

## ABSTRACT

Title of dissertation:           APPLYING OPERATIONS RESEARCH MODELS TO PROBLEMS IN HEALTH CARE

Stuart Price, Doctor of Philosophy, 2015

Dissertation directed by:    Professor Bruce Golden  
Decision, Operations, & Information Technologies

Intensity- modulated radiation therapy is a form of cancer treatment that directs high energy x-rays to irradiate a tumor volume. In order to minimize the damage to surrounding tissue the radiation is delivered from multiple angles. The selection of angles is an NP-hard problem and is currently done manually in most hospitals. We use previously evaluated treatment plans to train a machine learning model to sort potential treatment plans. By sorting potential treatment plans we can find better solutions while only evaluating a fifth as many plans. We then construct a genetic algorithm and use our machine learning models to search the space of all potential treatment plans to suggest a potential best plan. Using the genetic algorithm we are able to find plans 2% better on average than the previously best known plans.

Proton therapy is a new form of radiation therapy. We simulated a proton therapy treatment center in order to optimize patient throughput and minimize patient wait time. We are able to schedule patients reducing wait times between 20% and 35% depending on patient tardiness and absenteeism.

Finally, we analyzed the impact of operations research on the treatment of prostate cancer. We reviewed the work that has been published in both operations research and medical journals, seeing how it has impacted policy and doctor recommendations.

APPLYING OPERATIONS RESEARCH MODELS TO  
PROBLEMS IN HEALTH CARE

by

Stuart Price

Dissertation submitted to the Faculty of the Graduate School of the  
University of Maryland, College Park in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
2015

Advisory Committee:  
Professor Bruce Golden, Chair  
Professor David Bigio, Dean's Representative  
Professor Shawn Mankad  
Professor Ilya Ryzhov  
Professor Edward Wasil

© Copyright by  
Stuart P Price  
2015

## Contents

1	Introduction	1
2	Data Mining to Aid Beam Angle Selection for Intensity-Modulated Radiation Therapy	5
2.1	Introduction	6
2.2	Methods	10
2.2.1	Penalty Score to evaluate plan quality	10
2.2.2	Angle Set Evaluation	12
2.2.3	Models	12
2.3	Results	14
2.4	Discussion	23
	Appendix 2-1	26
	Appendix 2-2	27
3	Optimizing Intensity-Modulated Radiation Therapy Angle Selection using Genetic Algorithms	28
3.1	Introduction	29
3.2	Methodology	31
3.2.1	Penalty Function	32
3.2.2	Data	33
3.2.3	Machine Learning	34
3.2.4	Tumor Geometry	38
3.2.4	Results on Training Set	40
3.2.5	Geometric Results	42
3.3	Genetic Algorithm	43
3.3.1	Population Size and Stopping Criteria	43
3.3.2	Mutation, Crossover, and Local Search	49
3.4	Results	52

3.5	Conclusions	53
	Appendix 3-1	55
4	Scheduling Guidelines for a Multi-Room Proton Therapy Treatment Center	58
4.1	Introduction	59
4.2	Simulation Model	62
4.3	Simulation Results	65
4.4	Extending the Simulation	71
4.5	Extended Simulation Results	74
4.5.1	Mixtures of in-patients and out-patients	74
4.5.2	Block Scheduling	78
4.5.3	Equipment Failures	83
4.5.4	Special Needs	84
4.6	Conclusions	85
5	The Impact of Operations Research on Prostate Cancer Research	86
5.1	Introduction	87
5.2	Prostate Specific Antigen Screening	91
5.2.1	Quantifying the Effects of PSA Screening	91
5.2.2	Predicting Prostate Cancer from the PSA Level	97
5.2.3	PSA Screening Policy	103
5.3	Detecting Tumors	103
5.3.1	Improving Biopsies	109
5.3.2	Using an MRI to Detect Prostate Cancer	110
5.4	Prostate Cancer Treatment	114
5.4.1	Staging	114
5.4.2	Active Surveillance	117
5.4.3	Hormone Therapy	119
5.4.4	Brachytherapy	121

5.4.5	Patient Choice	125
5.4.6	Recurrence	128
5.4.7	Costs	133
5.5	Conclusions	135
6	Conclusions	136
7	References	140
7.1	References for IMRT	140
7.2	References for Proton Therapy Treatment Center	143
7.3	References for Prostate Cancer	145
7.4	Annotated Papers for Prostate Cancer Analysis	146

# 1 INTRODUCTION

In 2012, the United States spent 17.9% of its GDP on healthcare, more than any other nation (OBP 2014). With healthcare costs representing such a major part of the economy, there has been a great deal of research devoted to improving not only the treatment of patients, but the efficiency of healthcare delivery. The Agency for Healthcare Research and Quality (AHRQ) is tasked with producing evidence and guidelines to make healthcare safer, higher quality, more accessible, equitable, and affordable. It has repeatedly called for the use of operations research to increase the efficiency of healthcare delivery, reducing the cost, and improving the outcomes of patients (AHRQ 2011). The importance of operations research in healthcare extends beyond improving the efficiency of healthcare systems and to the treatment of patients themselves. Increasingly sophisticated treatments in medicine require optimization of everything from drug plans for cancer patients to imaging of computed tomography scanners.

In 2014, there are expected to be more than 1.6 million new cases of cancer diagnosed in the United States, with over half a million cancer deaths (ACS 2014). In 2009, the United States spent \$86.6 billion on medical treatments directly related to cancer, but indirect mortality costs raised the total cost of cancer to an estimated \$216.6 billion (ACS 2014). The treatment of cancer represents an important part of the healthcare industry, with specialized doctors, hospitals, facilities, and other resources dedicated to its treatment. This dissertation focuses on how operations research can be used to improve cancer treatment, from planning external beam radiation therapy to the efficiency of treatment delivery to patients. There are three parts to the research: improving the delivery of radia-

tion to tumors in intensity-modulated radiation therapy (IMRT), scheduling patients for treatment at a proton therapy treatment center, and analyzing the impact of operations research on the treatment of prostate cancer.

Currently, nearly two thirds of all cancer patients in the United States receive some form of radiotherapy (NIH 2012). IMRT is currently the most common form of external beam radiation therapy (Smith et al. 2011). IMRT, like all forms of radiotherapy, has side effects from radiation being delivered to healthy tissue surrounding the tumor and treatment volume. Improving treatment plans can reduce the incident radiation to healthy tissue, decreasing both acute and long term side effects. The importance of a good treatment plan leads to longer planning times. A good portion of the cost of IMRT is the planning time it takes to design each patient's treatment plan; IMRT takes radiation oncologists between 30% and 88% longer than other treatment types (Van de Werf et al. 2012). The planning cost per patient has been estimated to be £484 or about \$780 (Ploquin and Duscombe 2008). Treatment planning consists of identifying the important structures, deciding the beam angles, and optimizing delivery over those beam angles. The step of beam angle selection is currently manually performed at most hospitals with the final step of optimization being performed by commercial software. The manual search of beam angles is a time intensive iterative process. We develop a tool for sorting beam angle sets by potential plan quality that finds better plans while searching one fifth the number of plans. Such a tool could vastly decrease the time spent manually searching for angle sets in treatment planning. We extend this work by constructing a genetic algorithm that will search the space of beam angle sets for the plan it predicts will perform best. These tools

have the potential to improve both the treatment of the patient and to reduce the burden of treatment planning.

Proton therapy is a promising new form of external beam radiation therapy. Protons have the ability to better pinpoint the delivery of radiation by scattering less than x-rays and delivering almost no radiation to the tissue behind the treatment volume. Currently, the size and nature of the equipment needed for proton therapy necessitates a specialized facility be built. The cost of such a facility and equipment runs in excess of one hundred million dollars. The widespread adoption of proton therapy has been greatly limited by this high cost. There are currently only 40 facilities worldwide. When presented with a treatment that is both scarce and in high demand it is important to ensure that facilities are designed to see as many patients as possible and that all equipment is being optimally used. A proton therapy treatment center is a complex system with many patients able to be treated in parallel, each with stochastic treatment times and multiple treatment steps. Furthermore, even though some steps of treatment can be done in parallel, the cyclotron delivering the high energy protons can supply only one treatment room at a time, meaning that only one patient can receive radiation at a time. We show the throughput increase of having multiple treatment rooms and the diminishing returns of having too many. We show that when having multiple rooms relying on a common resource, scheduling patients with respect to their demand on that resource is important. We go further to analyze system failures and special needs of patients. We show how to maintain throughput and reduce patient wait times in the presence of high uncertainty in patient arrival.

Prostate cancer is the most common form of cancer in men, with one in seven developing prostate cancer within their lifetime (SEER 2012). It has been shown that early

detection of prostate cancer leads to improved treatment outcomes. In an effort to detect prostate cancer in its early stages it has been common practice to measure the prostate specific antigen of men over forty. However, screening with prostate specific antigen leads to a large number of false positive detections, resulting in unnecessary biopsies. Biopsies are intrusive and carry a risk of infection. Operation researchers have simulated policies in order maximize the welfare gain from early detection and minimize unnecessary procedures. Researchers are looking for better screening alternatives including MRIs. The data from MRIs is complex and humans do a poor job of analyzing the images for prostate cancer detection. Using machine learning, however, the accuracy of detection can be improved well beyond that of PSA screening and human interpretation of MRI. Optimization techniques have also been applied to the treatment of prostate cancer through seed placement in brachytherapy and the decision of when to start treatment. From screening to post-operative monitoring, operations research has had a significant impact on the treatment of prostate cancer.

The dissertation is ordered as follows. The first section focuses on using data mining and genetic algorithms to improve IMRT treatment plans. The second section is on the optimization of beam angle selection through genetic algorithms. The third section is on simulating a proton therapy treatment center to test scheduling procedures. Finally, the fourth section is an effort to measure the impact that operations research has had on the treatment of prostate cancer. Each section begins with its own introduction and literature review, followed by results.

## **2 DATA MINING TO AID BEAM ANGLE SELECTION FOR INTENSITY-MODULATED RADIATION THERAPY**

### **ACKNOWLEDGEMENTS:**

This is research conducted with:

Bruce Golden, University of Maryland  
Edward Wasil, American University  
Howard H. Zhang, University of Maryland School of Medicine

### **ABSTRACT**

The beam angle optimization problem for Intensity-Modulate Radiation Therapy (IMRT) has been well studied. Previous research has focused on developing optimization approaches to obtain a single high quality angle set for each patient. In this research, we propose a population-based method to aid the beam angle selection process with the goal of significantly improving the efficiency of the process while identifying high-quality angles. We used a database of 2,700 IMRT treatment plans for 10 patients with locally advanced head and neck cancer. We used machine learning to rank beam angle sets to determine those most likely to be of high quality. Logistic regression, neural network, and k-nearest neighbor algorithms were used with different parameters and feature sets. The ranking algorithms are able to find better angle sets by evaluating just three plans than would be found by randomly searching five times as many.

## 2.1 Introduction

Over the last 10 years, intensity-modulated radiation therapy (IMRT) has become one of the dominant forms of external beam radiation therapy. IMRT replaced 3D conformal radiation therapy for treating many types of cancer because it is better able to spare the surrounding healthy tissue while maintaining the prescription dose to the tumor. This increased precision in delivering radiation is particularly important in tumors surrounded by vulnerable tissue, such as cancer in an advanced stage. This increased precision comes with increased complexity, making the IMRT planning problem NP-hard (Saher and Sultan 2010). Current IMRT planning procedures have yet to use the vast amount of information collected from the optimizing of treatment plans for the tens of thousands of patients who undergo IMRT every year.

An oncologist and a medical physicist collaborate on producing plans for the IMRT planning problem (we refer to them as the planners). There are three primary stages in solving the planning problem. During the first stage, organs at risk (OARs) and the tumor are delineated on the CT scan (X-ray computed tomography). The tumor and tissue immediately adjacent to it are referred to as the planning target volume (PTV). The PTV and nodal volumes are assigned varying levels of radiation to ensure that the malignant cells are destroyed. An isocenter is selected and the radiation source (a linear accelerator in the case of X-ray IMRT) is pivoted on a gantry arm around the isocenter. Typically, the isocenter is chosen to be the center of mass of the tumor and varies from patient to patient. The isocenter may vary, to a lesser degree, over the course of treatment.

The second stage of the planning process involves the selection of beam angles to deliver the radiation relative to the patient's anatomy. First, the planners must decide how many beam angles to use. Research has shown that while more beam angles might spread

out the radiation, ensuring no individual organ receives too high a dose, there are diminishing returns for too many angles (Mohan and Ling 1995; Narayanan et al. 2012; Söderström and Brahme 1995; Stein et al. 1997). In addition, there are practical limitations. For example, each additional angle increases the amount of time needed to deliver the radiation, which can cause complications for very sick patients. Wu et al. (2010) suggested that the dosimetric quality of plans is dependent on the number of beam apertures used. No matter how many angles are used, a minimum number of apertures is required to ensure adequate plan quality.

After deciding on the number of angles to use, the values of the angles are chosen. Angle selection is usually done manually, despite research showing that intelligent beam angle selection can further reduce radiation to the OARs (Aleman et al. 2008, 2009; Pugachev et al. 2001). Much of the literature on optimizing beam angles focuses on beam view methods (Cho et al. 1999; Goitein et al. 1983; Myriantopoulos et al. 1992; Pugachev and Xing 2001a, 2001b, 2002). These methods evaluate one beam plan at regular intervals and create a vector of radiation delivered to both the tumor and the OARs. A set of angles is then constructed based on the performance of the individual beam plans. Methods of optimizing using the beam view method include simulated annealing (Aleman et al. 2008), mixed integer programming (D'Souza et al. 2004; Lee et al. 2006; Lim et al. 2008; Wang et al. 2003), genetic algorithms (Li et al. 2004; Schreiber et al. 2004), clustering (Bangert and Oelfke 2010) and heuristic approaches (Aleman et al. 2008; Bertsimas et al. 2013; Cao et al. 2012; Craft 2007; D'Souza et al. 2008; Lim and Cao 2012; Misic et al. 2010; Potrebko et al. 2007).

Artificial neural networks have been applied to the beam angle selection problem. In Rowbottom, et al. (1999), neural networks were trained with a single angle set for each patient. The angle set was generated using a heuristic that was previously shown to produce good angle sets (Rowbottom et al. 1998). By using only a single angle set for each patient, the information gained from the multiple sets that are typically evaluated during planning is lost. The authors used the neural network to generate an angle set given input parameters, instead of evaluating a beam angle set.

In the third stage, after the beam angles have been selected, beam intensity profiles are computed for each angle. The objective function for this stage ensures that the tumor receives the prescription dose, while minimizing the radiation to the surrounding tissue. There are many ways to model these objectives. Linear and quadratic forms are used to penalize delivering radiation to healthy tissue or delivering an insufficient dose of radiation to the treatment area. The penalty weights for the PTV and OARs are specified by the planners for a patient. The planners often adjust the penalty score equation as a way of modifying computer optimized plans without changing beam angles. From each angle, the intensity profile of the beam is divided into pixels or beamlets as variables to be optimized, which gives a flexible control of radiation dose delivered to a patient. This level of control is especially useful for concave and oddly shaped tumors. A comprehensive discussion of the IMRT planning problem can be found in four review papers (Bortfeld 2006; Hunt and Burman 2003; Romeijn and Dempsey 2008; Webb 2003).

In practice, for a given beam angle set, the beam intensity profiles are optimized using commercial software. Typically, the selection of beam angles is constructed manually and is based on the experience of the planners designing the plan. The optimization

of beam intensity profiles for a given set of beam angles is computationally expensive, taking up to 30 minutes per plan using the planning software. In current clinical practice, IMRT treatment planning is an iterative process. After the computationally expensive plan generation process is performed by clinicians, treatment plans are presented to physicians for approval. Frequently, physicians would like the clinicians to improve the plan quality by balancing the doses to different structures. This iterative process to achieve the final plan is even more time-consuming. The goal of this paper is to develop a model for sorting potential beam angle sets in order to identify those that are high quality, so that the number of plan evaluations can be limited during the computationally expensive third step of optimization. With the help of our method, top ranked plans that have different dose distributions to different OARs (but have similar overall quality) could provide physicians with more flexibility in plan selection and eliminate the time-consuming iterative process. A patient population-based method utilizing data mining tools is proposed in order to identify quality beam angles for a new incoming patient. Using the results of previously evaluated beam angle sets to aid in the selection of which sets to evaluate allows us to use the thousands of hours of computer processing that have already been spent optimizing plans for cancer patients. The purpose of this new tool is not to find the best of all possible plans, but to take advantage of prior knowledge to identify the most promising angle sets for full evaluation. It can be used along with any beam angle optimizers in order to speed up the process. In Section 2, we discuss the data used for the development of the data mining models. In Section 3, we fine tune the model parameters and present the results of each model. In Section 4, we provide concluding remarks and suggestions for future research.

## 2.2 Methods

Our population is built from 2709 evaluated treatment plans of 10 locally advanced head and neck cancer patients. Each plan has seven coplanar beam angles. Each patient has between 252 and 282 evaluated plans, 11 of which are rotations of seven evenly spaced angles rounded to the nearest five degrees. The remainder of the plans are randomly generated. The patients have different isocenters and tumor geometries, but all share the same set of OARs. The OARs includes the right and left parotid (salivary glands), oral mucosa, spinal cord, and brainstem. The treatment targets include primary PTV, high risk nodal volumes (high risk PTV), and low risk nodal volumes (low risk PTV).

