

COMP 396 Final Report - Fall 2014

Analyzing Radiation Oncology Data for Prediction of Radiotherapy Patient Wait Time

Alvin Leung, McGill University
Supervisor: Prof. Laurie Hendren, McGill University

Abstract

A common cause of anxiety for radiotherapy patients is the fact that they do not know how long they will have to wait for treatment to start following preliminary scans. In this study, we attempt to provide an initial estimate of radiotherapy patient wait time based on electronic health records. For this estimate, we extracted data for six key events from the initial CT scan to the start of treatment. The effect of patients' primary oncologist, diagnosis, age, priority, treatment season, and gender were considered in turn as potential explanatory variables and data trends for various combinations of patient factors were visualized. The oncologist, diagnosis, and priority of the patient were identified as the key factors affecting wait time. Using these factors as features, various regression algorithms were trained on part of the patient dataset and tested on the rest of the dataset. A support vector machine approach gave wait time estimates that were accurate to 3.97 ± 3.36 days, a substantial improvement over the default "two-week" estimate typically given to patients. Improving this estimate by adding informative features and applying the information to radiotherapy workflow scheduling will be considered in future studies.

Introduction

The effect of waiting time on the overall patient experience and patient satisfaction in radiotherapy treatment is well documented [1]. However, with Electronic Health Records (EHR) for patient data, the waiting time can be extrapolated for a given patient by giving an estimate based on past patients that are similar based on a variety of features. The goal of this project is to attempt to explore these features based on current available radiotherapy data and provide reliable estimates for personalized patient wait times during the radiotherapy treatment planning process in order to relieve patient anxiety. We believe that improving patient satisfaction in this way is as important as trying to reduce the wait times themselves. In addition to radiation oncology, the algorithm that we develop can also be extended to other hospital departments in which an estimation of patient waiting time is needed and EHRs are used to record patient data.

In this paper we present an analysis of radiotherapy patient data and planning wait times recorded in the Montreal General electronic database at the McGill University Health Centre (MUHC). We also present a predictive model for the wait times by testing on this database, which performs reasonably as an initial estimate. Already there is merit in this approach because the accuracy of this estimate is much better than the "approximately two weeks" estimate usually given, which corresponds approximately to the average waiting time across all patients. The paper is organized as follows. First, prior work on factors affecting patient wait times and the consequences of prolonged delays is discussed in the related work section. Then, brief backgrounds are given for a typical patient radiotherapy treatment course and the electronic health record database from when we extracted data. Our analysis is described in the data

extraction and trends sections, followed by an outline of our prediction method, and finally a discussion of the benefits and problems with our method is offered. In this paper, "wait time" and "treatment planning time" are used interchangeably to refer to how long it takes for the patient to start treatment from the time of initial scan.

Related Work

Over the past few decades the concern of waiting times for radiotherapy has been extensively analyzed in the literature. As early as 1989, before the commercial advent of electronic health records, a retrospective study [2] found that an unacceptable long wait time elapsed between radiotherapy consult and start of treatment. A large body of work focuses on studying the effect of prolonging overall treatment time on local tumour control and the possibility of metastasis. Mackillop et al. suggests that delays in radiotherapy can affect the outcome by permitting both local proliferation of clonogenic cells and spread of the cancer beyond the treatment volume [3]. In work by Waaijer et al., a tumour control probability model was applied and the expected control probability decreased from 66% to 47% factoring in 56 days of delay [4]. A metaanalysis of radiotherapy studies [5] concludes that risk of local recurrence increases markedly with increasing wait times. These studies and more provide much incentive to try to decrease waiting time for radiotherapy patients.

Studies also address the variation in wait times between different cancer treatment networks and different diagnosis types. In a large study in England [6], 12 cancer types were investigated and they attempted to classify patients based on the length of their waiting time, finding that there were significant differences in wait time and proportion of people treated within 60 days not only among different diagnoses but also across various cancer hospital networks in the UK. They also found an alarming trend that the median wait time for radiotherapy increased from 45 days to 76 days from 1992 to 2001, a trend that is mirrored elsewhere in the world. Arndt studied provider delay among breast cancer patients in Germany, and concluded that some important factors for wait time are tumor stage and time constraints for patients and physicians [7]. Afzelius et al. evaluated the effect of age, tumour size, anaplasia grade, and number of positive lymph nodes on treatment delay [8]. Other studies try to model delays and improve wait times directly - for example, the daily operation of the ambulatory care unit was simulated and the outcomes were a series of recommendation for more efficient use of rooms and a proper redistribution of workload [9]. Our project is situated in an area of active research but it provides a more systematic and computational approach to address the problem of radiotherapy wait times than many previous studies.

Radiation Oncology Patient Course

The patient's journey in radiation oncology at Montreal General Hospital begins with consultation requests sent to the booking center, where faxes come in from referring physicians. The incoming requests usually have two sections - one containing information on the referring institution and cancer type, and the other with detailed patient information including tissue biopsies, diagnosis, and date of diagnosis. The radiation oncology department operates on a "task" system - these tasks usually consist of a time issued, time expected, and time completed timestamps associated with them, along with a description of the work required, the name of the sender, and the name of the receiver, and are sent among staff, technologists, and doctors as a record of events. These are different from but similar to "appointments", which also include the same metadata time stamps, and usually include information on the attending physician or technologist, any rooms to be booked, and patient information. The booking staff tasks the doctors and provides a chart and chooses from 8 consultation types, based on whether treatment is to be curative or palliative,

the patient is hospitalized or not, and if they are a new or returning patient. They choose a time for an initial consult according to doctor availabilities, and the patient then arrives at a set time.

Patients have to check in at the registration desk initially (but as of October 2014 they can check-in via new self-check in machines for treatments). The room, time they are seen (actually called in), and attending physician is recorded in the system. A list of patients in the waiting room is displayed in multiple rooms. After the patient is seen, the time the appointment is completed is recorded in the system (this is not exact, but close - someone manually enters this when they see the patient/doctor leave the consult room). After the doctor sees the patient for the initial consult, he reviews the chart, goes over patient history, and ideally on the same day fills out the rad onc requisition (ROR). The ROR outlines a basic recommended treatment plan, including an idea of dosage, location of treatment, and how long the patient can wait to be treated. Either the patient does not need to be treated and is removed from the system, or the patient has tests missing in which case that test is ordered and the patient needs to followup, or the patient is undecided, or if treatment is required, a CT scan is tasked to the booking staff. The patient is matched to both a machine for CT scan and a radiation machine for the duration of their treatment. The liaison office communicates with other departments and finds out, for example, when chemotherapy finishes and treatment is ready (or sometimes chemotherapy and radiation can be given in parallel). The patient receives the CT Scan as soon as possible, and many steps - all mediated by logged tasks in the system - take place before the patient is contacted for the beginning of treatment.

If the radiotherapy is being done in the head or neck region, a thermoplastic mask is created, the dosimetry contourers create labels on the CT image for the doctors, then the MD contours the tumour region, CTV, planning target volume, etc. After the doctor is done, noting organ limits and other restrictions in the CT planning sheet, the plan is ready for dose calculation to ensure the tumour volume gets the full dose of radiation and surrounding tissues get minimal radiation. After dose calculation, the doctor approves the plan and writes a prescription note, and the plan is ready for the physicists to approve the practical aspects. At any step the plan may have to be changed and must be sent back in the pipeline. Finally, when the treatment plan is ready, the patient is contacted for treatments. The time between when the patient receives a CT scan and start of treatment can take many weeks - this is often the most uncertain and anxiety-inducing part of a patient's radiotherapy experience.

