

This is the author's post-print version of a manuscript accepted for publication in European Journal of Operational Research. This version does not include post-acceptance editing and formatting. Readers who wish to access the published version of this manuscript should go to <https://doi.org/10.1016/j.ejor.2018.03.040>. Those who wish to cite this manuscript should cite the published version.

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

Bruno Vieira^{a,b,c}, Derya Demirtas^{b,d}, Jeroen B. van de Kamer^a, Erwin W. Hans^{b,d}, Wim van Harten^{a,c,e}

^a*Department of Radiation Oncology, Netherlands Cancer Institute, Amsterdam, The Netherlands*

^b*Center for Healthcare Operations Improvement and Research (CHOIR), University of Twente, Enschede, the Netherlands*

^c*Department of Health Technology and Services Research, Faculty of Behavioural Management and Social Sciences, University of Twente, Enschede, The Netherlands*

^d*Department of Industrial Engineering and Business Information Systems, Faculty of Behavioural Management and Social Sciences, University of Twente, Enschede, The Netherlands*

^e*Rijnstate General Hospital, Arnhem, The Netherlands*

Abstract

As the number of people diagnosed with cancer increases, demand for radiotherapy (RT) services has been continuously growing. In RT, 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. The RT process, which involves imaging and treatment planning before treatment, is subject to complexities that hamper resource planning and control. On the demand side, the amount of workload in each operation depends on the highly variable patient inflow. On the supply side, radiation therapy technologists (RTTs) have multiple skills, rotation needs and partial availability, which makes the allocation of RTTs a complex task that often leads to situations of understaffing, jeopardizing the fulfillment of the patients' waiting time standards. In this paper, we propose a stochastic mixed-integer linear programming model that optimizes the allocation of RTTs to multiple operations in RT over a set of scenarios of patient inflow. The 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 the waiting time target. Results for a case study in the RT department of the Netherlands Cancer Institute show 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.

Keywords: OR in health services, Staff allocation, Mathematical Programming, Radiotherapy, Stochastic programming

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 (Stewart and Wild, 2014), and the global use of radiotherapy in cancer care is estimated to be around 50% (Delaney et al., 2005).

In RT, delays in the start of treatment increase the risk of local recurrence and tumor progression (Chen et al., 2008), and patients experience greater psychological distress and prolonged symptoms when subject to longer waiting times (Mackillop, 2007). For these reasons, standards for maximum waiting times ¹ have been set by the Dutch Society for Radiotherapy and Oncology (NYRO) (NYRO, 2017). 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 of the main causes for patient dissatisfaction regarding pain management in RT (Pignon et al., 2004). 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. (2016). In their review, it is shown that although most of the published studies focus on the scheduling of treatment sessions (such as Conforti et al. (2010), Sauré et al. (2012), and Legrain et al. (2015)), 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.

¹time between referral and first irradiation session.

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.

2. Problem statement and background

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 1 shows the maximum waiting times in the Netherlands, as recommended by the NVRO.

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

Table 1: Waiting time targets defined by the NVRO, in calendar days

Fig. 1 depicts a typical deployment flowchart of the main operations involved in external-beam RT. At referral, patients are scheduled for a consultation 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.

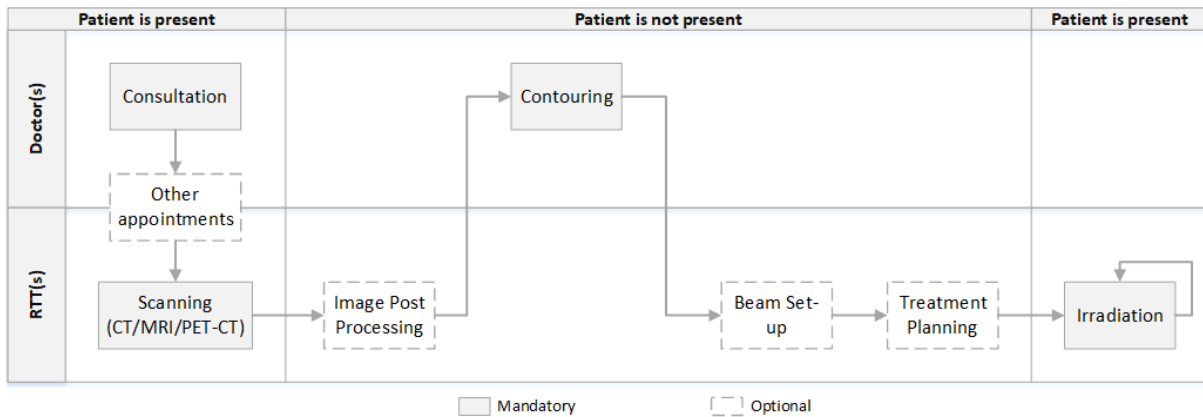


Figure 1: Deployment flowchart of the RT process

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 Fig. 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 we present our innovative solution methodology designed to solve the RTT planning problem. Section 4 presents the computational experiments designed to test our methodology using real data. In Section 5 we present the results obtained from the case study and elaborate a discussion of the results, and in Section 6 we draw the conclusions of this study and elaborate on future research lines.