### 2.2.1 Penalty Score to evaluate plan quality

The penalty score formula, which is widely used in the literature (Cao et al. 2012; Mistic et al. 2010; Romeijin and Dempsey 2012), is consistent across the 10 patients. Scores are weighted sums of violations of the desired dose values for both the OARs and the PTVs. A lower score corresponds to a better plan. Specifically, a score is given by

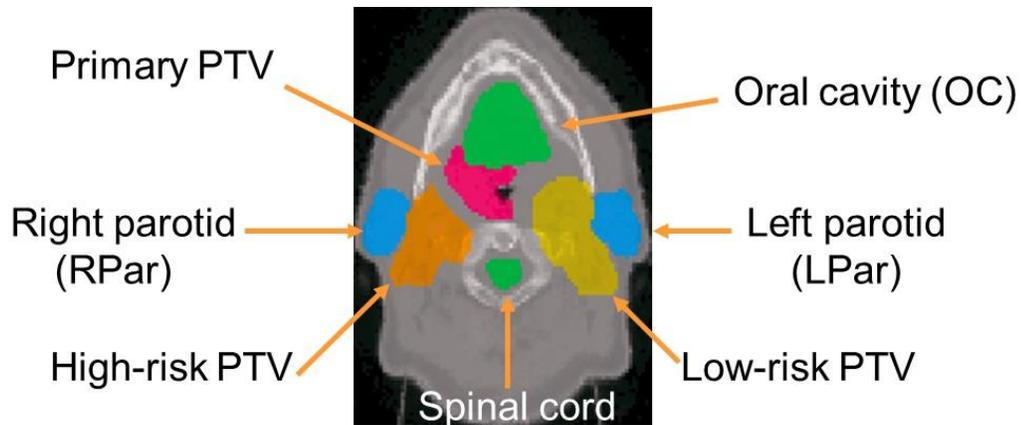
$$S = \sum_{i \in \text{OAR} \cup \text{PTV}} \alpha_i \max(A_i - d_i, 0) + \sum_{i \in \text{PTV}} \beta_i \max(d_i - A_i, 0),$$

where  $\alpha_i$  is the penalty associated with delivering excess radiation to OAR and PTV in the score calculation, which reflects the priority of satisfying the corresponding constraint of the OAR or the PTV during the planning process. Similarly,  $\beta_i$  is the penalty associated with failing to deliver sufficient radiation to the PTV. The specified desired dose value, measured in Grays (Gy), for each OAR and PTV ( $d_i$ ) is listed in Table 2-1 (note that multiple values can be specified for each OAR and PTV).  $A_i$  represents the actual dose achieved by the treatment plan at a percentage volume of the OAR or PTV. For the

OARs, the goal is to deliver as little radiation as possible. The most sensitive tissues, such as the spinal cord and brain stem, have a limit as to the maximum amount of absorbed dose that can be received by any part of the tissue. For these OARs, the  $d_i$  is measured in maximum Grays received. Before calculating the penalty score, the treatment plan is normalized so that the primary PTV receives the prescription dose (70 Gy) to ensure that enough dose is being delivered to the tumor. In Figure 2-1, we show a cross section of one patient demonstrating the difficulty of the cases considered. The three treatment targets were surrounded by OARs, and the spinal cord was sitting within the concavity of the targets.

Constraint	$d_i$	$\alpha_i, \beta_i$
Less than 66% of the left parotid receiving	26 Gy	3
Less than 33% of the left parotid receiving	32 Gy	3
Less than 66% of the right parotid receiving	26 Gy	3
Less than 33% of the right parotid receiving	32 Gy	3
Less than 90% of the oral mucosa receiving	30 Gy	8
Less than 30% of the oral mucosa receiving	40 Gy	8
Maximum spinal cord dose	45 Gy	15
Maximum brain stem dose	54 Gy	15
More than 95% of the low-risk PTV receiving	54 Gy	6
Less than 5% of the low-risk PTV receiving	59.4 Gy	6
More than 95% of the high-risk PTV receiving	59.4 Gy	6
Less than 5% of the high-risk PTV receiving	70 Gy	6

**Table 2-1:** Summary of dose and dose-volume constraints for locally advanced head and neck cases



**Figure 2-1:** A cross section of a CT for a patient depicting the PTV and OARs.

### 2.2.2 Angle Set Evaluation

The associated beam intensity profiles were optimized using P<sup>3</sup> IMRT, part of the Pinnacle<sup>3</sup>® Planning system by Phillips (2013). P<sup>3</sup> IMRT optimizes the dosage to the specified OAR and PTV using a sequential quadratic programming algorithm for solving general nonlinear optimization problems. There are 1,000 to 10,000 variables, and the process of optimizing the intensity profiles for a set of beam angles can take between 5 and 30 minutes. Each angle set had seven coplanar angles which is consistent with previous research (Bortfeld and Schlegel 1993; Liu et al. 2006; Takamiya et al. 2007) that has explored the trade-offs involved in using more beam angles. Also, in accordance with practice, no two beam angles were spaced less than 30 degrees apart or opposite between 170 and 190 degrees of another.

### 2.2.3 Models

Three different methods were compared: linear regression, neural networks (Haykin 1994), and *k*-nearest neighbor (Aha 1991). Each method used fully evaluated

beam angle sets with penalty score as observations, and was tested by predicting the penalty score of beam angle sets on which it had not been trained.

Linear regression allows for an interpretable model with no predetermined model parameters. The linear regression models were constructed using the R software package (R Core Team 2013). Both the linear regression and neural network models used the patient ID, normalized angle sets, normalized product of each pair of angles, and normalized cube of each angle as feature sets. Patient ID identifies the patient whose penalty score was evaluated from the beam angle set. The patient ID allows for patient fixed effects to be captured using nine dummy variables that allow for the identification of the 10 patients.

Neural networks require several model parameters to be specified prior to training, including activation functions, network topology, and a learning algorithm. A hyperbolic tangent activation function (a common activation function) is used for all nodes in the network. A neural network has an input layer, hidden layer, and output layer. We have to decide on the number of hidden layers and the number of neurons in the hidden layers. Too many neurons in the hidden layer will result in overfitting and long computation times. Too few neurons will result in underfitting. The network has one hidden layer with a number of nodes equal to one more than half the size of the feature set. The hidden layer is fully connected to both the input and output layers. Backpropagation is the learning algorithm. There are several parameters associated with backpropagation including a learning parameter and momentum. We set the learning parameter value at 0.01 and momentum at 0.90. MATLAB release 2012b and the Neural Networks Toolbox (MATLAB 2012) were used to construct and test the neural network models.

Finally,  $k$ -nearest neighbor is an instance-based learner that uses the Euclidean distance between the angle set being evaluated and the angle sets in the training set to determine the  $k$  nearest instances. The penalty score is then estimated, giving more weight to the penalty scores of the closer instances. The  $k$ -nearest neighbor algorithm requires no training time. Weka 3.6 (Hall et al. 2009) was used to run the  $k$ -nearest neighbor algorithm with  $k$  set to seven (based on predictive performance on test set) and the distance weight set to the inverse of the Euclidean distance between angle sets. The  $k$ -nearest neighbor algorithm used only the patient ID and normalized angle set as features.

### **2.3 Results**

Models were generated using 5-fold cross-validation on the 2,709 beam angle sets. Four folds are used as training data, with the fifth fold being withheld for testing. The data mining algorithms were trained to predict the penalty score for each beam angle set. For each patient, the plans were then sorted by their predicted penalty score and this ranking was compared to the actual ordering of the plans. For each 5-fold cross-validation, this process results in 50 ordered lists (5 folds each with 10 patients). Each list has between 50 and 56 beam angle plans for a specific patient. We used 5-fold cross-validation rather than the more common 10-fold cross-validation to allow for a larger test set. This results in an average list of 54 plans for each of the 10 patients to be sorted (10-fold cross-validation would have an average list size of only 27 plans). By using fewer folds, we double the size of the test set for each run, but reduce the size of the training set by less than 12%.

The model predictions correlate well with the actual penalty scores. The correlation coefficients are 0.85 for the linear regression model, 0.83 for the neural network model,

and 0.90 for the  $k$ -nearest neighbor model. To estimate time savings and performance improvements, however, we need to know how well the models predict the order of the lists they are sorting. In Table 2-2, we see that all the data mining methods outperform the baseline of taking the same number of elements from an unsorted list.

The probability of finding the best of the potential plans in a list is more than doubled at all levels by using simple machine learning methods. Evaluating the top three ranked plans by  $k$ -nearest neighbor produces better results, on average, than evaluating 15 plans from an unsorted list. If we were to evaluate only six plans, the average rank of the best plan found using a  $k$ -nearest neighbor sorted list is less than a quarter of that found by randomly evaluating six plans. The best plan is found 54% of the time, almost five times as often when evaluating six random plans. A top three plan is found 86% of the time (almost three times as often as evaluating six random plans) and a top five plan is found 94% of the time (over twice as often).

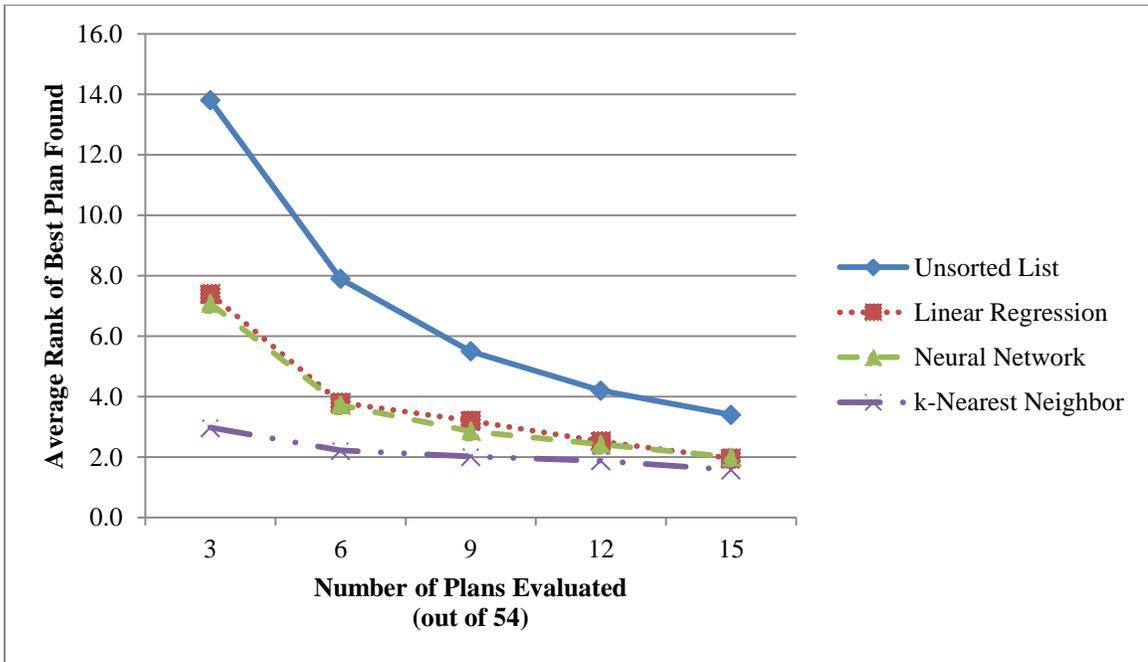
Plans searched from machine sorted list	Average rank of best plan found (out of 54)	Probability of finding best plan	Probability of finding a top 3 plan	Probability of finding a top 5 plan	Rank of bottom 10th percentile of best plan found (out of 54)
Baseline					
Random 3	13.8	5.6	16.0	25.7	30
Random 6	7.9	11.1	30.3	45.9	18
Random 9	5.5	16.7	42.8	61.4	13
Random 12	4.2	22.2	53.7	73.1	10
Random 15	3.4	27.8	63.2	81.8	8
Linear Regression					
Top 3	7.4	14	34	46	17
Top 6	3.8	22	56	80	8
Top 9	3.2	26	62	86	6
Top 12	2.5	38	74	94	4
Top 15	2.0	50	88	98	4
Neural Network					
Top 3	7.1	16	40	52	15
Top 6	3.7	32	64	84	10
Top 9	2.9	38	74	90	5
Top 12	2.4	50	78	94	5
Top 15	2.0	56	86	98	4
<i>k</i> -Nearest Neighbor					
Top 3	3.0	34	76	92	5
Top 6	2.2	54	86	94	4
Top 9	2.0	62	90	94	3
Top 12	1.9	68	92	94	3
Top 15	1.6	70	96	98	2

**Table 2-2:** Performance of linear regression, neural network, and *k*-nearest neighbor algorithms over a 5-fold cross-validation resulting in 50 ordered lists with an average of 54 elements per list. This is compared with the baseline performance of a randomly ordered list of 54 elements. All rankings are ordinal; each has a distinct integer rank (no ties).

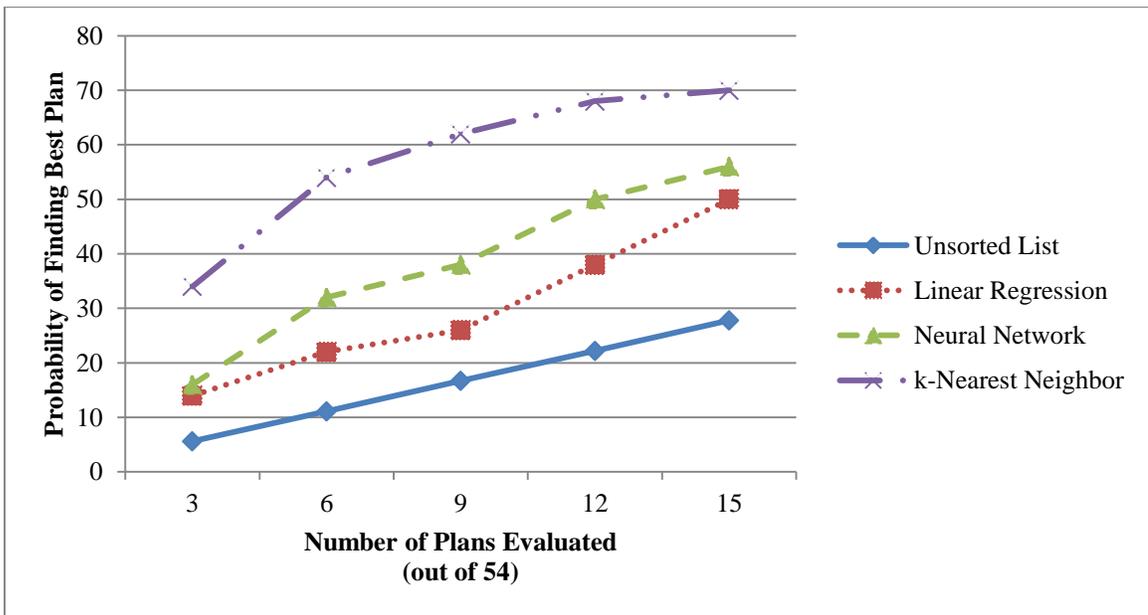
In Figures 2-2 to 2-5, we see that  $k$ -nearest neighbor strictly dominates both linear regression and neural network models, both of which dominate a randomly sorted list. However,  $k$ -nearest neighbor is more sensitive to the number of dimensions used (Beyer et al. 1999). If the dimensionality is too high, the performance declines. This property might limit the ability of  $k$ -nearest neighbor models to explicitly handle large numbers of patient variables found in larger databases (our models use dummy variables for each patient). Neural network and linear regression models both can effectively use a larger feature set than the  $k$ -nearest neighbor, so long as the number of instances is sufficiently large.

In Figure 2-5, we see that, by searching just six plans using  $k$ -nearest neighbor, on average, the penalty score of the best plan found is less than five percent above that for the best plan in the list. This result compared to a penalty score over 20% above that for the best plan using an unsorted list. Even searching 15 plans with an unsorted list still yields a penalty score that is over 10% above that for the best plan in the list. Given that the ultimate goal of any sorting or optimization technique is to improve the quality of the plans being used, we see that, by sorting the list prior to evaluation, we can evaluate far fewer plans and still find high-quality plans.