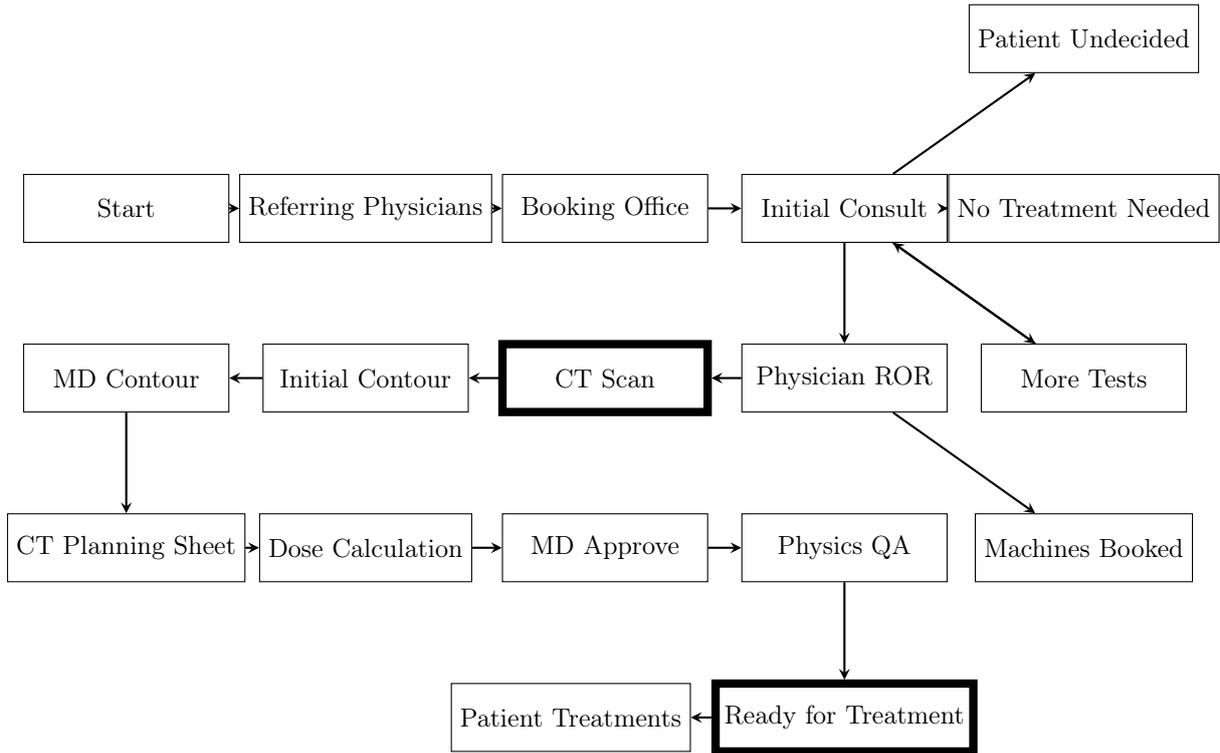


Figure 1: Overview of the Radiation Oncology Treatment Planning Process. We focus on events between the bolded boxes.

Data Background

The data for this project comes from a subset of ARIA, a relational database system developed by Varian Medical Systems (Palo Alto, California), that the Radiation Oncology department at the MUHC uses as an electronic health record system. Used in the clinical context, it ensures that patient charts have complete documentation, comply with standards, and that the hospital can be held accountable for any problems. Although ARIA is an excellent system for recording and managing patient health records, it does not come with any tools to analyze the collected data. The sheer amount of patient data that is available lends itself to the use of data mining algorithms to extract patterns and in turn improve quality and efficiency of treatment. The dataset we used for our analyses was extracted from the real ARIA database originally used [10] by John Kildea, a medical physicist at the MUHC, and constitutes a very small subset of the full database. The complexity of the MySQL schema for ARIA is displayed in figure 2 - it includes a large number of fields corresponding to records of the operation of medical physics machines and detailed patient histories that were not necessary or relevant for the current study. Due to confidentiality, all patient names were removed and the patient labels were anonymised. The ARIA subset used is shown in figure 3.

In the subset, most tables are connected to each other through the patient serial number key. The database includes individual data for each patient such as gender, age, and postal code, information about the doctors assigned to the patient, the priority and diagnosis for each treatment course the patient went through in the Radiation Oncology department, as well as events which are separated into tasks and appointments, as described above, and documents and plans which represent prescriptions and actual treatment days respectively. It is important to note that tasks and appointments were not referenced by their actual names (this was not extracted from ARIA) but instead by a system of aliases - each alias is linked to a series of

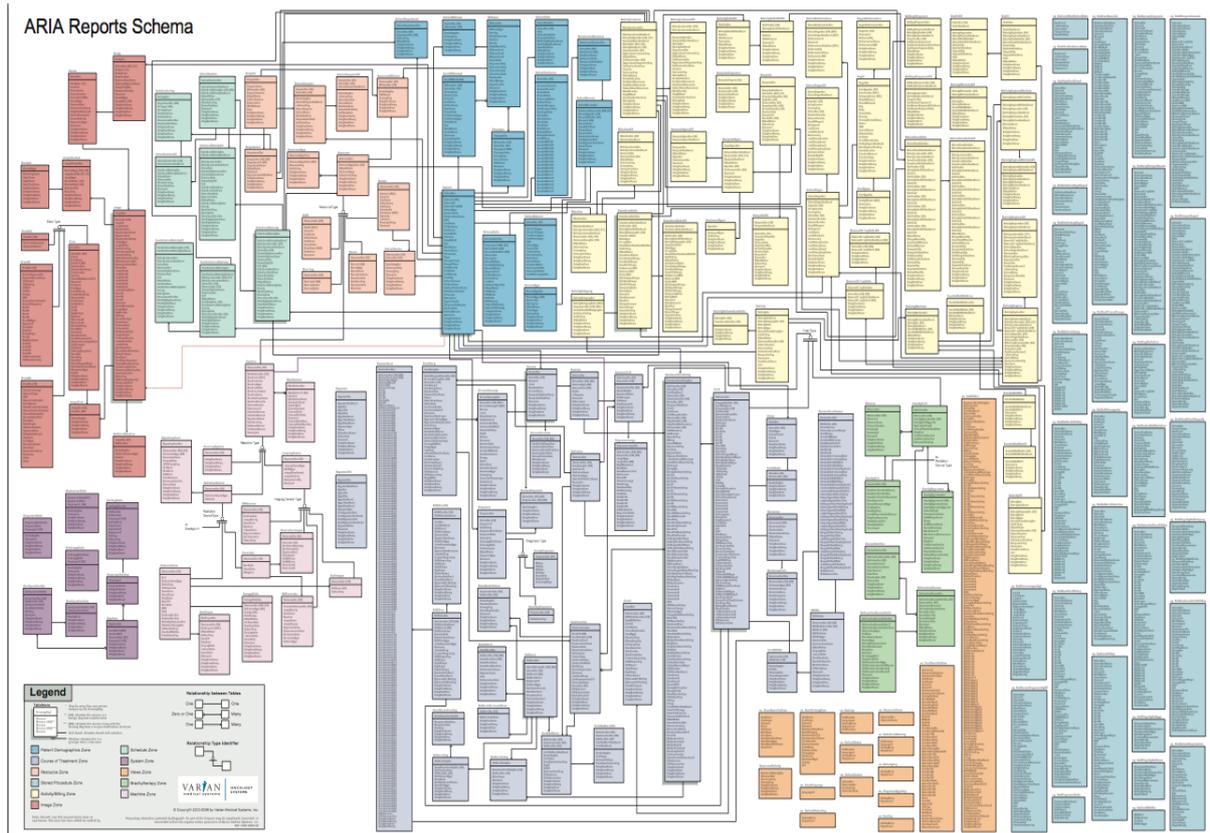


Figure 2: The Complex ARIA Database Schema

alias expressions, and represents grouping together equivalent event names into one descriptive category. For example, all initial consults, regardless of the type, were grouped into a single “Consult Appointment” category, all prescription types were grouped into a “Prescription Approved” field, and all radiotherapy treatment plans for a cancer type were grouped into one field, such as “Test Breast Plan”. We manually grouped together diagnoses into aliases by searching for a cancer type keyword in the diagnosis description. The data that we used was collected over a two-year period, from January of 2012 to October of 2014, and includes patient information for approximately 9000 patients that consulted with radiation oncologists or treated with radiation therapy at the MUHC during this time.