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 (Vieira et al., 2016). Bikker et al. (2015) 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 (2012) 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. (2015) 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. (2011) 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. (2009) 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. (2012) 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 health care settings (Rais and Viana, 2011). Hulshof et al. (2013) 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 health care settings in which waiting lists arise between care processes. Jerić and Figueira (2012) 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. (2012) 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. 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 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 pre-defined 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.

3.1. 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 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.

3.2. Notation

The sets, subsets and indices used in the MILP model can be found in Tables 2 and 3. The list of input parameters is presented in Table 4.

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 2: Sets and indices of the MILP model.

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 of binary variables and continuous variables. We let Y_{pit} be 1 if operation i of patient p

Subset	Description
\mathcal{P}^A	set of actual patients
\mathcal{P}^ω	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

Table 3: Subsets of the MILP model.

Parameter	Description
<i>Integer</i>	
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 , $\forall t \in \mathcal{T}$
$\text{tra}_{r,i}$	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 , $\forall p \in \mathcal{P}$
<i>Binary</i>	
rav_{rt}	1 if RTT r is available during time period t , 0 otherwise, $\forall r \in \mathcal{R}, \forall t \in \mathcal{T}$
$\text{cap}_{r,i}$	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, $\forall r \in \mathcal{R}$
<i>Continuous</i>	
mav_{imt}	time, in minutes, that machine m of operation i works during period t , $\forall i \in \mathcal{O}^c, \forall m \in \mathcal{M}^i, \forall t \in \mathcal{T}$

Table 4: Input parameters of the MILP model.

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 5.

Variable	Description
<i>Binary</i>	
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 RRT 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, $\forall p \in \mathcal{P}, \forall t \in \mathcal{T}$
<i>Continuous</i>	
T_{pit}	time, in minutes, spent on operation i of patient p during period t , $\forall p \in \mathcal{P}, \forall i \in \mathcal{O}, \forall t \in \mathcal{T}$

Table 5: Decision variables of the MILP model.

3.4. Constraints

3.4.1. 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 (Ruiz and Vázquez-Rodríguez, 2010). 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 (1) ensure that patient operations can only be scheduled once the patient is ready for RT. Constraints (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) and (4), respectively. Inequalities (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 (5). Constraints (6) force a delay c_p between the imaging planning processes for contouring of the tumor by the doctor(s). Finally, Constraints (7) and (8) bound the values of the continuous variables T_{pit} .

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

$$\sum_{t \in \mathcal{T}} Y_{pit} \leq 1, \forall p \in \mathcal{P}, \forall i \in \{\mathcal{O}^p \cap \mathcal{O}^C\} \quad (2)$$

$$\sum_{n=1}^t Y_{pi'n} \geq Y_{pit}, \forall p \in \mathcal{P}, \forall i \in \mathcal{O}^p, \forall i' \in \{\mathcal{O}^i \cap \mathcal{O}^p \cap \mathcal{O}^C\}, \forall t \in \mathcal{T} \quad (3)$$

$$\sum_{n=1}^t T_{pi'n} \geq Y_{pit} \cdot p_{gi'}, \forall g \in G, \forall p \in \mathcal{P}^g, \forall i \in \mathcal{O}^p, \forall i' \in \{\mathcal{O}^i \cap \mathcal{O}^p \cap \mathcal{O}^N\}, \forall t \in \mathcal{T} \quad (4)$$

$$\sum_{i \in \{\mathcal{O}^p \cap \mathcal{O}^N\}} T_{pit} + \sum_{i \in \{\mathcal{O}^p \cap \mathcal{O}^C\}} Y_{pit} \cdot p_{gi} \leq l_t, \forall g \in G, \forall p \in \mathcal{P}^g, \forall t \in \mathcal{T} \quad (5)$$

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

$$i = \min\{b \in \mathcal{O}^p \cap \{5..6\}\}$$

$$0 \leq \sum_{t \in \mathcal{T}} T_{pit} \leq p_{gi}, \forall g \in G, \forall p \in \mathcal{P}^g, \forall i \in \{\mathcal{O}^p \cap \mathcal{O}^N\} \quad (7)$$

$$Y_{pit} \cdot h_i \leq T_{pit} \leq Y_{pit} \cdot l_t, \forall p \in \mathcal{P}, \forall i \in \{\mathcal{O}^p \cap \mathcal{O}^N\}, \forall t \in \mathcal{T} \quad (8)$$

3.4.2. Constraints for RTT allocation

We aim to satisfy the restrictions encountered by RT managers when producing RTT master mid-term schedules. Constraints (9) assign each RTT to at most one operation, if he/she is available in the corresponding time period. Inequalities (10) and (11) ensure that an RTT is not assigned to an operation that he/she cannot perform. Relations (12) match demand and supply restrictions for non-continuous operations in terms of RTT availability, and Constraints (13) serve the same purpose but for continuous operations as a function of machines' availability. Constraints (12) and (13) ensure that there is enough supply to cover the estimated workload, per operation and time period, in all scenarios. Restrictions (14) and (15) assign the necessary number of RTTs to operate imaging machines and linacs, respectively.

