Data-Driven Two-stage Appointment Radiotherapy Scheduling Model for Resource Optimization at a Tertiary Cancer Center

Fan Jia  
University of Toronto

Michael Carter  
University of Toronto

Srinivas Raman (srinivas.raman@rmpuhn.ca)  
Princess Margaret Cancer Centre

Research Article

Keywords: radiotherapy scheduling, operations research, Mixed Integer Programming

Posted Date: March 20th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2693973/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Background: The timely delivery of radiotherapy (RT) is crucial to cancer care, and excessive delays in RT have been associated with detrimental oncological and psychological outcomes. Prior to receiving RT on treatment units (Linear accelerators), there are a few processes that need to take place including simulation (on CT simulators), radiotherapy plan generation/optimization and quality assurance. The assignment of patient schedules on CT simulators and Linear accelerators is currently done manually at most cancer centers. We propose that data-driven optimization of patient scheduling has the potential to improve wait-times, and optimize use of departmental resources.

Methods: A two-stage Mixed Integer Programming model was developed to optimize the patient appointment scheduling process and to forecast machine utilization. The model was tested with historical institutional data from Princess Margaret Cancer Center. By analyzing the dataset and simulating historical patient arrivals, the model output is evaluated by comparing patient wait time statistics and monthly machine utilization against what occurred during this time frame.

Results: Testing our model on data from 2019-06 to 2020-02, we found a reduction in average wait time from 11.2 to 6.7 business days for standard category patients. The number of standard patients exceeding the wait time target of 10 business days were reduced from 118 to 15 patients each month. In addition, our model could accurately estimate future machine utilization for both CT simulators and linear accelerators based on the model output appointments, which could facilitate better planning and utilization of departmental resources.

Conclusion: Our scheduling model has the potential to reduce the standard patient wait time for radiation treatment without compromising the wait time for urgent patients. The model can be also used to forecast department resources and machine utilization based on the output of the scheduling model. Radiotherapy departments could use this model to generate patient appointment schedules as well as to reduce machine idle time or appointment over-booking.

Background

1.1 Introduction

Cancer is one of the most significant health issues worldwide and a leading cause of mortality. Radiotherapy (RT) is a cornerstone of cancer care and more than half of the cancer patients will require RT during their lifetime (1) RT is most commonly delivered as external beam radiotherapy using linear accelerators (LINACs) to deliver high energy x-rays. There is significant evidence to support that timely cancer treatment is critical to optimizing oncological outcomes (2). Recognizing this, the Canadian Association of Radiation Oncology (CARO) has set a target that 90% of patients receive radiation treatment within 14 days (10 business days) of being ready to treat (Canadian Partnership Against Cancer, 2020). Accordingly, at the institutional level, it is critical to allocate adequate resources to ensure
that capacity can match the incoming demand RT on LINACs in order to prevent bottlenecks and excessive delays in initiating treatment.

The Radiation Medicine Program (RMP) at Princess Margaret Cancer Center is Canada's largest department and delivers approximately 11,000 courses of radiation treatment for over 8000 patients annually. A generalized RT workflow at RMP is represented by the flowchart in Fig. 1. After a patient consults with the radiation oncologist and patient is prescribed a course of RT, a computed tomography (CT) simulation scan is scheduled to allow the RT team to design a radiation treatment plan based on the individual patient's anatomy. The target site is outlined and distinguished from surrounding healthy issues, which are delineated as “organs at risk” (OAR) during the contouring process. Next, detailed treatment plans are created with the goal of delivering the prescribed radiotherapy dose to the target and minimizing the dose to the OAR's. Finally, after going through peer review and various quality checks, the RT plan is approved and the patient can be scheduled for their treatments, which may range from 1–40 sessions.

1.2 Objective and approach

The objective of this paper is to create a model that can schedule RT appointments according to the current scheduling policies at RMP and estimate resource utilization ahead of time. A non-block scheduling model is one that can assign a different treatment duration to each patient. By using a non-block scheduling model with estimation of the pre-treatment duration, the appointment scheduling process at RMP is recreated and the machine utilization can be computed with the model output, which is the appointment date and appointment duration for each machine resource, including CT simulators and LINACs.

Our approach to the resource planning problem is to build a two-stage scheduling model for the CT simulation and radiation treatment appointments. We used a hybrid flow shop approach in which each patient can be considered as a job that is going to be processed by two types of machines in the CT simulation stage and RT delivery stage respectively. This scheduling problem is formulated as a Mixed Integer Programming (MIP) model and solved using Gurobi in a Python 3.8 environment. The model will output the appointment dates for CT simulation and RT treatment sessions.