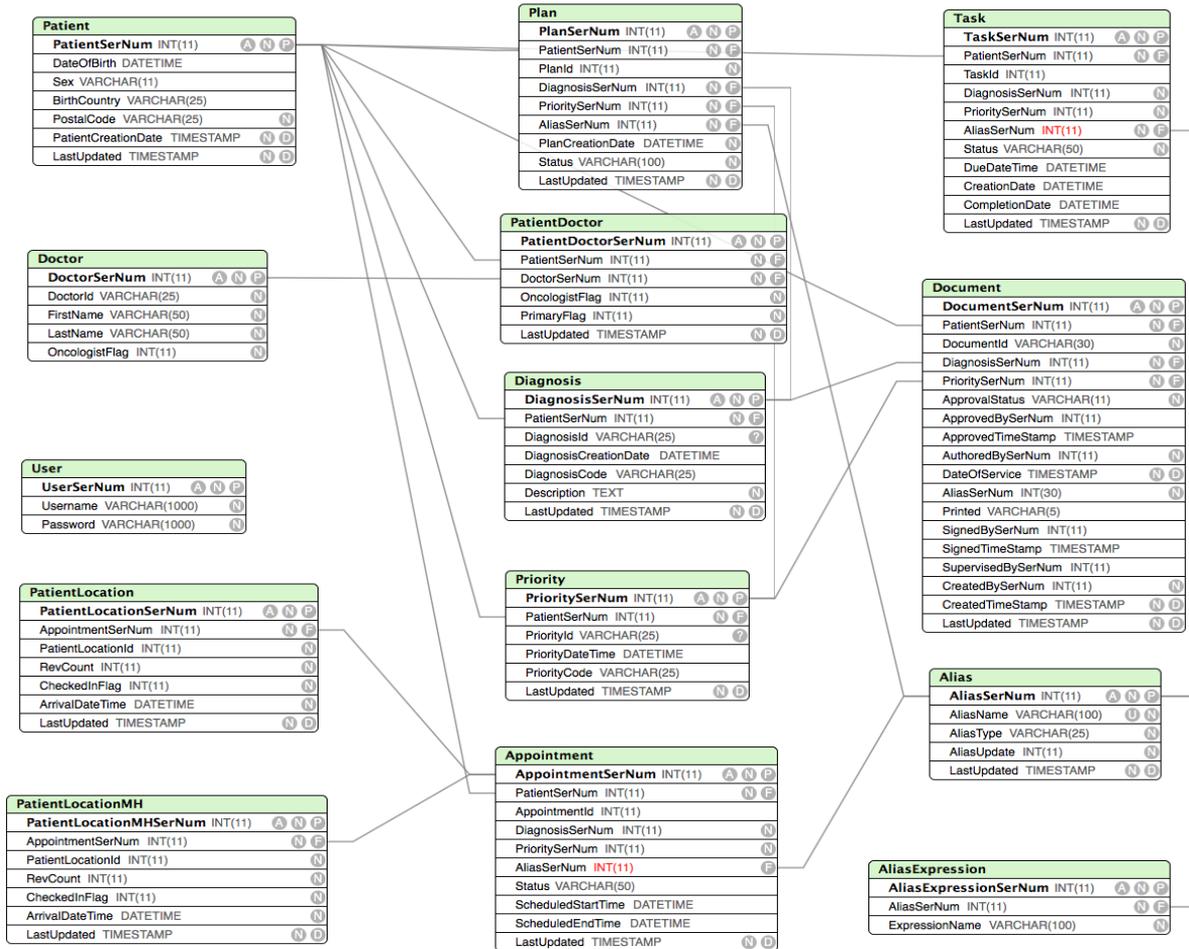


Figure 3: Database Schema for the ARIA Subset used

Data Extraction

From the ARIA subset database, we extracted the set of all radiation oncology events recorded in the system, sorted by patient. This was done by doing an inner join of the Appointment, Task, Plan, Document, and Patient tables by the PatientSerNum key that all of those tables have in common.

We define a time series as an array of consecutive time points that correspond to the start times for events in the radiotherapy planning process. Times for all events were recorded in the database to the second precision - the Creation Date field is a good approximation of the actual time the task was started (or more accurately, put in a queue of tasks for someone to complete). The Creation Date field was used instead of the Completed Date field because the latter is sometimes not present (having the default date value of zeros), probably due to doctors or other staff forgetting to manually input completion when the corresponding activity is actually complete. Initial data filtering was done to extract out all patients for whom a complete radiotherapy time course was completed - specifically, all patients who completed a CT-scan and proceeded through to start of treatment. Then, these patient time series were further filtered by only considering those containing the six essential steps in the radiotherapy process in the exact order: *Ct-Sim*, *Ready for MD Contour*, *Ready for Dosimetry*, *Prescription Approved*, *Ready for Physics QA*, and *Ready for Treatment*. An example time series for an anonymous patient is shown in figure 4, with the relevant rows highlighted in red. As can be seen, when

AliasName	PatientSerNum	CreationDate	CompletionDate	Status
Consult	21	14-07-19 11:18	14-07-16 16:36	A
Ct-Sim	21	14-07-30 8:30	14-07-30 9:00	Manually Completed
READY FOR CONTOUR	21	14-07-30 9:59	14-07-30 11:49	Completed
READY FOR MD CONTOUR	21	14-07-30 11:49	14-08-06 16:15	Completed
READY FOR CONTOUR	21	14-07-30 13:47	14-08-06 8:57	Completed
READY FOR DOSE CALCULATION	21	14-08-05 22:45	14-08-06 7:19	Completed
READY FOR DOSE CALCULATION	21	14-08-06 7:17	14-08-06 8:57	Completed
READY FOR DOSE CALCULATION	21	14-08-06 8:55	14-08-06 9:18	Completed
READY FOR DOSE CALCULATION	21	14-08-06 8:56	14-08-08 12:19	Completed
PRESCRIPTION APPROVED	21	14-08-08 11:58	14-08-08 11:29	A
READY FOR PHYSICS QA	21	14-08-08 12:20	14-08-08 14:11	Completed
READY FOR TREATMENT	21	14-08-08 14:12	14-08-08 14:45	Completed
READY FOR TREATMENT	21	14-08-08 14:19	14-08-08 15:59	Completed
End of Treatment Note Task	21	14-08-12 12:39	0000-00-00 00:00:00	Open

Figure 4: All events recorded in the database for anonymous patient 21. Events used in the analysis are highlighted in red.

constructing a time series we always select the first instance of the event in the case of duplicates.

These tasks were selected by creating graph in which each node represents a task and each edge is weighted by how many times a task is done after another in the data (see figure 5) in order to extract the most common possibilities in time course progression. This is much more apparent when considering only edges with weights (number of occurrences) > 2000 (figure 6)- we considered all steps present in this reduced graph (with the exception of the Ready for Contour step because this was absent for many otherwise valid time series). As well, these steps were chosen in particular because they are necessary for treatment planning and provide interesting partitions to study. Also, treatment-planning scenarios where further delays occur by treatment plans being revised and sent back to previous steps (a cycle in the figure 5) is taken into account, because all treatment planning courses must go through the essential steps in that order. Patients whose treatment planning occurred near the beginning and the end of the data collection period were thus not included, because not all essential steps were recorded in the dataset. Through the filtering process, the number of eligible patients went down from 7874 to 2238. These numbers are not particularly surprising, because the filtered data includes patients who, for example, were deemed unsuitable for radiotherapy treatment after an initial consult. Some patients had multiple time series, which correspond to that patient going through multiple treatment planning courses - the total number of time series analyzed as data samples was 2663.