$$\sum_{i \in \mathcal{O}} X_{rit} + L_{rt} \leq \text{rav}_{rt}, \forall r \in \mathcal{R}, \forall t \in \mathcal{T} \quad (9)$$

$$X_{rit} \leq \text{cap}_{ri}, \forall r \in \mathcal{R}, \forall i \in \mathcal{O}, \forall t \in \mathcal{T} \quad (10)$$

$$L_{rt} \leq q_r, \forall r \in \mathcal{R}, \forall t \in \mathcal{T} \quad (11)$$

$$\sum_{p \in \{\mathcal{P}^A \cup \mathcal{P}^\omega\}} T_{pit} \leq \sum_{r \in \mathcal{R}} X_{rit} \cdot l_t, \forall i \in \mathcal{O}^N, \forall t \in \mathcal{T}, \forall \omega \in \Omega \quad (12)$$

$$\sum_{g \in G} \left(\sum_{p \in \{\mathcal{P}^g \cap (\mathcal{P}^A \cup \mathcal{P}^\omega)\}} Y_{pit} \cdot p_{gi} \right) \leq \sum_{m \in \mathcal{M}^i} \text{mav}_{imt}, \forall i \in \mathcal{O}^C, \forall t \in \mathcal{T}, \forall \omega \in \Omega \quad (13)$$

$$\sum_{r \in \mathcal{R}} X_{rit} = \sum_{m \in \mathcal{M}^i} k_i \cdot \min\{\text{mav}_{imt}, 1\}, \forall i \in \mathcal{O}^C, \forall t \in \mathcal{T} \quad (14)$$

$$\sum_{r \in \mathcal{R}} L_{rt} = f_t, \forall t \in \mathcal{T} \quad (15)$$

Constraints (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 (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 .

$$p_{gi} - \sum_{n=1}^t T_{pin} \leq (1 - W_{pt}) \cdot p_{gi}, \quad \forall g \in G, \forall p \in \mathcal{P}^g, i = \max\{i' : i' \in \mathcal{O}^p\}, \forall t \in \mathcal{T} \quad (16)$$

$$\sum_{n=1}^t T_{pin} - p_{gi} \leq W_{pt} - 1, \quad \forall g \in G, \forall p \in \mathcal{P}^g, i = \max\{i' : i' \in \mathcal{O}^p\}, \forall t \in \mathcal{T} \quad (17)$$

3.5. Objective function

The objective function (18) aims to maximize the number of actual patients (\mathcal{P}^A) and the expected number of future patients (\mathcal{P}^ω) amongst scenarios $\omega \in \Omega$ finishing the pre-treatment stage within the waiting time target d_p .

$$\max \sum_{p \in \mathcal{P}^A} W_{p,d_p} + \frac{1}{|\Omega|} \sum_{\omega \in \Omega} \sum_{p \in \mathcal{P}^\omega} W_{p,d_p} \quad (18)$$

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 (19) and (20), which model the rotation needs for pre-treatment operations and linacs, respectively.

$$\sum_{t \in \mathcal{T}} X_{rit} \geq \text{tra}_{ri}, \forall r \in \mathcal{R}, \forall i \in \mathcal{O} \quad (19)$$

$$\sum_{t \in \mathcal{T}} L_{rt} \geq \text{lin}_r, \forall r \in \mathcal{R} \quad (20)$$

4. NKI case study

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 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. $h_i = 15$.

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

Table 6: Daily availability, in minutes, of the imaging machines of the RT department at the NKI.

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 7 shows the skills for a subset of the RTTs of the NKI.

	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

Table 7: RTT skills in the RT department of the NKI for a subset of RTTs.

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 (ELEKTA, 2017), 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 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.

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 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.

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

¹The mean difference is statistically significant below the 0.05 level (*).

²The mean difference is statistically significant below the 0.05 level (*).

(I) Weekday	(J) Weekday	Mean Diff. (I-J)	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

Table 8: ANOVA of the patient arrivals in the NKI during the year 2016².

Weekday	Sample size	Prob. Dist.	Mean	Std. Dev.	<i>p</i> -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

Table 9: Statistical analysis of the patient arrivals of the NKI, per weekday, during the year 2016.

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 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.

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%

Table 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.