1.3 Literature review

Previous studies on RT resource planning have focused on modeling patient flow using simulation models. For example, Vieira et al. (3) and Babashov et al. (4) used discrete event simulation (DES) to test the effect of scheduling strategies on LINAC appointments and changes of various resources on patients’ wait times. Proctor et al. (5) simulated patient flow in the pre-treatment stage. Lindberg et al. (6) developed a simulation model based on real data to predict LINAC utilization. While simulation models can generate different testing scenarios to evaluate various scheduling strategies, they do not optimize the scheduling process.
In the field of patient appointment scheduling for radiation treatment, CT appointments and RT appointments are most often studied separately. Gocgun et al. (7) developed a Markov Decision Process model to prioritize patients for CT appointments based on arrival patterns and queue length. Patient prioritization was also used in a case study by Vermeulen et al. (8) to adaptively allocate CT resources according to patient arrival patterns. Petrovic and Castro (9) developed a genetic algorithm (GA) for pre-treatment scheduling. GA is a metaheuristic approach inspired by natural selection that searches through an evolving population of possible solutions.

For patient scheduling on treatment machines (LINACs), Legrain et al. formulated an online stochastic optimization model so patients have their appointment booked after consultation (10). An infinite-horizon Markov decision process model is developed by Sauré et al. for assigning treatment capacity to patient groups on a tactical level (11). The scheduling problem was also tackled by advanced scheduling with both block and non-block scheduling techniques. In block scheduling models, each day is divided into a number of time slots of the same duration. In non-block scheduling models, each appointment to be scheduled can have different durations. Conforti et al. developed a block scheduling integer programming model (12), where as Frimodig and Schulte applied non-block integer programming and constraint programming models to the problem (13). Pham et al. proposed a two-phase approach to first assign each patients’ treatment sessions to a machine and a day using Integer Linear Programming, then determining the sequence of patients and their appointment time on each LINAC and each day using MIP and Constraint Programming models (14).

The model proposed in this paper is most similar to the non-block scheduling strategy developed by Conforti, Guerriero, & Guido, with additional consideration of scheduling the CT appointment at the same time as scheduling the LINAC (15). Conforti et al. explained that a non-block scheduling strategy is more representative of the real workload as the treatment durations vary a lot for each patient. Their model minimizes the mean wait time by maximizing the number of new patients scheduled, with the option to reschedule some appointments that are already booked. In comparison to their model, our proposed non-block scheduling model aims to predict resource utilization in the department, with the objective to minimize the number of patients exceeding their target wait time, without rescheduling the existing patients. Conforti et al. designed several realistic scenarios to evaluate their model on 2 machines and their results show that the LINAC machines are fully utilized in the middle of the week, which caused problems if emergency patients arrive in the system. Our model assumes emergency patients are immediately scheduled so forecasted machine hours are reserved each day for emergency arrivals. In addition, our test case is based on real world data with 4 CT simulators and 16 LINAC machines, and the predicted machine utilizations are compared with the real utilization to evaluate the resource planning ability of our model. To the best of our knowledge, this is the first time a mathematical scheduling model is tested on a large-scale treatment center with real-world patient arrival data. The proposed mathematical model is able to predict resource utilization, and at the same time optimize wait times for patients.
Another contribution of our model is to schedule both CT and treatment appointment at the same time by developing a Mixed Integer Programming (MIP) scheduling model. In the current literature, this problem is usually addressed using simulation modelling (16) or heuristic approaches (17, 18). Scheduling appointments in both the CT simulation stage and the treatment stage can be considered as a two-stage hybrid flow shop problem with multiple parallel machines at both stages. The application of a multi-stage flow shop problem in the healthcare industry is mostly studied in surgical units. Dekhici & Belkadi presented a two-stage hybrid flow shop problem without buffer for the operating rooms and post-anesthesia care unit (19). Hachicha and Mansour considered a three-stage elective surgery flow shop scheduling problem of a private healthcare facility, with the aim to minimize wait time in between stages in order to minimize the total length of stay (20). They proposed two Mixed Integer Linear Programming (MILP) formulations, network flow MILP and assignment MILP. The network flow MILP uses the same variable for patient sequencing and resource assignment, while the assignment MILP model uses two separate variables. By using real-world experimental instances, the performance of the two models is tested, and the assignment MILP model outperforms the network flow MILP. Our model is also based on assignment MILP variable definition, using separate variables for appointment dates and resource assignment.

Method

2.1 Scheduling problem for CT simulation and radiation treatment

There are 4 CT simulators and 16 LINACs at RMP. The current practice is to pre-allocate different units of machines to patients of different treatment groups. For example, Head and Neck patients are scheduled on CT simulator 3 or 4, and then LINAC treatment unit 1, 3, 4, or 14. The radiation treatment process consists of two stages, and each patient can be considered as a job to be processed sequentially by two types of machines, CT simulators and LINACs, with sufficient time estimated for the pre-treatment processes in between these stages. This patient appointment scheduling problem is considered a two-stage hybrid flow shop scheduling problem with a minimum delay constraint between stages.