Additional rules were defined in order to extract out time series that accurately represent the actual treatment planning process. Firstly, some patients with no Prescription Approved document instead had a Prescription (Fast Track) document instead. Secondly, sometimes the Ready for MD Contour event occurred before the Ct-sim event - this is most likely a glitch in the system because this always happens when the two events occur on the same day. In this case, I still included the time series, setting the start times of the two events to be equivalent. Thirdly, often in the treatment planning process the Ct-scan has to be redone for various reasons, and the patient is called in again. For these time series, I took the beginning of the series to be the second time the patient is scanned. Finally, sometimes the MD contour step and the prescription approved step were switched, and in these cases, instead of discarding valuable data, I inserted dummy time points for these two steps while keeping the other time points intact.

Even having defined these rules, some exceptions came up. For example, there were cases for

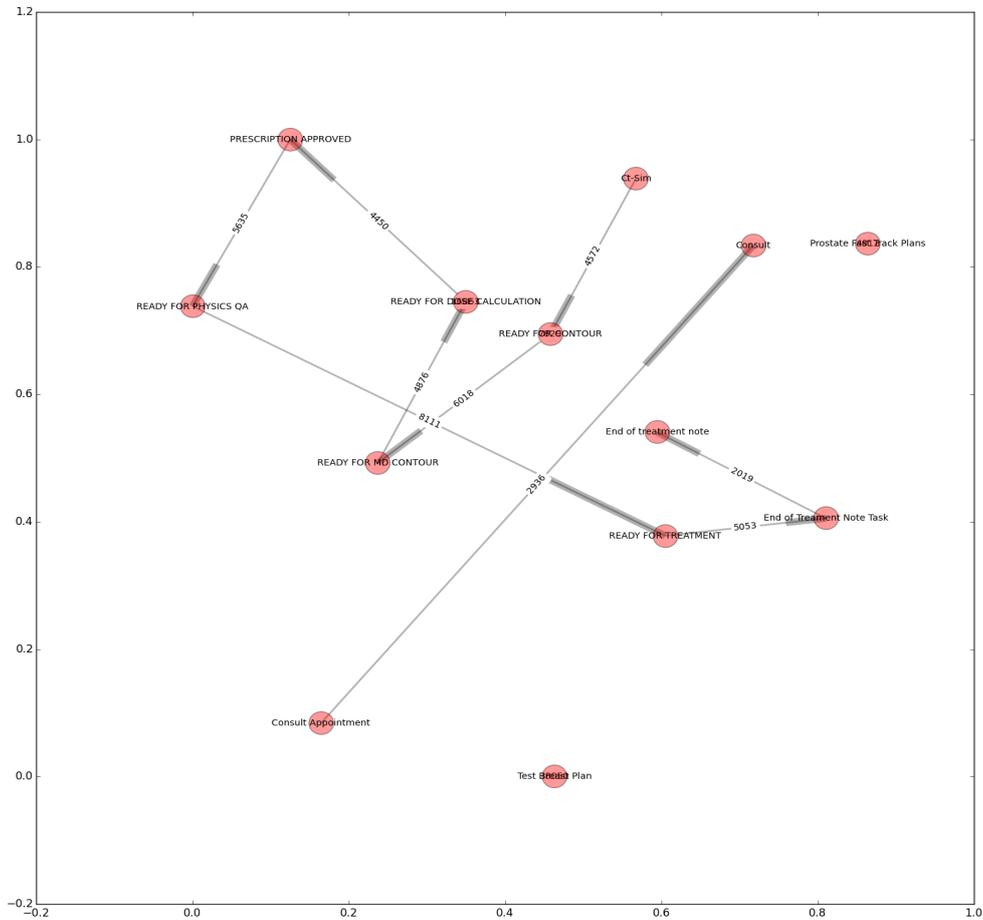


Figure 6: Reduced Task Dependency Graph for Radiation Oncology Events - This only includes edges with weights greater than 2000. The first seven nodes consecutive nodes, following the arrows from the Ct-Sim node (with the exception of the Ready for Contour event) correspond to our timeseries events.

Data Trends

For analysis of data trends, the potential factors that affect patient wait time investigated were the doctor, the diagnosis, the patient’s age, the priority of the diagnosis, the season that the treatment planning was initiated, and the patient’s gender. Each patient’s doctor was defined as being their primary oncologist registered in the database (note that for our predictive model, described below, we considered all oncologists as potential predictors because we do not have precise information on the degree of collaboration in the treatment planning process). Similarly, the patient’s diagnosis was defined as their primary diagnoses - subsequent metastases, for example, were not considered. Many diagnosis types are given similar priorities. For example, all breast cancer patients are assigned a priority of 3 or 4 (meaning lower priority). This indicates that factors are not statistically independent - multiple factors may be dependent on each other.

Firstly, all the time series were considered for each potential influencing factor in turn. The mean

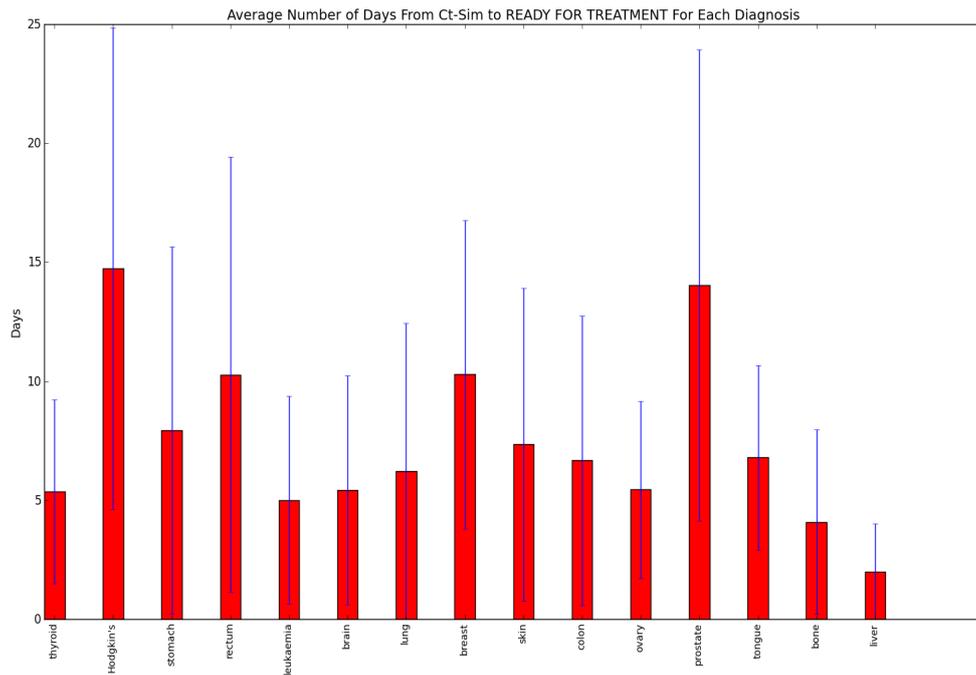


Figure 7: Average radiotherapy wait time separated by diagnosis. The standard deviation is indicated as the blue error bar.