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 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 in three sequential stages, as a hierarchical optimization approach (see Fig. 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:

$$\begin{aligned} \max z_1 &= \sum_{p \in \mathcal{P}^A} W_{p,|\mathcal{T}|} + \frac{1}{|\Omega|} \sum_{\omega \in \Omega} \sum_{p \in \mathcal{P}^\omega} W_{p,|\mathcal{T}|} \\ \text{subject to:} & \quad (1) - (16) \end{aligned} \tag{21}$$

In the second stage, we add the objective value z_1^* obtained by solving problem (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 1) minus the access time needed due to the scheduling on the linacs (dl_p). Thus, we solve the following problem:

$$\begin{aligned} \max z_2 &= \sum_{p \in \mathcal{P}^A} W_{p,d_p^1} + \frac{1}{|\Omega|} \sum_{\omega \in \Omega} \sum_{p \in \mathcal{P}^\omega} W_{p,d_p^1} \\ \text{subject to:} & \quad (1) - (16), z_1 \geq z_1^*. \end{aligned} \tag{22}$$

In the third stage, the objective value z_2^* obtained by solving the program (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:

$$\begin{aligned} \max z_3 &= \sum_{p \in \mathcal{P}^A} W_{p,d_p^2} + \frac{1}{|\Omega|} \sum_{\omega \in \Omega} \sum_{p \in \mathcal{P}^\omega} W_{p,d_p^2} \\ \text{subject to:} & \quad (1) - (16), z_1 \geq z_1^*, z_2 \geq z_2^*. \end{aligned} \tag{23}$$

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.

4.3.1. 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 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

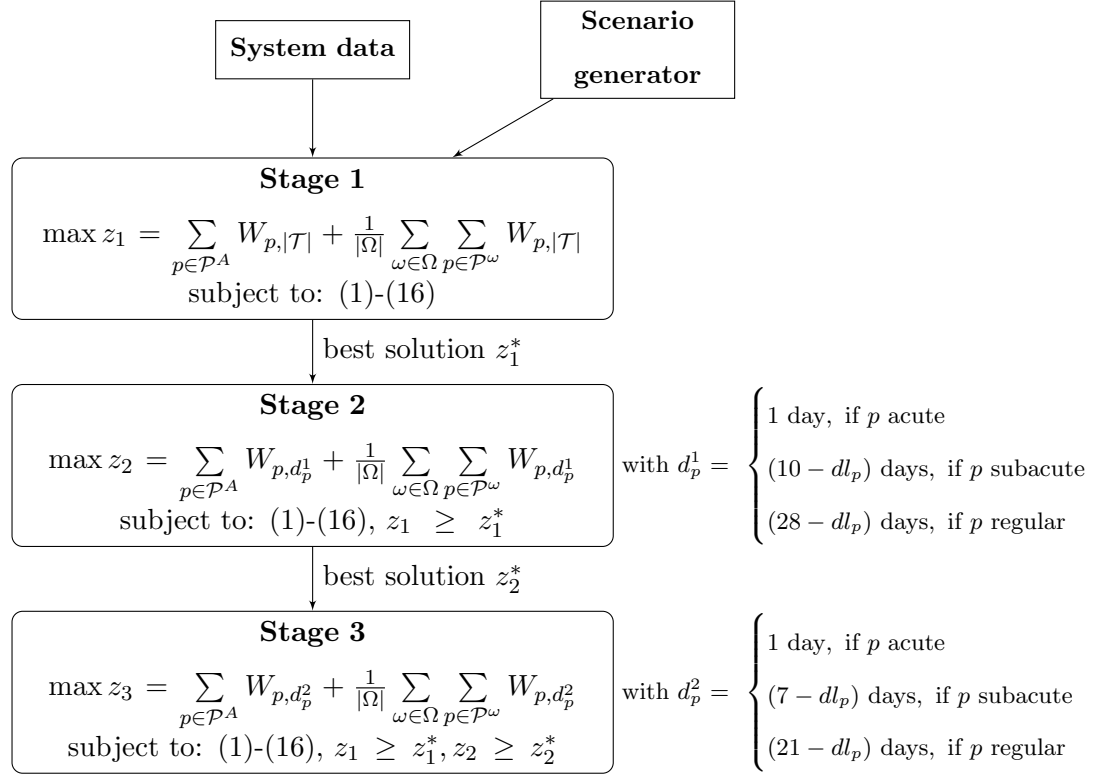


Figure 2: Overview of the solution approach designed for the NKI case study.

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 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 9 and 10). Fig. 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.

The number of patient arrivals and distribution of patients per urgency level can be seen in Table 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 11

