incorporating patient preferences while ensuring that all timeliness, medical and technical constraints are take into account. Since the modeled problem and corresponding assumptions are standard among RT centers and the patient mix at the NKI is representative of the patient population found in RT in general, our methodology can be generally applied to RT centers.

5.7 Appendix

Table 5.9 Pre-allocation of care plans to linacs in the RT department of the NKI (2017).

ID	Careplan	# patients	L1	L2	L3	L4	L5	L6	L7	L8
1	Anus +/- inguinal lymph node	22	X	X	X	X	X	X	X	X
2	Adrenal Stereotaxic	4	X	X		X				
3	Bladder	43		X						X
4	Bladder (partial)	22		X						X
5	Chest wall	45	X	X	X	X	X	X	X	X
6	Chest wall+Axilla	114	X	X	X	X	X	X	X	X
7	Chest wall+Axilla+Parasternal	15	X							
8	Chest wall+Parasternal	3	X							
9	Chest wall (bsu)	2							X	
10	Bone metastasis	1099	X	X	X	X	X	X		
11	Bone metastasis Stereotaxic	52	X	X		X				X
12	Uterus	42		X		X				
13	Endometrium	16	X	X		X				X
14	Neck (bsu)	9	X		X	X	X	X	X	X
15	Brain 1 fraction	107		X					X	X
16	Brain several fractions	86		X	X				X	X
17	Brain (whole)	159	X	X	X	X	X	X	X	X
18	Brain Electrons	6	X							
19	Lymph node Stereotaxic	17	X	X		X				X
20	Head-and-neck	230	X		X	X	X	X		
21	Head-and-neck (palliative)	4	X		X	X	X	X		
22	Larynx 2vs	2		X						
23	Liver	6	X	X		X				
24	Lung (palliative)	61	X	X	X	X	X	X	X	X
25	Lung	273	X	X	X	X	X	X	X	X
26	Lung (bsu)	47	X	X	X	X	X	X	X	X
27	Lung Stereotaxic	206	X	X		X	X	X		X
28	Lymphoma	44	X	X	X	X	X	X	X	X
29	Lymphoma (bsu)	19	X	X	X	X	X	X	X	X
30	Stomach	14	X	X	X	X	X	X	X	X
31	Breast	777	X	X	X	X	X	X	X	X
32	Breast+Axilla	179	X	X	X	X	X	X	X	X
33	Breast+Axilla+Parasternal	20	X	X	X	X	X	X	X	X
34	Breast+Parasternal	1	X	X	X	X	X	X	X	X
35	Esophagus	77	X	X	X	X	X	X	X	X
36	Esophagus (palliative)	22	X	X	X	X	X	X	X	X
37	Axilla (virtual)	5	X	X	X	X	X	X	X	X
38	Orbit (eye socket)	1		X						X
39	Ovaries	5	X	X		X				X
40	Others	183	X	X	X	X	X	X	X	X
41	Others (bsu)	60	X	X	X	X	X	X	X	X
42	PAO (+/- iliac single-sided)	1	X	X		X				X
43	Penis	11	X	X		X				X
44	Prostate	243	X	X	X	X	X	X	X	X
45	Prostate+Pelvic lymph nodes	60	X	X	X	X	X	X	X	X
46	Prostatic bed	36	X	X	X	X	X	X	X	X
47	Prostatic bed+Pelvic lymph nodes	25	X	X	X	X	X	X	X	X
48	Rectum / Sigmoid	87	X	X	X	X	X	X	X	X
49	Rectum 13 x 3 Gy	5	X	X	X	X	X	X	X	X
50	Rectum 5 x 5 Gy	56			X		X		X	
51	Sarcoma abdominal/thoracic wall	13	X		X	X	X	X	X	X
52	Sarcoma extremity	28	X		X	X	X	X	X	
53	Sarcoma retroperitoneal	9	X		X	X	X	X	X	
54	Spinal cord	33		X						
55	Vagina	5	X	X		X				X
56	Vulva +/- inguinal lymph nodes	9	X	X	X	X	X	X	X	X
Total		**4720**								

CHAPTER 6

Radiotherapy treatment scheduling: implementing operations research into clinical practice

6.1 Introduction

Radiotherapy (RT), as a treatment modality for cancer care, has been experiencing a rising global demand [14]. In Europe, the number of radiotherapy treatments has been estimated to increase by 16% until 2025 [8]. External beam RT treatments are delivered by a set of linear accelerators (linacs) in a series of (daily) radiation sessions of 10-30 minutes each. With the growing demand for RT services, the number of treatment sessions to be booked amongst the available machines has been continuously increasing. This makes the problem of scheduling RT sessions increasingly complex for RT centers, which aim at managing their resources in the most efficient manner in order to provide patient-centered care while keeping waiting times low. Therefore, several indicators, possibly distinct between RT centers, must be considered when designing a methodology for scheduling RT treatments, with timeliness, patient-centeredness, and staff satisfaction being amongst the most important ones [108]. Timeliness is crucial as long waiting times (time from referral to start of treatment) have been linked with, amongst others, higher risk of local recurrence [20], and prolonged psychological distress in cancer patients [63]. Patient-centeredness relates to not only providing the best care for each specific patient based on their personal and medical needs, but also relates to maximizing patient satisfaction levels regarding their treatment options. Results of a survey amongst six RT centers showed that patients have different preferences regarding the time of their treatment appointments, and that approximately 80% of the patients preferred a 2-week or shorter interval between referral and first oncologic consultation [70]. From a patient's perspective, research shows that the professional staffing standards and low waiting times for both diagnosis and treatment are the most important factors [71]. To keep quality of labor high, a predictable work schedule and an appropriate amount of assigned workload are necessary from a staff viewpoint. Every week, RT centers are faced with the problem of scheduling hundreds of treatment sessions on the available linacs

[72]. Given the high number of technical and medical constraints to be considered for each patient (e.g. start treatment within the due date, patient allocation to specific machines, etc.), the manual execution of such a schedule is a difficult, time-consuming task that often leads to solutions that are far from optimal. Therefore, the development, validation, and implementation of scheduling algorithms can be a solution to support RT centers to schedule radiation sessions in an optimized manner regarding the relevant performance indicators.

Operations research (OR) is a discipline that includes a range of techniques aimed at improving decision-making processes in many areas. Combining knowledge from applied mathematics, computer science, and industrial engineering, OR methods such as computer simulation [106] and mathematical programming [105] have been widely used to propose solutions for complex real-world problems, including the healthcare field [83, 89]. Recent studies have shown that there is very little scientific reporting on implementation of OR-based algorithms in clinical practice [13, 100]. Moreover, several OR-based models have been developed to solve RT treatment scheduling problems, as shown by a literature review by Vieira *et al*. [107]. However, their review also revealed that none of the 18 reviewed papers reported a (pre-)implementation of the model, suggesting that implementation rates of OR approaches in RT are rather low. The inherent complex nature of the optimization problems, the impact on organizational changes, the involvement of several specialized personnel (such as operations research specialists, IT, managers and clinicians) and the need for developing a stable, user-friendly and updated decision support system contribute to the challenging nature of the implementation process. A scoping review found that poor availability of representative data of sufficient quality, and a lack of collaboration between those who develop OR models and relevant internal stakeholders were found to be common challenges for effective OR modelling in global health [9].

In this paper, we study the implementation of a weekly schedule for radiotherapy treatments obtained by using an earlier developed mathematical programming model [104] in two different large Dutch RT centers. The model is iteratively adjusted together with planners and clinicians to meet the constraints and objectives of each RT center, with the resulting computer program being able to find a good quality schedule in reasonable computation time. We compare the performance of the final schedule output by the model with the solution that was manually constructed in practice and elaborate on the main limitations found during the model adjustment process. By having the resulting schedules validated by the actual decision-makers in the clinics, we show that it is feasible to actually bring OR models towards implementation in clinical practice.

6.2 Methods

In this study, we have used and adapted the mixed-integer linear programming (MILP) model proposed by Vieira *et al*. [104] to improve the RT treatment scheduling based on historic off-line data of two Dutch RT centers. MILP is a mathematical programming framework that involves the use of integer and continuous variables to model a decision problem by means of linear inequalities [95]. The process of adjusting the existent model to ensure that it produces a feasible schedule for each RT center was completed iteratively following a structured sequence of steps (Figure 6.1). We started by gathering the necessary patient data for running the MILP model, and discussion meetings were performed to check on the feasibility of each obtained schedule. If modifications were necessary due to infeasibilities found in the output schedule as a result from model deficiencies, the MILP model was corrected. When changes on the MILP model were not possible, the necessary changes were then completed manually by the OR specialist and the planner in order to ensure that all constraints are met for all patients, further guaranteeing the feasibility of the planned schedule.

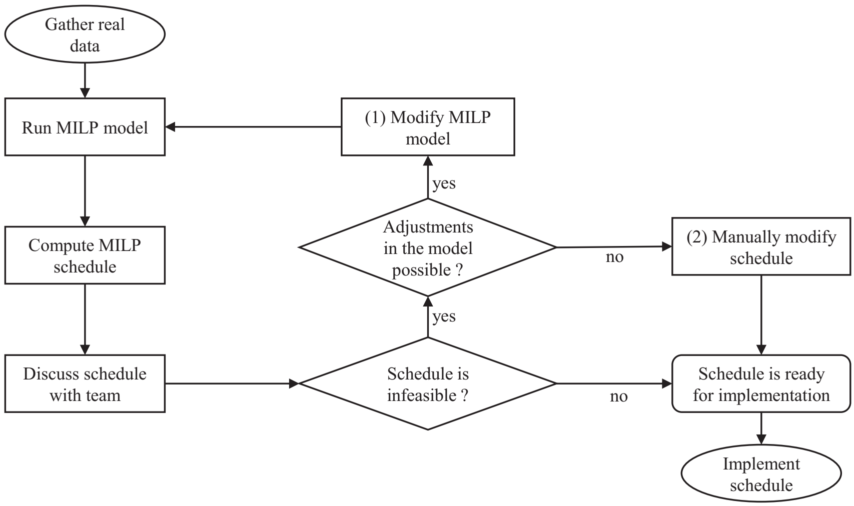


Figure 6.1 Schematic overview of the MILP-based schedule construction and validation phases before implementation.

6.2.1 The radiotherapy centers

Netherlands Cancer Institute–Antoni van Leeuwenhoek Hospital (NKI)

The NKI is a comprehensive cancer center that combines research and care, located in Amsterdam, the Netherlands. The RT department of the NKI treats approximately 5000 patients per year and is equipped with eight linacs divided over two locations, Amsterdam (six) and Hoofddorp (two). The department focuses its operations on patient-centered and personalized quality treatments, highly adapted to the individual. Therefore, the number of possible patient care pathways, as well as the number of different healthcare professionals (e.g. specialized RT technicians, dietitians, dentists, etc.) involved in the process is rather high when compared to an average RT center. By combining research and care, the RT department of the NKI aspires to be a front runner in the development and adoption of innovative (information) technologies.

Bernard Verbeeten Instituut (BVI)

The Bernard Verbeeten Instituut (BVI) is an independent RT group delivering RT treatments to patients living in the South of the Netherlands. The BVI provides around 4700 treatments per year. The main location, in Tilburg, is equipped with four linacs, while the satellite locations in Den Bosch and Breda have two linacs each. BVI is highly focused on patient care logistics and efficiency, with the goal of delivering timely, good quality treatments in a patient-friendly manner. Being early adaptors of emerging technologies, BVI has been active in fostering and implementing interventions for logistics improvement (e.g. Lean methodologies) in their clinic over the last decade. We take the Breda location as the subject of our study.

6.2.2 Information gathering

We gathered information regarding the problem specificities, boundary conditions, and objectives of each participating RT center via a series of meetings held with personnel from each institute. The NKI, being the institution that hosts the project, has allowed a higher number of professionals to be involved in the process. A group consisting of an OR specialist, a radiation oncologist, an appointment planner, the head of the appointment office, and a medical physicist were involved in the project, in which only the Amsterdam location (six linacs) was considered. In the BVI, a logistics specialist and the head of medical physicists have been working together with an OR specialist to adjust and validate the model for the Breda location (two linacs). A series of five face to face meetings with members from both centers, dozens of "on-site" interviews with managers and clinicians, and hundreds of e-mail exchanges clarifying details over the course of two years, facilitated the necessary steps (problem definition, data gathering, model development/adjustment, and schedule validation) for the application of the OR model.

6.2.3 Mathematical model description

The MILP model outputs a schedule for treatment sessions of a pre-defined number of patients amongst a set of linear accelerators with a certain daily capacity, divided in time slots of five minutes. The model decides upon the specific time slot(s) of each linac to allocate to each patient on each day of a planning horizon of one week (five working days). The full model, together with the description of the specific constraints, decision variables, and input parameters can be found in Appendix 6.5. The model has been adapted to represent the problem specificities of each RT center by formulating an objective function according to the main objective of the center, and by removing or modifying constraints. In the NKI case, patients are given the possibility of choosing a preferential time window for the time of their treatment sessions. The objective of the NKI is then to find a schedule that maximizes the fulfillment of those patient preferences requests [104]. Thus, we use objective function (6.1) of the MILP model (Appendix 6.5) to maximize the number of sessions that are scheduled within the time window given by patients as a preference. All constraints (6.3-6.20) are applicable to the NKI case. In the BVI, however, patient preferences are not explicitly asked to patients to be considered in the scheduling process. In their case, the main objective is to find a schedule that minimizes the number of gaps (15 minutes or longer) in the schedule in order to minimize staff idle times. Thus, objective function (6.2) is used instead. Moreover, constraints (6.6)-(6.8) and (6.18), as well as variables $\Delta_{it}^{-}, \Delta_{it}^{+}$ are discarded when running the MILP model for the BVI case.