This general radiation treatment process is complicated by patient prioritization, different treatment intent, and personalized treatment plans. Patients at RMP are categorized into four priorities: emergency, urgent, standard, and planned delay by the level of urgency. The most critical group is “emergency” patients, followed by “urgent” and “standard patients”. There is another category of “planned-delay” patients who are purposefully scheduled for a later start date. Appointments of emergency patients are usually scheduled immediately on the next available machine. As for patients with planned delays, there are special considerations when scheduling their appointments, such as waiting for recovery from surgery, and so the wait time for planned delayed patients is by definition longer than the 10-business-days wait time target set by CARO. Due to the lack of data on when planned-delay patients are ready for their appointment, our model does not schedule their appointments. However, machine hours are reserved
for emergency and planned-delay patients based on time series estimation so that when they arrive, there should be machine availability. Hence the target patient group of this scheduling model is the urgent and standard patients.

Treatment intent also influences the wait time for radiation treatments. There are three treatment intents at RMP namely palliative, complex palliative (palliative patients with more complex treatment planning), and curative patients. Palliative treatment aims to relieve symptoms and/or pain to improve quality of life, whereas curative treatment, which often uses higher doses of RT aims to eradicate the tumor or achieve durable local control. The palliative category patients usually have a shorter wait time goal compared to curative patients given that palliative RT plans are less complex and the palliative patients are more symptomatic. Also, each patient has a personalized care plan that consists of a certain number of treatment sessions and a prescribed total RT dose that will be delivered to the target site. The appointment scheduling model must be able to address different treatment requirements for each patient, characterized by patient categories, treatment intents and various treatment protocols.

The total number of radiation treatment sessions prescribed are spread out over a few days or a few weeks, leaving some time in between treatment sessions. The radiation treatment appointments for most patients are usually booked on consecutive business days, one session per day. The length and frequency of treatment is estimated by the radiation oncologist at the time of consultation based on each patient’s treatment requirements. Hence the appointment durations are known in advance and vary from patient to patient. Non-block scheduling techniques with known appointment durations are used to schedule CT simulation and LINAC appointments.

2.2 Model Formulation

2.2.1 Assumptions

The following assumptions are made for the patient appointment scheduling problem for CT simulation and LINAC treatments.

- As the appointment durations are personalized according to the patient’s treatment plan, non-block scheduling is deployed; appointments are booked on the allocated CT simulators and LINACs if the capacity of the respective machine is not exceeded.
- The scheduling model books all urgent and standard patients who arrive each day as a batch at end-of-day. Patients are prioritized based on their urgency level and treatment protocol. Different prioritizations have different wait time targets.
- The planning horizon is 20 business days, which is long enough to book the first treatment appointment for every urgent and standard patient. There is no wait list for patients not scheduled. If the first RT appointment is scheduled, all subsequent appointments are scheduled.
- RT treatments in RMP are booked on consecutive business days.
Due to the availability of a large amount of historical data that we can analyze, the number of patient arrivals each day can be forecasted, hence are assumed to be known (for example, the arrivals of emergency and planned delayed patients and their appointments on each machine can be estimated). In addition, the treatment length and frequency are estimated by radiation oncologists at the consultation stage and are assumed to be known in advance.

For each patient, the first treatment appointment is longer than the consecutive treatment sessions. The rest of the treatment appointments have the same duration.

The CT scanners are identical machines, but the LINACs have different capabilities. Treatments for different cancer sites are performed on a pre-allocated subset of LINACs.

The daily capacity of CT simulators and LINACs is fixed. Machine maintenance or down time is not considered in the model due to the lack of relevant data. The current total capacity; in other words, no machines are going to be added or removed.

The human resources related to the CT simulations and radiation treatments are assumed to be consistent with the machines’ availabilities.

Patient preferences, appointment cancelations, and patient no-shows are not considered in the model. Since we are assigning patient treatments to a day, we assume that a separate routine could be used to satisfy patient time-of-day preferences.

### 2.2.2 Notations

**Variable definitions:**

- $d_{1j}$ = day index of CT appointment booked for patient $j$
- $d_{2j}$ = day index of the first radiation treatment session booked for patient $j$
- $d$ = index for days in the planning horizon (0 to $D$)
- $W_j$ = the wait time (number of days) of patient $j$ for the first treatment session
- $O_j$ = the overage of wait time that is greater than the wait time target for patient $j$
- $U_j$ = the underage of wait time that is smaller than the wait time target for patient $j$
- $z_j = 1$, if patient $j$ waits more than the official wait time target for the first treatment session
- $X_{jdc} = 1$, if patient $j$ is booked for CT simulation on day $d$ on CT scanner $c$
- 0, otherwise
- $Y_{s jdl}^s = 1$, if patient $j$'s $s^{th}$ treatment session is booked on LINAC machine $l$ on day $d$
0, otherwise

\( L_{l_j} = 1 \), if patient \( j \) is assigned to LINAC machine \( l \)

0, otherwise

**Parameters:**

The following parameters can be obtained from the data or as a result of historical data analysis and parameter estimations.