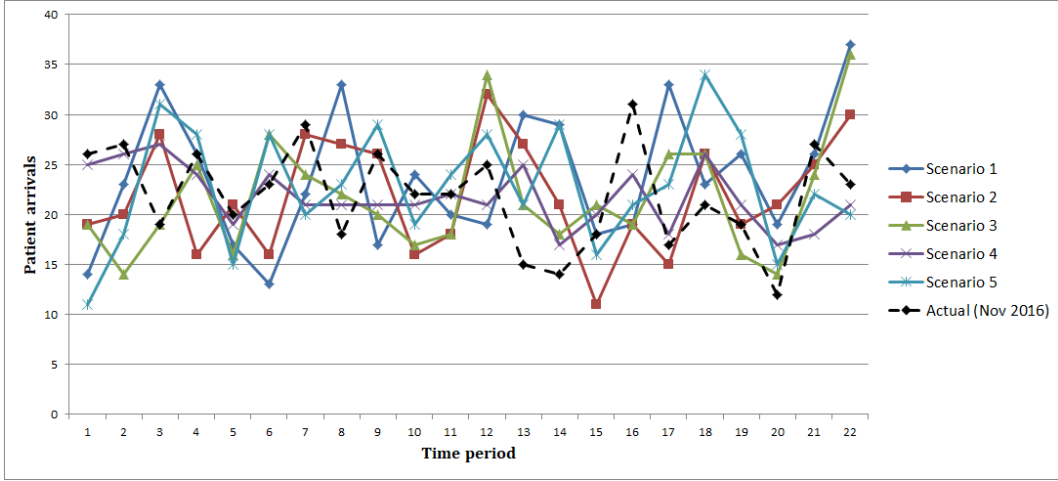


Figure 3: Patient arrivals, per time period, for the scenarios created by the scenario generator (solid lines) for generating the RTT allocation solution, and patient arrivals measured in practice during November 2016 (dashed line).

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 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.

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

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

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%

Table 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.

gap tolerance of 5% and a CPU time limit of 14400 seconds per stage.

Table 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 below 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 14.

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 13: Solver information for the solution generation.

4.4. Solution evaluation

To evaluate the generated RTT solution, we re-run the solution framework depicted in Fig. 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

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

Table 14: Results of the allocation of RTTs for the NKI case study.

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 Fig. 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 4.3.1, reaching similar conclusions. The remaining parameters were kept the same as in the solution generation experiments.

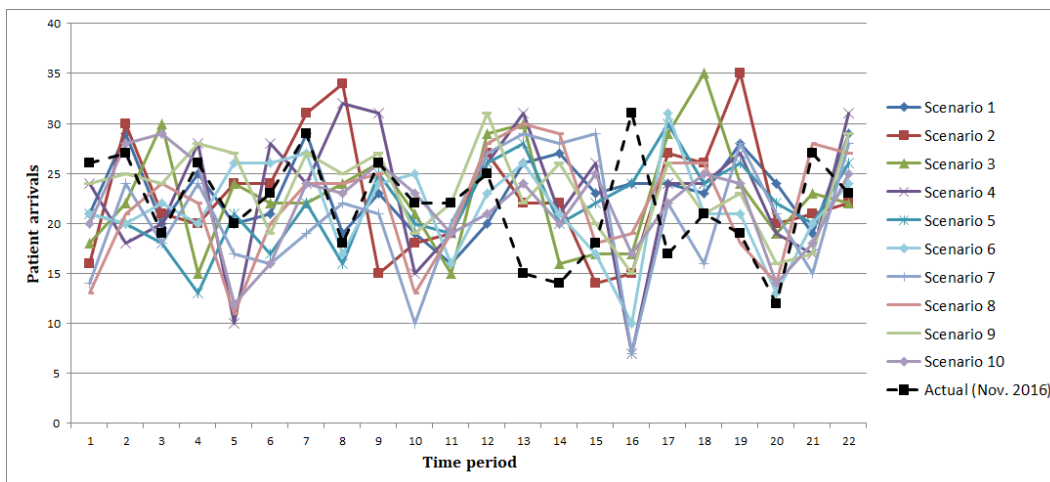


Figure 4: Patient arrivals, per time period, for the scenarios created by the scenario generator (solid lines) for evaluating the RTT allocation solution, and patient arrivals measured in practice during November 2016 (dashed line).

5. Results and discussion

Table 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 (Fig. 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 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 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.