6.2.4 Input data and planning horizon

Anonymized historical patient data has been extracted from the patient information systems of each center: MOSAIQ [33] in the NKI, and ARIA [103] in the BVI case. Table 6.1 presents the summary of the input data used to feed the MILP model for the targeted planning week in which the development and validation steps were performed. Patient data extracted from the databases correspond to the necessary input parameters needed to run the model, as presented in Appendix 6.5. All external-beam RT patients (regular and urgent) known by the Friday prior to the planning week were included in the patient group to be scheduled. The due date of each patient is set according to the national timeliness standards [68]. Linacs down times for maintenance were also considered for both centers, with an average of 2.5 hours for scheduled maintenance per linac, per week. In the NKI, distinct linac specifications required by certain patient groups exist. Therefore, a pre-allocation of patient groups to certain linacs was considered for their case (e.g. bladder patients can only be assigned to linacs number 2 and number 6). Since patient preferences regarding appointment times were not recorded in actual practice in the NKI, their values are unknown. Therefore, we have interviewed actual planners to estimate patient preferences according to their empirical knowledge, resulting in the proportion and time windows shown in Table 6.1. Since maximizing the satisfaction of patient preference requests is not an objective for the BVI, the corresponding data has not been collected or estimated.

Table 6.1 Data used as model inputs for the computational experiments of each center

	NKI	BVI
Week (planning horizon)	24-28 June 2019	9-13 December 2019
Number of patients to be scheduled	213	48
Number of patients starting treatment	48	6
Linacs	6	2
No. sessions scheduled	807	232
Linac opening times (daily capacity per linac)	07h30 – 17h30 (600 min)	08h00 – 17h30 (570 min)
Average session duration, in minutes (SD)	14.7 (3.8)	13.3 (3.1)
Patient preference distribution	07h30-09h30 (10%) 09h30-15h30 (60%) 15h30-17h30 (30%)	n.a.

6.2.5 Performance metrics

Key Performance Indicators (KPIs) have been assessed through interviews conducted with the actual planners and managers of each center. In both cases, the KPIs found to be monitored during the schedule construction method are related to patient-centeredness, timeliness, and staff satisfaction (Table 6.2). The identified KPIs were modeled either in the form of the objective function being optimized (e.g. minimize gaps in the schedule), or in the form of constraints (e.g. ensure that all patients start treatment within the target date according to the national timeliness standards).

6.2.6 Model validation

The schedules generated by the various versions of the MILP model have been validated by the planners of the corresponding center by cross-checking actual patient and resource information data to confirm that the final schedule is ready for implementation. Since the task of assessing all technical and medical constraints for all patients showed to be too complex and time-consuming, we have selected a random sample of 20 patients in the NKI, and 10 patients in the BVI, from several tumor sites, to perform the validation check. We continued to perform model iterations until all inconsistencies with daily practice were tackled and no further improvements were considered likely by the responding teams. This resulted in the final MILP model.

For the final test, the treatment requirements (e.g. number of sessions, frequency, due date, etc.) associated with each of the corresponding sample of patients have been validated by matching them with actual values. Moreover, the

Table 6.2 Key Performance Indicators defined by the RT centers.

KPI	Description	RT center (s)
Number of gaps in the schedule	Minimize the number of empty time slots between appointments to avoid staff idle times (only time intervals of 15 minutes or longer are considered)	BVI
Number of sessions within desired window	Maximize the fulfillment of patient requests regarding the time window indicated as the most preferred: 07h90-09h30, 09h30-15h30, or 15h30-17h30	NKI
Consistency of appointments	Minimize the (average) deviation between sessions, in order to allow the patient set up a routine for their sessions	BVI, NKI
Number of patients switching linacs during treatment	Minimize the number of patients switching between different linacs in order to have the patient receive treatment with the same personnel and facilities	BVI, NKI
Fulfillment of the maximum waiting time targets defined by the Dutch Society for Radiation Oncology [68]	Ensure that patients fulfill the national standards: acute patients should be treated within one day, subacute patients should start treatment within 10 days, and regular patients should start treatment within 28 days.	BVI, NKI

weekly schedule of each linac produced by the final model has been analyzed by the managers and planners of each center in order to look for possible infeasibilities regarding the distribution of appointment sessions, idle times, linacs' down times for maintenance, or the management of staff. The schedules were considered valid for implementation after this check.

6.3 Results

During the discussion meetings undertaken together with planners and clinicians throughout the schedule construction process (Figure 6.1), practical infeasibilities of the schedule output by the MILP model were identified (and solved) in an iterative manner. In order to overcome those limitations, the MILP model structure and/or input parameters were modified accordingly until an implementable final schedule could be obtained. Table 6.3 shows the main infeasibilities and limitations found during the schedule construction process, as well as the solutions found to overcome them. We use the notation and constraints of the MILP model (Appendix 6.5) to describe the solutions.

Table 6.4 and Table 6.5 show the performance of the final schedule obtained by the MILP model after it has been considered validated by each RT center, in comparison with the solution manually built and implemented in practice by the corresponding center on that same week. In either test case, all new patients started treatment within their target due date, both in practice and in the

Table 6.3 Main infeasibilities and limitations found during the schedule construction method and corresponding model adjustments undertaken to overcome them.

Unfeasibility / limitation	Solution found to overcome limitation	RT center(s)
Linacs' maintenance times	Set linac as unavailable during maintenance hours using parameters a_{kst}	NKI, BVI
Treatment sessions of some patient must fall within a restricted time window (e.g. 08h30-17h00) in order to guarantee that specialized staff are present during the sessions of the applicable patients	Add constraints (6.17)	NKI, BVI
Patients with 2 sessions per day (6h of interval is necessary between same-day sessions)	Create one "fictitious" patient and insert time window constraints between the daily sessions of those two patients (using constraints (6.17))	BVI, NKI
Certain linac(s) cannot be assigned to certain patient(s) (e.g. due to lack of cone-beam CT technology)	Update $\mathcal{K}^i$ accordingly	NKI
Staff breaks	Set linac as unavailable during staff breaks by setting parameters a_{kst} accordingly	BVI
Account for treatments that need to start on a Monday	Set due date d_i to Monday	NKI, BVI
Maximum number of patients starting treatment in the same linac, same day is limited to 6	Add constraints (6.6)-(6.8) with $C = 6$	NKI
5-min empty time slot at the end of every hour to accommodate possible delays	Set the last time slot of every hour as unavailable (using a_{kst}) in each linac	BVI

model's solution. Moreover, the number of patients switching between different linacs throughout the course of their treatment decreased from 71 to 0 (NKI), and from 43 to 2 (BVI) with our model. In the NKI case, all performance indicators improved substantially, although a fewer number of sessions have been scheduled by the model (795) than was done in practice (807). It should be noted that also in this case, the timeliness requirements were met, also for the patients that were to be scheduled in the succeeding week.

Although data regarding the fulfillment of patient preferences have so far not been recorded in practice, the model was able to schedule 98% of the treatment sessions within the desired time window. Moreover, time variation amongst appointment times for the same patient decreased by 51% from an average SD of 103.0 to 50.4 minutes. A total of 1 hour and 27 minutes of computation time was needed to obtain the final schedule for the NKI case.

Table 6.4 Performance comparison between the schedule built manually in practice and the schedule obtained by using the final MILP model in the NKI.

	Practice	Model
No. sessions scheduled	807	795
No. new patients starting treatment within national targets	48 (100%)	48 (100%)
No. sessions within the time window requested by patient	n.a.	776 (97.6%)
Average standard deviation between sessions (min)	103.0	50.4
No. patients switching linacs	71	0

Results for the BVI case show that the number of sessions scheduled by the final MILP model matched the ones scheduled in practice at 232. The solution obtained by using the model was able to provide a more compact schedule with as few as 5 gaps, as opposed to 18 gaps counted in the solution produced by the actual planners. The MILP model was also able to decrease the average variation between appointments times by 43% from 94.8 to 54.5, thus improving the consistency of appointment times. For the BVI case, the solution could be found in just 5 minutes of computation time.

6.4 Discussion

The use of a mathematical OR model allowed to find a solution for the RT scheduling problem in both centers that proved both feasible and acceptable to the local staff. Several adjustments, performed in a stepwise manner, were necessary in order to improve and validate the MILP model before the output solution could be considered implementable. We could improve the weekly schedule for radiation sessions while optimizing the measured KPIs in both centers. In the NKI case, fewer sessions have been scheduled by the model than in practice (795 vs 807). We found it impossible to enter this “fast planning” decisions that are based on personal judgement in the model.

Table 6.5 Performance comparison between the schedule constructed manually in practice and the schedule obtained by using the final MILP model in the BVI.

	Practice	Model
No. sessions scheduled	232	232
No. new patients starting treatment within national targets	6 (100%)	6 (100%)
No. gaps in the schedule	18	5
Average standard deviation between sessions (min)	94.8	54.5
No. patients switching linacs	43	2

Although actual patient preferences could not be obtained from the NKI clinic as they were not being recorded in practice, our model was able to fulfill the estimated preferences a total of 98% of the times. A sensitivity analysis on the impact of varying the preferential time window distributions may be useful to further assess the robustness of the model [104]. Moreover, the number of gaps in the schedule decreased by 72% in the BVI, with the daily number of treatment sessions being delivered earlier in the day during the whole week. This allows staff members to dedicate part of their time to other tasks. The consistency amongst the appointment times has also improved considerably with our MILP model, with the average standard deviation between appointment times decreasing by 51% and 43% for the NKI and BVI cases, respectively. This offers much more consistency for patients planning their daily calendar.

Our tool has proved effective in reducing the planning time needed to construct a schedule for RT treatment sessions. In practice, this task commonly takes 1.5 days of planning work, while with our tool this time could be lowered to a maximum of 1.5 hours (NKI) plus the time needed to manually adjust the schedule, which we estimate to be a maximum of 1-2 hours. In very large RT centers where problem size increases complexity to a significant degree, a heuristic procedure may be applied to pre-allocate patients to (subgroups of) linacs before the MILP model is applied to solve the problem for each subgroup of linacs [104]. By providing a decision-making tool that generates a schedule in an automated way, we not only save planners the burden of having to produce a feasible schedule in due time, but also allows them to prioritize their activities more easily.

Although the application of the mathematical programming model resulted in substantial performance improvements, the planned schedule can still be further (manually) improved *a posteriori*. This may in fact be necessary to address the needs of specific patients, such as the coordination between the radiation sessions and other appointments that patients may need. For instance, patients are often scheduled for a weekly follow-up consultation with his/her doctor in a one stop shop planning, preferably with a minimal time deviation from the start/end of the radiation session. Gradually, the coordination between treatment sessions and other appointment types (e.g. chemotherapy appointments, blood analysis, a follow-up CT scan, mouth hygienist, etc.) can be added to the MILP model for the development of a more holistic planning tool. We estimate that these changes would require an additional planning time ranging between one and two hours for manual rescheduling of some sessions.

While healthcare institutions have to strive for efficiency and stakeholders demand excellence in the delivery of care, we conclude that operations research tools can certainly be considered for implementation as they have demonstrated to be able to contribute to improved performance of treatment facilities. We achieved this by supporting planners with (theoretical) evidence-based tools and took them along in a stepwise and interactive implementation process, overcoming the use of traditional planning methods that provide sub-optimal solutions for both patients and resources [98]. By achieving a schedule that has been verified by the actual planners until it was considered ready for implementation, we take a significant step in the implementation of OR models in clinical practice compared to the current literature [100]. In our experience, a gradual development/adjustment of the OR models, in small steps, is recommended for a smooth translation of those models into the clinic. Moreover, the engagement of all stakeholders from the very beginning of the study has allowed to create a collaborative environment based on constant communication and mutual confidence that were crucial for the realization of the implementation steps. Bringing the OR specialists, planners, managers and clinicians together as part of a project team with regular meetings, where model adaptations and new results have been made easy to visualize and interpret, has helped bridge the gap between the different professionals involved.

Since the RT scheduling processes designed for this study are rather standard, we believe that our intervention would produce similar results in other RT centers. Moreover, the process of adjustment and validation of OR models proposed in this work can be used by healthcare institutions to promote implementation efforts of OR-based methodologies into clinical practice.

The application of an OR model for RT treatment scheduling has proven to be capable of supporting RT planners produce a high-quality schedule that satisfies all medical and technical constraints in a much faster way than currently done in practice. In our study, the early engagement of managers and clinicians, staying in close contact with stakeholders, and performing stepwise model adaptions in small steps proved to be the most important factors to facilitate the implementation of an OR model in the practice setting.

6.5 Appendix

In this appendix, we present the mixed-integer linear programming (MILP) model used to find a feasible and optimized schedule for the RT scheduling problem. The capacity of each linac is divided in time slots $s = 1, \ldots, |\mathcal{S}|$ of a pre-defined, fixed duration l. Patients' sessions, when scheduled, are assigned to a certain starting time slot on a certain linac and day. If a certain starting slot is assigned to a patient, we prevent the next slots needed to achieve the corresponding patient's session duration on that same linac and day from being assigned to other patients.

6.5.1 Input parameters

We use the notation and input parameters of Table 6.6 to formulate the problem.

Table 6.6 Notation and input parameters of the MILP model.

Parameter	Description
$\mathcal{P}$	set of patients to be scheduled ($i, j \in \mathcal{P}$)
$\mathcal{K}$	set of linear accelerators ($k \in \mathcal{K}$)
$\mathcal{S}$	set of time slots per linac ($s \in \mathcal{S}$)
$\mathcal{T}$	set of time periods (days) in the planning horizon ($t \in \mathcal{T}$)
$\mathcal{P}^n$	set of patients who have not started treatment ($\mathcal{P}^n \subseteq \mathcal{P}$)
$\mathcal{P}^m$	set of patients who must start treatment on monday ($\mathcal{P}^m \subseteq \mathcal{P}$)
$\mathcal{P}^f$	set of patients with restricted time frame for treatment sessions ($\mathcal{P}^f \subseteq \mathcal{P}$)
$\mathcal{K}^i$	set of feasible linacs for treating patient i ($\mathcal{K}^i \subseteq \mathcal{K}$)
T	maximum number of patients starting treatment in the same linac and same day
l	time slot duration, in minutes, in each linac, each day
a_{kst}	1 if slot s of linac k is available on time period t, 0 otherwise
$\underline{f}^t, \overline{f}^t$	lower and upper bound of the restricted time frame set for time period t
I_i	number of total remaining sessions to be delivered to patient i
d_i	due date: time period by which patient i must start treatment
p_i	duration, in number of time slots, of each session of patient i
b_i	number of time periods needed between sessions of patient i (1 for consecutive daily sessions)
$t_i^{\min}, t_i^{\max}$	lower and upper bound of the time window preference for patient i
c_i	linac in which patient $i \notin \mathcal{P}^n$ is currently undergoing treatment

6.5.2 Decision variables

We use a set of binary variables x_{iks}^t, which take the value 1 if patient i is scheduled for a session starting on time slot s of linac k, in day t, and 0 otherwise. Real variables Δ_{it}^- and Δ_{it}^+ are used to represent the deviations from the intended sessions' starting time for each patient, in each day. Binary variables y_{ik}^t are auxiliary variables, which will be equal to 1 if a new patient starts his/her treatment in period t and linac k, and 0 otherwise. An overview of the decision variables is presented in Table 6.7.