\( C_j = \) consultation date of patient \( j \)

\( p_j = \) the minimum number of days required for processing the pre-treatment steps, i.e., contouring, planning, and reviewing before proceeding to the treatment stage

\( S_j = \) the number of treatment sessions prescribed for patient \( j \)

\( n_j = \) the number of days needed in between two radiation treatment sessions

\( pr_j = \) priority weight assigned to patient \( j \) based on category of patient \( j \). (Urgent patients are given a higher priority.)

\( Target_j = \) the wait time target of patient \( j \)

Machine capacities:

\( m_{ld} = \) LINAC \( l \)'s daily capacity (available working time)

\( m_{cd} = \) CT scanner \( c \)'s daily capacity (available working time)

Appointment durations:

\( D_{cj} = \) Duration of CT appointment for patient \( j \)

\( D_{tjs} = \) Duration of \( s^{th} \) treatment session for patient \( j \)

**Sets:**

\( P = \) the set of patients to be booked

\( M_c = \) the set of CT simulators in the facility

\( M_l = \) the set of LINACs in the facility

\( M_{lj} = \) the set of LINACs capable to treat patient \( j \)
2.2.3 Mixed integer model formulation

The problem is formulated as an MIP with the objective function to minimize the weighted number of patients exceeding the wait time targets based on priority, while ensuring the requirements of scheduling guidelines and treatment protocols are satisfied.

Objective: \( \min \sum_{j=1}^{J} z_j \times pr_j \)

s.t.

\[ W_j = d_{2j} - C_j, \forall j \]
\[ Target_j = W_j - O_j + U_j, \forall j \]

(1) This constraint set defines the wait time, overage, and underage of wait time.

\[ O_j \leq z_j \times M, \forall j \]

(2) Constraints for the number of patients exceeding the wait time target:

if \( O_j > 0 \), \( z_j = 1 \); if \( O_j = 0 \), \( z_j = 0 \).

\( M \) is a very large number to ensure the correct definition of \( z_j \).

\[ d_{2j} - d_{1j} \geq p_j, \forall j \]

(3) The time between CT simulation and first treatment appointment should be sufficient for processing the intermediate pre-treatment steps.

\[ \sum_{l=1}^{M_l} \sum_{d=0}^{D} Y_{jdl}^1 \times d = d_{2j}, \forall j \]

(4) Constraint for booking the patient j's first treatment session on day \( d_{2j} \).

\[ \sum_{l=1}^{M_l} \sum_{d=d_{1j}+p_j}^{D} Y_{jdl}^1 = 1, \forall j \]

(5) Only one first treatment appointment should be booked.

\[ \sum_{c=1}^{M_c} \sum_{d=0}^{D} X_{jdc} = 1, \forall j \]

(6) Only one CT appointment should be booked.
(7) Every patient is assigned to one treatment machine.

\[
\sum_{l=1}^{M_l} L_{lj} = 1, \forall j
\]

(8) All treatment sessions are booked on the LINAC unit patient \( j \) is assigned to.

\[
\sum_{d=0}^{D} Y_{jdl}^s = L_{lj}, \forall j, l, s
\]

(9) The patient \( j \) should only be treated on LINAC units that are capable for this site group.

\[
Y_{jdl}^s = 0, \text{ for } l \notin M_{ij}, \forall j, d, s
\]

(9) The patient \( j \) should only be treated on LINAC units that are capable for this site group.

\[
Y_{jdl}^s = Y_{jdl'}^{s'}, \forall j, l
\]

\[
s' = s + 1, \forall s = 1, \ldots, S_j - 1
\]

\[
d' = d + 1, \forall d
\]

(10) All treatment appointments are booked 1 business days apart.

\[
\sum_{j=1}^{J} D_{ej} X_{jdc} \leq m_{cd}, \forall d, c
\]

\[
\sum_{j=1}^{J} \sum_{s=1}^{S_j} D_{ij} s Y_{jdl}^s \leq m_{ld}, \forall d, l
\]

(11) All the appointments booked on a CT or LINAC machine on a certain day should be within the machine capacity on that day.

\[
Y_{jdl}^s \in \{0, 1\}, \forall j, d, l
\]

\[
X_{jdl} \in \{0, 1\}, \forall j, d, l
\]

\[
z_j \in \{0, 1\}, \forall j
\]

\[
L_{lj} \in \{0, 1\}, \forall j, l
\]

\[
d_{1j} \geq 0, \forall j
\]

\[
d_{2j} \geq 0, \forall j
\]

\[
o_j \geq 0, \forall j
\]
The variable domains.

\[ U_j \geq 0, \forall j \]

(12) The variable domains.