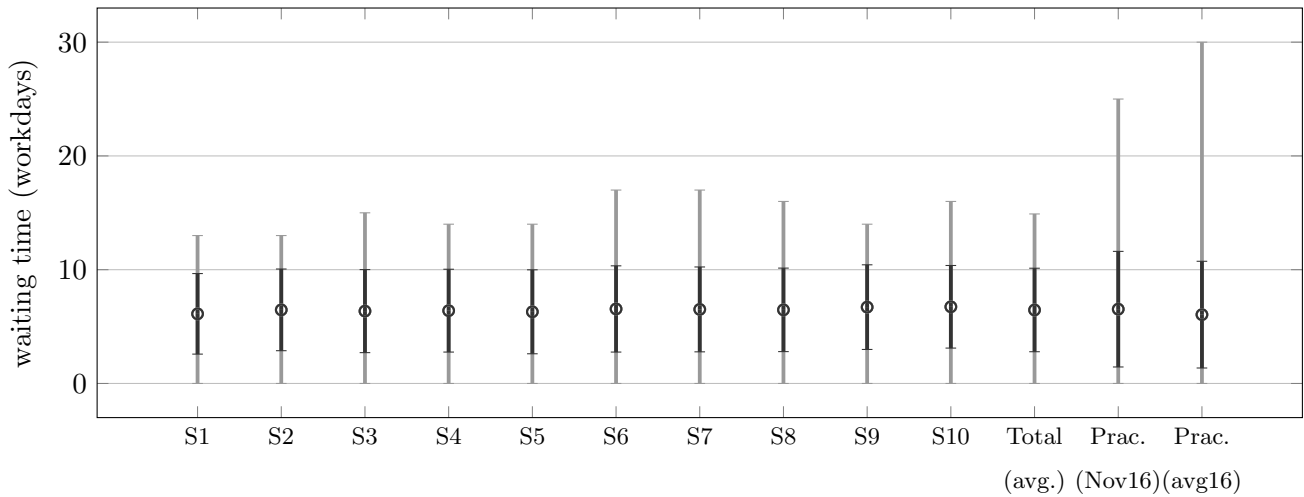


Figure 5: Distribution of average waiting times (in workdays) and corresponding standard deviation and maximum/minimum values of the computational results and the samples measured in practice (November 2016 and the average throughout 2016).

Sample	Patients	Within		Within	
	completed (z_1)	100-target (z_2)		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%

Table 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).

Figures 6, 7, and 8 show the results regarding the fulfillment of the waiting time standards for acute, subacute and regular patients, respectively. As shown in Fig. 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 (Fig. 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 average 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 (Fig. 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%.

The number of RTTs allocated to each operation in the final solution and the total RTT availability

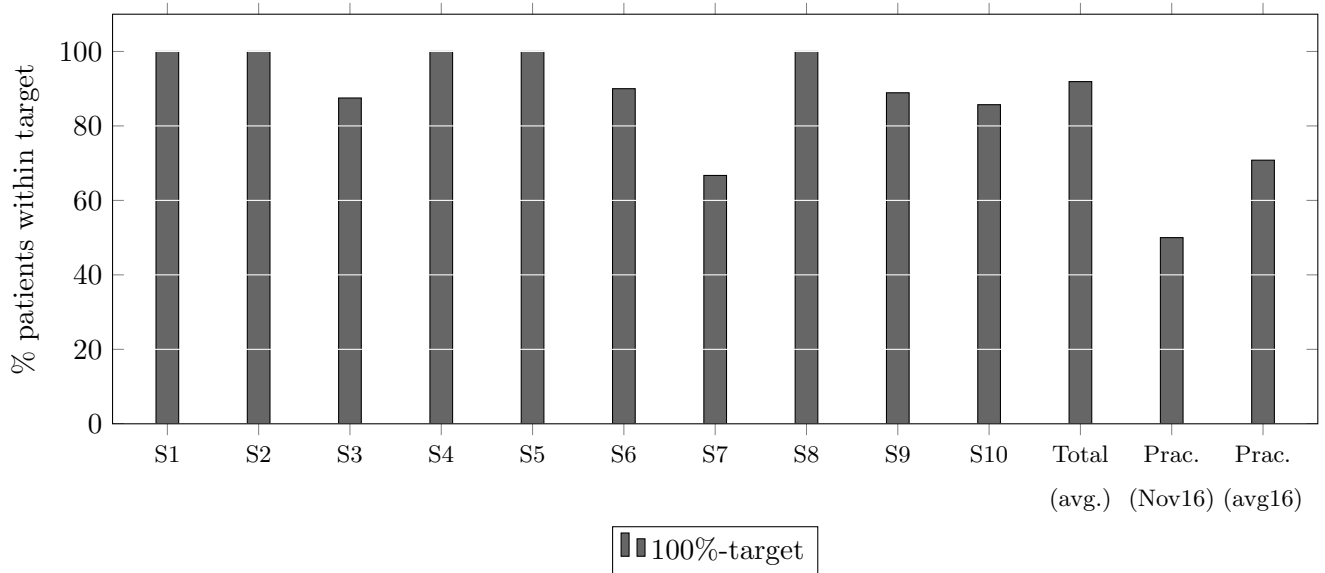


Figure 6: Comparison of the theoretical results per scenario and the corresponding average, with the performance measured in practice (November 2016 and the average throughout 2016) in terms of target satisfaction for acute patients.