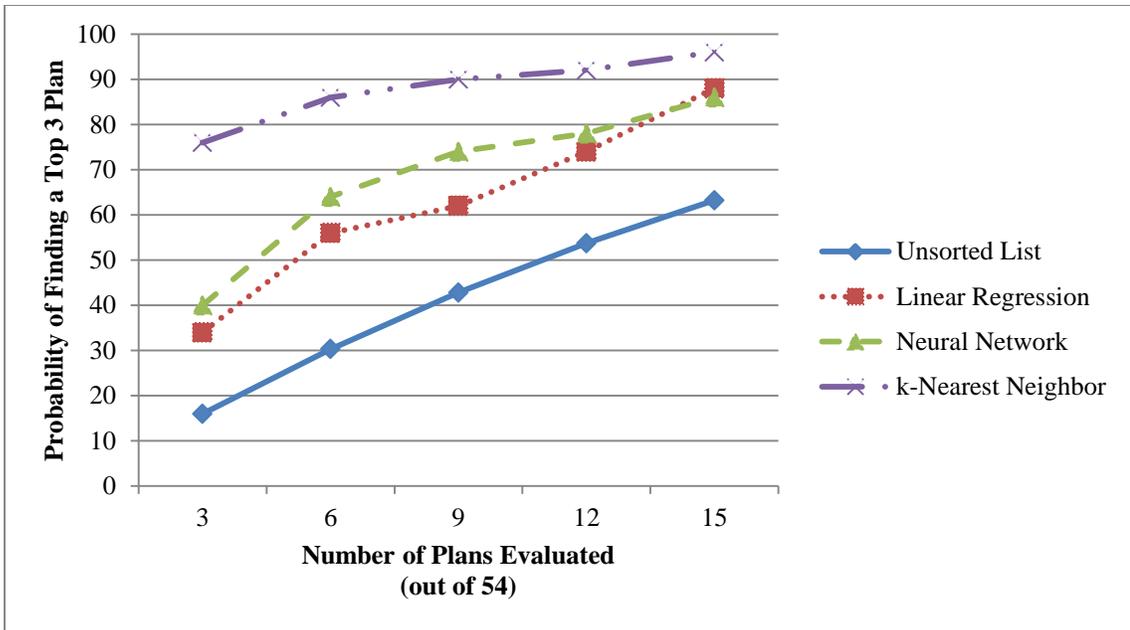
The penalty score of the top ranked plan from each of the 50 lists of plans averages between 4% and 5% less than the plan used in clinical experience for both linear regression and neural networks. The  $k$ -nearest neighbor algorithm achieves a 17% reduction. Similarly, when comparing the penalty score of the average top ranked plan in each list with the results of commercial software, we find an 11% reduction for the linear regression and neural network models, and a 22% reduction for  $k$ -nearest neighbor model.



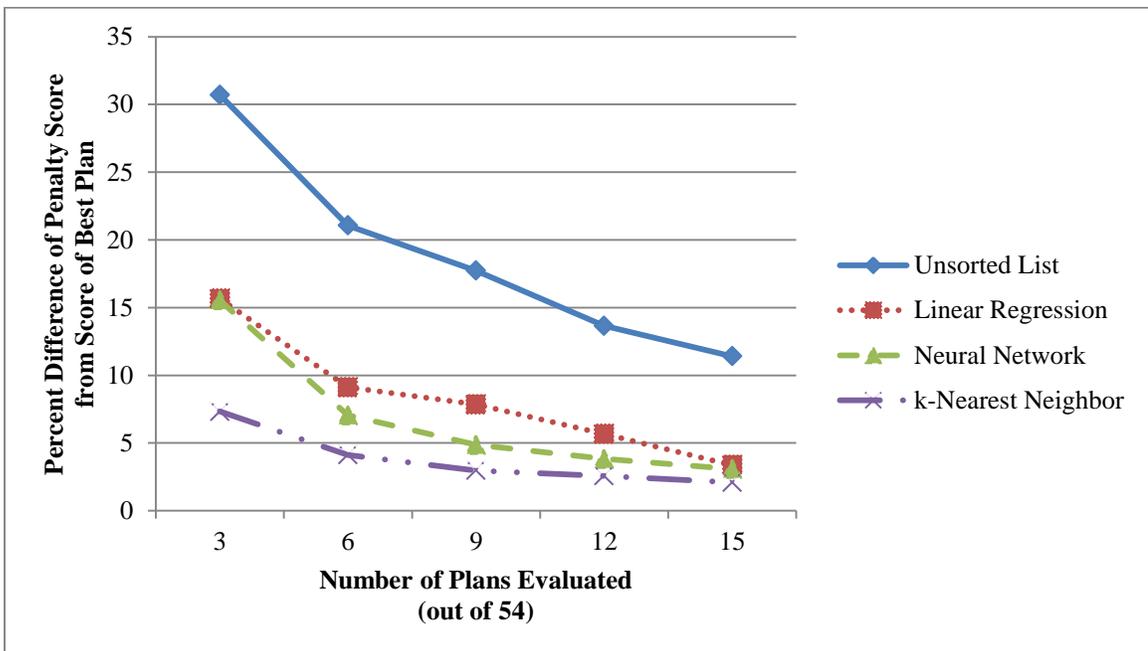
**Figure 2-2:** The average rank of the best plan found by evaluating a fixed number of plans from a list of 54 elements.



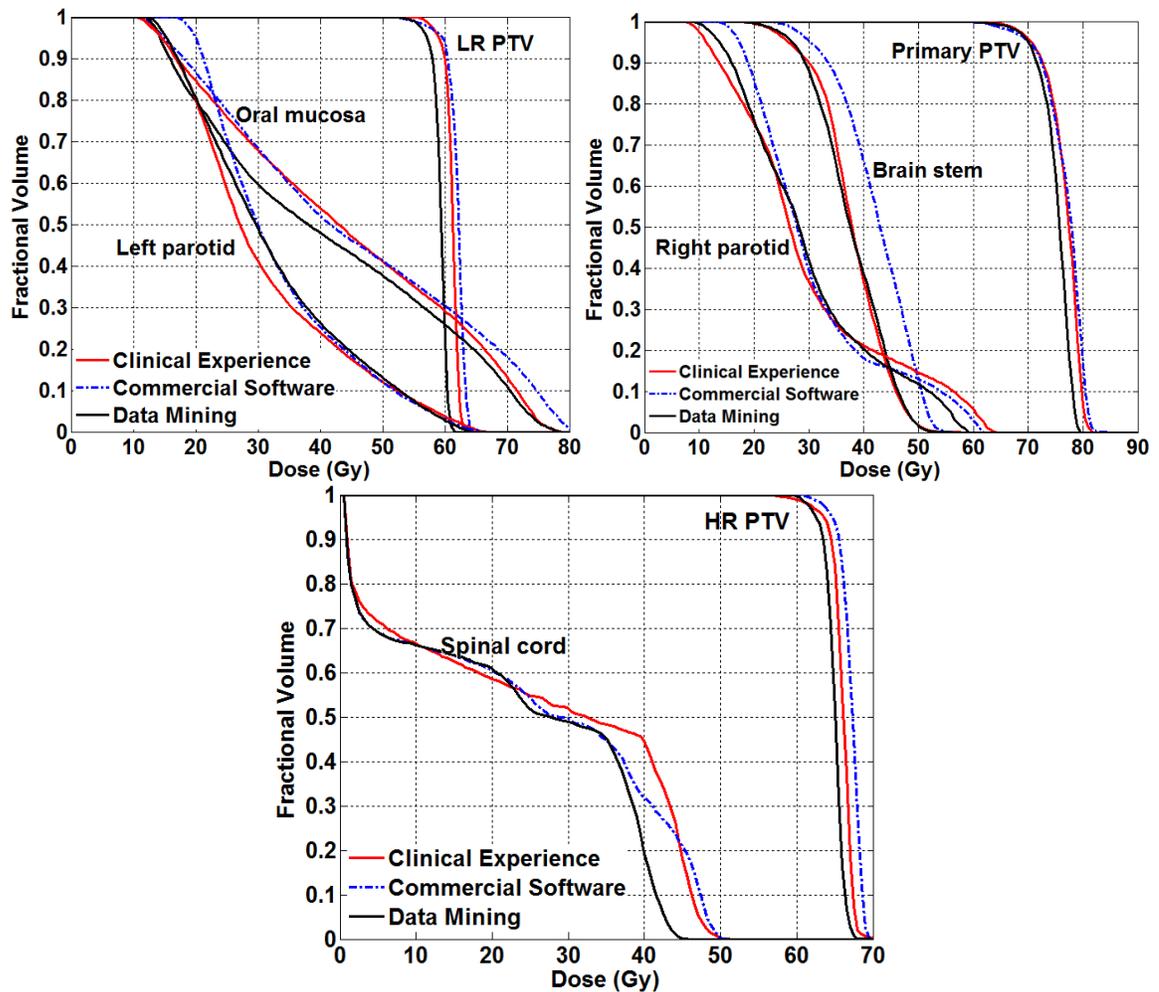
**Figure 2-3:** The probability of finding the best plan in a list of 54 elements by evaluating a fixed number of plans.



**Figure 2-4:** The probability of finding one of the three best plans in a list of 54 elements by evaluating a fixed number of plans.



**Figure 2-5:** The percent difference between the penalty score of the actual best plan of a list of 54 and the best plan found having evaluated a fixed number.



**Figure 2-6:** A comparison of the DVHs for one patient comparing plans based on clinical experience (red), commercial software (blue), and a sample run using the top ranked plan from a list sorted using linear regression. The OARs and PTVs are separated for presentation purposes.

The overall beam angle quality obtained using either clinical experience, commercial software, or the top plan from a list sorted using data mining is demonstrated using the dose-volume histograms (DVHs) in Figure 2-6. The DVH presents the fractional volume of each OAR or PTV that receives at least a certain absorbed dose (in Gy) of radiation. All three approaches provided the required coverage for PTVs (primary, high risk (HR) and low risk (LR)). However, the data mining approach obtained the most uniform

dose for all three PTVs. The bigger differences were achieved for spinal cord, oral mucosa, and brain stem. At least a 10% improvement (reduction in maximum dose) was achieved by the data mining approach for sparing of the spinal cord compared to the other methods. A 15% improvement of oral mucosa sparing was achieved as well. We can see that data mining successfully identified beam angle sets that yielded significantly improved treatment plans relative to clinical plans and commercially implemented optimization techniques. Appendix 2-1 gives an example of an alternate plan (with slight changes to the beam angles) developed using  $k$ -nearest neighbor that improves upon the plan based on clinical experience with respect to penalty score.

In Table 2-3, we compare the performance of the three data mining algorithms (linear regression, neural network,  $k$ -nearest neighbor) to the plans produced by clinical experience, the Eclipse treatment planning system (Varian 2013), and a nested partitions optimization algorithm (D'Souza et al. 2008). The optimization algorithm has been benchmarked against other methods found in the literature (Lim et al. 2012). The achieved dose (in Gy) at each constraint setting level is listed in the table, which were averaged for the 10 patients. The plans for the data mining methods are based on the best of the top three ranked plans in the fold, from 5-fold cross validation, where the test set contained the best-known plan for that patient. The linear regression model reduced the penalty from 90% of the oral mucosa receiving more than 30 Gy by 12% compared to clinical and more than 15% compared to the Eclipse system. Each data mining method reduced the delivery to the brain stem compared to the clinical and Eclipse plans, reducing the penalty by 6% to 9%. The top-ranked plans from our data mining methods performed well compared to angle sets from a good angle selection optimizer. The plan from

Constraint	Clinical Experience	Eclipse System	Linear Regression	Neural Network	<i>k</i> -Nearest Neighbor	Nested Partitions
More than 95% of the high-risk PTV receiving 59.4 Gy	61.6	61.6	60.4	61.1	60.7	60.6
Less than 5% of the high-risk PTV receiving 70.0 Gy	69.1	69.3	68.0	68.2	68.0	68.6
More than 95% of the low-risk PTV receiving 54.0 Gy	56.5	56.2	55.0	55.4	54.9	55.2
Less than 5% of the low-risk PTV receiving 59.4 Gy	62.4	62.8	61.5	61.7	61.4	62.1
Maximum spinal cord dose 45.0 Gy	49.3	51.0	48.7	48.1	47.9	48.5
Less than 66% of the left parotid receiving 26.0 Gy	17.0	18.2	17.4	17.8	17.4	17.5
Less than 33% of the left parotid receiving 32.0 Gy	42.2	42.1	40.0	42.0	40.0	41.1
Less than 66% of the right parotid receiving 26.0 Gy	18.6	18.7	18.4	18.3	18.3	19.0
Less than 33% of the right parotid receiving 32.0 Gy	44.4	44.6	43.8	44.2	44.1	44.0
Maximum brain stem dose 54.0 Gy	38.2	38.1	35.1	34.6	35.8	36.3
Less than 90% of the oral mucosa receiving 30.0 Gy	17.5	18.2	15.4	17.0	16.9	16.5
Less than 30% of the oral mucosa receiving 40.0 Gy	45.7	45.2	43.1	43.6	43.0	43.6

**Table 2-3:** The actual dose (in Gy, averaged over the 10 patients) achieved at each constraint setting level for plans produced by clinical experience, the Eclipse system, linear regression, neural network, *k*-nearest neighbor, and nested partitions.

the optimizer was consistently ranked among the top five plans of at least one data mining method.

As the length of the list being sorted grows, we cannot say for certain how performance will scale. Due to the approximate limit of 270 plans per patient that we have to work with, we cannot test the performance on lists of arbitrary lengths. Once trained, both linear regression and neural network models can be run quickly to evaluate many potential treatment plans. To evaluate each new plan simply requires a few matrix multiplications. The  $k$ -nearest neighbor algorithm requires no training time, which would be an advantage when selecting a training set that most closely resembles a patient.

## **2.4 Discussion**

This research has the potential to use the vast amount of information that is gathered during the IMRT planning process. Every plan that was evaluated for the thousands of patients undergoing treatment each year could potentially be used. As the population database grows, it will become possible to include more patient specific variables such as tumor geometry and location, information that models trained on a small number of patients cannot utilize.

Alternatively, a subset of patients with similar conditions could be selected as the training set for each new patient, thereby eliminating the need of adding new variables to the model. A repository of tumor shapes and anatomies of patients for each disease site, along with all plans calculated during a patient's treatment planning phase, could serve as the training set for new patients with the same disease site. Although there are databases of medical imaging of tumors that are currently available, they do not include treatment information. When a new patient considers treatment, we could, with our tools, identify

potential high-quality beam angle sets. Treatment information could easily be attached to such a database and previous tumors could be simulated for treatment plans as well.

Research (Dink et al. 2012) has shown that using multiple beam angle sets over a series of treatments has the potential to improve outcomes. Unlike previous work that focused on finding a single best angle set, by sorting, we can find multiple good angle sets; these would allow for a greater distribution of angles to be used with time-varying plans. Appendix 2-2 gives an example of two plans (obtained from linear regression) with disparate angle sets that have a similar penalty score and spare different OARs. This also gives the physician more flexibility to choose a treatment plan based on the trade-offs between the OARs.

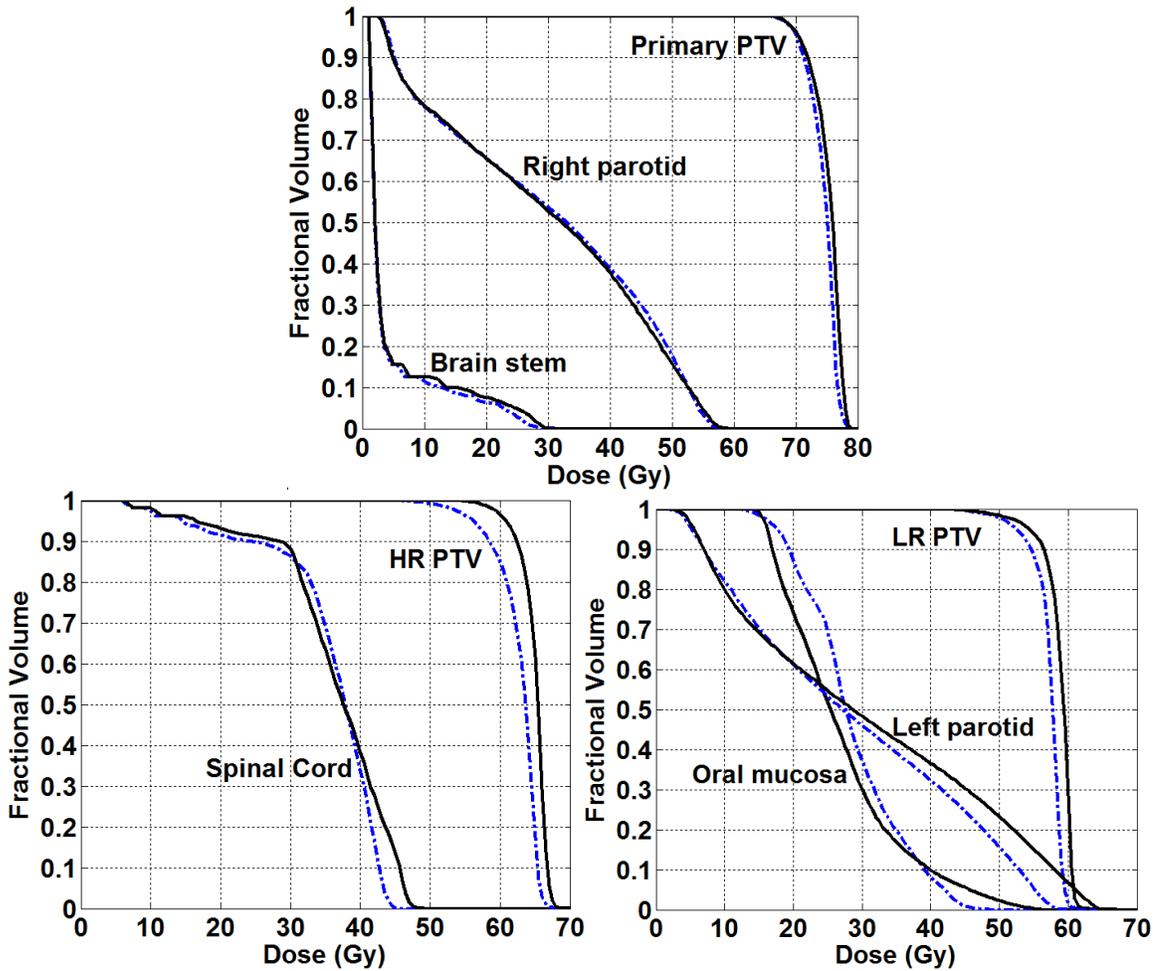
Our research has similar implications for use with adaptive treatment planning. Current treatment plans are static throughout the treatment process. Adaptive treatment seeks to anticipate response to treatment and adjust accordingly. Given that the tumor will likely change slowly over the course of the treatment, plans developed for preceding stages could serve as training instances for screening future sets and distinct angle sets could be found to better distribute the radiation.