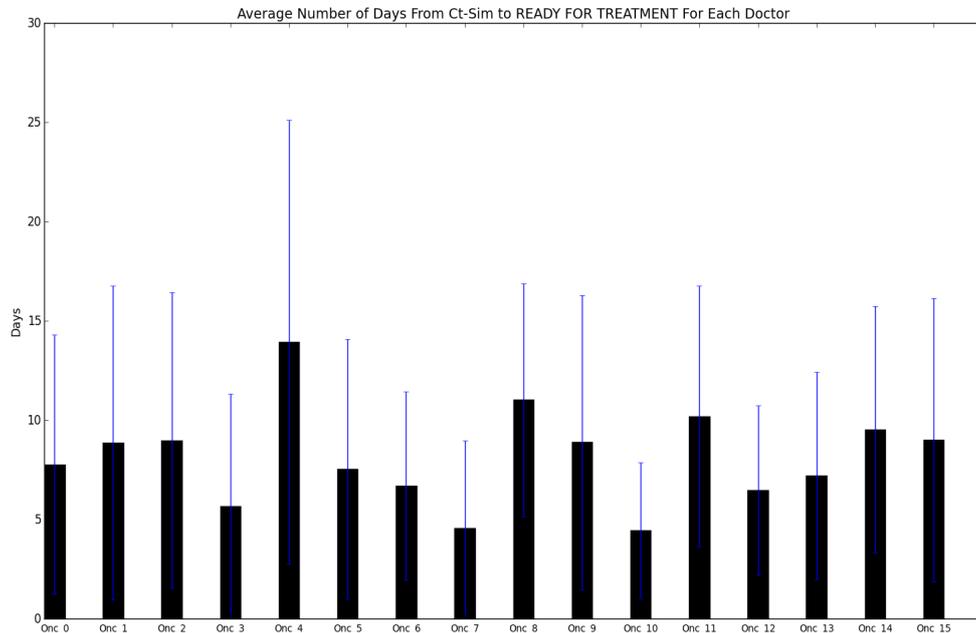


Figure 8: Average radiotherapy wait time separated by primary oncologist. The standard deviation is indicated as the blue error bar.

wait time definitely varies depending on the patient's doctor, diagnosis, or priority. As can be seen in figure 7, some doctors' patients complete the treatment planning process faster than

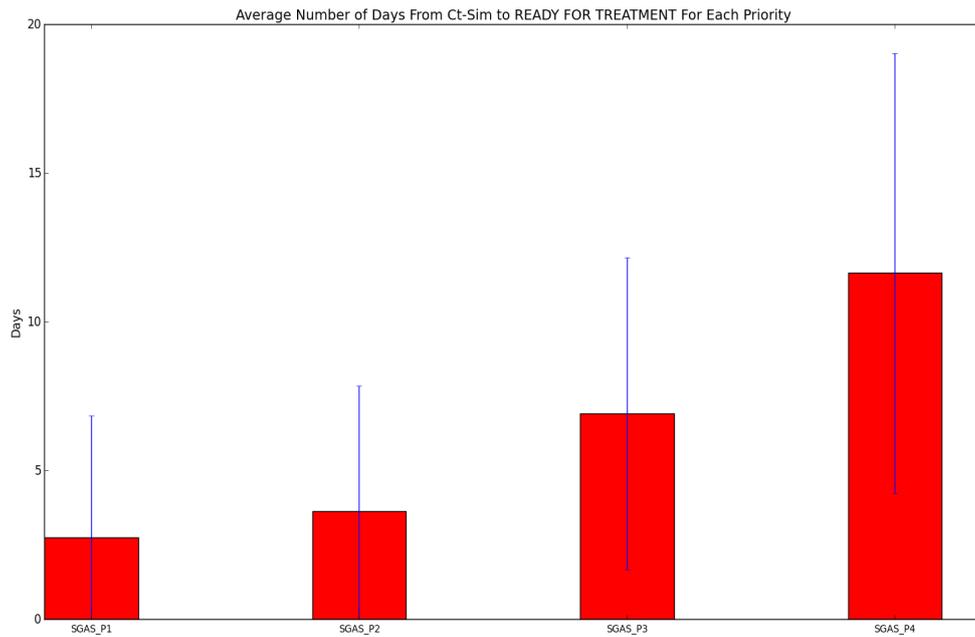


Figure 9: Average radiotherapy wait time separated by priority - priorities range from SGAS_P1 (highest) to SGAS_P4 (lowest). The standard deviation is indicated as the blue error bar.

others, presumably due to factors such as their cancer specialty, how busy that doctor is, and their efficiency. Average treatment planning times are stratified according to diagnosis (figure 8), due to a variety of factors including difficulty of treatment planning for that cancer type and the usual priority of that diagnosis (as described above). Cancers such as breast and prostate are easy to plan but very common and low priority, and so the average treatment planning time is longer than, for example, brain cancer and leukemia, which are often higher priority. Just looking at the impact of policy on wait time, we can see an obvious and expected upward trend as the priority decreases (figure 9). The policy for maximum allowed delay for priorities are same day for P1, <3 days for P2, <15 days for P3, and < 28 days for P4. Looking at average wait time per priority, we can see that all wait deadlines are met except for priority 1 - this indicates that a lot of the time radiation oncology is unable to properly address patients that need to have treatment planning done on the same day. A study of the impact of season showed no significant differences between months of the year for treatment planning time (figure 10) - because the dataset contains patient time series from January 2012 to October 2014, we have less samples for the months of November and December. It is interesting to note that there is a slight increase in waiting time until December, and January has the shortest waiting time - this could be an effect of the winter holidays on clinic efficiency. Similarly, males and females had insignificant differences in average treatment planning time. Plotting the age of patients against their wait times reveals a very slight upward trend that waiting times become longer as the patient's age increases (figure 11), justifying its inclusion as feature in our predictor (see below).

A plot of the average delay incurred by each essential treatment-planning step (figure 12) revealed that the two most significant delays in the process are the MD Contour step and the dosimetry step, while the CT-scan and Physics QA steps are basically trivial in length. The difference in planning time between oncologists is particularly apparent when only considering the average

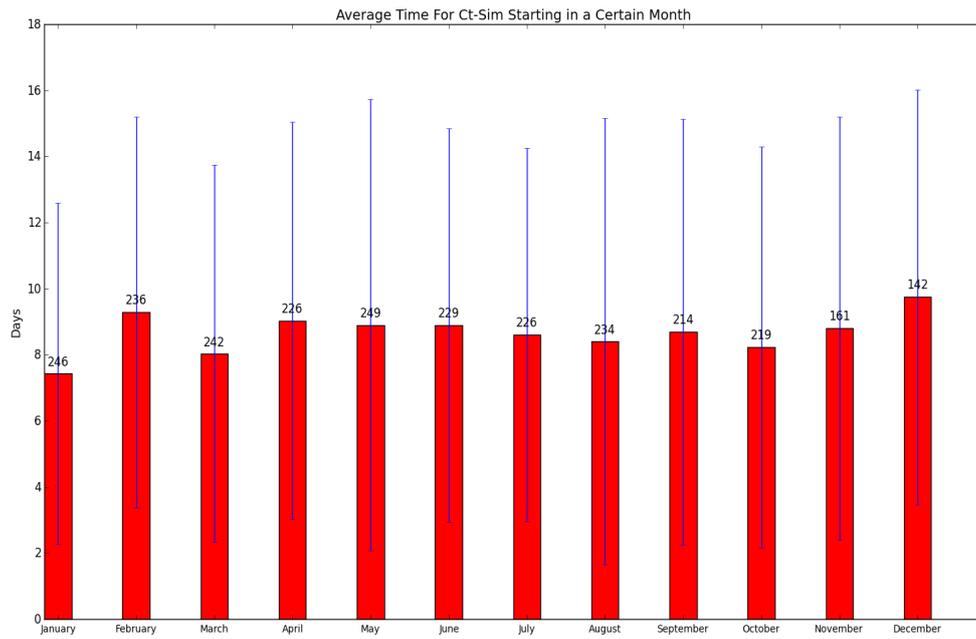


Figure 10: Average radiotherapy wait time separated by each month of the year. The number on top of each bar represents the number of patients who started treatment in that month in our dataset.