Table 6.7 Decision variables of the MILP model.

Variable	Description
x_{iks}^t	1 if patient i is scheduled a session starting on time slot s of linac k in day t, 0 otherwise
y_{ik}^t	1 if new patient i starts treatment in period t and linac k, 0 otherwise
$\Delta_{it}^-, \Delta_{it}^+$	lower and upper deviation, in minutes, from preference time window of patient i in time period t

6.5.3 Objective function(s)

The objective function to be used in the MILP model will depend on the goal(s) of each RT center:

Minimize the overall deviation between the bounds of the preferred time window $[t_i^{\min}, t_i^{\max}]$ *given by patients and the starting time of their appointments:*

$$\min \sum_{i \in \mathcal{P}} \sum_{t \in \mathcal{T}} \left(\Delta_{it}^- + \Delta_{it}^+\right) \tag{6.1}$$

Minimize the overall starting time of irradiation sessions (thus scheduling them together and as early as possible):

$$\min \sum_{i\in\mathcal{P}} \sum_{k\in\mathcal{K}} \sum_{s\in\mathcal{S}} \sum_{t\in\mathcal{T}} s.x_{iks}^{t} \tag{6.2}$$

6.5.4 Constraints

The values of the decision variables are bounded by a set of linear inequalities, which represent the practical constraints of the problem, as follows:

Limit the number of sessions that each patient can receive to a maximum of one per day:

$$\sum_{k\in\mathcal{K}} \sum_{s\in\mathcal{S}} x_{iks}^{t} \leq 1, \forall i \in \mathcal{P}, \forall t \in \mathcal{T} \tag{6.3}$$

Ensure that each (available) slot of each linac is scheduled at most one session per day:

$$\sum_{i\in\mathcal{P}} x_{iks}^{t} \leq a_{kst}, \forall k \in \mathcal{K}, \forall s \in \mathcal{S}, \forall t \in \mathcal{T} \tag{6.4}$$

Each patient is assigned to a feasible linac, by preventing sessions of being assigned to slots of linacs that do not belong to $\mathcal{K}^i$:

$$\sum_{s\in\mathcal{S}} \sum_{t\in\mathcal{T}} x_{iks}^{t} \leq 0, \forall i \in \mathcal{P}, \forall k \in \mathcal{K} \setminus \{\mathcal{K}^i\} \tag{6.5}$$

Force y_{ik}^t variables to take the value 1 if a new patient i starts treatment on linac k and day t:

$$y_{ik}^{t} \geq \sum_{s\in\mathcal{S}} x_{iks}^{t} - \sum_{s\in\mathcal{S}} x_{iks}^{t'}, \forall i \in \mathcal{P}^n, \forall k \in \mathcal{K}, \forall t = 2, .., \mathcal{T},$$
$$t' = \max\{1, t - b_i\} \tag{6.6}$$

$$y_{ik}^{1} \geq \sum_{s\in\mathcal{S}} x_{iks}^{1}, \forall i \in \mathcal{P}^n, \forall k \in \mathcal{K} \tag{6.7}$$

Limit the number of patients starting treatment in the same linac and same day to the pre-defined threshold C:

$$\sum_{i\in\mathcal{P}\setminus} y_{ik}^{t} \leq C, \forall k \in \mathcal{K}, \forall t \in \mathcal{T} \tag{6.8}$$

Restrict the number of sessions delivered during the planning horizon to the number of remaining sessions for that patient:

$$\sum_{k\in\mathcal{K}} \sum_{s\in\mathcal{S}} \sum_{t\in\mathcal{T}} x_{iks}^{t} \leq I_i, \forall i \in \mathcal{P} \tag{6.9}$$

Ensure that patients receive their sessions on the same linac and with the required frequency b_i until the number of sessions or the end of planning horizon is reached:

$$\sum_{s\in\mathcal{S}} x_{iks}^{t} - \sum_{s\in\mathcal{S}}\sum_{t'=1}^{t-1} x_{iks}^{t'} \leq \sum_{s\in\mathcal{S}} x_{iks}^{n}, \forall i \in \mathcal{P}, \forall k \in \mathcal{K}, \forall t = 2, .., \mathcal{T},$$
$$\forall n = t + b_i, t + 2b_i, \ldots, \min\left\{ |\mathcal{T}|, t + b_i (I_i - 1)\right\} \tag{6.10}$$

$$\sum_{s\in\mathcal{S}} x_{iks}^{1} \leq \sum_{s\in\mathcal{S}} x_{iks}^{n}, \forall i \in \mathcal{P}, \forall k \in \mathcal{K},$$
$$\forall n = b_i + 1, 2b_i + 1, \ldots, \min\left\{ |\mathcal{T}|, b_i (I_i - 1) + 1\right\} \tag{6.11}$$

Force the all the necessary sessions to be booked, at least every b_i days, as soon as a first session is scheduled:

$$1 - \sum_{s\in\mathcal{S}} x_{iks}^{t} \geq \sum_{s\in\mathcal{S}} x_{iks}^{n}, \forall i \in \mathcal{P}, \forall k \in \mathcal{K}, \forall t = 1, .., |\mathcal{T}| - b_i,$$
$$\forall n = t + 1, \ldots, t + b_i - 1, b_i \geq 2 \tag{6.12}$$

Avoid unnecessary sessions from being scheduled in days occurring between the days of the sessions booked by constraints (6.10-6.11) when $b_i > 1$:

$$1 - \sum_{s\in\mathcal{S}} x_{iks}^{t} \geq \sum_{s\in\mathcal{S}} x_{iks}^{n}, \forall i \in \mathcal{P}, \forall k \in \mathcal{K}, \forall t = 1, .., |\mathcal{T}| - b_i,$$
$$\forall n = t + 1, \ldots, t + b_i - 1, b_i \geq 2 \tag{6.13}$$

Impose that every patient starts treatment before their due date d_i (note that for patients starting on Monday we can set $d_i = 1$):

$$\sum_{k\in\mathcal{K}}\sum_{s\in S}\sum_{t=1}^{d_i} x_{iks}^{t} \geq 1, \forall i \in \mathcal{P} \tag{6.14}$$

Prevent the remainder slots needed to achieve the session duration p_i after the chosen starting slot x_{iks}^{t} from being assigned to other patients on the same linac and day:

$$x_{iks}^{t} \leq 1 - \sum_{i'\in\mathcal{P}} x_{i',k,s'}^{t}, \forall i \in \mathcal{P}, \forall k \in \mathcal{K}, \forall s = 1, \ldots, |\mathcal{S}| - p_i + 1, \forall t \in \mathcal{T},$$
$$\forall s' = s + 1, \ldots, s + p_i - 1, p_i \geq 2 \tag{6.15}$$

Ensure that the starting slot of sessions with a duration of two or more slots are not assign to the last slot(s) of the day:

$$x_{iks}^{t} = 0, \forall i \in \mathcal{P}, \forall k \in \mathcal{K}, \forall s = |\mathcal{S}| - p_i + 2, \ldots, S, \forall t \in \mathcal{T}, p_i \geq 2 \tag{6.16}$$

Ensure that sessions of each patient fall within the restricted time frame set by the department due to the need of ensuring that specialized staff are present during the sessions of the applicable patients ($\mathcal{P}^f$):

$$x_{iks}^t \leq 0, \forall i \in \mathcal{P}^f, \forall k \in \mathcal{K}, \forall s \in \mathcal{S}, \forall t \in \mathcal{T}, s < \underline{f}^t, s > \overline{f}^t \tag{6.17}$$

Force variables Δ_{it}^- and Δ_{it}^+ to take a non-zero value if a session's starting time deviates from the desired lower and upper bounds, respectively:

$$t_i^{\min} x_{iks}^t - \Delta_{it}^- \leq l\,(s-1)\, x_{iks}^t \leq t_i^{\max} x_{iks}^t + \Delta_{it}^+, \\ \forall i \in \mathcal{P}, \forall k \in \mathcal{K}, \forall s \in \mathcal{S}, \forall t \in \mathcal{T} \tag{6.18}$$

Non-negativity constraints associated with the real variables:

$$\Delta_{it}^- \geq 0, \Delta_{it}^+ \geq 0, \forall i \in \mathcal{P}, \forall t \in \mathcal{T} \tag{6.19}$$

Binary variables can only take the value 0 or 1:

$$x_{iks}^t, y_{ik}^t \in \ 0, 1, \forall i \in \mathcal{P}, \forall k \in \mathcal{K}, \forall s \in \mathcal{S}, \forall t \in \mathcal{T} \tag{6.20}$$

The MILP model was coded in C++ using Visual Studio 2017 and the Concert Technology of CPLEX v12.8.0, which was used as a solver. All experiments were conducted on a desktop computer with a processor Intel i7 3.6 GHz and 16 GB of RAM using up to 8 threads, running on a 64-bit version of Windows 10. A time limit of 28800 seconds (8 hours) of CPU was set for each computational experiment.

CHAPTER 7

General discussion

In external-beam RT, delays in the start of treatment can negatively affect the patient's outcome and quality of life. Fluctuations in the inflow of new patients combined with highly dedicated care paths per tumor site can cause delays and under/over utilization of resources. This effect can be diminished by the way operations are organized in RT departments. RT operations are tightly and time-wise connected, personnel are highly specialized, and equipment may be technologically dedicated to patients with specific tumor sites. Managing patient flow in such an environment is challenging and is further complicated by the high levels of uncertainty in the process. As the number of patients diagnosed with cancer increases and treatments get more personalized, efficient process management approaches become increasingly complex for RT centers. In this thesis, we have developed several OR models for the logistical optimization of RT processes to support decision-makers plan and use their resources more efficiently. We have proposed innovative approaches for solving logistical problems encountered in the whole RT chain of operations, from pre-treatment to treatment, while optimizing for the most important KPIs related to timeliness and patient-centeredness. A collaboration with six Dutch RT centers has allowed to adjust, validate and test models using real-world information and data.

7.1 Main contributions

There are several contributions provided by this thesis work. We started by performing the first extensive literature review on OR models developed specifically for RT logistics, assessing the current state on especially the poor extent of implementation of those models in practice (Chapter 2). Having identified the most common planning and scheduling problems which had not been addressed in the literature, we propose several OR approaches to solve those "open" logistics problems (Chapter 3, 4, and 5). Chapter 3 presented the first OR model for the optimal allocation of RTTs considering natural variations in patient inflow. In Chapter 4 we perform an innovative study on the impact of different policies (push or pull) for scheduling the start of treatment of RT patients after their arrival, and in Chapter 5 we develop the first OR model where patient preferences regarding the time of their treatment sessions are included in the optimization process. In Chapter 6 we take the current extent of implementation of OR models in RT one step closer to clinical practice by developing, validating and testing the OR model together with planners and managers of two Dutch RT centers. Moreover, the OR models developed and presented in this dissertation address the most relevant KPIs found in RT logistics. Chapters 3 and 4 focus mainly on timeliness, by increasing the percentage of patients starting treatment within the waiting time targets through enhanced staff allocation and improved scheduling policy making. Chapters 5 and 6 improve patient-centeredness by increasing the fulfillment of patient preferences for treatment sessions' times while helping planners find high-quality solutions in a faster way. The minimization of waiting times and the fulfillment of patient preferences are accounted for in the objective function of the mathematical programming models, while other objectives are included in the form of constraints (e.g. have patients receive treatment sessions in the same linac).

7.2 Main findings and implications for policy making

OR for RT logistics

In Chapter 2 we reviewed the current state of literature on existing OR models for resource planning in RT. We found that many models addressed the problem of scheduling patients on treatment machines (in 12 of 33 papers), while little focus (3 publications) has been given to pre-treatment scheduling, despite its impact on timeliness. The lack of studies focusing on the pre-treatment workflow is likely related to the difficulty of modelling certain components such as the non-standardized behavior of some agents (e.g. doctors), or the many small steps and actions that occur in RT clinics (e.g. review of treatment plans, work that needs to be re-done), as demonstrated by Chapters 3 and 4. We have also found that the development of OR methods for RT macro-planning is rather low and can be further extended. Although OR may provide powerful tools for the cost-effective allocation of funds in decisions regarding, for instance, the long-term capacity needs of RT centers [91], we could only find one study regarding facility planning on a macro level.

Although the literature on OR applications in RT covers a wide range of approaches from strategic capacity management to operational scheduling levels, none of the 33 analyzed papers reported a full implementation of the model. This indicates that, unless there is lack of reporting, implementation rates of OR models in RT are low. Despite the potential improvements demonstrated by the results obtained by the OR models, these are still not easily accepted by clinicians [11]. This may indicate that the theoretical gains need to be considerably high in order to trade-off the effort, behavioral change, costs, and time needed to perform the necessary changes in practice. These may require the involvement and collaboration of several professionals with different backgrounds (OR experts, process leaders, software developers, etc.), which further hampers implementation. As we found in the implementation part of our projects, it is important that those stakeholders are kept informed and updated on all milestones of the implementation project. Moreover, we found that 9 of the 33 analyzed studies were tested using fictitious data rather than real data. This indicates that data availability may be another reason that complicates the verification, validation, and consequently the acceptance of OR models that should be further considered.