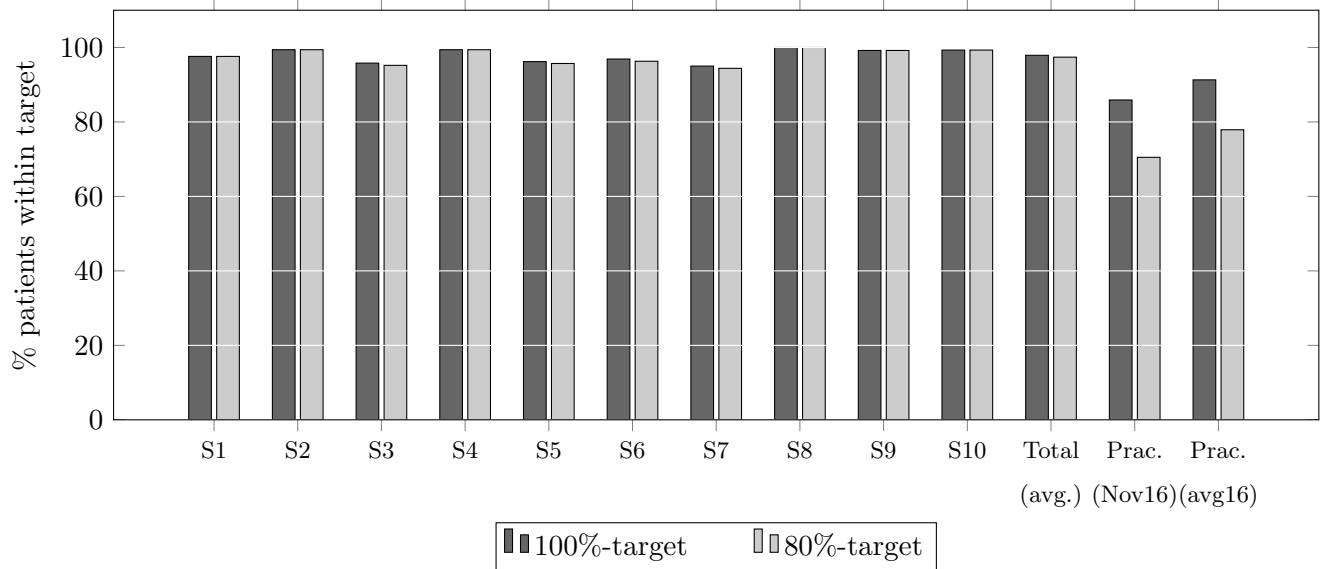


Figure 7: Comparison of the theoretical results per scenario and the corresponding average, with the performance measured in practice (November 2016 and the average throughout 2016) in terms of target satisfaction for subacute patients.

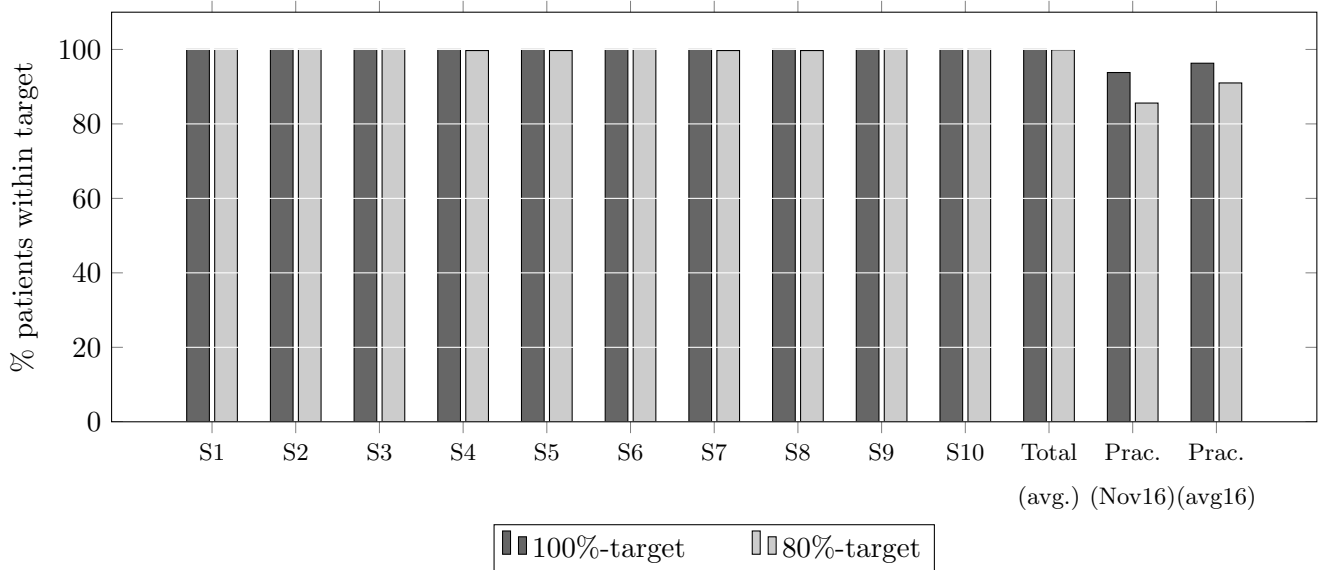


Figure 8: Comparison of the theoretical results per scenario and the corresponding average, with the performance measured in practice (November 2016 and the average throughout 2016) in terms of target satisfaction for regular patients.

in the clinic are presented in Table 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.

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 4.2 and 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 4.4. Results are presented in Table 16. Comparing these with the results for the base case (Table 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.

Sample	Patients	Within		Within	
	completed (z_1)	100-target (z_2)		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%

Table 16: Results for the NKI case study including RTT training needs constraints.

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 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%).

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.