The objective is to minimize the sum of the number of patients with excessive wait time (defined as over their wait time targets. The primary output of the model is going to be a daily schedule of the appointment dates for CT simulations and LINAC treatment sessions. The MIP model is run daily to estimate the machines utilization in the following weeks, based on the scheduling results on patient’s appointments. Patients who were scheduled on previous days reduce machine capacity in the future. The model schedules all new patients that arrived today.

2.2.4 Parameter Estimations

Data was obtained from the RMP’s appointment booking and treatment record systems, including patient diagnosis, pre-treatment processes and appointment data. The data is used to estimate model parameters as well as to create test cases for the machine utilization estimation for urgent and standard patients.

2.2.4.1 Pre-treatment Duration Estimation

The pre-treatment process takes place after CT simulation and before the first treatment appointment. Based on the historical data from RMP, the majority of the pre-treatment time takes place during the contouring and planning stages. The contouring and planning time of the treatment plans are related to cancer sites, categories, and treatment intent, which is related to the total RT dose prescribed, and the number of treatment fractions. To test for feature importance, an ANOVA test is used for categorical features (site groups, categories, and treatment intents) and Pearson correlation is used for numeric features (total dosage and number of fractions). Based on feature importance testing, all the above-mentioned features are correlated to the contouring and planning durations.

Contouring and planning in the cancer center are manual processes and the exact time needed to complete contouring and planning may be affected by other factors not captured in the data. For example, it is possible that the planning process was complete, but the timestamp of completion is recorded a few days later, causing a prolonged planning time duration being recorded. Due to these reasons and other associated human resources factors, the contouring and planning time did not fit logically to any regression models. So, we decided to use the 80th percentile in each combination of the main feature groups (site groups, patient categories, and treatment intents) to represent the estimated contouring and planning time. In 80% of the time, the contouring and planning processes should be finished in the estimated period using 80th percentile. As input to the scheduling model, the estimations will ensure that reasonable time is left between CT simulation and first LINAC treatment session.

2.2.4.2 Wait Time Targets Estimation
At RMP, there are no specific wait time targets for patients of different categories and treatment intents, except that the standard patients have the CARO wait time target of 10 business days. Instead, RMP have scheduling guidelines specifying which day to book patient appointments for CT and radiation treatments, according to their site group, category, and treatment intent. As a result, the wait time targets, $Target_j$, in the model, are not specified for urgent and palliative patients who are higher priority. In order to formulate a quantitative objective function for the model, we derived the average wait time for the high priority patient groups with recent data (2018–2019) and used the averages as the parameters for the wait time targets. Patients are classified based on their category and treatment intent, as these factors affect the patient prioritization at RMP.

The respective wait time target parameters are summarized in Table 1. For urgent palliative, other urgent, as well as standard palliative patients, their internal wait time targets are estimated by the average wait time for the patient group, because these patients are high priority at RMP. Hence their average wait time rounded to the nearest integer can be estimated as their wait time targets. For other standard patients, the wait time target is the official wait time target of 10 business days, as suggested by CARO.

<table>
<thead>
<tr>
<th>Patient categories and intents</th>
<th>Wait time targets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urgent palliative</td>
<td>2 days</td>
</tr>
<tr>
<td>Other urgent (complex palliative and curative)</td>
<td>3 days</td>
</tr>
<tr>
<td>Standard palliative</td>
<td>4 days</td>
</tr>
<tr>
<td>Other standard (complex palliative and curative)</td>
<td>10 days</td>
</tr>
</tbody>
</table>

2.2.4.3 Machine Capacity Calculation

The total machine capacity is provided by RMP. However, as Emergency and Planned Delay patients are not included in the current model, their appointment time on the machines is considered to be “reserved” and excluded from the total machine capacity. Time series forecasting is used to estimate the total appointment time to be booked on every machine for these two types of patients, so that sufficient time on the machines is reserved every day. Different time series forecasting methods, including Seasonal Naïve, Moving Average (MA), Simple Exponential Smoothing (SES) and Autoregressive Integrated Moving Average (ARIMA) were used to estimate the appointment time for Emergency and Planned Delay patients. SES forecasting model had the best performance for model selection with cross validation. Hence SES is used to estimate the total time reserved on each machine for emergency and planned delayed patients every month. Figure 2 illustrates two examples of forecasting number of minutes booked on LINAC unit 'EA07' for emergency patients and planned delay patients respectively.
In the graphs, the dark blue lines are the forecasted values in each month. The dark and light colored regions are the 80% and 90% confidence intervals for the estimations respectively.