In future work, we will incorporate patient specific anatomical features (Chanyanavich et al. 2011; Magome et al. 2013) to improve the prediction of the penalty score and ranking accuracy. After improving the accuracy, we will incorporate the data mining models within an optimization method to identify the optimal angle set with limited quality evaluations.

In conclusion, we have shown that previously evaluated beam angle sets can be used to help identify the potential of new angle sets. Sorting plans shows great promise in

speeding the search for good angle sets by helping find plans with low penalty scores first. Future models might recommend how many plans should be evaluated based on the relative performance of the new plans to the existing training plans. This would allow for fewer plans to be evaluated if their predicted penalty score was higher than the plans that have already been evaluated.

## Appendix 2-1

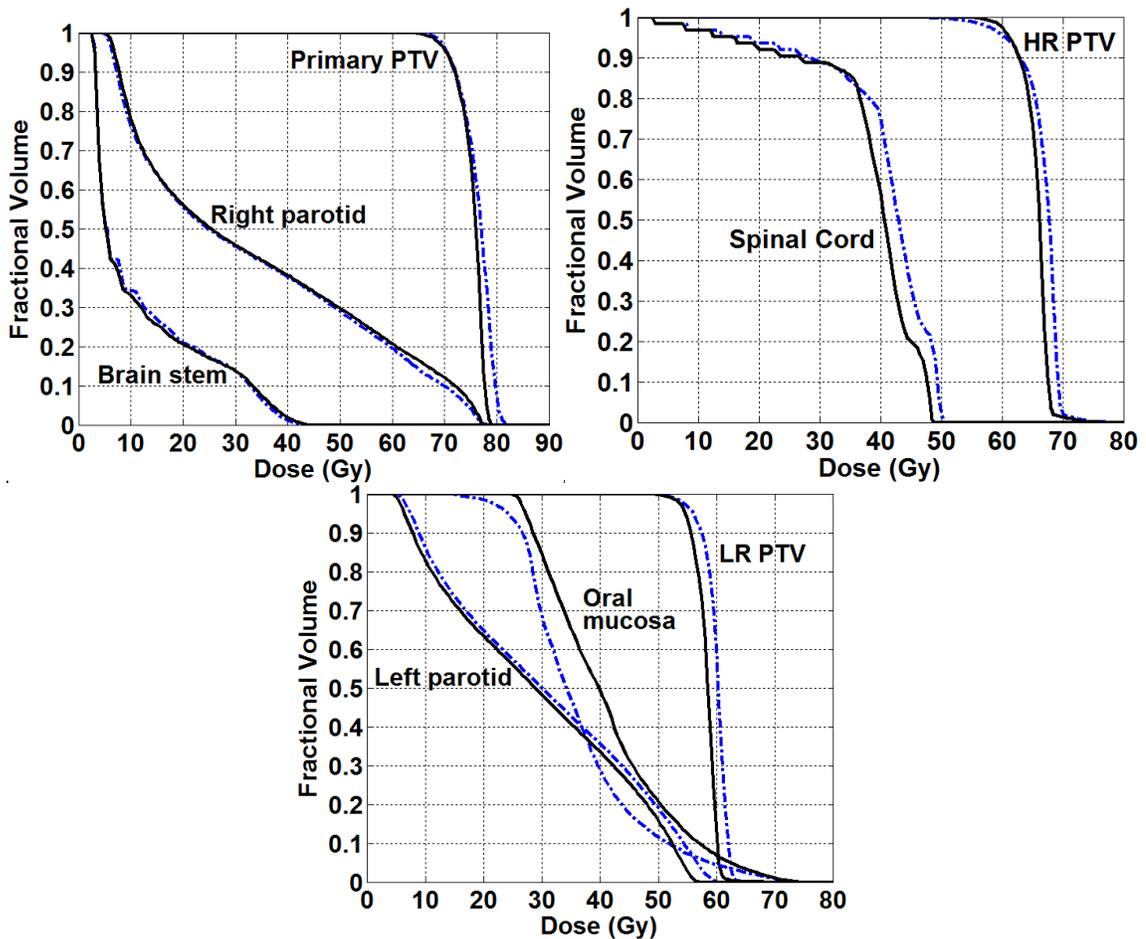


**Figure 2-7:** A comparison of the DVHs for one patient comparing plans based on clinical experience (black) and a sample run using the top ranked plan from a list sorted using  $k$ -nearest neighbor (blue). The OARs and PTVs are separated for presentation purposes.

The  $k$ -nearest neighbor model was able to identify a plan similar to the one used in clinical practice, that differed by more than five degrees on only two angles, that improved the penalty score by 24%. In Figure 2-7, we see that while the clinical plan delivers less radiation to the oral mucosa, the  $k$ -nearest neighbor plan reduces radiation delivered to the left parotid and delivers a high dose to a smaller fraction of the brain stem. The plans have similar characteristics for the brain stem, right parotid, and primary PTV.

## Appendix 2-2

The top two plans from a list sorted using a linear regression model were compared. The plans penalty score differed by less than 1%, but four of the angles used differed by at least 15 degrees. Figure 2-8 shows similar dosing for the brain stem, left and right parotid, and PTV. One plan spares the spinal cord, the other spares the oral mucosa. Having multiple plans with the same penalty score allows for plans to be compared using objectives not included in the penalty score.



**Figure 2-8:** A comparison of the DVHs for one patient comparing two top ranked plans found using linear regression. The OARs and PTVs are separated for presentation purposes.

### **3 OPTIMIZING INTENSITY-MODULATED RADIATION THERAPY ANGLE SELECTION USING GENETIC ALGORITHMS**

#### **ACKNOWLEDGEMENTS:**

This is research conducted with:

Bruce Golden, University of Maryland

Edward Wasil, American University

Howard H. Zhang, University of Maryland School of Medicine

**Abstract:** Developing treatment plans for intensity-modulated radiation therapy (IMRT) involves first selecting beam angles and then optimizing the delivery of radiation from those angles. Optimizing the delivery of radiation for a set of beam angles is a time intensive process and finding the beam angles from which to best deliver the radiation is an NP-hard problem. We use machine learning to construct models to quickly estimate the quality of optimized plans for new angle sets. These models are used as objective/fitness functions in a genetic algorithm (GA) that searches the space of feasible beam angle sets. Our GA quickly finds solutions for new patients based on the models trained on previously evaluated treatment plans for other patients. The solutions found are more than 2% better than the previously best known solutions for patients.

### **3.1 Introduction**

Intensity-modulated radiation therapy (IMRT) is a form of external beam radiation therapy that uses x-rays to irradiate a tumor. Since the radiation also affects the healthy tissue surrounding the tumor, the x-rays are projected from multiple angles with the tumor at the center. A linear accelerator is used to generate the x-rays and is mounted on a gantry arm that can be rotated to deliver the radiation from multiple angles. Using multiple angles allows the radiation to be delivered in multiple doses, distributing the dose to healthy tissue to minimize damage, while delivering a concentrated lethal dose to cancerous tissue. A multi-leaf collimator is used to control the precise delivery of radiation. The collimator can be moved while the radiation is being delivered, blocking the x-rays for a period of time to control the exposure. This further minimizes damage to healthy tissue, by controlling the amount of radiation being delivered to each pixel from each angle to ensure lower dosage to areas with more sensitive tissue.

Constructing a treatment plan for a patient is a three stage process. In the first stage, the patient is imaged so that all important tissue in the treatment area can be identified. An oncologist identifies the tumor and tissue at risk of cancer, the planned treatment volume (PTV), and prescribes radiation levels for these areas that will kill the cancerous tissue. Next, the organs at risk (OAR) of damage from radiation are assigned thresholds of radiation, below which, damage from radiation is considered acceptable. Plans are evaluated using an objective function that penalizes plans for delivering either too little radiation to the PTV or too much radiation to the OAR.

The second stage involves selecting angles from which to deliver the radiation. Using multiple angles allows for the radiation to be distributed and these angles can be chosen to avoid delivering radiation to more sensitive tissue. The beam angle selection problem has been shown to be NP-hard (Saher and Sultan 2010). Using more angles has diminishing returns and leads to longer treatment times. The selection of angles for treatment is still primarily done manually by trial and error.

The third stage of the planning process is to decide the precise way in which to deliver the radiation from each angle. The PTV creates a cross section when viewed from each angle. The cross section is divided into pixels. Using multi-leaf collimators, the delivery of radiation can be precisely controlled to each pixel of the cross section. Given the large number of pixels and the interaction between their deliveries, the optimization of this planning phase is done using commercial optimization software. The software can take up to 20 minutes to optimize the delivery of radiation for an angle set. For a more in depth discussion of IMRT and the treatment planning process, literature reviews are available (Bortfeld 2006; Hunt and Burman 2003; Romeijn and Dempsey 2008; Webb 2003).

IMRT is currently the most popular form of external beam radiation therapy in the United States. Each year thousands of treatment plans are developed for patients undergoing IMRT. These treatment plans have the potential to form the basis of a database that can be used to guide future treatment plan construction. Previous work (Price et al. 2014) has shown that using machine learning to sort potential beam angle sets could aid treatment planners in deciding which angle sets to examine. However, sorting or evaluating all possible angle sets is infeasible. An efficient way to search the space of angle sets is

needed in order to generate good solutions. Dias et al. (2014) use neural networks trained for individual patients to serve as the fitness function for a beam angle search procedure, but such a method requires that new neural networks be trained for each patient and multiple beam angle sets be evaluated for these patients to train the neural networks. Our work aims to use a genetic algorithm (GA) to search the space of potential beam angle sets for the set of angles, which when radiation delivery from these angles has been optimized, will minimize the objective function value, thus minimizing damage to OAR and delivering an appropriate level of radiation to the PTV. Instead of optimizing the radiation delivery for each angle set, we use machine learning trained on previous patients to develop models that predict the objective function value of optimized beam angle sets. Our GA is able to improve upon the best known solutions by more than 2%.

The paper is organized as follows. Section 2 explores the methods used by the machine learning models and Section 3 describes the details of the GA. Section 4 presents the results of the plans identified by the GA using the different fitness functions from the machine learning algorithms. Section 5 provides conclusions.

### **3.2 Methodology**

This section covers the steps taken in constructing the GA. First, we discuss the penalty function used to optimize the plans. Our work is applied to head and neck cancer patients and a customized penalty function is used. Next, we examine the data used to train the machine learning algorithms. For this data set, the beam angle sets are evaluated for patients using commercial software. This data is then used to make models to predict the penalty score using machine learning. Finally, the machine learning models are part of a genetic algorithm used to search the space of beam angles.

### 3.2.1 Penalty Function

We use a linear penalty score formula commonly used in the literature (Cao et al. 2012; Misic et al. 2010; Romeijn and Dempsey 2012). The penalty score evaluates the relative quality of the treatment plans based on how they deviate from a series of OAR and PTV target values. The penalty score is a weighted sum of violations from the desired dose values for both the OARs and the PTVs. Better plans have fewer violations resulting in lower penalty scores. The score is given by

$$S = \sum_{i \in \text{OAR} \cup \text{PTV}} \alpha_i \max(A_i - d_i, 0) + \sum_{i \in \text{PTV}} \beta_i \max(d_i - A_i, 0),$$

where  $\alpha_i$  and  $\beta_i$  are the weights associated with each OAR and PTV in the score calculation. The weights reflect the seriousness of violating the constraint of the OAR or the PTV during the planning process, with more important structures receiving higher weights. Table 3-1 lists the desired dose values ( $d_i$ ), measured in Grays (Gy), for each OAR and PTV. OARs more sensitive to radiation will have lower desired dose levels, while the PTV at highest risk will have larger desired dose levels. There can also be multiple constraints each OAR or PTV.  $A_i$  represents the actual dose achieved by the treatment plan at a percentage volume of the OAR or PTV. For the OARs, the goal is to deliver as little radiation as possible. The most sensitive tissues, such as the spinal cord and brain stem, have a limit on the maximum amount of absorbed dose that can be received. For these OARs, the  $d_i$  is measured in maximum Grays received. The treatment plan is normalized so that the primary PTV receives the prescription dose (70 Gy) prior to calculating the penalty score to ensure that an adequate dose is being delivered to the tumor.

Constraint	$d_i$	$\alpha_i, \beta_i$
Less than 66% of the left parotid receiving	26 Gy	3
Less than 33% of the left parotid receiving	32 Gy	3
Less than 66% of the right parotid receiving	26 Gy	3
Less than 33% of the right parotid receiving	32 Gy	3
Less than 90% of the oral mucosa receiving	30 Gy	8
Less than 30% of the oral mucosa receiving	40 Gy	8
Maximum spinal cord dose	45 Gy	15
Maximum brain stem dose	54 Gy	15
More than 95% of the low-risk PTV receiving	54 Gy	6
Less than 5% of the low-risk PTV receiving	59.4 Gy	6
More than 95% of the high-risk PTV receiving	59.4 Gy	6
Less than 5% of the high-risk PTV receiving	70 Gy	6

**Table 3-1:** Summary of dose and dose-volume constraints for locally advanced head and neck cases

### 3.2.2 Data

The associated beam intensity profiles were optimized using P<sup>3</sup> IMRT, part of the Pinnacle<sup>3</sup>® Planning system by Phillips (2013). P<sup>3</sup> IMRT optimizes the dosage to the specified OAR and PTV using a sequential quadratic programming algorithm for solving general nonlinear optimization problems. There are 1,000 to 10,000 variables, and the process of optimizing the intensity profiles for a set of beam angles can take between 5 and 30 minutes. Each angle set had seven coplanar angles which is consistent with previous research (Bortfeld and Schlegel 1993; Liu et al. 2006; Takamiya et al. 2007) that has explored the trade-offs involved in using more beam angles. Also, in accordance with practice, no two beam angles were spaced less than 30 degrees apart or opposite (between 170 and 190 degrees of one another).

We use two populations to train our machine learning algorithms. The first population is a set of treatment plans that deliver radiation from five beam angles. The set

consists of 1,740 treatment plans evaluated for 10 locally advanced head and neck cancer patients. A total of 174 plans are evaluated for each patient with 160 of them being randomly generated and 14 representing equally spaced angle sets. The average penalty score of the plans is 303, while the average of the best plan known for each patient is 179. Five angle plans take less time to deliver the radiation than seven angle plans. This allows for faster treatment times for patients, meaning less time immobilized for their daily treatments.

Our second population is built from 2,709 evaluated treatment plans of 10 locally advanced head and neck cancer patients. Each plan has seven coplanar beam angles. Each patient has between 252 and 282 evaluated plans, 11 of which are rotations of seven evenly spaced angles rounded to the nearest five degrees. The remainder of the plans are randomly generated. The average penalty score of the plans is 266, while the average of the best plan known for each patient is 163. These scores reflect the fact that seven beam plans can better spread the radiation across healthy tissue, allowing for better penalty scores than the five beam plans, about 10% on average. The patients have different isocenters and tumor geometries, but all share the same set of OARs. The OARs includes the right and left parotid (salivary glands), oral mucosa, spinal cord, and brainstem. The treatment targets include primary PTV, high risk nodal volumes (high risk PTV), and low risk nodal volumes (low risk PTV).

### **3.2.3 Machine Learning**

The penalty function for evaluating plan quality is based on the levels of radiation being delivered to the OAR and PTV of the patient. The radiation being delivered is based on the optimal plan arrived at by P<sup>3</sup> IMRT planning software, which was optimized

over a set of beam angles. It is reasonable to assume that arbitrarily small changes in beam angle will bound the magnitude of the changes in the penalty function. We assume that the penalty score of the optimized treatment plan for a patient is a continuous function of the beam angles used in the plan. To estimate the shape of this function, we use machine learning and regression methods based on previously optimized beam angle sets for each patient. To begin, we use a variety of basic techniques and then use combinations of these techniques to improve performance. Techniques used include regression, neural networks, radial basis functions, and k-nearest neighbors.

Assuming that the penalty function is a continuous function of the beam angles implies that the function can be approximated with any desired level of accuracy by a sufficiently high order polynomial. Unfortunately, the number of coefficients that needs to be estimated for higher dimension polynomials proliferate quickly with an increase in the dimensionality. Starting with a feature set that includes the beam angles, first order interaction terms, the square of beam angles, and the cubes, stepwise regression is used to select features with high predictive value. Four stepwise regressions are run; forward regression starting with no features and adding those with high predictive value, backward stepwise regression starting with all features and removing those with little predictive value, and stepwise regression with both forward and backward movements beginning from the empty and full feature sets.

Neural networks are a form of machine learning that can be trained to identify complex, non-linear patterns. A neural network is defined by its topology, activation function, and learning algorithm. We use a three layer system composed of input, hidden, and output layers. Each layer is fully connected with the next. An inverse arctangent activation

function is used, having been shown to work well with the resilient propagation (RPROP) learning algorithm (Riedmiller 1993). RPROP is an adaptive form of back-propagation that adjusts learning parameters in response to the error function during the training phase. After it has been trained, a neural network can be run quite quickly.

Radial basis function networks are a special type of neural networks that use radial basis functions as activation functions. The basis function centers are chosen as a random sample of the training set. Radial basis function networks are a popular method for function interpolation.