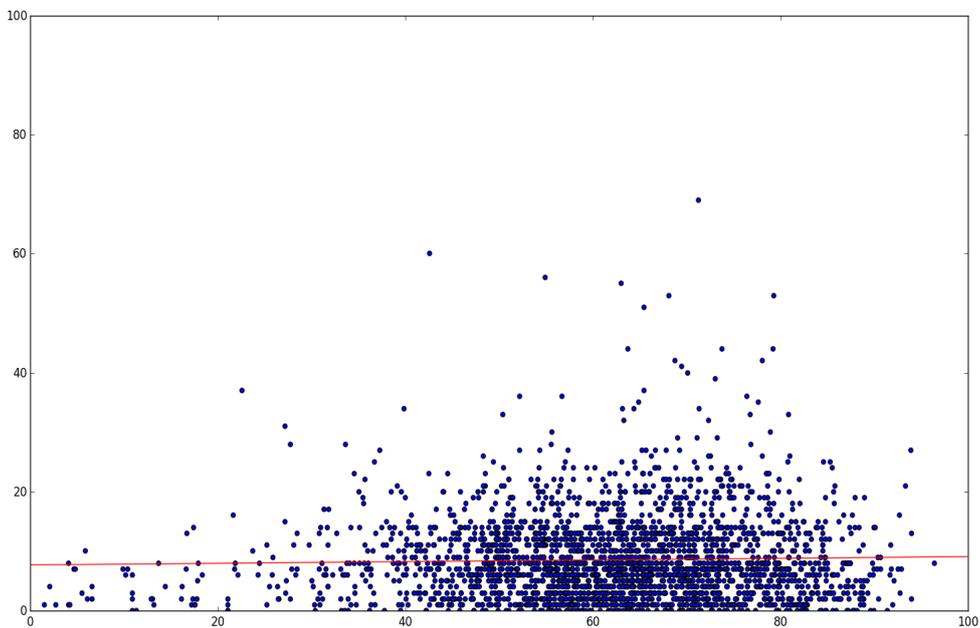


Figure 11: Plot of age (in years) versus waiting time (in days) for all samples in the dataset. The red line represents a linear regression of all the points.

length of the MD Contour step (figure 13).

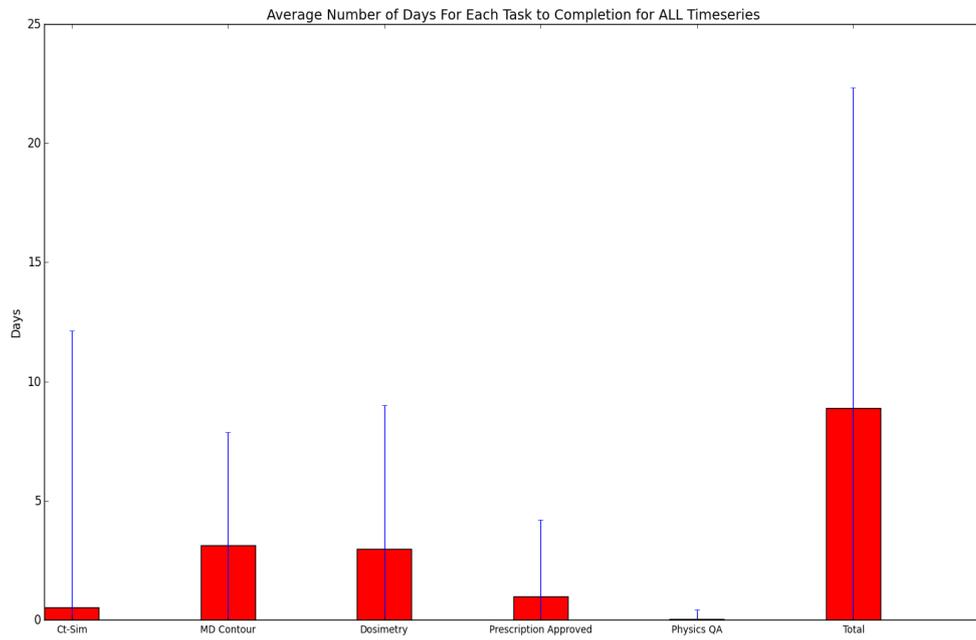


Figure 12: The average delay (in days) for each essential step described.

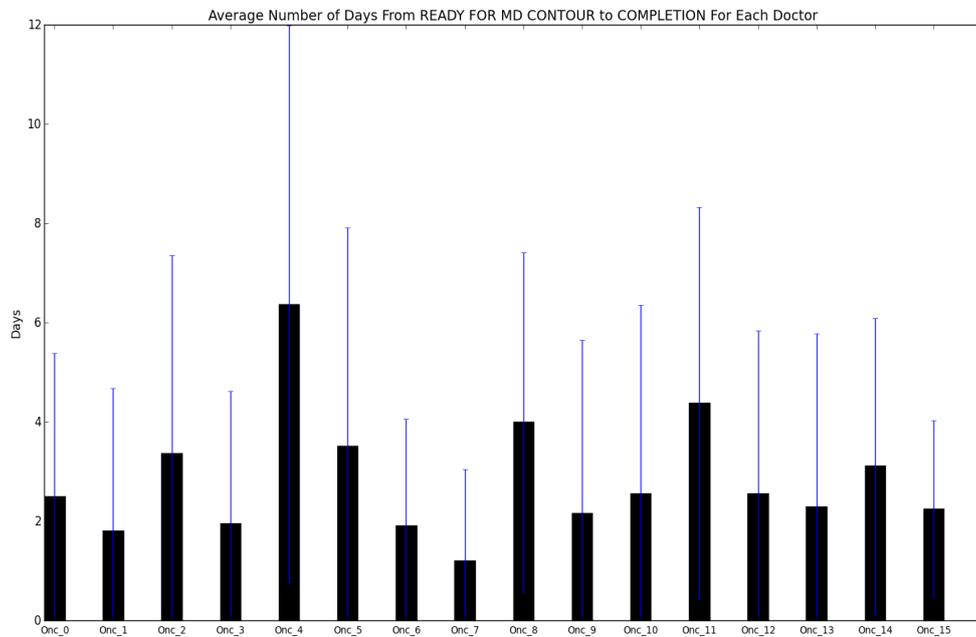


Figure 13: Average MD Contour times for each oncologist across all samples.

In order to study how the useful predicting factors interact with each other, a utility function

was defined in order to extract all time series from that data that correspond to any combination of priority, diagnosis, and priority, in order to visualize combined effects. Particularly interesting cases for analysis are the average wait times of patients for each doctor by diagnosis (figures 14 and 15) and average wait times of patients for each priority by diagnosis. For the first case, the data was visualized by additive delays for each step adding up to the total delay. This can be seen as a "track record" for each oncologist, stratified by the different diagnoses. Figure 14 represents a "good" track record, where MD Contour times are typically short, and the overall average waiting time for each diagnosis is also low. Figure 15 represents a track record that is not as optimal, with longer MD Contour times for the same diagnosis compared to the "good" track record. A comparison between diagnoses for different priorities shows that for the same diagnosis, the average wait time differs significantly between priorities (all priority 1 cases have a much lower average than all priority 4 cases). The combinations of these factors are taken into consideration when we constructed an initial model for prediction.