### 2.2.4.4 Weighting Factors for Patient Groups

In the current scheduling practice RMP, urgent patients are prioritized for both CT and LINAC treatment appointments. Based on the analysis of historical data, their average wait time for the first radiation treatment is usually 2 to 3 days after consultation (with average wait time for CT of 0–1 day). Our model aims to reproduce this prioritization by penalising extra wait time for urgent patients. This is achieved by assigning a larger weighting factor to urgent patients if they wait longer than their wait time targets. This set of weighting factors for urgent and standard patients are calibrated with the wait time statistics of the model output for urgent patients. The priority weights assigned for urgent and standard patients are 1000 and 1 respectively.

### 2.2.5 Simulation

Historical data is used as test cases for the arrival of urgent and standard patient categories. The testing period is from 2019-06 to 2020-02 with a three-month warm-up period prior to the test period. Monthly utilizations for CT and LINAC are computed and compared with the actual machine utilization as recorded in the dataset.

The total capacity of the two types of machines is the product of the number of machines in RMP and the average monthly CT and LINACs capacity respectively according to the CT and LINAC machine schedule. The CT and LINAC capacity is adjusted by subtracting the time reservation for emergency and planned-delay patients from the machines’ total capacity. We assume there are 20 business days in each month and the average daily capacity of each machine is assumed to be the same throughout the testing period as the recent LINAC average daily capacity in 2019 to 2020. The formulas used to compute machine utilization rate are shown below:

**Average CT Simulator utilization rate:**

\[
Average\text{utilization}\,(\%) = \frac{\text{TotalappointmenttimebookedonCTs}}{\text{TotalCTs’capacity}} \times 100
\]

**Average LINAC utilization rate:**

\[
Average\text{utilization}\,(\%) = \frac{\text{TotalappointmenttimebookedonLINACs}}{\text{TotalLINACs’capacity}} \times 100
\]

### Results

By scheduling CT and LINAC appointments for patients, the machine utilization can be computed for better resource planning and resource allocation. Table 2 and Fig. 3 show the CT utilization in each
month of the testing period, comparing the machine utilization from the model output with the historical data. Likewise, Table 3 and Fig. 4 summarize the results for LINAC utilization.

### Table 2
CT utilization forecasting results

<table>
<thead>
<tr>
<th>Month</th>
<th>Real data</th>
<th>Model estimate</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>2019-06</td>
<td>0.590</td>
<td>0.570</td>
<td>-0.020</td>
</tr>
<tr>
<td>2019-07</td>
<td>0.599</td>
<td>0.593</td>
<td>-0.006</td>
</tr>
<tr>
<td>2019-08</td>
<td>0.515</td>
<td>0.511</td>
<td>-0.004</td>
</tr>
<tr>
<td>2019-09</td>
<td>0.581</td>
<td>0.596</td>
<td>0.015</td>
</tr>
<tr>
<td>2019-10</td>
<td>0.620</td>
<td>0.628</td>
<td>0.008</td>
</tr>
<tr>
<td>2019-11</td>
<td>0.619</td>
<td>0.594</td>
<td>-0.025</td>
</tr>
<tr>
<td>2019-12</td>
<td>0.544</td>
<td>0.564</td>
<td>0.020</td>
</tr>
<tr>
<td>2020-01</td>
<td>0.676</td>
<td>0.661</td>
<td>-0.016</td>
</tr>
<tr>
<td>2020-02</td>
<td>0.567</td>
<td>0.586</td>
<td>0.020</td>
</tr>
</tbody>
</table>

### Table 3
LINAC utilization forecasting results

<table>
<thead>
<tr>
<th>Month</th>
<th>Real data</th>
<th>Model estimate</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>2019-06</td>
<td>0.673</td>
<td>0.621</td>
<td>-0.052</td>
</tr>
<tr>
<td>2019-07</td>
<td>0.774</td>
<td>0.793</td>
<td>0.019</td>
</tr>
<tr>
<td>2019-08</td>
<td>0.761</td>
<td>0.682</td>
<td>-0.080</td>
</tr>
<tr>
<td>2019-09</td>
<td>0.620</td>
<td>0.673</td>
<td>0.053</td>
</tr>
<tr>
<td>2019-10</td>
<td>0.804</td>
<td>0.783</td>
<td>-0.022</td>
</tr>
<tr>
<td>2019-11</td>
<td>0.748</td>
<td>0.746</td>
<td>-0.002</td>
</tr>
<tr>
<td>2019-12</td>
<td>0.688</td>
<td>0.749</td>
<td>0.061</td>
</tr>
<tr>
<td>2020-01</td>
<td>0.763</td>
<td>0.734</td>
<td>-0.029</td>
</tr>
<tr>
<td>2020-02</td>
<td>0.653</td>
<td>0.661</td>
<td>0.008</td>
</tr>
</tbody>
</table>

As shown in the above tables and figures, our model is able to give reasonable machine utilization estimations during the testing period. In addition to the machine utilization, our model objective is to minimize the number of patients with excessive wait time for their radiation treatments. The monthly wait
time statistics for urgent and standard category patients are shown in Tables 4 and 5. Compared to the real-world situation represented in the data, our model is able to reduce the wait time and decrease the number of patients with prolonged wait time, especially for the standard patient category. As urgent patients are prioritized in the current scheduling practice, the model significantly reduced the wait time for standard patients from average wait time of 11.2 to 6.7 business days in the testing period, and the number of standard patients exceeding the wait time target of 10 business days are reduced from 118.1 to 14.8 patients each month.