Finally,  $k$ -nearest neighbor is an instance based method that requires no training. The  $k$  closest neighbors in the training set to a new instance are found and a weighted sum of their penalty score is used to predict the penalty score of the new instance. While no training is required, it requires that the distance to each observation in the training set is calculated for each new instance. Based on the performance of using different numbers of neighbors, we use seven neighbors weighted by the inverse distance to the new instance. Nearest neighbor is a popular method for multidimensional interpolation, but can be slow as the training set grows.

A shortcoming of using an interpolation method like  $k$ -Nearest Neighbor ( $k$ NN) to predict the value of a new beam angle set is that the predicted value of any beam angle set is bounded by the maximum and minimum values in the training set. Therefore, when using  $k$ NN to search the space of all beam angle sets, the predicted optimum would be an extreme point from the training set. Extrapolation methods, on the other hand, are vulnerable to extreme predictions, well outside anything in the training set, especially when searching a large space. For example, regressions with polynomial terms of second or

third order can have extreme values at the boundary of the space. This can result in predicted values which are several standard deviations from anything seen in the training set.

To balance the shortcomings of the two methods, we propose a method of combining predictions that gives greater weight to interpolation methods when extrapolation methods have values many standard deviations from the norm of the training set. Let  $\mu$  be the mean of the training set and  $\sigma$  be the standard deviation. Letting  $\Phi(x)$  represent the inverse function of the normal distribution. Then for  $k(a)$  being the kNN prediction for an angle set,  $a$ , and  $E(a)$  being the extrapolation prediction for the same angle set then

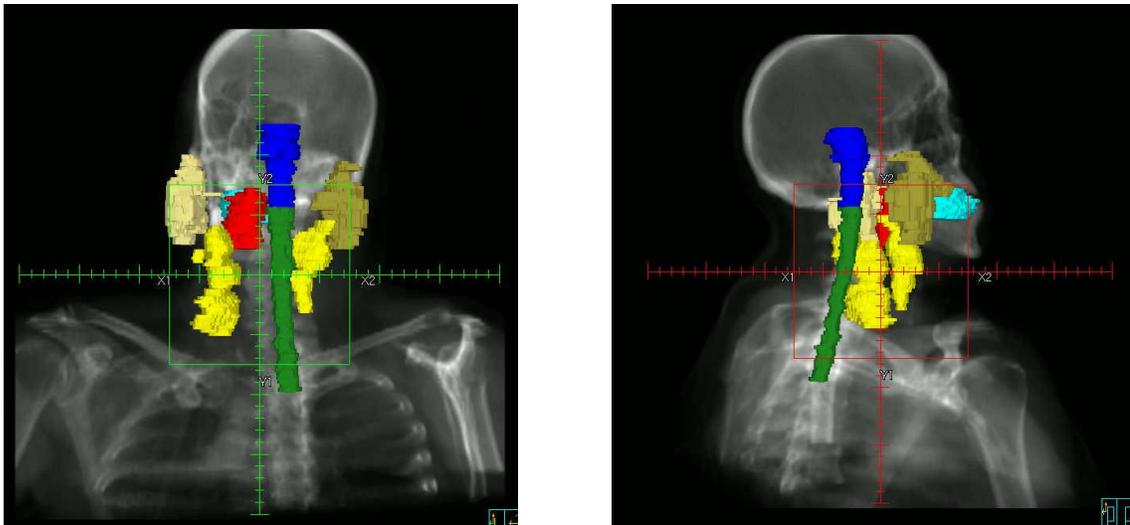
$$\Phi\left(\frac{|E(a)-\mu|}{\sigma}\right)E(a) + \left(1 - \Phi\left(\frac{|E(a)-\mu|}{\sigma}\right)\right)k(a).$$

Note that when the predicted value of the extrapolation is too many standard deviations from the norm, the more stable interpolation term dominates the extreme extrapolation term. Appendix 3-1 justifies the use of the inverse normal distribution.

There are multiple methods for combining machine learning models. Methods that use a simple or arbitrary rule decided a priori, such as weighting the models equally or weighting based on predictive accuracy in the training set, can use a single training set without overestimating the predictive value. If, however, one wishes to learn how best to combine multiple machine learning models, it is important to use two training sets. The first training set is used to train each of the basic model methods (in our case kNN, RBF, Linear Regression, Neural Networks). The second training set is used to learn how to combine the predictions of each of the machine learning methods to better predict the penalty score. This is particularly important to avoid overfitting. We use multiple methods to combine methods including the previously mentioned probabilistic method, equal weight, and proportional weight.

### 3.2.4 Tumor Geometry

An important piece of information that our earlier data mining methods (Price et al. 2014) did not include is tumor geometry. When mining plans are performed on a single patient, tumor geometry is unimportant because the geometry information is constant across all the beam angle sets. Since we are working with a small number of patients, but a large number of evaluated plans for each patient, we used fixed effects. Fixed effects capture that the geometry is the same within a patient, but different between patients. Fixed effects, however, have limited use in capturing any similarities or differences between patients that might affect beam angle performance and improve the machine learning models predictive performance on new patients. Figure 3-1 shows how the geometry as seen from each beam angle is different, this allows for different geometric data to be included for each beam angle set.



**Figure 3-1.**Picture illustrating a beam's eye view of patient specific geometry from two angles. The tumor (red) and lymph nodes (yellow) are part of the PTV, while the brain stem (dark blue), spinal cord (green), oral mucosa (light blue), and parotids (beige and olive) are the OARs. The exact size, shape, and location of each patients OARs and PTV will be unique.

The ten patients in our data set have locally advanced stage III and IV tumors. These tumors have grown quite large and are irregularly shaped. The number of variables to fully encode the information on the tumors size, shape, and location relative to the OARs exceeds the number of patients we have in our data set. Because there are more geometry variables than patients, attempts to explicitly describe the tumor will fail to have statistical meaning. We developed a method for encoding the geometric data from the beam angles used for each patient.

To integrate some geometric data into our angle sets, we use a method described in D'Souza et al. (2004). For each of the 72 beam angles that can be a part of a beam angle set, a fixed dose of 2 Gy of radiation was simulated firing at the entire area of the beam's eye-view projection of the PTV. The radiation delivered to each of the OAR is recorded and the mean organ dose (MOD) is calculated. In this way, coarse geometric information is included by capturing the overlap of the projections of the OAR and PTV when viewed from each angle. The MODs, however, fail to capture all the geometric information and present a worst case scenario, since the radiation delivery from each beam angle can be modulated to further reduce the radiation received. The beam angle sets can now be viewed as a set of MOD vectors that hint at how each angle in the set will contribute to the final penalty score. For example Figure 3-1 shows the tumor in red, when delivering a fixed dose of radiation from the angle depicted on the left, the unobstructed PTV (red and yellow) to the right parotid (olive) receives a MOD of 69.7. Delivering the same dose to the obstructed parotid from the second angle (right), it receives a MOD of 274.2.

### 3.2.4 Results on Training Set

Table 3-2 presents the results of the basic models using ten-fold cross validation. The correlation coefficient is the most important metric for purposes of searching the space for optimal values as keeping the relative order of the solutions is necessary for identifying the proper minimum. The other metrics hint at how far the model differs from the actual value on average (mean absolute error, MAE, and relative absolute error, RAE) or weighting large variations more (root mean square error, RMSE, and relative root square error, RRSE). All methods are dominated by k-nearest neighbors, with linear regression coming in second. Neural networks perform similarly to linear regression in maintaining correlation but have larger MAE and RMSE values, implying that neural networks result in more extreme solutions.

	Correlation Coefficient	MAE	RMSE	RAE	RRSE
Linear Regression	0.8468	39.6027	55.7426	46.6201	54.728
kNN	0.8978	31.0624	45.0694	36.5665	44.249
Neural Network	0.8285	44.6507	64.5706	52.1623	62.7801
Radial Basis Function	0.7962	46.9434	66.082	54.8407	64.2496

**Table 3-2:** The correlation coefficient, mean absolute error (MAE), root mean square error (RMSE), relative absolute error (RAE), and relative root square error (RRSE) of the basic methods from ten-fold cross validation.

Table 3-3 shows the correlation between the error terms of the basic models. The error terms are positively correlated which limits the potential of strictly additive methods. As we can see in Table 3-4, the performance of the hybrid methods offer marginal improvements for most metrics. Our proposed probability blended method has the best performance, outperforming the simple additive methods and the best of the basic models.

	RBF	NN	LR	kNN
RBF	1	0.39815	0.67373	0.5572
NN	0.39815	1	0.54376	0.5924
LR	0.67373	0.54376	1	0.6654
kNN	0.5572	0.5924	0.6654	1

**Table 3-3:** The correlation of the error terms of the different methods. The positive correlation of the errors implies that simply additive methods will not necessarily outperform the base methods.

	Correlation Coefficient	MAE	RMSE	RAE	RRSE
Equal Weights	0.8894	32.3872	47.6961	37.5603	45.428
Proportionally Weighted	0.8938	31.4654	45.9395	36.9264	44.747
Probability Weighting	0.9098	28.3979	42.9551	33.1753	41.764

**Table 3-4:** The correlation coefficient, mean absolute error (MAE), root mean square error (RMSE), relative absolute error (RAE), and relative root square error (RRSE) of the combined methods from ten-fold cross validation.

### 3.2.5 Geometric Results

Using the geometric data described in Section 2.3.1 may encourage methods that will better adapt to new patients, requiring no plans to be evaluated prior to searching with the GA. Table 3-5 shows that using only aggregate MOD data, without using fixed effects for patients or angle data, we are able to perform comparably to models that use fixed effects and several angle based variables. Since no fixed effects are used for patients, the expected accuracy on new patients should be the same as the results from the 10-fold cross validation. Expanding the feature set to include angle variables and interaction terms further improves the accuracy as seen in Table 3-6. Linear regression, radial basis functions, and neural networks all have similar results with k-nearest neighbor out performing all other methods. The k-nearest neighbor algorithm performed best when run with the sine and cosine functions of the beam angles and the sum of the MOD values.

	Correlation Coefficient	MAE	RMSE	RAE	RRSE
Linear Regression	0.6588	59.0513	77.3239	68.9852	75.1797
kNN	0.8775	33.5294	49.5358	39.1701	48.1622
Neural Network	0.7768	50.3131	65.5260	58.7773	63.7089
Radial Basis Function	0.8307	41.1039	57.2227	48.0188	55.6359

**Table 3-5:** The correlation coefficient, mean absolute error (MAE), root mean square error (RMSE), relative absolute error (RAE), and relative root square error (RRSE) of the basic methods using only MOD from ten-fold cross validation.

	Correlation Coefficient	MAE	RMSE	RAE	RRSE
Linear Regression	0.8447	39.5234	55.0196	46.1724	53.4939
kNN	0.9124	28.0211	42.3556	32.7351	41.1807
Neural Network	0.8402	41.4411	56.3871	48.4128	54.8235
Radial Basis Function	0.8428	39.5441	55.3516	46.1966	53.8168

**Table 3-6:** The correlation coefficient, mean absolute error (MAE), root mean square error (RMSE), relative absolute error (RAE), and relative root square error (RRSE) of the basic methods using full feature set from ten-fold cross validation.

### 3.3 Genetic Algorithm

Genetic algorithms are a meta-heuristic used to search for optimal solutions to combinatorial optimization problems. Genetic algorithms maintain a population of potential solutions and generate new solutions over successive generations in order to search for an optimal solution. Solutions to the optimization problem are represented as chromosomes. We use the ordered beam angle set as a chromosome. The fitness function is used to judge the quality of each chromosome; it determines which chromosomes are passed from one generation to the next. To generate new solutions, mutations, crossovers, and other search techniques can be employed. Finally, stopping criteria are set to determine when to terminate the search.

#### 3.3.1 Population Size and Stopping Criteria

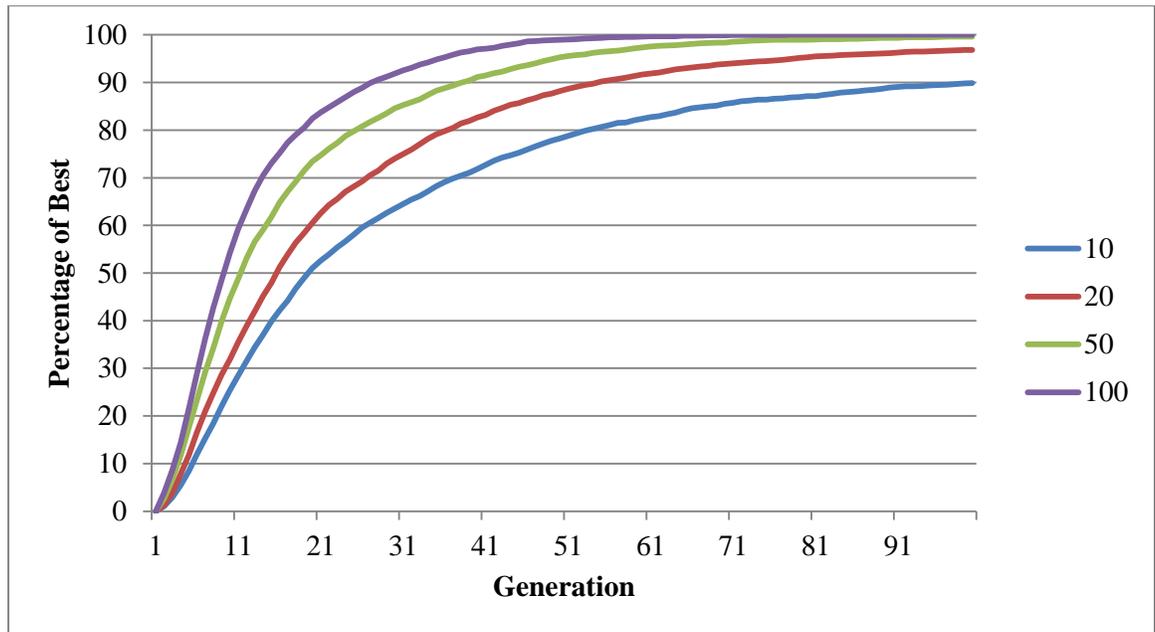
The running time of a genetic algorithm is proportional to the population of solutions that are evaluated each generation and the number of generations. Small populations have a higher likelihood of crossover between any two chromosomes in the population and quickly discard low performing solutions. Small populations, however, have the po-

tential of being stuck at locally optimum solutions if the population becomes too homogenous. Larger populations are slower to become homogenous. They also may store an excessive numbers of solutions with little promise of leading to the optimal solution. To test which would be most appropriate to search the space of beam angles, the performance of combinations of population size and generations were used to evaluate ideal population size and stopping criteria. Table 3-7 shows the population size and generation counts used to determine the stopping criteria.

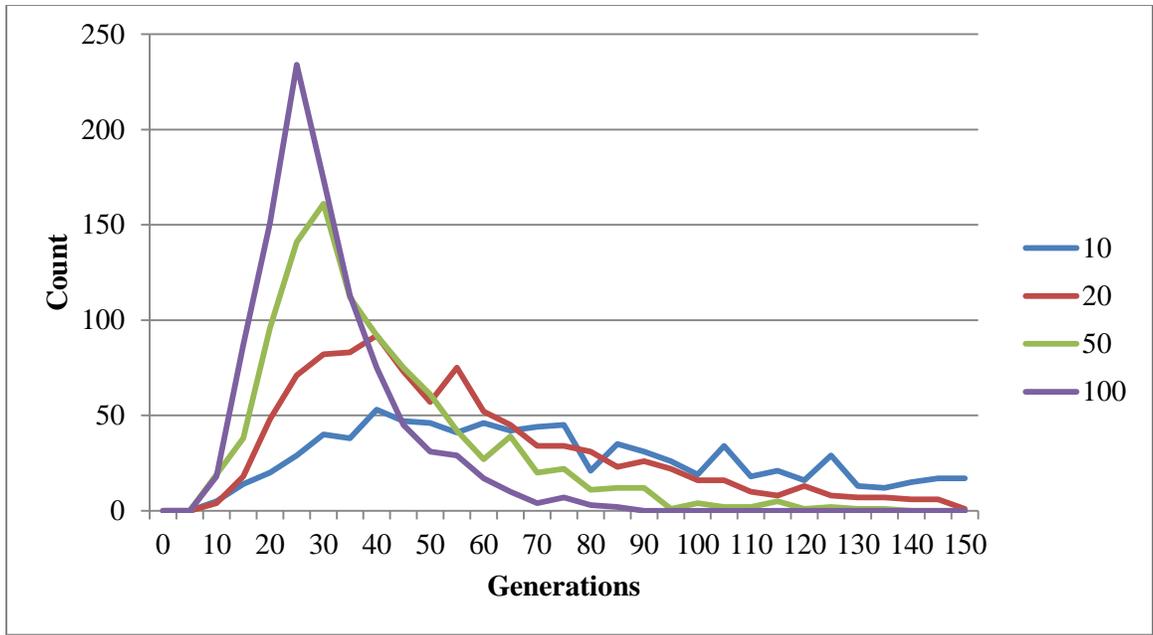
A polynomial objective function with multiple local optimums was used to calibrate a set of parameters used for all the machine learning objective functions in the genetic algorithm. While running very quickly, compared to the other machine learning fitness functions, the polynomial objective function can be used to calibrate the GA to search the space. Figure 3-2 shows the average convergence of the different population sizes to the optimal solution. For each population size, the GA was run 1,000 times and the results were averaged. Every run of the GA converged to the optimal solution of the polynomial. As expected, larger population sizes converge more quickly. Figure 3-3 shows the generation in which the optimal solution was found for each of the 1,000 runs. This information can be used to determine how many runs would have failed to find the optimal solution had the termination criteria been set lower. The goal is to determine the population size that minimizes the total number of chromosomes evaluated by calculating the product of max generation size and population size.