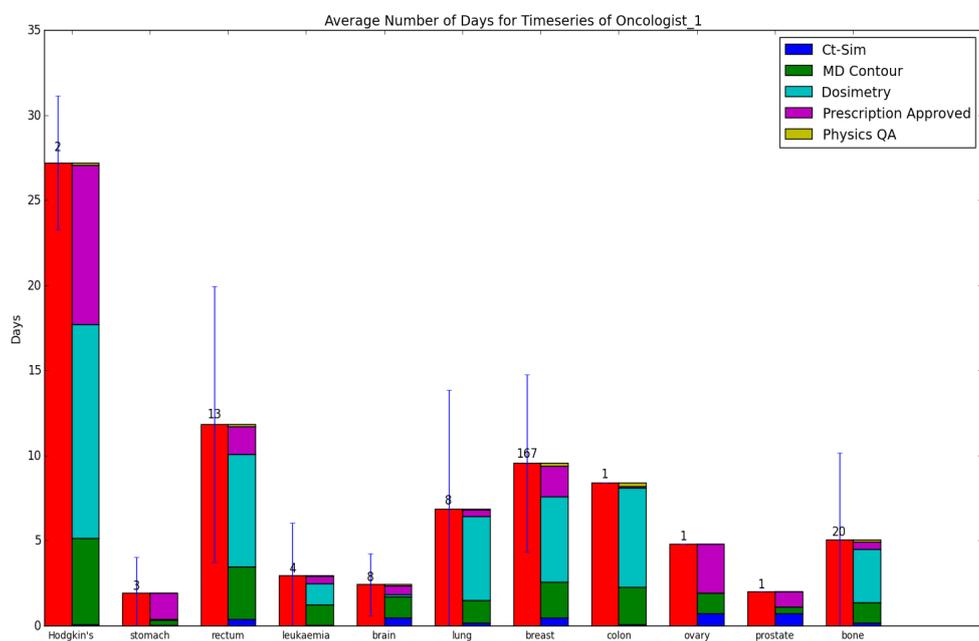


Figure 14: A visualization of average wait times separated by diagnosis for an efficient oncologist

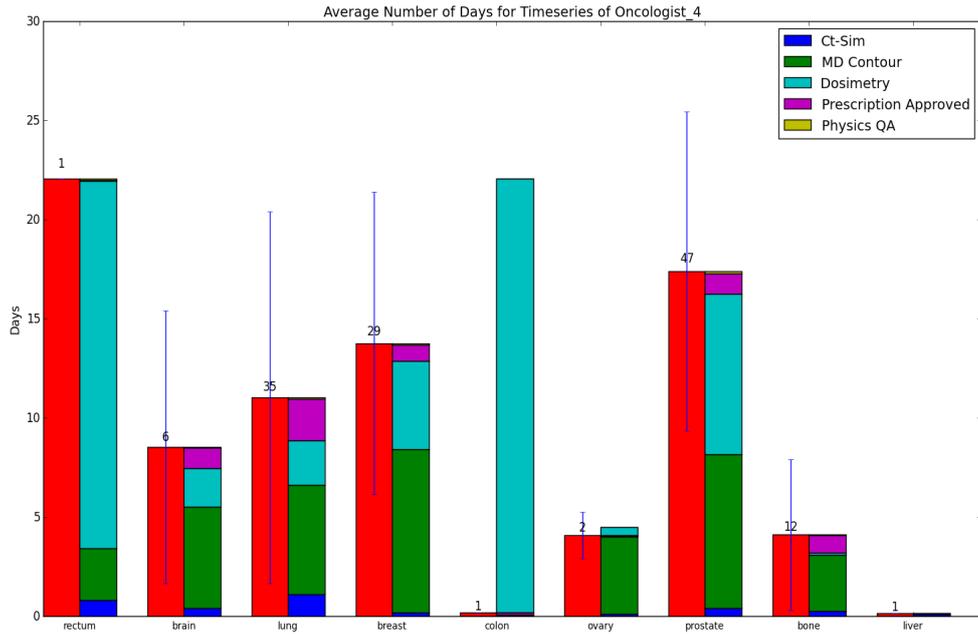


Figure 15: A visualization of average wait times separated by diagnosis for an inefficient oncologist

Initial Prediction Model

Based on our analysis of the trends of factors affecting radiotherapy-planning times, we proposed an initial predictive model for machine learning to predict the wait time, effectively as a supervised regression problem. The features used in the regression included the diagnosis (cancer type) of the patient, the patient’s oncologist(s), the patient’s age, and the priority level of the diagnosis. Of the 2663 time series data samples initially extracted, only 2046 of those were used, because the rest had missing data for one or more of the regression features. The parameter matrix was defined by having a single column for each diagnosis type, oncologist in the clinic, and priority; each training sample (patient time series), had a value of 1 under the appropriate column if that diagnosis or oncologist or priority corresponds to that treatment course, and a zero otherwise. Age was the only non discrete variable used. The training set was constructed by randomly choosing 4/5 of the whole dataset, and the remaining 1/5 was assigned to the testing set. A number of regression models were considered in turn by fitting it to the training set, and evaluating its performance on the testing set. This was done using the implementations found in the scikit-learn python library for data mining and analysis [11]. The accuracy of the regression models was compared by calculating the average of the absolute value errors of the prediction from the actual value as well as the standard deviation of these errors. A comparison of the error rates of the different regressors used is shown in figure 16. A comparison of our estimated waiting time (red) versus the actual recorded waiting time (green) is shown in figure 17 - one interesting property is that when our predictor is inaccurate, it tends to over-predict the actual wait time rather than provide an earlier estimate - this is arguably beneficial because it is better to give a patient a conservative estimate rather than an optimistic one.

In simple linear regression, a straight line with n parameters through n -dimensional space is optimized by minimizing the least squares error to all the training data points. A regularized

Regression Algorithm	Mean Error (Days)	Standard Deviation
Linear	4.13	3.29
Ridge (Regularized)	4.13	3.26
SVM (Linear Kernel)	3.97	3.35
Linear RFE	4.12	3.29
SVM RFE	4.03	3.39

Figure 16: Average Prediction Error for various regression algorithms.

(or ridge) regression method tries to address the potential problem of overfitting by adding a regularization term to the least squares cost function, in order to limit the size of parameter values and have a “simpler” hypothesis for prediction. Support Vector Machine (SVM) regression is another supervised learning method that, like support vector machine classification, attempts to make predictions that by having a maximal classification margin. In this context, the SVM with a linear kernel fits a line through the dataset, but only by considering all points within a specific margin and ignoring all data points outside this margin in order to minimize the effect of outliers. Next, recursive feature elimination (RFE) was used for better feature selection - given a classifier (in our case the linear and SVM regressors), RFE first trains the estimator on the initial set of features and then, based on the parameter weights obtained, RFE recursively repeats the process of pruning the parameters with lowest weights and training the estimator on the new set of features until a desired number of parameters is reached. Here we ran RFE by testing parameter weights using 5 fold cross validation at each recursive step until no further improvement could be made. This yielded a slight improvement in the performance of the linear regressor, but did not improve the SVM regressor.

It is important to note that not all features that appeared to differentiate the time series improved the regression model. For example, when a feature was added for whether the time series corresponded to the first of a patient’s treatment courses or not, the linear and SVM predictors performed worse on the testing set. This may be because this new predictor over fitted the testing data - the average error after fitting the entire dataset was slightly better than without this new feature (3.85 days versus 3.88 days). Overfitting is a common regression problem when you add more and more features to the model, fitting the training set better but failing to generalize to new data. This illustrates how simply putting in all informative features may not provide the best estimate of wait time - proper selection of combinations of model features is required. Model selection is also very important - for example, an SVM with a quadratic kernel instead of a linear one yielded much poorer results than the rest of the regressors, probably due to significant overfitting.