Pre-treatment workflow: improving timeliness via optimal staff allocation and improved workflow control

Chapter 3 introduced a stochastic MILP model that optimizes the allocation of RTTs to multiple (pre-treatment) operations over a set of scenarios of patient inflow. Given that the amount of workload in each operation (e.g. CT scanning, treatment planning, etc.) depends on the highly variable patient inflow and that RTTs have multiple skills, rotation needs and partial availability, building a schedule for RTTs over a mid-term planning horizon (1-6 months) has shown to be a complex task. If not proven to be valid in practice, this may lead to situations of understaffing, jeopardizing the fulfillment of the patients' waiting time standards. In our approach, several scenarios of patient inflow were generated from historical patient data, and the final RTT allocation covered the workload associated with all scenarios maximizing the (expected) number of patients completing pre-treatment within their waiting time target. Results for a case study in the RT department of the NKI showed that, on average, the number of patients able to start treatment within the maximum waiting time standards may increase from 91.3% to 97.9% for subacute patients, and from 96.3% to 100.0% for regular patients. Although the obtained data showed that the NKI was already close to fulfilling the national waiting time targets for all patients, the study was able to provide an estimation of the best achievable outcome when RTTs are optimally allocated. We can safely presume that the actual results of an implementation of the proposed methodology would be between the fulfillment percentages measured in current practice and the theoretically optimal outcomes. However, the benefits obtained by using our model can go beyond timeliness improvements. In the NKI, the allocation of the 109 RTTs to the modelled operations over a mid-term planning horizon is presently done manually, requiring the involvement of several team leaders and process leaders that need several weeks to produce a feasible schedule. Using the proposed model by means of a software tool that is integrated in the planning software, a high-quality roster can be obtained in an automated way in just over an hour of computational time, releasing staff members to dedicate their time to other (managerial) tasks within the department.

Besides aiding decision making for the direct allocation of the currently available group of RTTs, our model can also provide support in re-dimensioning workforce. RT managers can use the model for running experiments for hypothetically fewer or larger number of RTTs and evaluate the impact on timeliness. In our experiments we verified that the RTT capacity needed to cover the expected demand could be reduced by 5.5%. This means that six RTTs could theoretically be assigned to other tasks that do not belong to the modeled process (e.g. patient education sessions, research, etc.) or the related costs be reduced. Both or either efficiency and effectiveness could be improved in this way.

In Chapter 4 discrete-event simulation was used to model the pre-treatment patient flow of the RT department of the NKI. A staff survey, interviews with managers, and historical data from 2017 were used to generate model inputs considering fluctuations in patient inflow and resource availability. Two strategies were compared: one booking the start of treatment sessions right after consultation (pull), while the other strategy only schedules the starting session when pre-treatment is complete (push). Results showed that a hybrid (40% pull / 60% push) strategy representing the current practice leads to 12% lower average waiting times and 48% fewer first appointment rebooks when compared to a full pull strategy, which in turn leads to 41% fewer patients breaching the waiting time targets. This indicates that, despite a slight increase in waiting times, there is room to schedule all patients on a pull strategy without breaching timeliness targets. Several benefits may emerge from a pull strategy. By having due dates earlier on, workflow control increases by creating the need of tracking the progress of each treatment pathway and constantly ensuring that resources will be available when needed. Moreover, patient satisfaction levels increase by informing them about the start of treatment date right at consultation, possibly allowing patients to participate in the definition of the start date of their treatment according to their preferences.

The process of increasing the number of patients working on a pull fashion can be gradual, possibly per tumor site. For instance, a center may start booking all regular breast patients using a pull strategy as a first trial, as suggested by our experiments. The performance of the implemented changes would certainly give a clearer indication of the characteristics of patient groups which are more susceptible to benefit from such a strategy before choosing the next patient group. Ultimately, our simulation model can be used by radiotherapy policy makers to identify the optimal balance between push and pull strategies to ensure timely treatments while providing patient-centered care adapted to their specific conditions.

Treatment scheduling: from theory to clinical practice

Chapter 5 presented a MILP model that solves the problem of scheduling and sequencing RT sessions considering time window preferences given by patients. With the increase in the incidence of cancer and the utilization rates of linear accelerators, the problem of manually scheduling RT sessions while satisfying patient preferences regarding the time of their appointments becomes increasingly difficult for RT planners. The proposed MILP model alone can solve the problem to optimality in reasonable time for small size instances up to 66 patients and 2 linacs per week. For larger centers, we propose a heuristic method that pre-assigns patients to linacs to decompose the problem in subproblems (clusters of linacs) before using the MILP model to solve each of the subproblems to optimality in a sequential manner. We tested our methodology using real-world data from the RT department of the NKI (8 linacs). Results showed that, combining the heuristic with the MILP model, the problem can be solved in reasonable computation time with as few as 2.8% of the sessions being scheduled outside the desired time window. We found, however, that the size of the (smallest) time window must be enlarged (from 90 up to 150 minutes) when the percentage of patients requesting that same time window increases (from 25% to 50% of the patients) in order to keep the computation times reasonably low. Based on our sensitivity analysis, we concluded that RT centers should constantly monitor the percentage of patients asking for the same time window and, when the “competition” goes above a certain threshold, re-dimension the time window size being offered to future patients accordingly. Alternatively, the RT center may offer the patient an alternative time window when the requested one appears to be particularly popular.

The pre-assignment of patients to subsets of linacs allowed to solve the problem more efficiently by using the MILP model for solving subproblems associated with smaller groups of linacs. RT centers may split the set of available linacs in subgroups based on their physical location (satellites), technological specifications (e.g. cone-beam CT embedded or not) or the interchangeability of RTTs between linacs. When partitioning the problem, our model can be effective in generating a (sub-)optimal schedule in due time for most real-world RT centers. Not only the solutions found by our combined approach were close to optimal with a maximum of 6.2% gap amongst instances between 1 and 8 linacs, but also the methodology can save staff members the planning time required to produce a solution manually. This makes our tool especially interesting for RT centers, which are more interested in finding a good, feasible solution in a quick way over highly detailed methods that take extreme computational effort to prove optimality. Especially when, as in most cases, manual adjustments in the final solution will have to be performed to cope with particular characteristics of patients and staff.

In Chapter 6 we used the MILP model from Chapter 5 for performing adjustments towards an implementation in clinical practice. The model was adapted to generate schedules for the RT treatment scheduling problem in a specific location of two Dutch RT centers: NKI (213 patients per week and 6 linacs) and BVI (51 patients per week and 2 linacs). Patient data was collected for the planning horizon of one week, and the feasibility of the obtained schedules was verified by the relevant staff members and planners of each center. The model was iteratively adjusted to fulfill the technical and medical constraints of each center until a valid model was attained, with the optimized solutions being compared with the ones manually developed in practice. Results showed that the weekly schedule improved with our model in both centers. For instance, in the NKI the average SD between sessions' starting times decreased by 51%, and in the BVI the number of gaps in the schedule was reduced by 72%. A more compact schedule minimizing idle times allowed all treatment sessions to be delivered earlier on the day, potentially leading to cost savings equivalent to 0.45 FTEs, or releasing RTTs from the treatment machines to perform other tasks within their skills, such as treatment planning or image post-processing.

Besides ensuring that timeliness requirements for the start of treatment of new patients are fulfilled, this study mainly focused on patient-centeredness. With our model, very few patients would be switching linacs when comparing to practice, decreasing from 71 to 0 in the NKI case, and from 43 to 2 in the BVI. This allows patients to experience the same facilities and RTTs throughout the whole treatment, allowing them to build empathy and trust with these professionals. Moreover, for the NKI case, besides ensuring that scheduled sessions are delivered at more approximate times between each other, our model estimated that patient preferences regarding appointment times would be fulfilled 98% of the times. Although real data regarding patient preferences was unavailable for comparison, results indicated that patient preferences would be largely fulfilled with the validated model further increasing their overall satisfaction.

The computational effort needed to solve the problem required 5 minutes and 1.5 hours of CPU time for the two studied centers, as opposed to 1.5 days needed when performed manually. Although manual modifications may be needed after a solution is found, the practical application of the OR model proved to be able to help RT planners produce a high-quality schedule in a faster way.

7.3 Methodological considerations and future directions

Uncertainty management and number of scenarios

One of the most important factors to consider when developing OR models for RT logistics is uncertainty management. In Chapter 3 we account for the stochasticity inherent to patient inflow by solving the staff allocation problem for a number of patient arrival scenarios. In Chapter 4, a number of replications are run for each computational experiment to neutralize the impact of randomly generated data. However, defining the optimal number of scenarios and replications of stochastic models is not straightforward, indicating that more objective methods of assessing the optimal number of scenarios for stochastic programming models are needed. Another way to minimize the impact of uncertainty could be to shorten the length of the planning horizon whenever possible. As opposed to Chapter 3 where schedules are produced for a one-month planning horizon, in Chapters 5 and 6 we apply a MILP model for solving a scheduling problem on a weekly basis. This allowed us to build a deterministic model where all (regular) patients to be scheduled are known by the beginning of the planning horizon. In case the planning horizon of the MILP model of Chapter 3 could be shortened to one-week (schedules would be delivered to RTTs weekly instead of monthly), the computational performance of the proposed MILP model could probably be improved.

Apart from dealing with uncertainty in constructing our models, matching the mathematical value with actual willingness and practical options for change in practice is an ultimate uncertainty that needs to be confronted when aiming to actually implement the results of our research. This requires input and organizing commitment from all relevant stakeholders from the start of this type of projects and qualities with researchers to deal with sociodynamic aspects of work processes.

Increase of computational complexity

With the continuous increase in the number of patients receiving RT, it is expected that RT centers will continue to grow in capacity to be able to cover the rising demand. Thus, the decision problems addressed by the mathematical programming models proposed in this thesis may become computationally too complex to be solved in useful time using the proposed methodologies. In those cases, other OR methods with demonstrated ability to handling large scale instances may be applied. For instance, decomposition techniques such as column or row generation may be used to divide for the MILP problem of Chapter 5 in subproblems (groups of linacs) and solve the original problem to optimality iteratively. Markov Decision Processes and Approximate Dynamic Programming, on the other hand, may be valuable tools for tackling the complexity introduced by the increase in the number of scenarios used in the stochastic model of Chapter 3. Depending on the structure and complexity of the problem, alternative mathematical methods may be used when computational effort has proved to be an issue. Yet the added value for practice should be evident when striving for more advanced mathematical methods.

Healthcare simulation modeling

Creating a simulation model that mimics the pre-treatment workflow in RT with the currently available tools proved to be a complex and difficult task. RT operations, which are performed in sequential steps triggering new events and changing state variables, can be modeled using DES modeling. However, RT resources involved in the realization of these operations do not always have a standardized working routine that can be represented by sequences of discrete events. RTTs, for instance, may dedicate extra time to finalize treatment plans and avoid delaying the process of more urgent cases. Doctors often work late or skip certain meetings when their clinical workload for contouring is high. Clinicians' behavior, also in RT, is therefore better represented by system dynamics and/or agent-based simulation modeling. Software tools combining features of

DES and other simulation methods are therefore deemed necessary for a more accurate representation of the individual components of the RT processes. Based on our experience with the work of Chapter 4, DES features can help modeling patients' care pathways as a sequence of operations that are triggered after each other, and mimicking patient scheduling routines with a high degree of accuracy. Representing the behavior of doctors and RTTs, however, proved to be limited by the available features of the software tool used in our study. Modeling resource availability by means of decisions made by individual agents based on, e.g. utility functions, would allow the model to become even more realistic. This can be accomplished by using features available in agent-based modeling tools [7].

Added value of OR in RT

The OR methodologies proposed in this thesis have demonstrated the potential to provide considerable added value for operations management in RT. The algorithm of Chapter 3, for instance, showed that the RTT capacity could be reduced by 5.5% by optimally allocating RTTs to cover the demand. According to estimations derived from the NVRO's technical report in 2012 [67], there are around 1254 FTEs of RTT capacity in the Netherlands, which means that our model could potentially lead to cost savings of 69 FTEs nationwide. Similarly, by obtaining more compact schedules for the BVI case in Chapter 6, our OR model showed that around 15% of each linac's operating time could theoretically be saved. Assuming that there are 3 RTTs working daily in each linac, that would correspond to overall savings of around 188 FTEs in case RT centers operate in similar circumstances to the BVI. Moreover, waiting times could be largely reduced if our models were to be implemented on a national scale. If we extrapolate the performance obtained for the NKI case to the whole country, we would be able to meet waiting time targets for up to 99% of the patients (Chapters 3 and 4), as opposed to 91% measured in practice at the NKI (2016). Furthermore, and according to results of Chapter 5, 98% of all regular patients could be receiving their sessions within their preferred time window. Estimating that there are over 70 thousand RT treatment pathways in the Netherlands annually [67], this would provide considerable gains in terms of both patient-centeredness and efficiency, increasing the overall quality of treatment with clear added benefits for cancer patients.

7.4 Collaboration network and generalization of results

The set-up of a collaboration network with several RT centers actively participating in the project allowed for the development of broader and more generally applicable OR models. RT centers have different sizes, value propositions and strategic goals. For example, logistics tend to be more complex, i.e. have more diversified treatment pathways, in centers focusing on highly specialized treatments such as the NKI. By contrast, other cancer centers may have more standardized care pathways with more consistent sequences of operations. Integrating information collected from six Dutch RT centers with diversified value propositions and objectives allowed to design robust methodologies that are representative of several centers' operational procedures. This makes the proposed models easily adjustable and useable by RT centers, avoiding the need to develop a model for each center specifically as seen in other healthcare fields where, for instance, over a thousand simulation models can be found for emergency departments alone [12].

During the course of this thesis work, regular meetings were held with members of all collaborating centers. Lively discussions and exchange of information were promoted between the managers and planners of the involved institutes, allowing them to learn from each other by being exposed to alternative ways of thinking regarding resource planning and use in RT. We verified that setting up a network of collaboration and encouraging the participants to expose and explain their (recurring) managerial dynamics to each other led to the identification of potential improvement interventions in some of the centers. Solutions for the logistics problems that are common amongst RT centers were discussed through brainstorming, which generated new research ideas that originated gradual methodological improvements. This collaborative approach is also described in the literature by Hulscher *et al*. [41]. Their literature study showed that teamwork and participation in specific collaborative activities improved the success of the implementation of quality improvement projects in the short-term, as seen with our project. In the long run, the authors found that keeping projects teams intact and gathering data continuously over time increases the chances of success in the implementation process.

7.5 Implementation of OR models for RT logistics

The full implementation of OR models, however, is not straightforward. The steps towards a clinical implementation performed in this thesis (Chapter 6) showed that performing model adaptations gradually, in small steps, is a major contributor. A step-by-step approach on the adjustment of the OR models allowed for an early identification of pitfalls that could be immediately corrected, promoting a smoother transition into a clinically valid solution. Conducting the research of this thesis in an actual RT center (NKI) has been another important facilitator. Working on the premises of an actual clinic made it possible to keep constant communication with RT clinicians and managers that promoted quicker and more accurate model adaptations. Moreover, keeping constant communication with stakeholders both from the NKI and collaborating centers from the very beginning of the project was crucial for the realization of the implementation steps. By setting regular meetings, model adaptations and the corresponding solutions could be discussed frequently for the identification of pitfalls and/or infeasibilities that led to new model adaptations.