Population Size	Maximum Generations
10	1000
20	500
50	200
100	100

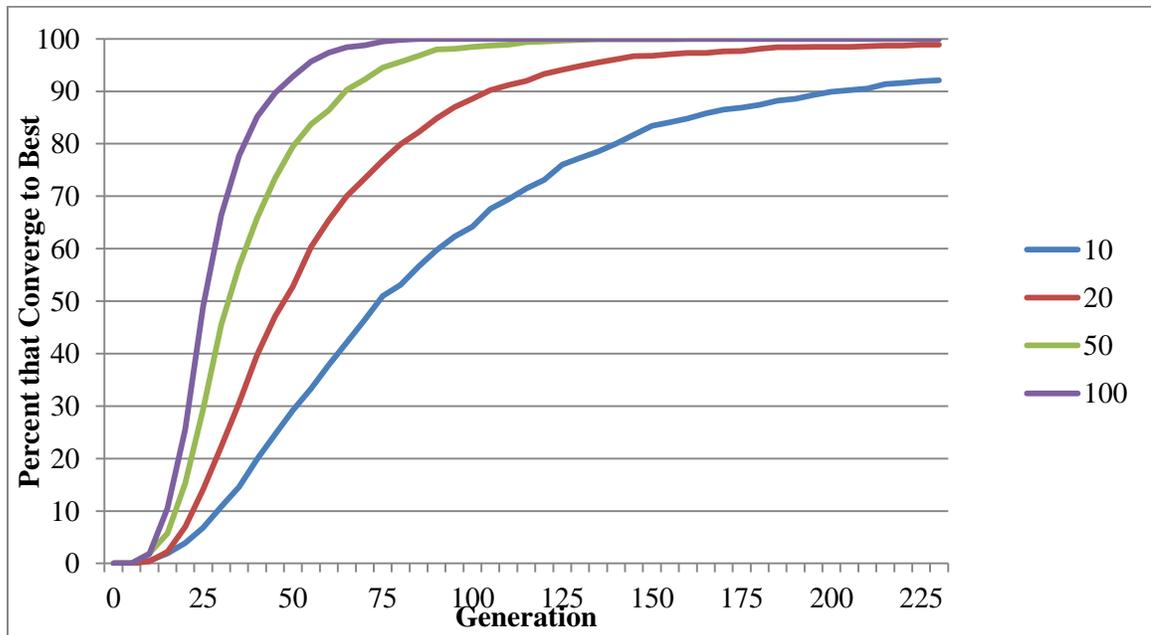
**Table 3-7:** All experiments evaluated a total of approximately 20,000 chromosomes and took on average 1.7 seconds to run on an Intel® Core™ i5 CPU with 2.4 GHz and 6 GB of RAM.



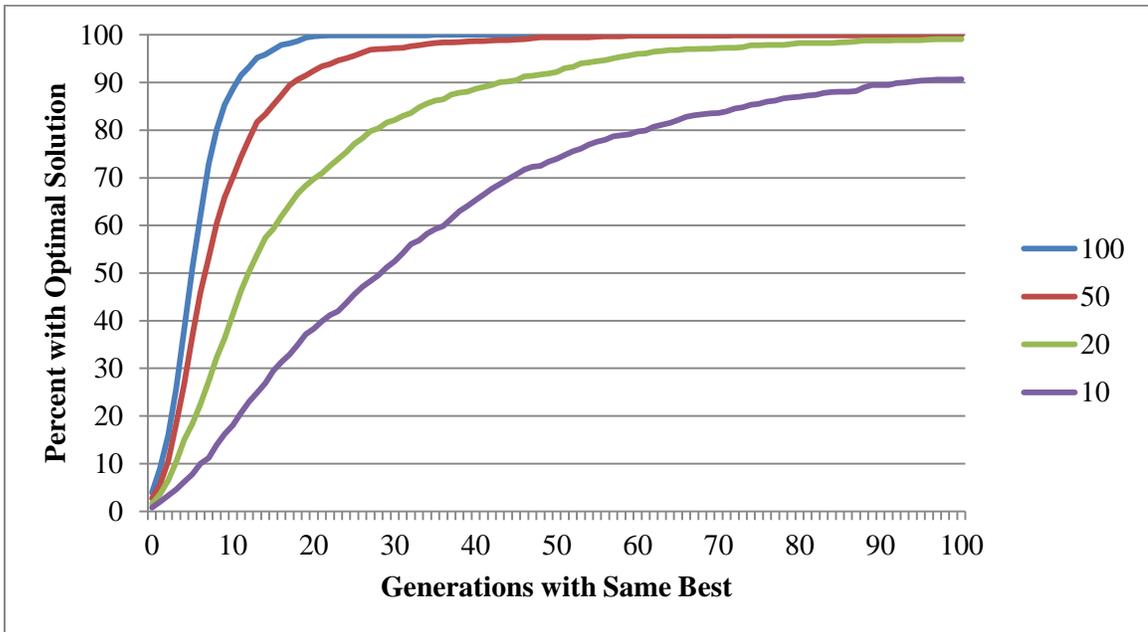
**Figure 3-2:** Plot of percentage of best solution found by nth generation. Each population was averaged over 1,000 runs of the genetic algorithm.



**Figure 3-3:** Histogram of which generation the best plan was found during. If the total number of generations was set to less than any of these numbers, then all runs in which the best generation was found to the right would not find the optimal solution.



**Figure 3-4:** Percent of runs that would converge to best solution if the run were truncated at a given generation.

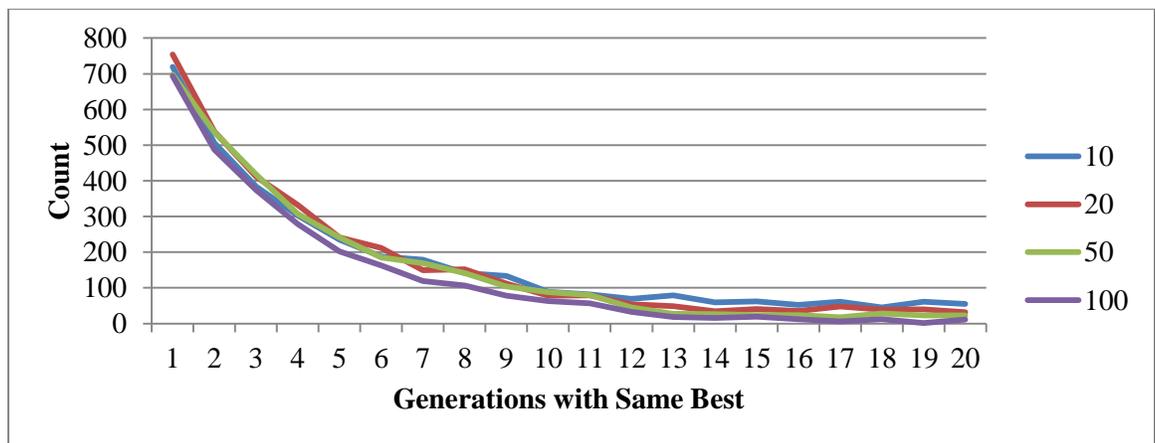


**Figure 3-5:** Percent of runs that would find optimal solution if the run were set to terminate after a fixed number of generations with the same best solution.

Population Size	95 <sup>th</sup> Percentile of Best Solution	Solutions Evaluated	Maximum Generation Best Solution Found	Solutions Evaluated
10	263	5,260	641	12,820
20	129	5,160	277	11,080
50	76	7,600	158	15,800
100	54	10,800	84	16,800

**Table 3-8:** The total runtime of the Genetic Algorithm (GA) is proportional to the number of solutions evaluated. A population size of 20 had a minimum maximum number of solutions evaluated.

In addition to stopping after a fixed number of generations, many GAs have a secondary stopping criteria based on when the best solution found has been the same for many generations. This is used to terminate in case the search is stuck in a local optimum. To determine where to set this stopping criteria, we calculated the length of every streak of the same best solution, prior to finding the actual best solution. If the stopping criteria was set too low, the optimal solution would not be found as the search would have been terminated prematurely.



**Figure 3-6:** Plot of the length of consecutive generations with same best solution, prior to finding new best solution. If the stopping criteria had been set shorter than a streak of a

given length, then all runs with streaks longer would have not found the optimal solution. Zero was excluded from the figure as the number of cases that new best solutions were found consecutively was an order of magnitude greater due to local search.

### 3.3.2 Mutation, Crossover, and Local Search

A mutation function takes a single angle set and mutates it to return a new valid angle set. A crossover function takes two different angle sets and returns a single new valid angle set that somehow combines the two. Mutations help to maintain population diversity, while crossovers help to blend solutions to combine the best features of each. Table 3-9 shows the mutation and crossover functions that were used by our GA. The second chromosome for crossover functions can either be chosen randomly with uniform probability from the population, selected randomly with probability proportional to the fitness function value, or the chromosome with the best known fitness function value. We used a roulette system which favored higher ranked chromosomes for crossover.

Title	Description of Function
Mutation 1	Replaces a random angle from an angle set with a new angle to create a new valid angle set
Mutation 2	Each angle is shifted by either -5, 0, or 5 degrees
Mutation 3	Rotates all angles by the same amount
Mutation 4	Replaces a random angle with the angle 180 degrees from that angle
Crossover 1	Replaces a random angle from the first chromosome with one from the second chromosome
Crossover 2	Replace a random angle from the first chromosome with one between it and the nearest angle from the second chromosome (including end points)
Crossover 3	Shifts all ordered angles in chromosome one in the direction of the same ordered angle in chromosome two

**Table 3-9:** A list of the five mutation and three crossover functions tried as part of our genetic algorithm. Algorithms are detailed in Table 3-10.

Mutation 1:

Step 1: Randomly select a beam angle from the beam angle set.

Step 2: Replace with angle from the set of angles valid with respect to other angles in set.

Step 3: Order new angle set from smallest to largest.

Mutation 2:

Step 1: Randomly add -5, 0, or 5 to each angle.

Step 2: If the resulting beam angle set is valid accept, else return to Step 1.

Mutation 3:

Step 1: Randomly select an angle at 5 degree intervals from 5 to 355 and shift all angles in the angle set by this amount.

Step 2: Take all angles Mod 360 and order new angle set from smallest to largest.

Mutation 4:

Step 1: Randomly order the beam angle set into a list.

Step 2: Replace the next angle in the list with the angle 180 from the original angle. If the angle set is invalid replace with original set, and repeat Step 2.

Step 3: Take all angles Mod 360 and order new angle set from smallest to largest.

Crossover 1:

Step 1: Randomly order Chromosome A and Chromosome B.

Step 2: Replace first angle in Chromosome A with the first from Chromosome B, if valid accept, else take the next from Chromosome B. If no valid replacements for first angle in Chromosome A found, then repeat process for next angle in Chromosome A.

Step 3: Order new angle set from smallest to largest.

Crossover 2:

Step 1: Choose a random angle from Chromosome A.

Step 2: Create a set of angles valid with the remaining angles in Chromosome A between the chosen angle and the corresponding angle from Chromosome B. Replace angle with randomly selected angle from the list.

Crossover 3:

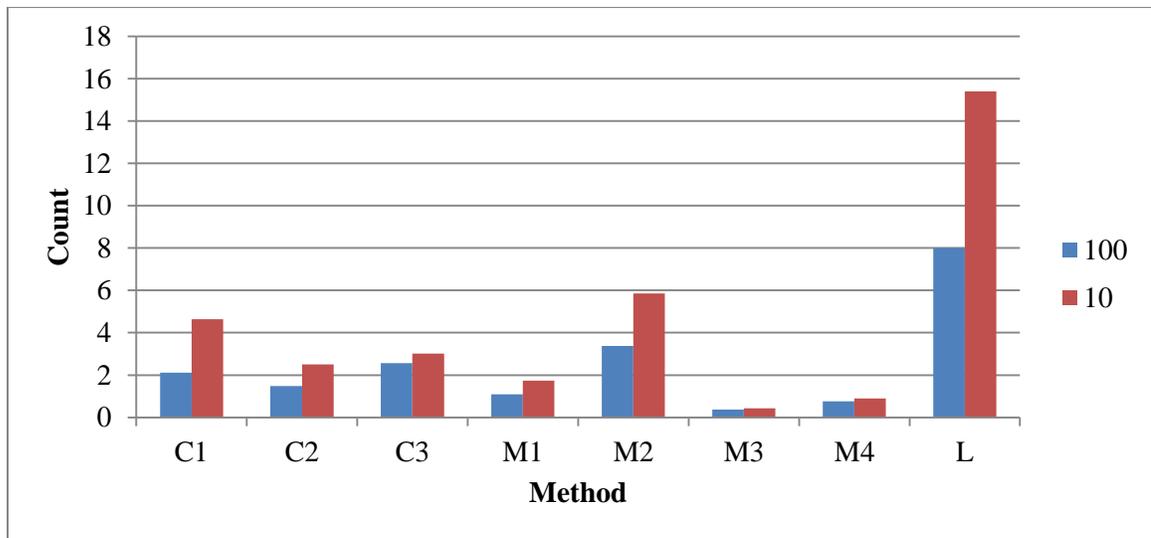
Step 1: Pair each angle in Chromosomes A and B. Randomly order pairs.

Step 2: For each angle pair; If corresponding angle in Chromosome B is larger, then add 5 degrees to angle from Chromosome A. If the corresponding angle in Chromosome B is smaller, then subtract 5 degrees to the angle from Chromosome A. If the resulting angle set is valid, accept change, otherwise reject.

**Table 3-10:** Outline of algorithms

In addition to mutation and crossover functions, a local search was implemented. The local search was performed each time a new best solution was found, examining all angle sets within 5 degrees of the new best solution.

Prior to optimizing the probability of each mutation and crossover function, we examined which mutation, crossover, and search procedures were contributing to finding the best solution. Starting with the random construction of chromosomes in the first generation, every new solution generated records the methods used to generate its parent and adds the method used to generate it. Each mutation, crossover, and local search is coded with a unique letter. Figure 3-7 compares the distribution of crossovers and mutations used to generate the best solutions for population sizes of 10 and 100 chromosomes. For a population size of 100, an average of 19.7 steps were used to generate the optimal solution compared to 34.5 steps for a population size of 10.



**Figure 3-7:** Corresponding to the methods in Table 3-9 shows the average number of mutations, crossovers, and local searches involved in finding the optimal solution for populations of 10 and 100.

Larger populations are typically associated with lower crossover and mutation probabilities and smaller populations with higher probabilities (Srinivas and Patnaik 1994). This corresponds with the need to keep small populations diverse in order to prevent getting stuck at a local optimum. Experiments were done to determine the best combination of mutation and crossover functions to use with each fitness functions. Each mutation and crossover function has a probability of occurring and parameter optimization was used to choose the best parameters for each fitness function. The steps of the parameter optimization are listed in Table 3-11. The performance of each set of parameters in step 2 will be determined by averaging the results of 1,000 runs of the GA.

Step 1	Initialize with all mutations and crossovers having an equal probability
Step 2	Compare the performance of raising the probability of each mutation or crossover by the step size, while reducing the probability of all others by step size / (number of mutation/crossover -1)
Step 3	If performance improves, then choose the step that results in maximum improvement and repeat step 2, Else If the step size has reached stopping criteria, then end Else reduce step size by half and repeat step 2

**Table 3-11:** List of steps when optimizing mutation/crossover parameters for GA

### 3.4 Results

The genetic algorithm was run 100 times for each of the machine learning models, for each of the patients. Each unique solution was then evaluated using the Pinnacle Planning software. The average results were calculated by weighting the penalty score by the number of times the genetic algorithm returned each solution. Table 3-12 shows the results of the different machine learning methods normalized to k-nearest neighbor, which returns the best known solution for each patient from the observation set. The entire set of observations was used to train each of the models.

Method	Average	Best	Average Runtime
Linear Regression	117.4	117.4	<b>9.4s</b>
Radial Basis Function	110.3	102.2	14.5s
Neural Network	105.8	98.3	13.2s
k-nearest neighbor	100.0	100.0	328.7s
Equal Weights	99.6	97.5	342.3s
<b>Probability Weighted</b>	<b>97.9</b>	<b>96.1</b>	356.8s

**Table 3-12.** Performance of GA results with respect to the previously best known plan. Lower scores are better. Runtimes of the GA on an Intel® Core™ i5 CPU with 2.4 GHz and 6 GB of RAM. Bold indicates the best performance in each column.