<table>
<thead>
<tr>
<th>Month</th>
<th>Real data</th>
<th>Model output</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>S.D.</td>
</tr>
<tr>
<td>2019-06</td>
<td>2.956</td>
<td>3.164</td>
</tr>
<tr>
<td>2019-07</td>
<td>2.488</td>
<td>2.400</td>
</tr>
<tr>
<td>2019-08</td>
<td>2.096</td>
<td>2.376</td>
</tr>
<tr>
<td>2019-09</td>
<td>2.681</td>
<td>2.879</td>
</tr>
<tr>
<td>2019-10</td>
<td>2.750</td>
<td>3.300</td>
</tr>
<tr>
<td>2019-11</td>
<td>2.516</td>
<td>2.615</td>
</tr>
<tr>
<td>2019-12</td>
<td>1.882</td>
<td>1.986</td>
</tr>
<tr>
<td>2020-01</td>
<td>2.547</td>
<td>3.219</td>
</tr>
<tr>
<td>2020-02</td>
<td>2.278</td>
<td>2.597</td>
</tr>
</tbody>
</table>
### Table 5
Wait times for standard category patients

<table>
<thead>
<tr>
<th>Month</th>
<th>Real data</th>
<th></th>
<th></th>
<th></th>
<th>Model output</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>S.D.</td>
<td>Median</td>
<td># exceeds</td>
<td>Mean</td>
<td>S.D.</td>
<td>Median</td>
<td># exceeds</td>
</tr>
<tr>
<td>2019-07</td>
<td>11.451</td>
<td>7.233</td>
<td>11</td>
<td>167</td>
<td>6.201</td>
<td>2.530</td>
<td>6</td>
<td>14</td>
</tr>
<tr>
<td>2019-08</td>
<td>10.599</td>
<td>6.905</td>
<td>9</td>
<td>97</td>
<td>6.143</td>
<td>2.595</td>
<td>6</td>
<td>12</td>
</tr>
<tr>
<td>2019-09</td>
<td>11.923</td>
<td>7.949</td>
<td>10</td>
<td>86</td>
<td>6.328</td>
<td>2.596</td>
<td>6</td>
<td>16</td>
</tr>
<tr>
<td>2019-10</td>
<td>10.978</td>
<td>7.463</td>
<td>10</td>
<td>133</td>
<td>7.132</td>
<td>2.444</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>2019-12</td>
<td>12.089</td>
<td>8.253</td>
<td>10</td>
<td>100</td>
<td>7.435</td>
<td>3.245</td>
<td>7</td>
<td>18</td>
</tr>
<tr>
<td>2020-01</td>
<td>10.516</td>
<td>7.877</td>
<td>10</td>
<td>139</td>
<td>7.134</td>
<td>2.800</td>
<td>7</td>
<td>17</td>
</tr>
<tr>
<td>2020-02</td>
<td>10.896</td>
<td>7.854</td>
<td>10</td>
<td>108</td>
<td>7.060</td>
<td>3.280</td>
<td>7</td>
<td>18</td>
</tr>
</tbody>
</table>

### Discussion

This two-stage appointment scheduling model aims to reduce the number of patients exceeding their wait time targets. The effect is most significant for “standard” patients who are scheduled with reduced priority compared to “urgent” patients. With this scheduling model, the standard patients experience shorter wait times for their first RT treatment appointment without compromising the wait time for urgent patients. In addition, based on the analysis on the historical data, the model can also be used to estimate the departmental machine utilization so that resources can be better planned and utilized. However, the daily optimization models disadvantaged some urgent patients who could get radiation treatment on the same day or the next day of their consultation, because the model is only run at the end of every day, so neither CT simulation nor treatment appointments can be booked on the same day. The detriment to urgent patients with the model is on average < 0.25 days and does not exceed 1 day for any given patient, so this impact is likely minimal. Moreover, the current model assumes all “standard” patients have the same target wait time, but in reality, some “standard” patients would benefit more from starting treatment earlier than others. Since this information is not readily available, our model does not sub-prioritize patients who are in the same category.