This study also showed that there are hurdles that need to be overcome for a practical application of OR tools in RT. First, the involvement of several specialized personnel with different backgrounds is needed. OR specialist, appointment planners, process managers, software developers, etc. need to be involved in the design, development, and testing of OR-based software tools. A simple yet effective way of communicating new developments and results are therefore critical for the successful interpretation and implementation of those. In addition, we verified that the acceptance and validation of OR models can be easier to obtain when large amounts of quality data are available. In this thesis work, the use of complete and updated datasets allowed for more reliable testing that increased the stakeholders' perception of validity regarding the results being presented. Thus, recording and collecting as much logistics' data as possible is highly recommended when attempting the development and application of OR models in RT, which can be hard to achieve. In fact, four out of the six RT centers of our collaboration network could not be included in our implementation project (Chapter 6) due to lack of updated data for timely model testing. Moreover, the establishment of input and output data relations between the patient information system (PIS) software and the OR model can be challenging. A possible facilitator can be the integration of the OR tool into a patient information management system such as MOSAIQ [33]. This would potentially allow input and output data to be automatically transferred from and to the internal databases in an automated manner, saving planners the time needed for manual data handling. Furthermore, the implementation of the several models and/or interventions proposed in this thesis must be done sequentially on a step-by-step approach. After each change, actual performance should be evaluated and compared with the theoretical results to assess the validity of the predicted results and reflect upon eventual differences. Next, the new circumstances should be analyzed and considered before proceeding to the next implementation project. The corresponding OR model(s) can be re-used to test the system under the new circumstances in order to verify whether the previously obtained results remain valid.

Concluding, the research models proposed in this thesis are intended to provide decision-making support such that RT centers organize their processes more efficiently for the benefit of patients and staff. Although several hurdles hamper their application in practice, OR has proven effective in solving highly constrained problems where feasible (near-optimal) solutions are difficult to achieve manually. Advanced analytical methods based on OR have proven capable of optimizing the logistics of radiotherapy treatments for a better quality of care.

Bibliography

[1] A. Aitkenhead, D. Bugg, C. G. Rowbottom, E. Smith, and R. I. Mackay. Modelling the throughput capacity of a single-accelerator multitreatment room proton therapy centre. *The British journal of radiology*, 85 1020: e1263–72, 2012.

[2] G. Akin, J. S. Ivy, T. R. Huschka, T. R. Rohleder, and Y. N. Marmor. Capacity management and patient scheduling in an outpatient clinic using discrete event simulation. In *2013 Winter Simulations Conference (WSC)*, 2215–2226, 2013.

[3] V. Babashov, I. Aivas, M. Begen, J. Cao, G. Rodrigues, D. D'Souza, M. Lock, and G. Zaric. Reducing patient waiting times for radiation therapy and improving the treatment planning process: a discrete-event simulation model (radiation treatment planning). *Clinical Oncology*, 29 (6): 385 – 391, 2017.

[4] S. Bangsow. Manufacturing Simulation with Plant Simulation and Simtalk: Usage and Programming with Examples and Solutions. *Springer Publishing Company, Incorporated*, 1st edition, 2010.

[5] I. A. Bikker, N. Kortbeek, R. M. van Os, and R. J. Boucherie. Reducing access times for radiation treatment by aligning the doctor's schemes. *Operations Research for Health Care*, 7: 111 – 121, 2015.

[6] C. Blum and A. Roli. Metaheuristics in combinatorial optimization: Overview and conceptual comparison. *ACM Comput. Surv.*, 35 (3): 268–308, September 2003.

[7] E. Bonabeau. Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences*, 99 (suppl 3): 7280–7287, 2002.

[8] J. M. Borras, Y. Lievens, M. Barton, J. Corral, J. Ferlay, F. Bray, and C. Grau. How many new cancer patients in Europe will require radiotherapy by 2025? an estro-hero analysis. *Radiotherapy and Oncology*, 119 (1): 5 – 11, 2016.

[9] B. D. Bradley, T. A. Jung, A. Tandon-Verma, B. Khoury, T. C. Y. Chan, and Y.-L. Cheng. Operations research in global health: a scoping review with a focus on the themes of health equity and impact. In *Health Research Policy and Systems*, 2017.

[10] S. P. Bradley, A. C. Hax, and T. L. Magnanti. Applied Mathematical Programming. *Addison-Wesley Publishing Company*, Reading, MA, 1st edition, 1977.

[11] S. Brailsford. Overcoming the barriers to implementation of operations research simulation models in healthcare. *Clinical and investigative medicine. Médecine clinique et experimentale*, 28: 312–5, 01 2006.

[12] S. Brailsford. Tutorial: Advances and challenges in healthcare simulation modeling. In *2007 Winter Simulation Conference*, 1436–1448, Dec 2007.

[13] S. Brailsford, P. Harper, B. Patel, and M. Pitt. An analysis of the academic literature on simulation and modelling in health care. *Journal of Simulation*, 3: 130–140, 2009.

[14] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal. Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 68 (6): 394–424, 2018.

[15] E. K. Burke, P. L. Rocha, and S. Petrovic. An integer linear programming model for the radiotherapy treatment scheduling problem. *CoRR*, abs/1103.3391, 2011.

[16] B. Cardoen, E. Demeulemeester, and J. Beliën. Operating room planning and scheduling: A literature review. *European Journal of Operational Research*, 201 (3): 921 – 932, 2010.

[17] J. P. Cares, M. C. Riff, and I. Araya. LS2R: A local search algorithm to solve scheduling radiotherapy problems. In *13th International Conference on Hybrid Intelligent Systems (HIS 2013)*, 256–261, 2013.

[18] E. Castro and S. Petrovic. Combined mathematical programming and heuristics for a radiotherapy pre-treatment scheduling problem. *Journal of Scheduling*, 15 (3): 333–346, 2012.

[19] J. Caudell, J. Torres-Roca, R. Gillies, H. Enderling, S. Kim, A. Rishi, E. Moros, and L. Harrison. The future of personalised radiotherapy for head and neck cancer. *The Lancet Oncology*, 18, 2017.

[20] Z. Chen, W. King, R. Pearcey, M. Kerba, and W. J. Mackillop. The relationship between waiting time for radiotherapy and clinical outcomes: A systematic review of the literature. *Radiotherapy and Oncology*, 87 (1): 3–16, 2008.

[21] J. Coelho and M. Vanhoucke. Multi-mode resource-constrained project scheduling using rcpsp and sat solvers. *European Journal of Operational Research*, 213 (1): 73 – 82, 2011.

[22] D. Conforti, F. Guerriero, and R. Guido. Non-block scheduling with priority for radiotherapy treatments. *European Journal of Operational Research*, 201 (1): 289 – 296, 2010.

[23] D. Conforti, F. Guerriero, and R. Guido. Optimization models for radiotherapy patient scheduling. *4OR*, 6: 263–278, 2008.

[24] D. Conforti, F. Guerriero, R. Guido, and M. Veltri. An optimal decision-making approach for the management of radiotherapy patients. *OR Spectrum*, 33: 123–148, 2011.

[25] E. Crawford, P. Parikh, N. Kong, and C. Thakar. Analyzing discharge strategies during acute care: A discrete-event simulation study. *Medical decision making : an international journal of the Society for Medical Decision Making*, 34, 2013.

[26] F. Crop, T. Lacornerie, X. Mirabel, and E. Lartigau. Workflow optimization for robotic stereotactic radiotherapy treatments: Application of constant work in progress workflow. *Operations Research for Health Care*, 6: 18 – 22, 2015.

[27] G. Delaney, S. Jacob, C. Featherstone, and M. Barton. Radiotherapy in cancer care: estimating the optimal utilisation from a review of evidencebased clinical guidelines. *Technical Report*, *Cancer Outcomes Research and Evaluation (CCORE)*, 2003.

[28] G. Delaney, S. Jacob, C. Featherstone, and M. Barton. The role of radiotherapy in cancer treatment. *Cancer*, 104 (6): 1129–1137, 2005.

[29] C. E., S. Dahrouge, R. Samant, A. Mirzaei, and J. Price. Radical radiotherapy for cervix cancer: The effect of waiting time on outcome. *International Journal of Radiation Oncology*Biology*Physics*, 61 (4): 1071 – 1077, 2005.

[30] M. Ebert, W. Li, L. Jennings, R. Kearvell, and S. Bydder. Utilitarian prioritization of radiation oncology patients based on maximization of population tumour control. *Physics in Medicine and Biology*, 58 (12): 4013–4029, 2013.

[31] M. Ebert, W. Li, and L. Jennings. An analytical solution to patient prioritisation in radiotherapy based on utilitarian optimisation. *Australasian Physical & Engineering Sciences In Medicine*, 37 (1): 53–57, 2014.

[32] M. Ehrgott and A. Holder. Operations research methods for optimization in radiation oncology. *Journal of Radiation Oncology Informatics*, 6 (1): 1–41, 2009.

[33] ELEKTA. MOSAIQ radiation oncology. Available at `https://www.elekta.com/software-solutions/care-management/mosaiq-radiation-oncology.html`, 2017. [Accessed 20/01/2018].

[34] A. T. Ernst, H. Jiang, K. Mohan, and D. Sier. Staff scheduling and rostering: A review of applications, methods and models. *European Journal of Operational Research*, 153 (1): 3 – 27, 2004. Timetabling and Rostering.

[35] R. Famiglietti, E. Norboge, V. Boving, J. Langabeer, T. Buchholz, and O. Mikhail. Using discrete-event simulation to promote quality improvement and efficiency in a radiation oncology treatment center. *Quality Management in Health Care*, 26: 184–189, 2017.

[36] J. Feldman, N. Liu, H. Topaloglu, and S. Ziya. Appointment scheduling under patient preference and no-show behavior. *Operations Research*, 62 (4): 794–811, 2014.

[37] D. Gartner, M. Frey, and R. Kolisch. Hospital-wide therapist scheduling and routing: Exact and heuristic methods. *IISE Transactions on Healthcare Systems Engineering*, 8 (4): 268–279, 2018.

[38] L. V. Green. Using Operations Research to Reduce Delays for Healthcare, chapter Chapter 1, 1–16. INFORMS Annual Meeting, 2008.

[39] D. Gupta and B. Denton. Appointment scheduling in health care: Challenges and opportunities. *IIE Transactions*, 40 (9): 800–819, 2008.

[40] E. W. Hans, M. van Houdenhoven, and P. J. Hulshof. A Framework for Healthcare Planning and Control, 303–320. International Series in Operations Research & Management Science. Springer, 2012.

[41] M. Hulscher, L. Schouten, R. Grol, and H. Buchan. Determinants of success of quality improvement collaboratives: what does the literature show? *BMJ Quality & Safety*, 22, 08 2012.

[42] P. J. H. Hulshof, R. J. Boucherie, E. W. Hans, and J. L. Hurink. Tactical resource allocation and elective patient admission planning in care processes. *Health Care Management Science*, 16 (2): 152–166, 2013.

[43] P. J. Hulshof, R. J. Boucherie, T. van Essen, E. W. Hans, J. L. Hurink, N. Kortbeek, N. Litvak, P. T. Vanberkel, E. van der Veen, B. Veltman, I. Vliegen, and M. E. Zonderland. Orchestra: an online reference database of or/ms literature in health care. *Health care management science*, 14 (4): 383–384, 2011.

[44] P. J. Hulshof, N. Kortbeek, R. J. Boucherie, E. W. Hans, and P. Bakker. Taxonomic classification of planning decisions in health care: a structured review of the state of the art in OR/MS. *Health systems*, 1 (2): 129–175, 2012.

[45] D. Hutton, C. Beardmore, I. Patel, J. Massey, H. Wong, and H. Probst. Audit of the job satisfaction levels of the uk radiography and physics workforce in uk radiotherapy centres 2012. *The British journal of radiology*, 87: 20130742, 2014.

[46] IAEA. Dirac (directory of radiotherapy centres). Available at `https://dirac.iaea.org/`, 2020. [Accessed 26/03/2020].

[47] R. H. Jack and L. Holmberg. Waiting times for radiotherapy after breast cancer. *BMJ*, 340, 2010.

[48] S. Jacobson, S. Hall, and J. Swisher. Discrete-event simulation of health care systems, 273–309. International Series in Operations Research and Management Science. Springer New York LLC, 2013.

[49] Y. Jacquemin, E. Marcon, and P. Pommier. Towards an improved resolution of radiotherapy, scheduling. In *2010 IEEE Workshop on Health Care Management (WHCM)*, 1–6, 2010.

[50] Y. Jacquemin, E. Marcon, and P. Pommier. A pattern-based approach of radiotherapy scheduling. *IFAC Proceedings Volumes*, 44 (1): 6945 – 6950, 2011.

[51] S. V. Jerić and J. R. Figueira. Multi-objective scheduling and a resource allocation problem in hospitals. *Journal of Scheduling*, 15 (5): 513–535, 2012.

[52] P. E. Joustra, R. Kolfin, N. M. van Dijk, C. C. E. Koning, and P. J. M. Bakker. Reduce fluctuations in capacity to improve the accessibility of radiotherapy treatment cost-effectively. *Flexible Services and Manufacturing Journal*, 24 (4): 448–464, 2012.

[53] P. E. Joustra, E. van der Sluis, and N. M. van Dijk. To pool or not to pool in hospitals: a theoretical and practical comparison for a radiotherapy outpatient department. *Annals of Operations Research*, 178: 77–89, 2010.

[54] T. Kapamara, K. Sheibani, D. Petrovic, O. Haas, and C. Reeves. A simulation of a radiotherapy treatment system: A case study of a local cancer centre. In *Proceedings of ORP3*, 29–35. EURO, 2007.

[55] T. Kapamara and D. Petrovic. A heuristics and steepest hill climbing method to scheduling radiotherapy patients. In *Proceedings of the 35th International Conference on Operational Research Applied to Health Services*, Catholic University of Leuven, Belgium, 2009.