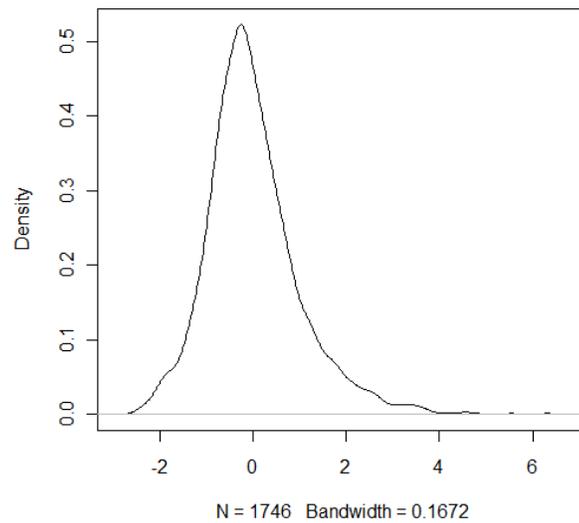
As seen in Table 3-12, the basic methods did not tend to return solutions that were better than the known solutions. By using combinations of the basic machine learning methods, we were able to improve upon the best known solution. Some methods such as linear regression and k-nearest neighbor converged to a single solution in all runs of the GA. Other methods had less consistent results, sometimes finding better solutions and othertimes failing to match the best known solution. Table 3-12 also shows the runtime of the different methods in finding a solution. As can be seen, the methods that do not call on k-nearest neighbor have a significantly faster runtime. To improve the runtime of methods that call on k-nearest neighbor, we could limit calls to evaluate k-nearest neighbor to when the other machine learning models predicted an angle set that was of very high quality.

### 3.5 Conclusions

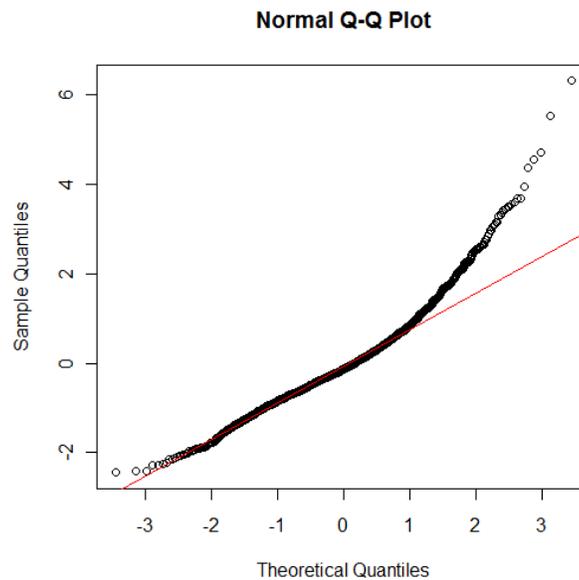
The improvement in the treatment plan quality by 2% over the best known plan implies that using models developed by machine learning to evaluate potential treatment plans is a viable method for beam angle selection. The quality of the predictions can be expected

to improve as the number of plans and patients in the training set increases. For methods that include k-nearest neighbor, this might result in unacceptably long search times when using the full training set. Given that similarities in tumor geometry and location are inferred based on the MOD of the 72 beam angles, identifying patients with similar tumors is simply a matter of checking the distance between vectors of MOD instead of a complex image analysis problem. Using a subset of patients with similar tumor geometries would likely speed the search and further improve the results. Methods that incorporate multiple machine learning methods performed the best in our research, but as the set of observations grows, it is likely that the volatility of extrapolation techniques will decrease.

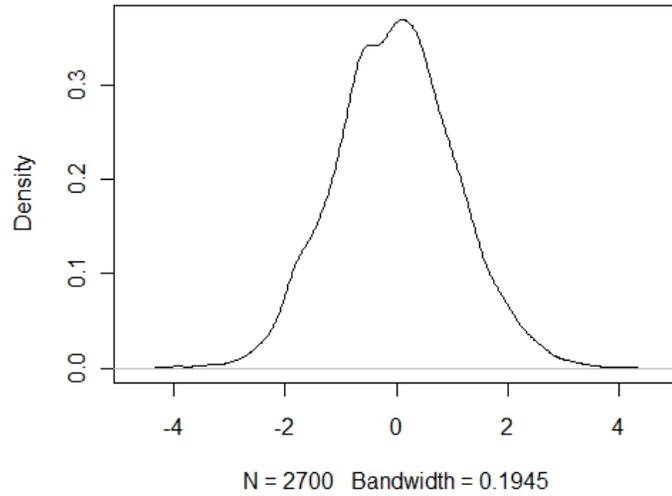
## APPENDIX 3-1



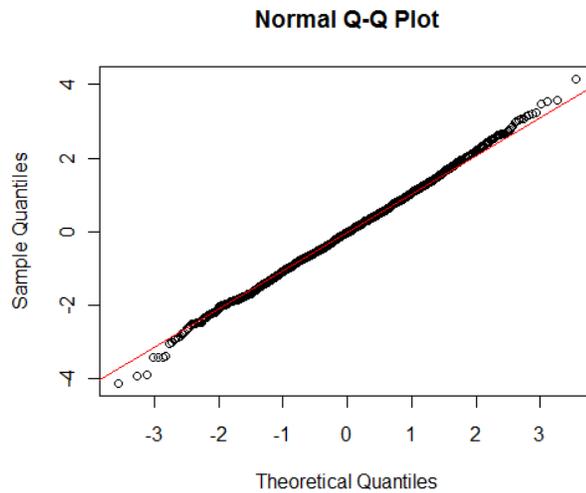
**Figure 3-8:** Plot of probability density function of the penalty scores for the seven beam plans.



**Figure 3-9:** Plot matching the sample quantile values with those of a normal distribution. The figure shows the deviations from normal distribution of the penalty scores for the seven beam plans. Note that the quantiles closely resemble normal for plan with lower than median penalty score. This implies that good plans resemble the first half of the normal distribution, while the underperforming plans are skewed right.



**Figure 3-10:** Plot of probability density function of the penalty scores for the five beam plans.



**Figure 3-11:** Plot matching the sample quantile values with those of a normal distribution. The figure shows the deviations from normal distribution of the penalty scores for the five beam plans.

Figures 3-8 to 3-11 show that the distribution of penalty scores follows a right skewed normal distribution. The close fit of the distribution for values below the median penalty score implies that the formula for weighting multiple machine learning methods,

$$\Phi\left(\frac{|E(a)-\mu|}{\sigma}\right)E(a) + \left(1 - \Phi\left(\frac{|E(a)-\mu|}{\sigma}\right)\right)k(a),$$

is valid for use for low penalty score methods. Given that we are searching for plans that have penalty scores at the left tail of the distribution, the formula should be valid for competitive plans. The right skew of the distribution implies that there will be an over-correction that underestimates the penalty score of the worst performing plans.

## **4 SCHEDULING GUIDELINES FOR A MULTI-ROOM PROTON THERAPY TREATMENT CENTER**

### **ACKNOWLEDGEMENTS:**

This is research conducted with:

Bruce Golden, University of Maryland  
Edward Wasil, American University  
Howard H. Zhang, University of Maryland School of Medicine

### **ABSTRACT**

Proton therapy is a form of radiation therapy with the potential to reduce the amount of radiation exposure received by healthy tissue during the course of cancer treatment. This leads fewer acute and long-term side effects. Due to the costs involved in constructing a proton therapy facility, there are only 10 facilities operating in the United States currently. These facilities need to treat as many patients as possible, while minimizing the time a patient spends waiting in the system. There are stochastic elements in the treatment time, and planners need to take into account all relevant patient information in constructing a schedule. Important information that is known prior to scheduling includes the design of the treatment plan, residential status of the patient, and special needs of the patient. We simulate the operations of a proton therapy treatment center in order to generate rules to help produce efficient and robust schedules.

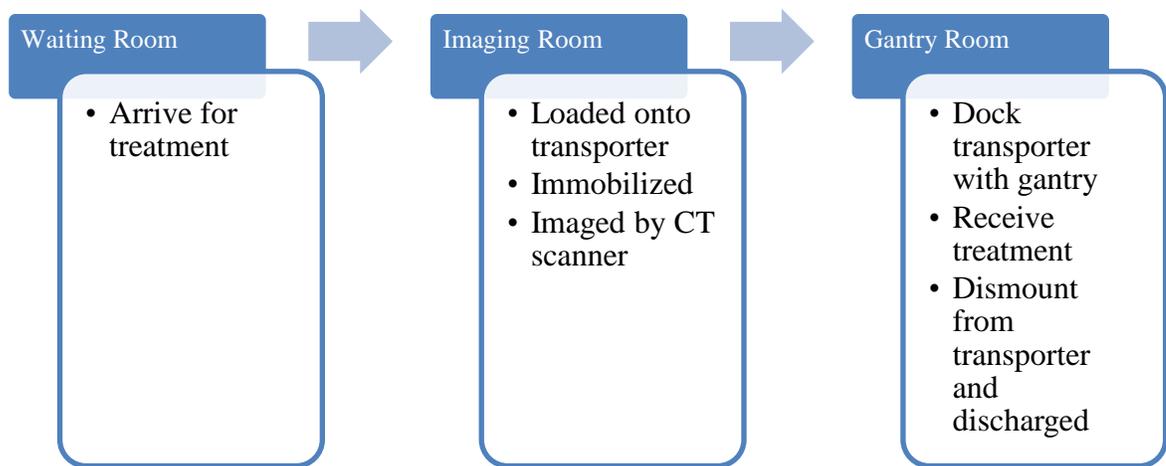
## 4.1 Introduction

The Maryland Proton Treatment Center involves an investment of more than \$200 million to build a proton therapy treatment center that includes imaging equipment, offices, and accommodations for patients (Roycastle 2010). This large initial investment is what makes proton therapy so expensive for patients. Due to the large initial investment, only 10 proton therapy treatment centers are operating in the United States currently (there are more than 10,000 linear accelerators used for x-ray intensity modulated radiation therapy (IMRT) worldwide) with eight more in development (National Association for Proton Therapy 2013). A cyclotron is needed to accelerate protons to the speed required for treatment. It is a much larger piece of equipment than the linear accelerators used for photons in x-ray IMRT. Typically, a new facility is built to house the equipment needed for PT, whereas linear accelerators can be installed in existing hospitals. When building an entirely new facility, there are few limitations that are imposed on placing the equipment. A wide array of layouts and configurations of the equipment can be considered. Our goal is to optimize the layout for patient throughput, while maintaining reasonable patient wait times. First, we need to determine how many gantry and imaging rooms are needed to ensure the cyclotron is used at full capacity. Second, we need to determine if the imaging rooms are fully connected to the gantry rooms or are there dedicated imaging rooms. Third, given a layout that can utilize the cyclotron at or near capacity, we need to determine whether or not the patient arrival rate affects patient wait time and machine idle time. Finally, we need to examine the effect of patient tardiness and absenteeism on patient wait time and machine idle time.

Patients undergoing proton therapy typically receive treatment five days a week for four to five weeks. Daily treatments take less than an hour and do not typically require anesthesia or other drugs that might affect a patient's daily routine. Some patients will come from out of town for treatment and will have an extended stay. The small number of side effects allows for local patients to continue working while receiving treatment. Tardiness of outpatients has been observed in other settings (Alexopoulos et al. 2008) and is likely to occur among local patients.

In Figure 4-1, we show the flow of patients through the treatment center. Patients enter the center and proceed to a waiting room where they wait for an available imaging room to begin treatment. Each daily treatment begins with a detailed imaging process to ensure the radiation is delivered to the precise treatment location. The biggest advantage of PT over IMRT is its ability to deliver a precise, highly concentrated dose of radiation. When accurately delivered, this results in less damage to the surrounding tissue and fewer side effects (Lundkvist et al. 2005). Accurate delivery is important because the higher radiation levels delivered by PT will easily destroy healthy tissue if accidentally irradiated (Schulz-Ertner and Tsujii 2007). Errors in imaging and patient immobilization have been cited as the cause of adverse reactions in the treatment of prostate cancer (Wroe, Rosenfeld, and Schulte 2007). In remote positioning, the patient is positioned in the imaging room and moves between rooms immobilized on a transporter. Upon entering the imaging room, the patient rests on the transporter and is immobilized and prepared for imaging. A computer tomography (CT) scanner is used to identify the planned treatment volume. Marks are made on the patient's skin that will later help align the proton beam. The transporter with the patient is undocked from the CT scanner and moved to the gan-

try room. The transporter is docked with the gantry arm and aligned in preparation for treatment. The patient receives radiation from one, two, or three different beam angles. For each beam angle, the radiation is delivered over the cross section of the planned treatment volume from that angle and divided into pixels. Each pixel receives a different dose of radiation. After the patient has been treated, technicians dismount the patient from the transporter and return the transporter to the imaging room. The patient is discharged.



**Figure 4-1:** Patient flow through a PT treatment center. Patients may have to wait between rooms.

There is very little published research on the scheduling and operations of PT treatment centers. The Paul Scherrer Institute (PSI) in Switzerland was the first to use remote patient positioning for proton therapy. The process used by PSI is described in the paper by Bolsi et al. (2008). The authors showed that remote positioning provided fast patient transport and maintained high accuracy in the delivery of radiation. The benefits of remote patient positioning when compared to traditional in-room positioning in PT treatment centers was examined in a follow-up paper based on simulations using data gathered from PSI (Fava et al. 2012). The authors ran simulations to compare patient

throughput of remote and in-room positioning with a variety of transport speeds, imaging times, and beam delivery times. The authors imposed a strict three-minute waiting time limit for patients, so that a limited range of gantry and imaging room configurations were considered.

In this paper, we compare the performance of dedicating an imaging room to each gantry room to fully connecting imaging rooms to gantry rooms. We examine the relationship of patient inter-arrival rate (total daily throughput) to patient wait time and machine idle time.

#### **4.2 Simulation Model**

We built our model using NetLogo 5.0.3, a specialized agent-based simulation language (Wilensky 1999). We chose NetLogo because it has all the capabilities that we require for this project and it is available to the public at no charge. In addition, using an agent-based modeling language, such as NetLogo, will allow the simulation to be extended to include more complex patient, facility, and personnel interactions in future versions of our model. Currently only patient and facility (i.e., the cyclotron) agents are implemented. Our model is run in discrete time with each tick representing five seconds. The treatment times that are used in our model come from an analysis of a PT treatment center using remote positioning due to Fava et al. (2012). The treatment steps and times are shown in Table 4-1. Treatment times for each step are drawn from a symmetric triangular distribution and are rounded to the nearest tick. A symmetric triangular distribution is used to ensure that all times are bounded and positive. A symmetric triangular distribution provides a clear peak and a wide range of values.

Treatment Steps	Mean Time (minutes)	Range of Values [min, max]
(1) Patient enters the facility	0.00	[0.00]
(2) Imaging	16.83	[8.50, 25.17]
(3) Move to gantry room	1.33	[1.33]
(4) Prepare for beam angles	6.67	[4.25, 9.08]
(5) First beam angle	1.25	[1.25]
(6) Move gantry arm	1.50	[1.50]
(7) Second beam angle	1.25	[1.25]
(8) Move gantry arm	1.50	[1.50]
(9) Third beam angle	1.25	[1.25]
(10) Discharge patient and reset gantry room	4.83	[2.83, 6.83]

**Table 4-1:** Mean time and standard deviation of the steps used in our model based on treatment sessions at the Paul Scherrer Institute (taken from Fava et al. (2012)). Step 2 occurs in the CT scanning room while steps 4 through 10 take place in the gantry room. Beam delivery time based on manufacturer estimate.

In our simulation runs, a patient is randomly assigned a treatment plan with one, two, or three beam angles. For each patient, we record the amount of time waiting for entry into the imaging room and gantry room, and the amount of time waiting in the gantry room for the cyclotron.

The waiting room is where a patient enters the system at the scheduled arrival time. If all imaging rooms are in use, then a count is recorded of the time a patient spends in the waiting room. Otherwise, the patient passes immediately into an available imaging room. For experiments that include a distribution of arrival times about a patient's scheduled appointment time, the wait time only includes time spent in the waiting room after the time of the scheduled appointment.

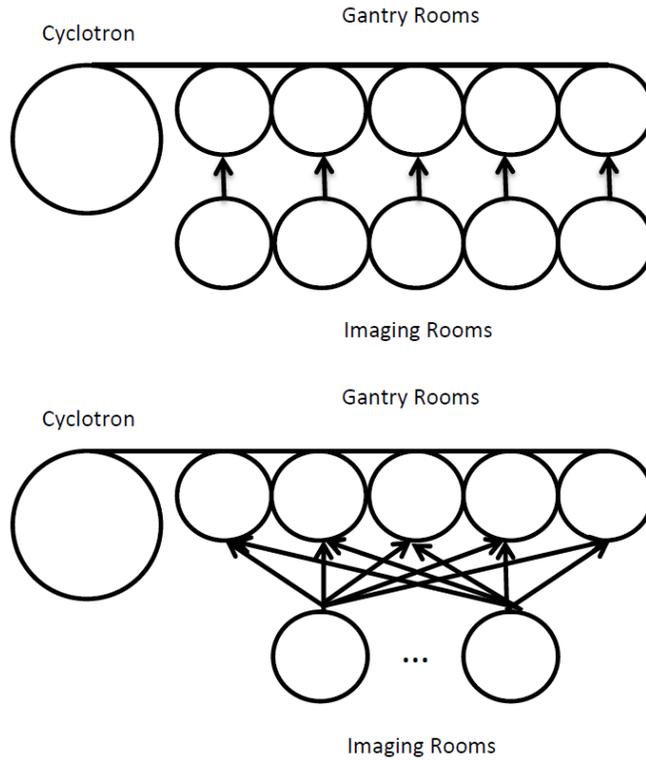
Each imaging room can service only one patient at a time. The patient undergoes a series of procedures in the imaging room. These procedures are the same for each patient and each room. The amount of time taken for each step of the imaging process is drawn from a triangular distribution and uses the means and ranges listed in Table 4-1. For each imaging room, we record the time spent without a patient and the time spent with a patient waiting to transfer to a gantry room.