Another limitation of this study is related to data quality. Due to the fact that completion of pre-treatment processes and their timestamps may not have occurred synchronously, some estimations had to be made for the contouring and planning durations resulting in less precise parameter estimations of the time between CT and first RT appointments. Currently, the model uses 80th percentile in each patient group based on patient category, site group and treatment intent as the estimated parameter for the time between CT simulation and treatment appointment. In reality, there would be patients whose pre-
treatment process is longer than the current 80th percentile values, which could result in longer wait time and less accurate resource planning. While this impact could be mitigated with patients whose pre-treatment time is shorter than the current estimations, the model performance should be tested again with varying pre-treatment durations in each patient group.

The test case presented in this paper is based on one scenario only, which is a replication of what happened during this timeframe and was recorded in the database. To further test and improve the scheduling model, different test scenarios could be created based on real-world data. For example, we could increase the number of patient arrivals to test the model's performance with increased demand. We could also generate test cases with varying machine availabilities. Currently, the model only considers the current availability of the CT simulators and LINAC units. With varying resource availability, the model could be tested against uncertain circumstances such as unexpected machine break-down or staffing shortages. In addition, if the machine availability can be forecasted, the scheduling model would be able to generate more accurate treatment schedules and the department could better plan their resources.

In the next stage of planned work, more complete and detailed timestamp data for pre-treatment time will be collected, including contouring, planning, reviewing and possible replanning time, so that the estimation of number of days in between CT simulation and first treatment appointment can be more accurate. Additionally other data-driven models, such as machine learning models will be tested to estimate the individual pre-treatment duration for patients with different characteristics.

Currently, the durations of the appointments are estimated by the oncologists based on standard treatment protocols, hence are assumed to be known to the model. However, we do not know how accurate the estimators are. If the estimated appointment duration is longer than the actual time a patient spends on the machine, there will be extra idle time on the machines and a lower utilization rate. On the other hand, if the estimated duration is too long, it will affect the appointments of other patients. Hence, a scheduling model with stochastic appointment durations could be studied in the future to optimize this patient scheduling problem.

This is the first study that patient appointment scheduling for both radiation pre-treatment and treatment is formulated into an entire mathematical model, and it is generalizable to fit the scheduling strategies in different cancer centers. Most existing studies consider the two stages of radiation treatment separately, which could ignore the impact of CT simulation, contouring and planning time on the treatment total wait time. In addition, the output of our model can be helpful in departmental resource planning and reduce the number of patients with excessive wait time. However, the model is relatively computationally expensive, especially for large centers with large number of machines.

**Conclusion**

This paper formulated a two-stage appointment scheduling problem for CT simulation and LINAC appointments using historical data at a tertiary cancer center. Patients are classified by site groups, categories, and treatment intents to estimate the 80th percentile pre-treatment time for each patient
group. The objective of the model is to reduce the number of patients with prolonged wait time based on the current scheduling practice for urgent patients and the wait time targets for standard patients. The other objective of this model is to better allocate departmental resources by forecasting machine utilization based on model output. This model was tested using real-world patient data and simulated over a 9-month testing period to ensure consistent performance. As the simulation results demonstrate, the model can accurately estimate monthly CT and LINAC utilization, and at the same time potentially reducing wait time for most patients in the standard category. To the best of our knowledge, this is the first time that the CT simulation and LINAC appointments are modeled by a two-stage flow shop scheduling problem and formulated by a mathematical program. For future work, more accurate data for each of the pre-treatment processes will be collected to improve the estimation on pre-treatment time of patients with different characteristics, as well as predicting the actual appointment durations to reduce machine idle time or overlapping appointments.

**Declarations**

**Ethics approval and consent to participate**: The study did not involve individual human subjects and was completed with institutional approval through the University Health Network Quality Improvement Review Committee.

**Consent for publication**: Not applicable. Only aggregate scheduling data is presented in this manuscript.

**Availability of data and materials**: The datasets generated and/or analysed during the current study are not publicly available due the terms of the Quality Improvement Review Committee approval.

**Competing interests**: None

**Funding**: This work was supported by the Robert L. Tundermann and Christine E. Couturier philanthropic funds.

**Authors’ contributions**: F.J wrote the main manuscript text and performed the analyses.

M.C and S.R co-supervised the project. All authors reviewed the manuscript.

**Acknowledgements**: We would like to thank Lucy Lu for her efforts in data collection, and Anita Vloet for reviewing the manuscript and providing feedback on the project.

**Conflict of interest statement**: There are no conflict of interests related to this work.

**References**


**Figures**

**Figure 1**

Generalized RT workflow at RMP

![Diagram](image1)

**Figure 2**

Example charts for forecasting utilization on LINAC unit ‘EA07’

a). The exponential smoothing forecast for emergency patients on LINAC unit ‘EA07’

b). The exponential smoothing forecast for planned delayed patients on LINAC unit ‘EA07’
Figure 3

CT utilization forecasting plot
Figure 4

LINAC utilization forecasting plot