[56] T. Kapamara, K. Sheibani, O. C. L. Haas, C. R. Reeves, and D. Petrovic. A review of scheduling problems in radiotherapy. In *Proceedings of the International Control Systems Engineering Conference*, 201–207, 2006.

[57] G. Kaplan, M. H. Lopez, and J. M. McGinnis. Transforming health care scheduling and access: getting to now. *The National Academies Press*, 2015.

[58] A. M. Law. Simulation Modeling & Analysis. *McGraw-Hill*, New York, NY, USA, 4 edition, 2007.

[59] A. Legrain, M.-A. Fortin, N. Lahrichi, and L.-M. Rousseau. Online stochastic optimization of radiotherapy patient scheduling. *Health Care Management Science*, 18 (2): 110 – 123, 2015.

[60] A. Legrain, M. Fortin, N. Lahrichi, L.-M. Rousseau, and M. Widmer. Stochastic optimization of the scheduling of a radiotherapy center. In *Journal of Physics Conference Series* 616, 1 – 6, May 2015.

[61] S. Li, N. Geng, and X. Xie. Radiation queue: Meeting patient waiting time targets. *IEEE Robotics Automation Magazine*, 22 (2): 51–63, June 2015.

[62] S. Li, X. Xie, and N. Geng. A queuing approach for radiotherapy treatment capacity planning. In *2014 IEEE International Conference on Automation Science and Engineering (CASE)*, 540–545, 2014.

[63] W. J. Mackillop. Killing time: The consequences of delays in radiotherapy. *Radiotherapy and Oncology*, 84 (1): 1–4, 2007.

[64] W. J. Mackillop, J. Bates, B. O'Sullivan, and H. Withers. The effect of delay in treatment on local control by radiotherapy. *International Journal of Radiation Oncology*Biology*Physics*, 34 (1): 243 – 250, 1996.

[65] J. Maschler and G. R. Raidl. Particle therapy patient scheduling with limited starting time variations of daily treatments. *International Transactions in Operational Research*, 27 (1): 458–479, 2018.

[66] Mathwave. Easyfit :: Distribution fitting made easy. Available at `http://www.mathwave.com/easyfit-distribution-fitting.html`, 2017. [Accessed 20/02/2020].

[67] NVRO. Brancherapport 2011. *Technical Report*, *Nederlandse Vereniging voor Radiotherapie en Oncologie*, 04 2012.

[68] NVRO. Waiting times, standards and maximum waiting times for radiotherapy (in dutch). Available at `http://www.nvro.nl/kwaliteit/indicatoren/`, 2017. [Accessed 26/03/2020].

[69] S. N. Ogulata, M. O. Cetik, E. Koyuncu, and M. Koyuncu. A simulation approach for scheduling patients in the department of radiation oncology. *Journal of Medical Systems*, 33: 233–239, 2008.

[70] I. A. Olivotto, J. Soo, R. A. Olson, L. Rowe, J. H. French, B. Jensen, A. Pastuch, R. M. Halperin, and P. Truong. Patient preferences for timing and access to radiation therapy. *Current Oncology*, 22 (4): 279–286, 2015.

[71] G. S. Petersen, J. L. Knudsen, and M. M. Vinter. Cancer patients' preferences of care within hospitals: a systematic literature review. *International Journal for Quality in Health Care*, 27 (5): 384–395, 2015.

[72] D. Petrovic, E. Castro, S. Petrovic, and T. Kapamara. Radiotherapy Scheduling, 155–189. Springer Berlin Heidelberg, 2013.

[73] D. Petrovic, M. Morshed, and S. Petrovic. Genetic algorithm based scheduling of radiotherapy treatments for cancer patients. In C. Combi, Y. Shahar, and A. Abu-Hanna, editors, *Artificial Intelligence in Medicine*, 101–105. Springer Berlin Heidelberg, 2009.

[74] D. Petrovic, M. Morshed, and S. Petrovic. Multi-objective genetic algorithms for scheduling of radiotherapy treatments for categorised cancer patients. *Expert Systems with Applications*, 38 (6): 6994 – 7002, 2011.

[75] S. Petrovic and P. Leite-Rocha. Constructive and grasp approaches to radiotherapy treatment scheduling. In *Advances in Electrical and Electronics Engineering - IAENG Special Edition of the World Congress on Engineering and Computer Science 2008*, 192–200, 2008.

[76] S. J. Petrović, W. Leung, X. Song, and S. Sundar. Algorithms for radiotherapy treatment booking. In *25th Workshop of the UK Planning and Scheduling Special Interest Group*, 10 – 112, Nottingham, UK, 2006.

[77] S. Petrovic and E. Castro. A genetic algorithm for radiotherapy pre-treatment scheduling. In *Applications of Evolutionary Computation*, 454–463. Springer Berlin Heidelberg, 2011.

[78] S. Petrovic and L.-R. Pedro. Constructive approaches to radiotherapy scheduling. In *Proceedings of the World Congress on Engineering and Computer Science*, San Francisco, USA, 2008.

[79] T. Pignon, L. Fernandez, S. Ayasso, M.-A. Durand, D. Badinand, and D. Cowen. Impact of radiation oncology practice on pain: A cross-sectional survey. *International Journal of Radiation Oncology*Biology*Physics*, 60 (4): 1204–1210, 2004.

[80] S. Price, B. Golden, E. Wasil, and H. H. Zhang. Optimizing throughput of a multi-room proton therapy treatment center via simulation. In *2013 Winter Simulations Conference (WSC)*, 2422–2431, Dec 2013.

[81] S. Proctor, B. Lehaney, C. Reeves, and Z. Khan. Modelling patient flow in a radiotherapy department. *OR Insight*, 20: 6–14, 2007.

[82] M. L. Puterman. Markov Decision Processes: Discrete Stochastic Dynamic Programming. *John Wiley & Sons, Inc.*, USA, 1st edition, 1994.

[83] A. Rais and A. Viana. Operations research in healthcare: a survey. *International Transactions in Operational Research*, 18 (1): 1–31, 2011.

[84] J. Rajgopal. Principles and application of operations research. In K. B. Zandin, editor, *Maynard's Industrial Engineering Handbook*, chapter 11, 27–44. McGraw-Hill, New York, 2001.

[85] R. T. Rockafellar. Lagrange multipliers and optimality. *SIAM Rev.*, 35 (2): 183–238, June 1993.

[86] E. Rosenblatt, J. Izewska, Y. Anacak, Y. Pynda, P. Scalliet, M. Boniol, and M. Boniol. Radiotherapy capacity in european countries : an analysis of the directory of radiotherapy centres (DIRAC) database. *Lancet Oncology*, 14 (2): e79–86, 2013.

[87] R. Ruiz and J. A. Vázquez-Rodríguez. The hybrid flow shop scheduling problem. *European Journal of Operational Research*, 205 (1): 1–18, 2010.

[88] A. Sauré, J. Patrick, S. Tyldesley, and M. L. Puterman. Dynamic multi-appointment patient scheduling for radiation therapy. *European Journal of Operational Research*, 223 (2): 573 – 584, 2012.

[89] C. E. Saville, H. K. Smith, and K. Bijak. Operational research techniques applied throughout cancer care services: a review. *Health Systems*, 8 (1): 52–73, 2019.

[90] A. Shtiliyanova, F. Feschet, and P. Pommier. Scheduling model for a tool evaluating new radiotherapies. In *Proceedings of the 2012 Symposium on Theory of Modeling and Simulation - DEVS Integrative M&S Symposium*, TMS/DEVS '12, San Diego, CA, USA, 2012. Society for Computer Simulation International.

[91] N. Shukla, R. Wickramasuriya, A. Miller, and P. Perez. An approach to plan and evaluate the location of radiotherapy services and its application in the new south wales, australia. *Computer Methods and Programs in Biomedicine*, 122 (2): 245 – 256, 2015.

[92] R. L. Siegel and K. D. Miller. Cancer statistics, 2017. *CA: A Cancer Journal for Clinicians*, 67 (1): 7–30, 2017.

[93] C. Solnon and N. Jussien. Constructive Heuristic Approaches, chapter 6, 85–92. John Wiley & Sons, Ltd, 2013.

[94] B. Stewart and C. Wild. World cancer report 2014. *Technical Report*, *International Agency for Research on Cancer*, 2014.

[95] C. Sung. Mixed-integer linear programming (MILP) methods for integration of production planning and scheduling. *University of Wisconsin–Madison*, 2009.

[96] H. A. Taha. Operations Research: An Introduction (8th Edition). *Prentice-Hall, Inc.*, USA, 2006.

[97] S. Thomas. Capacity and demand models for radiotherapy treatment machines. *Clinical Oncology*, 15 (6): 353 – 358, 2003.

[98] L. van Bodegom-Vos, F. Davidoff, and P. Marang-van de Mheen. Implementation and de-implementation: Two sides of the same coin? *BMJ Quality & Safety*, 26, 08 2016.

[99] W. van Lent, R. de Beer, and W. H. van Harten. International benchmarking of specialty hospitals. a series of case studies on comprehensive cancer centres. *BMC health services research*, 10 (253), 2010.

[100] W. van Lent, P. Vanberkel, and W. H. van Harten. A review on the relation between simulation and improvement in hospitals. *BMC medical informatics and decision making*, 12: 18, 2012.

[101] W. A. van Lent, J. W. Deetman, H. J. Teertstra, S. H. Muller, E. W. Hans, and W. H. van Harten. Reducing the throughput time of the diagnostic track involving {CT} scanning with computer simulation. *European Journal of Radiology*, 81 (11): 3131 – 3140, 2012.

[102] J. van Sambeek. Smarter imaging management: operations management for radiology. PhD thesis, University of Twente, Netherlands, 2018.

[103] VARIAN. ARIA OIS for medical oncology. Available at `https://www.varian.com/products/software/information-systems/aria-ois-medical-oncology`, 2020. [Accessed 31/01/2020].

[104] B. Vieira, D. Demirtas, J. B. van de Kamer, E. W. Hans, L.-M. Rousseau, N. Lahrichi, and W. H. van Harten. Radiotherapy treatment scheduling considering patient preference requests. *Under review*, 19, 2020.

[105] B. Vieira, D. Demirtas, J. B. van de Kamer, E. W. Hans, and W. H. van Harten. A mathematical programming model for optimizing the staff allocation in radiotherapy under uncertain demand. *European Journal of Operational Research*, 270 (2): 709 – 722, 2018.

[106] B. Vieira, D. Demirtas, J. B. van de Kamer, E. W. Hans, and W. H. van Harten. Improving workflow control in radiotherapy using discrete-event simulation. *BMC medical informatics and decision making*, 19 (199), 2019.

[107] B. Vieira, E. W. Hans, C. van Vliet-Vroegindeweij, J. van de Kamer, and W. H. van Harten. Operations research for resource planning and -use in radiotherapy: a literature review. *BMC Medical Informatics and Decision Making*, 16 (1): 149, 2016.

[108] A. Vitoux, C. Grenier, and E. Lartigau. 147 improvement in the quality of practices in radiotherapy: the regular measurement of indicators. *BMJ Quality & Safety*, 19 (Suppl 1): A170–A171, 2010.

[109] P. Vogl, R. Braune, and K. Doerner. Scheduling recurring radiotherapy appointments in an ion beam facility. *Journal of Scheduling*, 1–18, 07 2018.

[110] G. Werker, A. Sauré, J. French, and S. Shechter. The use of discrete-event simulation modelling to improve radiation therapy planning processes. *Radiotherapy and Oncology*, 92 (1): 76–82, 2009.

[111] K. M. Winkfield and D. Gabeau. Why workforce diversity in oncology matters. *International Journal of Radiation Oncology*Biology*Physics*, 85 (4): 900 – 901, 2013.

[112] W. L. Winston. Operations research: applications and algorithms. Business Statistics. *PWS-Kent Pub. Co.*, 1991.

[113] R. W. Wolff. Stochastic Modeling and the Theory of Queues. *Prentice Hall*, 1st edition, 1989.

Acronyms

ANOVA	Analysis of variance
ARTI	Radiotherapiegroep - location Arnhem
AUMC	Amsterdam Universitair Medische Centra
BCCA	British Columbia Cancer Agency
BSU	Beam set-up
BVI	Bernard Verbeeten Instituut
CHOIR	Center for Healthcare Operations Improvement and Research
CH	Constructive heuristics
CICL	Centré Integré de Cancérologie de Laval
CPU	Central processing unit
CS	Computer simulation
CT	Computed tomography
DES	Discrete event simulation
DIRAC	Directory of radiotherapy centers
EDD	Earliest due date
EIG	Enhanced iterated greedy
FTE	Full-time equivalence
IAEA	International Atomic Energy Agency
IPP	Image post-processing
KPI	Key performance indicator
LINAC	Linear accelerator
MDP	Markov decision process
MH	Metaheuristics
MILP	Mixed-integer linear programming
MP	Mathematical programming
MRI	Magnetic Resonance Imaging
NKI	Netherlands Cancer Institute
NVRO	Dutch Society for Radiation Oncology
OAR	Organs-at-risk
OR	Operations research
PET-CT	Positron emission tomography - Computed tomography
PT	Particle therapy
QT	Queuing theory
RIF	Radiotherapeutisch Instituut Friesland
RISO	Radiotherapiegroep - location Deventer
RT	Radiotherapy

RTT	Radiation therapy technologist
SD	Standard deviation
TCP	Tumor control probability
TP	Treatment planning
WT	Waiting time

Summary

The delivery of radiotherapy (RT) involves the use of rather expensive resources and multi-disciplinary staff. In RT, timeliness is crucial, and literature shows that delays in the start of treatment have shown to increase the risk of tumor progression in various cancer types, and patients experience greater psychological distress when subject to longer waiting times. As the number of cancer patients receiving RT increases, timely delivery becomes increasingly difficult due to the complexities related to, among others, variable patient inflow, specialized patient routing, and the joint planning of multiple resources. In this thesis, we studied, designed, and developed several Operations Research (OR) models for the logistical optimization of RT processes to support decision-makers use their resources more efficiently. We propose innovative approaches for solving logistical problems encountered in the whole RT chain of operations, from pre-treatment to treatment, while optimizing for the most important KPIs related to timeliness and patient-centeredness. The developed research work is practice-oriented, with models being built, validated, and tested using real-world information and data provided by six collaborating RT centers.