Acknowledgements

The authors would like to thank to Floris Pos, Hein Opdam, Bram van den Heuvel, and Jacqueline van der Geest from the department of radiation oncology of the NKI for their involvement in the development and validation of the model, and to Iris Walraven for her help on running the ANOVA and performing the statistical analysis of the corresponding results.

This work was supported by the Alpe d'Huzes/KWF under the ALORT project (2014-6078).

References

- Bikker, I. A., Kortbeek, N., van Os, R. M., Boucherie, R. J., 2015. Reducing access times for radiation treatment by aligning the doctor's schemes. *Operations Research for Health Care* 7, 111 – 121.
- Castro, E., Petrovic, S., 2012. Combined mathematical programming and heuristics for a radiotherapy pre-treatment scheduling problem. *Journal of Scheduling* 15 (3), 333–346.
- Chen, Z., King, W., Pearcey, R., Kerba, M., Mackillop, W. J., 2008. The relationship between waiting time for radiotherapy and clinical outcomes: A systematic review of the literature. *Radiotherapy and Oncology* 87 (1), 3–16.
- Conforti, D., Guerriero, F., Guido, R., 2010. Non-block scheduling with priority for radiotherapy treatments. *European Journal of Operational Research* 201 (1), 289 – 296.
- Delaney, G., Jacob, S., Featherstone, C., Barton, M., 2005. The role of radiotherapy in cancer treatment. *Cancer* 104 (6), 1129–1137.
- ELEKTA, 2017. MOSAIQ radiation oncology. Available at <https://www.elekta.com/software-solutions/care-management/mosaiq-radiation-oncology.html>, [Accessed 20/01/2018].

- Hulshof, P. J. H., Boucherie, R. J., Hans, E. W., Hurink, J. L., 2013. Tactical resource allocation and elective patient admission planning in care processes. *Health Care Management Science* 16 (2), 152–166.
- Jerić, S. V., Figueira, J. R., 2012. Multi-objective scheduling and a resource allocation problem in hospitals. *Journal of Scheduling* 15 (5), 513–535.
- Joustra, P. E., Kolfin, R., van Dijk, N. M., Koning, C. C. E., Bakker, P. J. M., 2012. Reduce fluctuations in capacity to improve the accessibility of radiotherapy treatment cost-effectively. *Flexible Services and Manufacturing Journal* 24 (4), 448–464.
- Legrain, A., Fortin, M.-A., Lahrichi, N., Rousseau, L.-M., 2015. Online stochastic optimization of radiotherapy patient scheduling. *Health Care Management Science* 18 (2), 110 – 123.
- Legrain, A., Fortin, M.-A., Lahrichi, N., Rousseau, L.-M., Widmer, M., May 2015. Stochastic optimization of the scheduling of a radiotherapy center. In: *Journal of Physics Conference Series*. Vol. 616. pp. 1 – 6.
- Mackillop, W. J., 2007. Killing time: The consequences of delays in radiotherapy. *Radiotherapy and Oncology* 84 (1), 1–4.
- NVRO, 2017. Waiting times, standards and maximum waiting times for radiotherapy (in dutch). Available at <http://www.nvro.nl/kwaliteit/indicatoren/>, [Accessed 20/01/2018].
- Petrovic, D., Morshed, M., Petrovic, S., 2011. Multi-objective genetic algorithms for scheduling of radiotherapy treatments for categorised cancer patients. *Expert Systems with Applications* 38 (6), 6994 – 7002.
- Pignon, T., Fernandez, L., Ayasso, S., Durand, M.-A., Badinand, D., Cowen, D., 2004. Impact of radiation oncology practice on pain: A cross-sectional survey. *International Journal of Radiation Oncology*Biological*Physics* 60 (4), 1204–1210.
- Rais, A., Viana, A., 2011. Operations research in healthcare: a survey. *International Transactions in Operational Research* 18 (1), 1–31.
- Ruiz, R., Vázquez-Rodríguez, J. A., 2010. The hybrid flow shop scheduling problem. *European Journal of Operational Research* 205 (1), 1–18.
- Sauré, A., Patrick, J., Tyldesley, S., Puterman, M. L., 2012. Dynamic multi-appointment patient scheduling for radiation therapy. *European Journal of Operational Research* 223 (2), 573 – 584.

- Stewart, B., Wild, C., 2014. World cancer report 2014. Tech. rep., International Agency for Research on Cancer.
- van Lent, W. A., Deetman, J. W., Teertstra, H. J., Muller, S. H., Hans, E. W., van Harten, W. H., 2012. Reducing the throughput time of the diagnostic track involving {CT} scanning with computer simulation. *European Journal of Radiology* 81 (11), 3131 – 3140.
- Vieira, B., Hans, E. W., van Vliet-Vroegindewij, C., van de Kamer, J., 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.
- Werker, G., Sauré, A., French, J., Shechter, S., 2009. The use of discrete-event simulation modelling to improve radiation therapy planning processes. *Radiotherapy and Oncology* 92 (1), 76–82.