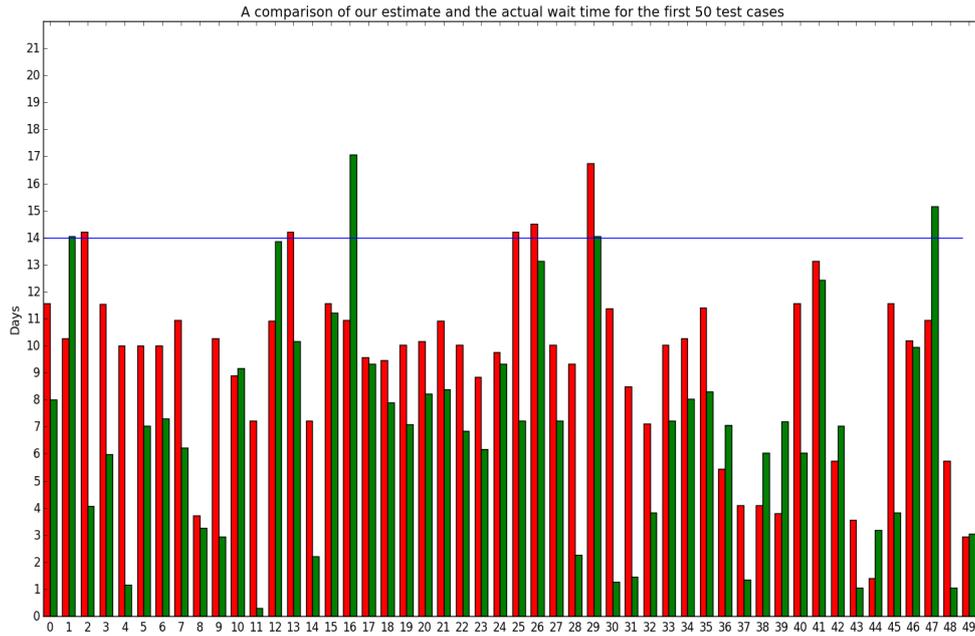


Figure 17: A comparison of the first 50 testing datapoints - our estimate is shown in red, whereas the actual patient wait time is shown in green. The blue line represents the two week estimate typically given to patients.

Conclusions and Future Work

We have shown that with our regression algorithms along with relevant patient information, we can estimate the true patient waiting time before the start of radiation therapy on average to within 4 days, with a standard deviation of about 3 days. This is a significant improvement on the generic two week estimate usually given to patients, and also on the estimate you would obtain by giving the average of all patient wait times - an estimate of 8.89 days, with a very large standard deviation of 13.45 days. These initial predictions are a promising start, but there is much future work to be done to validate and improve our model. Although the accuracy of our analysis and prediction was improved through extensive filtering with the rules stated above, there is potentially a lot of unused data in the database that could be useful to expand the set of data to be trained on. For example, there are some treatment planning courses that lack one or two of the five essential steps, and these are simply tossed out in our algorithm, even though they may provide some valuable insight in abnormal treatment planning courses and how they could potentially affect the waiting time. Also, when the last three essential steps occur on the same day, there are some treatment planning courses for which the events occur out of order - again, our algorithm discards this data. It would be helpful to define further rules to make use of data that does not conform to our filtering steps.

From our data, there is clearly a great deal of unexplained variance, even considering all the features we have used so far. All patients with the same doctor, diagnosis, and priority can have quite different waiting times. As an illustrative example, all breast cancer patients with priority 4 whose primary oncologist is Dr. Tarek Hijal had an average waiting time of 9.91 days, but with a wide variance (a standard deviation of 5.11 days). The wait time can range from 4 days to 30 days for all patients in this very specific category. Obviously there is a need to include

finer grained patient information in order to make a better prediction. Probably combinations of other factors are at play that we have not accounted for.

Some additional factors to consider include more specific information about diagnoses. Each breast cancer treatment can differ in many ways, for example by whether or not the cancer has regional nodes, and whether the treatment requires tomography unit plans; capturing this variance could improve our prediction. Another important potential factor affecting wait time is the impact of overall “busy-ness” of the radiation therapy clinic, a factor that needs to be better defined and analyzed; through discussion with radiation oncology staff, it is apparent that there are cycles in which work is accumulated in a queue-like fashion until many items can no longer be delayed, at which time all items are completed. As well, planning courses should be classified not only as whether or not it is the first one for the patient, but also whether it is before a treatment, during a treatment, or after completion of a first treatment.

Ultimately, we also need to ensure that our predictive model can be generalized to a completely new dataset once the MUHC Radiation Oncology Department moves to its new location at the Glen in 2015. Because protocols and facilities will be completely different, a useful future predictor will mostly have to be learned from new electronic health record data being collected there. We suspect, however, that many of the same factors will not change significantly, notably the priority system and the impact of the oncologist and diagnosis. In addition to alerting patients the amount of time they should expect to wait to begin radiotherapy treatment, a further step could be to implement a more optimal scheduling system, based on our analyses of factors affecting longer wait times. If all factors, such as difficulty of planning for each type of diagnosis, can be properly specified, it would then become possible to more optimally plan the sequence of events completed by oncologists, dosimetrists, and medical physicists. These possibilities and more will be explored in future work.

References

1. Paul, C., M. Carey, A. Anderson, L. Mackenzie, R. Sanson-Fisher, R. Courtney, and T. Clinton-Mcharg. "Cancer Patients' Concerns regarding Access to Cancer Care: Perceived Impact of Waiting times along the Diagnosis and Treatment Journey." *European Journal of Cancer Care* (2012): No. Print.
2. Klausen, OG, J. Olofsson, and B. Rosengren. "A Long Waiting Time for Radiotherapy. Not Acceptable for Patients with Neoplasms." *Tidsskr Nor Laegeforen* (1989). Print.
3. Mackillop, William J. "Killing Time: The Consequences of Delays in Radiotherapy." *Radiotherapy and Oncology* (2007): 1-4. Print.
4. Waaijer, Annet, Chris H.j. Terhaard, Human Dehnad, Gerrit-Jan Hordijk, Maarten S. Van Leeuwen, Cornelius P.j. Raaymakers, and Jan J.w. Lagendijk. "Waiting times for Radiotherapy: Consequences of Volume Increase for the TCP in Oropharyngeal Carcinoma." *Radiotherapy and Oncology* (2003): 271-76. Print.
5. Chen, Zheng, Will King, Robert Pearcey, Marc Kerba, and William J. Mackillop. "The Relationship between Waiting Time for Radiotherapy and Clinical Outcomes: A Systematic Review of the Literature." *Radiotherapy and Oncology* (2008): 3-16. Print.
6. Robinson, D., T. Massey, E. Davies, RH Jack, A. Sehgal, and H. Moller. "Waiting times for Radiotherapy: Variation over Time and between Cancer Networks in Southeast England." *British Journal of Cancer* (2005). Print.
7. Arndt, V. "Provider Delay Among Patients With Breast Cancer in Germany: A Population-Based Study." *Journal of Clinical Oncology* (2003): 1440-446. Print.
8. Afzelius, Pia, Karin Zedeler, Hanne Sommer, Henning T. Mouridsen, and Mogens Blichert-Toft. "Patient's and Doctor's Delay in Primary Breast Cancer: Prognostic Implications." *Acta Oncologica* (1994): 345-51. Print.
9. Santibáñez, Pablo, Vincent S. Chow, John French, Martin L. Puterman, and Scott Tyldesley. "Reducing Patient Wait times and Improving Resource Utilization at British Columbia Cancer Agency's Ambulatory Care Unit through Simulation." *Health Care Management Science* (2007): 392-407. Print.
10. A. Joseph, J. Seuntjens, C. Freeman, W. Parker, J. Kildea. "An analysis of the effectiveness of automated pre-, post- and intratreatment auditing of electronic health records." Department of Medical Physics and Radiation Oncology (2014). Poster.
11. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E. "Scikit-learn: Machine Learning in Python." *Journal of Machine Learning Research* (2011): 12:2825-2830.

Appendix - Patient Wait Times for Each Oncologist at MGH Stratified by Diagnosis