Chapter 2 presents a literature search of OR methods for resource planning in RT in six databases covering journal papers in the medical and mathematical domains. Data extraction included, amongst others, the subject of research and methods applied, the extent of implementation according to a six-stage model, and the (potential) impact of the results in practice. The review shows that most models addressed the problem of scheduling patients on treatment machines, while little focus has been given to the pre-treatment phase of the process, despite its high impact on timeliness. We also found that the development of OR methods for RT macro-planning is rather low and can be further extended. Furthermore, we verified that none of the 33 analyzed papers reported a full implementation of the model indicating that, unless there is lack of reporting, implementation rates of OR models in RT are rather low.

In Chapter 3 we introduce a mathematical approach for the optimal allocation of radiation therapy technologists (RTTs) to the several operations they perform throughout the (pre-)treatment chain of operations. Besides performing multiple activities, RTTs have specific rotation needs to maintain specialized skills and partial availability, which makes the allocation of RTTs a complex task especially due to the highly variable patient arrivals and care content. In our approach, we use a novel stochastic mixed integer linear programming (MILP) model to optimize the allocation of RTTs over a set of scenarios of patient inflow

generated from historical patient data. The goal is to maximize the (expected) number of patients completing the pre-treatment phase within the waiting time target. Results for a case study in the RT department of the Netherlands Cancer Institute (NKI) showed that, on average, the number of patients able to start treatment within the maximum waiting time standards may increase from 91.3% to 97.9% for subacute patients, and from 96.3% to 100.0% for regular patients.

Radiotherapy pre-treatment workflow is either driven by the scheduling of the first irradiation session, which can be set right after consultation (pull strategy) or after the pre-treatment operations have been completed (push strategy). In Chapter 4 we assess the impact of using pull and push strategies and explore alternative interventions for improving timeliness in radiotherapy using discrete-event simulation modeling. Staff surveys, interviews with managers, and historical data from the NKI (2017) were used to generate model inputs, in which fluctuations in patient inflow are considered. Results showed that a pull strategy allows for 41% fewer patients breaching the waiting time target than a hybrid strategy (40% pull / 60% push), in spite of slightly longer waiting times (12%).

Chapter 5 focuses on the problem of automatically scheduling RT sessions while satisfying patient preferences regarding the time of their appointments. While most literature focuses on the timeliness of treatments, collaborating RT centers have expressed their need to include patient preferences when scheduling appointments for irradiation sessions. Therefore, we propose a MILP model to solve the problem to optimality, scheduling all sessions within the desired window in reasonable computation time for small size instances up to 66 patients and two linacs per week. For larger centers, we propose a heuristic method that pre-assigns patients to linacs to decompose the problem in subproblems (clusters of linacs) before using the MILP model to solve the subproblems to optimality in a sequential manner. We tested our methodology using real-world data from the NKI (eight linacs) and found that, with our combined approach, the problem can be solved in reasonable computation time (3.3 hours) with as few as 2.8% of the sessions being scheduled outside the desired 90-min time window.

In Chapter 6 we adapted the MILP model of Chapter 5 to generate weekly schedules for the RT treatment scheduling problem for two Dutch RT centers. The model was iteratively adjusted to fulfill the technical and medical constraints of each center until a valid solution was attained. Patient data was collected from the internal databases, and the feasibility of the obtained schedules was verified by staff members of each center. We verified that the practical application improved the weekly schedule in both centers by decreasing the average SD between sessions' starting times from 103.0 to 50.4 minutes (51%) in one center, and by reducing the number of gaps in the schedule from 18 to 5 (72%) in the other. The process required 5 minutes respectively 1.5 hours of computation time for the two centers at test, as opposed to 1.5 days when performed manually by both centers.

The OR models proposed in this thesis have demonstrated considerable potential benefits when results are extrapolated on a national scale. The algorithm of Chapter 3, for instance, showed that our model could potentially lead to cost savings of 69 FTEs nationwide. Similarly, by obtaining more compact schedules in one of the test cases (Chapter 6), our model showed that savings of around 188 FTEs could be achieved overall. However, the implementation of OR tools in RT practice is not straightforward. We found that matching the mathematical value with actual willingness and practical options for change in practice is an ultimate uncertainty that needs to be confronted when aiming to actually implement the theoretical results. This requires input and organizing commitment from all relevant stakeholders and researchers from the start of the project in order to deal with sociodynamic aspects of work processes from an early stage. Implementation efforts performed within this project also revealed that gradual model adaptations performed in small steps and constant communication are needed to ensure the translation OR models into clinical practice.

Concluding, the research models proposed in this thesis proved to be able to provide decision-making support for RT centers aiming to organize their processes more efficiently for the benefit of patients and staff. Although several hurdles hamper the application of those models in practice, OR has proven effective in solving highly constrained logistics problems in RT where feasible (near-optimal) solutions are difficult to achieve manually. Overall, advanced analytical methods based on OR have proven capable of optimizing the logistics of radiotherapy treatments for a better quality of care.

Samenvatting

Het leveren van radiotherapie (RT) behandelingen vereist tamelijk kostbare middelen en de inzet van multidisciplinair personeel. Voor een succesvolle RT behandeling is een snelle start van de behandeling (tijdigheid) cruciaal en uit de literatuur blijkt dat vertragingen het risico op tumorprogressie bij verschillende soorten kanker kan vergroten. Daarnaast blijkt dat patiënten meer psychologische problemen ondervinden wanneer ze langer moeten wachten. Naarmate het aantal patiënten met kanker dat RT krijgt toeneemt, wordt op tijd beginnen met de behandeling steeds moeilijker door de complexiteit van onder andere de variabele instroom van patiënten, het gespecialiseerde zorgpad dat patiënten doorlopen en de gezamenlijke planning van personeel en middelen. In dit proefschrift hebben we verschillende mathematisch besliskunde modellen, in het Engels "Operations Research" (OR) models, bestudeerd en ontwikkeld voor de logistieke optimalisatie van RT processen. De modellen kunnen managers ondersteunen bij het efficiënter inzetten van middelen. We hebben innovatieve benaderingen voorgesteld voor het oplossen van logistieke problemen die zich voordoen in de hele radiotherapieketen, van voorbereiding tot behandeling, terwijl we de belangrijkste indicatoren (Key Performance Indicators, KPIs) hebben geoptimaliseerd: tijdigheid en patiëntgerichtheid. De behaalde resultaten zijn breed inzetbaar doordat de modellen zijn ontwikkeld, getest en gevalideerd met behulp van realistisch informatie en gegevens, aangeleverd door zes samenwerkende RT-centra .

Hoofdstuk 2 beschrijft een literatuuronderzoek naar OR-methoden voor resourceplanning in RT met behulp van zes databases met wetenschappelijke artikelen op medisch en wiskundig gebied. De dataextractie omvatte onder andere het onderwerp van het onderzoek en de toegepaste methoden, de mate van implementatie volgens een zes fasen model en de (potentiële) impact van de resultaten in de praktijk. Uit de resultaten bleek dat de meeste modellen het probleem van het inplannen van patiënten op behandelingsmachines (lineaire versnellers, linacs) aanpakken, terwijl er weinig aandacht is besteed aan de voorbereidingsfase van het RT traject, ondanks de grote invloed van deze fase op de doorlooptijd. We vonden ook dat de ontwikkeling van OR-methoden voor RT planning op macro (strategisch) niveau vrij beperkt is en verder kan worden uitgebreid. Bovendien hebben we geverifieerd dat geen van de 33 geanalyseerde artikelen een volledige implementatie van het model meldden. Dit wijst erop dat, tenzij er een gebrek aan rapportage is, de implementatiegraad van OR modellen in RT vrij laag is.

In hoofdstuk 3 hebben we een innovatieve wiskundige aanpak voorgesteld voor de optimale indeling van radiotherapeutisch laboranten in de verschillende operaties die zij uitvoeren over de gehele keten. Naast het uitvoeren van meerdere activiteiten, moeten laboranten regelmatig wisselen van taak om hun gespecialiseerde vaardigheden op peil te houden. Daarnaast werken ze vaak in deeltijd. Dit en het zeer variabele aanbod van patiënten met variërende zorginhoud maakt het indelen van laboranten een complexe zaak. In onze aanpak gebruiken we een nieuw stochastisch "mixed integer linear programming" (MILP) model dat de toewijzing van laboranten optimaliseert, gegeven verschillende patiëntinstroom scenario's die gebaseerd zijn op historische data. Het doel is om het (verwachte) aantal patiënten dat de voorbereidingsfase binnen de nagestreefde wachttijd afmaakt, te maximaliseren. De resultaten van een casestudy op de RT afdeling van het Nederlands Kanker Instituut (NKI) laten zien dat het aantal patiënten dat binnen de maximale wachttijd kan starten, gemiddeld kan toenemen van 91.3% naar 97.9% voor subacute patiënten, en van 96.3% naar 100.0% voor reguliere patiënten.

Het radiotherapietraject wordt ofwel gestuurd door de planning van de eerste bestralingssessie, direct na het eerste contact tussen de patiënt en de behandelaar (pull-strategie), of na afloop van het volledige voorbereidingstraject (push-strategie). In hoofdstuk 4 beoordelen we de impact van het gebruik van pull- en push-strategieën en onderzoeken we met behulp van discrete simulatiemodellen het effect van alternatieve strategieën om de tijdigheid van de bestraling te verbeteren. Met behulp van personeelsenquêtes, interviews met teamleiders en historische gegevens van het NKI (2017) is input voor het model gegenereerd, rekening houdend met fluctuaties in de instroom van patiënten en de beschikbaarheid van personeel en middelen. Het blijkt dat bij een pull-strategie 41% minder patiënten de beoogde wachttijd overschrijden dan een hybride strategie (40% pull / 60% push, de klinische praktijk in het NKI), ondanks een iets langere gemiddelde wachttijd (12%).

Hoofdstuk 5 onderzoekt het automatisch inplannen van bestralingssessies waarbij zo goed mogelijk rekening wordt gehouden met de voorkeuren van de patiënt met betrekking tot het tijdstip van de bestralingsafspraak. Terwijl de meeste literatuur zich richt op het optimaliseren van de doorlooptijd, hebben de samenwerkende RT-centra kenbaar gemaakt dat ze graag de voorkeuren van de patiënt hierin willen betrekken. In deze studie stellen we een MILP-model voor dat er naar streeft dat alle sessies binnen het gewenste venster worden ingepland. Voor kleinere centra, tot 66 patiënten per week verdeeld over twee linacs, is dit probleem in een redelijke rekentijd op te lossen. Voor grotere centra stellen we een heuristische methode voor waarmee patiënten vooraf aan een bepekte set uitwisselbare linacs worden toegewezen zodat het probleem in deelproblemen (clusters van linacs) wordt opgedeeld. Deze deelproblemen worden vervolgens geoptimaliseerd met het MILP-model. Deze sequentiële aanpak is getest met behulp van realistisch data van het NKI. Het bleek dat het probleem kan worden opgelost in een redelijke rekentijd (3.3 uur) met slechts 2.8% van de geplande sessies buiten het gewenste 90-minuten tijdsvenster .

In hoofdstuk 6 hebben we het MILP-model van hoofdstuk 5 aangepast om voor twee Nederlandse RT-centra weekschema's te genereren voor hun RT behandelingen. Het model werd iteratief aangepast om te voldoen aan de technische en medische randvoorwaarden van elk centrum totdat een geldige oplossing werd bereikt. Patiëntengegevens werden verzameld uit de interne databases en de haalbaarheid van de verkregen schema's werd geverifieerd door de medewerkers van elk centrum. Het bleek dat met dit MILP-model een haalbare, kwalitatief hoogwaardige planning gemaakt kan worden op een betere en snellere manier dan de huidige aanpak. De wekelijkse planning is in beide centra verbeterd: de gemiddelde standaardafwijking tussen de starttijden van de sessies is in het ene centrum (NKI) met 47% afgenomen, terwijl het aantal gaten in de planning in het andere centrum (Bernard Verbeeten Instituut) met 72% is afgenomen. Het proces vereiste respectievelijk 1.5 uur en 5 minuten CPU-tijd voor de 2 centra, onvergelijkbaar snel ten opzichte van de 1.5 dag die in beide instituten nodig is bij de huidige, handmatige planning.

De in dit proefschrift voorgestelde OR-modellen laten aanzienlijke potentiële voordelen zien indien de resultaten op nationale schaal kunnen worden geëxtrapoleerd. Het algoritme van hoofdstuk 3 liet bijvoorbeeld zien dat ons model potentieel kan leiden tot een kostenbesparing van 69 FTE's op nationaal niveau. Ook heeft ons model, door het verkrijgen van compactere schema's voor een van de testcases (hoofdstuk 6), laten zien dat in totaal ongeveer 188 FTE's kunnen worden bespaard. De implementatie van OR-tools in de RT-praktijk is echter niet eenvoudig. Het is lastig de theoretische verbeteringen, die komen van modellen die de werkelijkheid slechts deels beschrijven, klinisch te implementeren. Eén van de oorzaken is dat de te behalen resultaten niet altijd intuïtief zijn en de benodigde veranderingen, in de ogen van de medewerkers, complex zijn. Een succesvolle implementatie vereist input en organisatorische verbinding van alle relevante belanghebbenden en onderzoekers vanaf het begin van het project. Eventuele sociodynamische aspecten van werkprocessen moeten in een vroeg stadium onderkend worden. Uit de implementatie-inspanningen binnen dit project is ook gebleken dat geleidelijke modelaanpassingen in kleine stappen wenselijk is, gecombineerd met constante communicatie om de vertaling van de OR-modellen naar de klinische praktijk te bewerkstelligen. Tot slot bleek dat de in dit proefschrift voorgestelde onderzoekmodellen in staat zijn om de besluitvorming te ondersteunen voor RT-centra die hun processen efficiënter willen organiseren ten behoeve van patiënten en personeel. Hoewel verschillende hindernissen de toepassing van deze modellen in de praktijk belemmeren, is gebleken dat OR methoden effectief zijn in het oplossen van overzichtelijke logistieke problemen in RT, waar haalbare (bijna-optimale) oplossingen moeilijk handmatig te realiseren zijn. Over het algemeen zijn geavanceerde analysemethoden op basis van OR in staat gebleken om de logistiek van radiotherapiebehandelingen te optimaliseren voor een betere kwaliteit van zorg.

Acknowledgements

More people have contributed directly or indirectly to this thesis than I can mention. I am deeply grateful to all of them. Some I like to thank in particular, below.

First of all I want to thank my promotors, prof. dr. **Wim van Harten** and prof. dr. ir. **Erwin Hans**. **Wim**, thank you for your guidance and support throughout the project. You have always been there for everything I needed, always with comforting words that helped me gain confidence to overcome adversities. **Erwin**, it has been a pleasure to work with you all these years. I won't forget our long and deep conversations that transcend the valuable scientific insights that I have learned with you. Thank you so much for the encouraging words you had for me every time I needed them, and for being an inspiration to be not only a better researcher, but most importantly a better human being. A word of gratitude to my daily supervisors dr. **Jeroen van de Kamer** and dr. **Derya Demirtas**. **Jeroen**, I will not forget the countless discussions and brainstorming sessions we've had. You have backed me up when I needed the most and helped me stay on track to fulfill the demanding activities of the project on time. **Derya**, you have inspired me greatly with your ever-positive attitude and motivating words, and I am very grateful for your contributions to the project. It is a pity that circumstances did not allow us to work more closely for a longer period of time, I know I could have learned even more from you.

I would also like to give a special word of thanks to my former professor and supervisor dr. **Ana Viana**. **Ana**, you know the very important role you have had in the definition of my career path. Apart from all the technical training you have provided me during the last decade, you have inspired me to find the right motivation, work ethic, and attitude to grow, both as a professional and as a person, over these years. I will always be extremely grateful to you for that. I express a word of gratitude to the remaining members of my graduation committee: prof. dr. **Nico Verdonschot**, prof. dr. **Richard Boucherie**, prof. dr. **Coen Rasch** and prof. dr. **Corrie Marijnen**, thank you for the time and effort you have spent evaluating and commenting on the content of this thesis.

The results obtained in this thesis would not be achievable without the contribution of many of my colleagues from the radiotherapy department of the NKI. Although we could not work together as much as I would have liked, I am very grateful to dr. **Corine van Vliet-Vroegindeweij** and dr. **Roel de Boer** for their valuable contribution to the project. **Corine** and **Roel**, thank you for your engagement and leadership, and for guiding me during the early stages of my trajectory. I am also specially grateful to **Maarten Broekhof**, who provided me with all the necessary datasets needed to conduct experiments and test my models. I also thank **Floris Pos, Joost Knegjens, Peter Visser, Eva Euser, Pauline Roose, Anke van Mourik, Anja Betgen, Monique Gigengack, Mirelle de Valk, Jacqueline van der Geest, Bram van den Heuvel, Hein Opdam, Maarten Wolfrat, Iris Walraven, Clasca Daalman-van der Pol,** and **Herman Vijlbrief** for their availability to meet and provide data and information that were crucial for development of this research work. **Jasper ten Dam**, thank you for your collaboration on building the simulation model used as a basis for Chapter 4 of this thesis. **Diedie van Dinten** and **Patricia Fewer**, thank you for making my life so much easier as a researcher in the department.

To the members of the collaborating radiotherapy centers, **Willy de Kruijf, Willem Jongste, Rob van Os, Kees Visscher, Marc Hoogendoorn, André Minken,** thank you for your cooperation and for the sterling contribution you have made for the development and implementation of this research work.

During four months of my PhD trajectory, I had the opportunity to travel to the snowy Montreal, Canada, to work together with prof. dr. **Louis Martin-Rousseau** and dr. **Nadia Lahrichi** at the Polytechnique Montreal. That was a fruitful research visit where I could learn a lot from my local supervisors and colleagues. **Louis-Martin** and **Nadia**, I have greatly enjoyed working together with you and thank you so much for giving me this opportunity.

To my family: thank you very much for all your support in my life decisions and for providing me with the means I needed for moving forward with my career choices. I will be eternally grateful to **Gaspar Vieira** and **Rosa Soares** for making me the person that I am today. To my brothers **André Vieira** and **Ricardo Vieira**: I have been very lucky by having grown up with you, and by having you on my side when I needed your support the most.

A very special thanks to my paranymphs, **João Neto** and **Zeno Gouw**. Thanks for accepting the challenge, and for being there to cheer me up whenever I needed. My time in Amsterdam has been great and much of that is thanks to your friendship and fellowship. **João**, you have been my soulmate during this time, I could always count on you to either join my plans or take me with you in your crazy adventures. **Zeno**, I have always greatly admired you for your kind personality, for being the best listener I have ever met, and for always having a sound and authentic advice to give. I feel very lucky for having had the chance to share your house with you during the last year. Special words for my other Amsterdam "brother" and adventure companion **Iban Torres Xirau**. **Iban**, you helped me integrate in the hospital and the city from the very first moment we met, making me feel like home my entire time in the hospital. More than a friend, you were the best colleague, party partner, soccer teammate, conference roommate, etc. that I could ask for. I cannot find the words to thank you enough for that. **Mar Soto**, you will also hold a special place in my memories of the time I have spent in town. Thanks for worrying about me and cheer me up when I appeared to be down, and for being always so sweet and supportive to me.

My gratitude to my friends and colleagues from the radiotherapy depart-

ment of the NKI, who made me have a really good time in and outside of the hospital: **Tessa van de Lindt, Tom Konert, Kleopatra Pirpinià, Uros Stankovic, Roel Rozendaal, Patrick Gonzalez, Igor Olaciregui-Ruiz, Theo Driever, Nikita Moriakov, Sander Uilkema, Natasja Jansen, Chris Beekman, Anke Wiersma, Maddalena Rossi, Barbara Stam, Artem Knmelinskii, Ghazaleh Ghobadi, Celia Juan, Pavlos Papaconstadopoulos, Lennert Ploeger, Natascha Bruin, Kai Lonning, Lukas Schroder, Ernst Kooreman, Yenny Szeto, Vineet Mohan, Deji Adams, Scott Marshal, Priyanka Ojha, Minh Ann,** it has been a pleasure to spend time with you. Our borrels, dinners, parties, retreats, conferences, will not be forgotten. To my friends from the Portuguese crew within the institute: **Estela Cessa, Catarina Fernandes, Suzana Teixeira,** thank you so much guys for never letting me feel home sick!

Special thanks to my colleagues and friends from the PSOE department and Wim van Harten research group: **Melanie Lindenberg, Nora Franzen, Danalyn Byng, Jacobien Kieffer, Laura Kooij, Ann-jean Beck, Hester van de Wiel, Joost Verbeek, Valesca Rétel, Wim Groen, Lisanne Hummel, Anna Micquel Cases, Wilma Kuijpers, Willeke Naaktgeboren**. It has been a pleasure to meet you all and spend time together in my days at the PSOE department. Our outings, dinners and borrels were all super joyful for me. I feel a very proud member of the group. A special word of thanks to **Anke Wind** who has helped me go through the hard times when I most needed.

To my friends and colleagues from the CHOIR group and the University of Twente: **Maartje van de Vrugt, Thomas Schneider, Gréanne Leeftink, Joost van Sambeeck, Sem van Brummelen, Aleida Braaksma, Eline Tsai, Jasper Bos, Maarten Otten, Nardo Borgman, Ingrid Vliegen, and Ingeborg Bikker**. Thank you very much for making my day trips to Enschede more exciting. I have truly enjoyed the time we have spent together in our conference and course trips, group outings, and seminars.

To my Canada friends **Bruno Vasconcelos, Tiago Alves, Catarina Pimentel, Jaysika Loureiro, Patricia Patou Perrier, Dina Bentayeb,** and **Margarida Carvalho**, thank you so much for making me feel like home during my stay in Montreal. My time in Canada became very joyful and the cold winter became bearable thanks to your kindness and friendship.

I am thankful to all my friends whom I have spent quality time with during the course of my PhD trajectory. A special word of gratitude and nostalgia to my former roommates **Živa Pogacar, Alessanda Giustozzi** and **Ronak Shah**. I have truly enjoyed sharing the house with you guys, we shared many fun, awkward, but memorable moments together. I will keep those with me. **Isabel Mayayo Peralta**, thank you for all your support and for all the good times we spent together. **Antonio Mulero, Anna Gonzalez, Jeremy Tessier, Mariana Paes, Eric Pinto, Louise Janssen, Lorenzo Bombardello, Maria Escala Garcia, Jonne Raaijmakers, Marcelo Leite, Hayra Vieira, Danielle Carvalho, Veronica della Chiara, Pien van der Loo, Giulia de Conti, Renato Teixeira, Georgina Ferrer, Laura Ferrer,** I will not forget all the borrels, house parties, dinners,

festivals, and the countless events we have spent together, which gave me so much fun while discovering the city. You are part of the best memories I will bring back from the Netherlands. **Pablo Soria, Pedro Campos, João Ribeiro, Sérgio Silva, Sara Nogueira, Mário Couto,** and **Marco Ribeiro**, I will not forget your comforting support as well as the trips we made together during my time abroad. You have always kept in contact with me and made sure everything was alright which I truly appreciate. The same goes for **Augusto Alves, Bruno Ferreira, Pedro Sevilha, João Pereira**, and **Evelyn Soto**: thank you for your loyal support, and for being there for me when I needed the most during my time in the Netherlands. A last word of thanks to **Thaís Sales**, whom I have had the fortune of having on my side to support me during the difficult last stages of my PhD.

Bruno,
Amsterdam, April 2020

About the author

Bruno Vieira was born in Porto, Portugal, on November 18, 1987. He received a Bachelor's degree in Electrical Engineering and Computer Science in 2011, and a Master's degree in Systems Engineering and Industrial Planning in 2013, both from the School of Engineering of the Polytechnic Institute of Porto. During his university studies, Bruno participated in the ERASMUS programme in two separate occasions. In 2009/2010 he studied at the University of Maribor (Slovenia) for one academic year, and in 2013 he performed part of his Master's thesis work at the University of Bremen (Germany). From 2012-2015, Bruno collaborated with the Centre for Industrial Engineering and Management of the Institute for Systems and Computer Engineering, Technology and Science (INESC) as a research assistant. At INESC, Bruno designed, developed, and applied Operations Research (OR) methods to solve decision-making problems in two different projects. One focused on electricity generation planning with wind-hydro-thermal coordination, and in the other project Bruno designed a hybrid method that combined a genetic algorithm and a mathematical optimizer to solve sequential ordering problems found in, e.g., crane scheduling.

In February 2015, Bruno joined the Netherlands Cancer Institute (NKI) and the Center for Healthcare Operations Improvement and Research (CHOIR) of the University of Twente as a PhD researcher. He then started a 5-year project called "Advanced Logistical Optimization of the Radiotherapy Treatment", funded by the Dutch Cancer Society (KWF). During his PhD trajectory, Bruno participated as a logistics advisor in several small projects of the RT department of the NKI and gave tutoring talks about OR for the department researchers. During his time at the NKI, Bruno gave over a hundred public presentations split amongst conferences, seminars, department meetings, and research clubs. From September to December of 2018, Bruno visited prof. dr. Louis-Martin Rousseau and dr. Nadia Lahrichi from the CIRRELT group of the Polytechnique Montreal, Canada. The research work performed in Montreal was used as the basis for Chapter 5.

Bruno has been a competitive futsal player, having helped his team (Os Lusitanos) promote from the third to the first division of the North Holland regionals during the 5 seasons (2015-2020) he has played for the Amsterdam club.

As for the future, Bruno is interested in using the acquired knowledge and skills to develop OR-based tools for decision-making support in a more practical setting.

List of publications

Included in this thesis:

Vieira B., Hans E.W., van Vliet-Vroegindeweij C., van de Kamer J.B., van Harten W. (2016) Operations research for resource planning and -use in radiotherapy: a literature review. *BMC Medical Informatics and Decision Making* 16(1):149.
(Basis for Chapter 2.)

Vieira B., Demirtas D., van de Kamer J.B., Hans E.W., van Harten W. (2018) A mathematical programming model for optimizing the staff allocation in radiotherapy under uncertain demand. *European Journal of Operational Research* 270:709–722.
(Basis for Chapter 3.)

Vieira B., Demirtas D., van de Kamer J.B., Hans E.W., van Harten W. (2019) Improving workflow control in radiotherapy using discrete-event simulation. *BMC Medical Informatics and Decision Making* 19(1):199.
(Basis for Chapter 4.)

Vieira B., Demirtas D., van de Kamer J.B., Hans E.W., Rousseau L.M., Lahrichi N., van Harten W. (2020) A Mathematical programming model for radiotherapy scheduling with time Windows. In: Bélanger V., Lahrichi N., Lanzarone E., Yalçındağ S. (eds) Health Care Systems Engineering. HCSE 2019. *Springer Proceedings in Mathematics & Statistics* 316:241–249.

Vieira B., Demirtas D., van de Kamer J.B., Hans E.W., Rousseau L.M., Lahrichi N., van Harten W. (2020) Radiotherapy treatment scheduling considering time window preferences. *Submitted*.
(Basis for Chapter 5.)

Vieira B., Demirtas D., van de Kamer J.B., Hans E.W., Jongste W., van Harten W. (2020) Radiotherapy treatment scheduling: implementing operations research into clinical practice. *Submitted*.
(Basis for Chapter 6.)

Not included in this thesis:

Vieira B., Viana A., Matos M., Pedroso J.P. (2016) A multiple criteria utility-based approach for unit commitment with wind power and pumped storage hydro. *Electric Power Systems Research* 131:244–254.

In external-beam radiotherapy, delays in the start of treatment can negatively affect the patient's outcome and quality of life. Radiotherapy pre-treatment operations (CT, MRI, tumor delineation, treatment planning, etc.) are tightly and time-wise connected, personnel are highly specialized, and the required equipment may be dedicated to patients with specific tumor sites. Managing patient flows in such an environment is challenging and is further complicated by the high levels of uncertainty in the process, such as a highly fluctuating patient inflow. As the number of patients diagnosed with cancer increases and treatment pathways become more and more personalized, avoiding delays in the start of treatment and under/over utilization of resources becomes increasingly difficult for radiotherapy centers. This thesis proposes innovative approaches developed using Operations Research and Management Science methods for the logistical optimization of radiotherapy processes to support decision-makers plan and use their resources more efficiently for a better quality of care.

ISBN: 978-90-365-4997-4