Each gantry room can service only one patient at a time. Radiation is delivered to each patient from one, two, or three different beam angles. In addition to the time taken to deliver the radiation from each angle, the gantry arm must be repositioned between each angle. The time spent to rotate the gantry arm and treat a beam angle is fixed at 90 and 75 seconds, respectively, as these are purely automated steps. Treatment steps 4 and 10 are performed in the gantry room and their times are drawn from triangular distributions. For each beam angle, the gantry room must determine if the cyclotron is ready to fire. For each room, we record the time spent empty and the time spent waiting for the cyclotron.

The cyclotron provides protons for the gantry rooms. In addition to the 75 seconds a cyclotron spends treating a beam angle, it must wait an additional 45 seconds before servicing a different gantry room. When several rooms have requests for protons, they are serviced first come, first served. For the cyclotron, we record total time spent idle and time spent with one or more outstanding requests for protons.

All runs of our model have one cyclotron. The number of gantry and imaging rooms varies from run to run. Patients arrive at regular intervals that vary from run to run. Imag-

ing rooms were either dedicated to specific gantry rooms or were fully connected to all gantry rooms as shown in Figure 4-2.

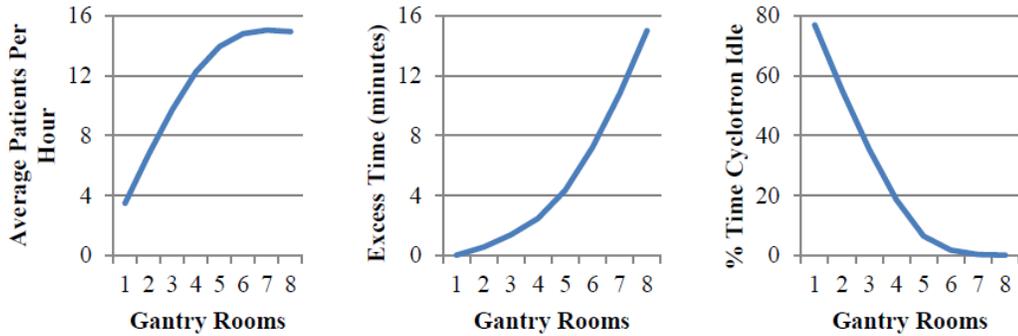


**Figure 4-2:** Configurations with dedicated imaging rooms (top) versus shared imaging rooms (bottom).

### 4.3 Simulation Results

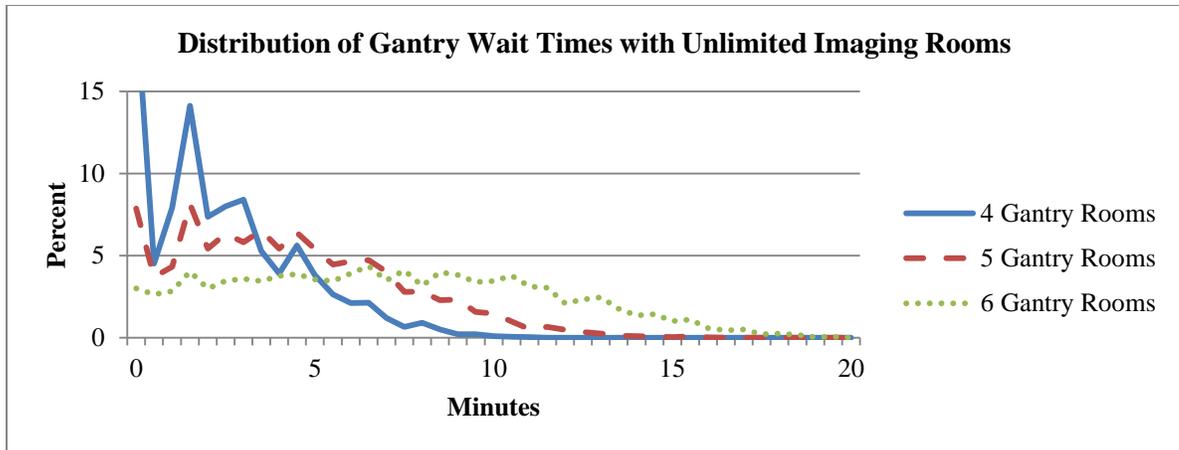
In order to determine the steady-state throughput, we ran simulations for 250,000 ticks (about 340 hours of simulated operation), where the first 50,000 ticks were treated as a burn-in period. The simulations ran with patients arriving whenever an imaging room became available (this could be thought of as an infinite waiting room scenario). The resulting throughput (expressed in patients per hour) provides an upper bound on throughput for one to seven gantry rooms because there is always a patient ready to enter the gantry room. In Figure 4-3, we see a 26% increase in throughput by adding a fourth

room, another 14% for a fifth room, but only 6% for a sixth room. With six rooms, the cyclotron is idle less than 2% of the time. Therefore, additional rooms can do little to increase throughput and only result in longer patient wait times in the gantry room.



**Figure 4-3:** Average patients per hour, excess time in a gantry room (minutes), and percent of time that the cyclotron is idle as a function of the number of gantry rooms for a one cyclotron system. In this system, there is always a patient ready to enter an available gantry room.

The average excess time spent in the gantry room is an aggregate measure. We need to look at the distribution of wait times. From Figure 4-3, for four gantry rooms, there is an average of 2.46 minutes of excess time spent in the gantry room at maximum capacity. In Figure 4-4, we see that, under these same conditions, many patients still spend more than three minutes waiting for the cyclotron in the gantry room. As the number of gantry rooms increases, both the mean and variance of wait time increase. The increased variance makes it difficult to bring maximum wait time below three minutes, since the maximum wait time grows faster than average wait time (see Figure 4-4).



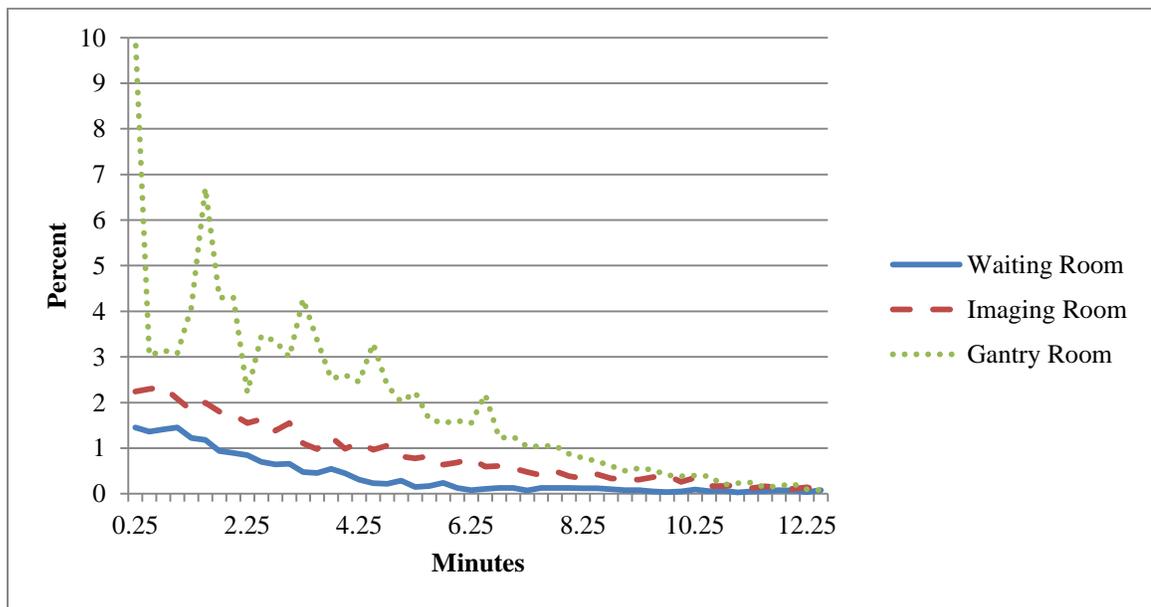
**Figure 4-4:** Percentage of patients experiencing various wait times for 10,000 patients.

The remaining experiments are performed on 100 days of simulated facility performance. Including daily startup effects is more realistic than using only steady-state behavior. In Table 4-2, we see that, by decreasing the arrival rate (given by patients per hour), we can reduce the wait times of patients. For example, by increasing the time between patient arrivals in a four gantry room and five imaging room system from 300 seconds (12.00 patients per hour) to 320 seconds (11.25 patients per hour), we reduce the average total wait time by over 55%. In Figure 4-5, we see that, even when operating below maximum throughput and having an average wait time of 2.72 minutes, 47% of the patients have wait times longer than three minutes in the gantry room.

There are long wait times in the gantry room; in contrast, 80% of patients spend no time waiting in the waiting room, 56% spend no time waiting in the imaging room, and 1.2% spend no time waiting in the gantry room. The multi-modal nature of the time spent in the gantry room (see Figure 4-5) is a result of the different treatment plans (one, two, or three beam angles) for patients. We constructed Figure 4-5 with patients arriving randomly and not taking into account the number of beam angles in the treatment plan.

Number of Rooms		Number of Patients		Time Spent (minutes)			
Gantry	Imaging	Per Hour	Per Day	Waiting	Imaging	Gantry	Total
4	4	11.25	157.50	0.51	1.33	1.87	3.71
4	5	11.25	157.50	0.03	1.42	1.83	3.28
4	5	12.00	168.00	1.39	3.78	2.21	7.38
5	5	12.00	168.00	0.00	0.40	2.63	3.03
5	5	13.33	186.67	0.68	1.72	3.33	5.73

**Table 4-2:** The average wait time by room for six combinations of arrival rates, gantry rooms, and imaging rooms for 100 simulated days of operation.



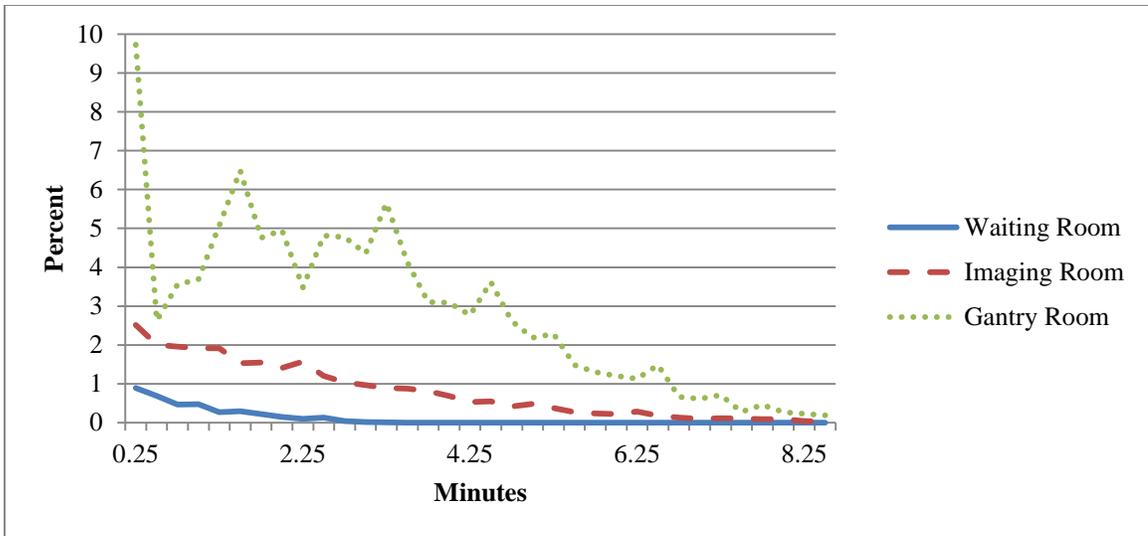
**Figure 4-5:** Percent of patients experiencing wait times for a system with five gantry rooms and five imaging rooms using an arrival rate of 13.33 patients per hour (given in Table 4-2) for 100 simulated days.

When scheduling patients, it would be important to distribute patients of each type throughout the day, because more beam angles means a longer time in the gantry room on average. For five imaging and gantry rooms, by simply alternating among the three plan types, we reduce the average total wait time by over two minutes and the time spent in

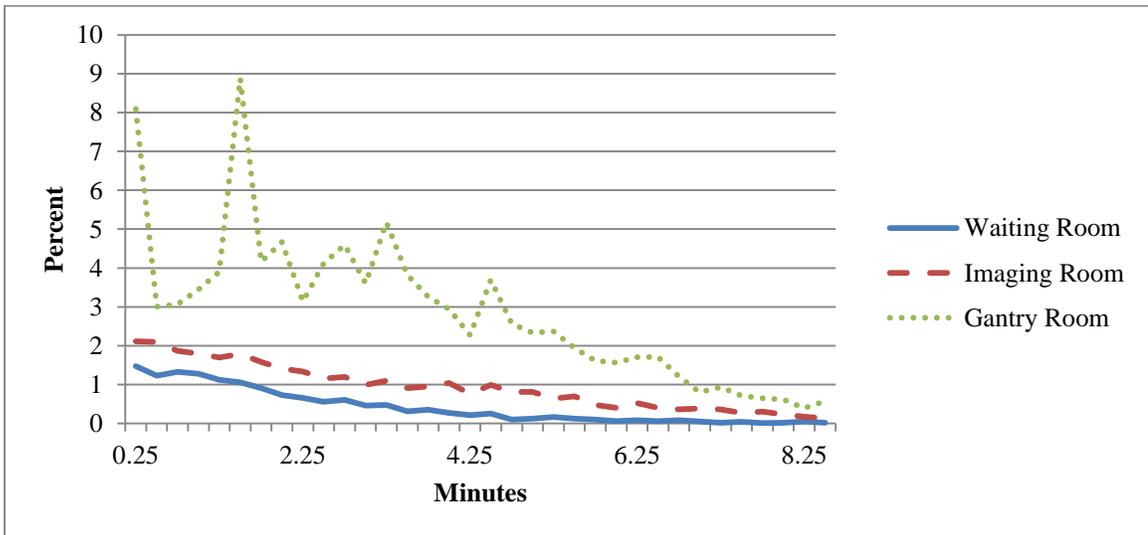
the gantry room by 37 seconds. The percent of patients spending more than three minutes in the gantry room falls from 47% to 39%. The multi-modal nature of the distribution has not changed in Figure 4-6, but the long tail from consecutive three beam plans (which take more time in the gantry room) now falls off much more quickly.

It is important to determine the sensitivity of waiting times to real-world factors such as patient tardiness and absenteeism. Both of these factors play important roles in scheduling outpatient procedures (Liu, Ziya, and Kulkarni 2010). The arrival times for patients not staying at the facility are modeled by a symmetric triangular distribution centered on the patient's scheduled arrival time and is 18 minutes in either direction. The 18-minute triangular distribution is an approximation of the fitted distribution found in the empirical study by Liu, Ziya, and Kulkarni (2010). Absenteeism rates are set at 5% for patients not staying at the facility. Half of the patients are estimated to stay at the facility and thus would be far less likely to be absent or tardy.

In Figure 4-7, we see that the distribution of waiting times look very similar. Despite seeing 2.5% fewer patients due to absenteeism, the wait times are longer in each part of the system when patient tardiness is included in the model. There is a 10% increase in waiting time in the gantry room and a 27% increase in the total time spent waiting in the system. To address the problem of patient tardiness, we note that patients staying at the treatment center will have greater flexibility in scheduling their treatment time. Patients not staying at the treatment center could be scheduled first, allowing them to choose times when they would be less likely to arrive late. Given that many patients actually arrive early for their appointments in outpatient settings (Liu, Ziya, and Kulkarni 2010), a queue might be created based on patient arrival time instead of patient appointment time.



**Figure 4-6:** Percent of patients experiencing wait times for a system with five gantry rooms and five imaging rooms with an arrival rate of 13.33 patients per hour for 100 simulated days of patients with alternating treatment plans.



**Figure 4-7:** Percent of patients experiencing wait times for a system with five gantry rooms and five imaging rooms with an arrival rate of 13.33 patients per hour for 100 simulated days of patients with alternating treatment plans and late arrivals.

#### 4.4 Extending the Simulation

We extend the work of Price et al. (2013) by adding patients with different treatment time distributions, late arrival distributions, and waiting room prioritization. The treatment times are based on the estimates from Fava et al. (2012) and manufacturer estimates have been updated reducing the beam delivery time and gantry arm rotation time. Treatment times are modeled using a symmetric triangular distribution with mean time and range given in Table 4-3. A symmetric triangular distribution has a single peak and wide range of values while ensuring that all times are bounded and positive.

Treatment Steps	Mean Time (minutes)	Range of Values [min, max]
(1) Patient enters the facility	0.00	[0.00]
(2) Imaging	16.83	[8.50, 25.17]
(3) Move to gantry room	1.33	[1.33]
(4) Prepare for beam angles	6.67	[4.25, 9.08]
(5) First beam angle	1.00	[1.00]
(6) Move gantry arm	1.00	[1.00]
(7) Second beam angle	1.00	[1.00]
(8) Move gantry arm	1.00	[1.00]
(9) Third beam angle	1.00	[1.00]
(10) Discharge patient and reset gantry room	4.83	[2.83, 6.83]

**Table 4-3:** Mean time and range of values for the steps used in our model based on treatment sessions at the Paul Scherrer Institute (taken from Fava et al. (2012)). Step 2 occurs in the CT scanning room while steps 4 through 10 take place in the gantry room. The delivery time and rotation between each of the beam angles is based on manufacturer estimate.

Patients have three basic characteristics: inpatient status, treatment plan, and scheduled arrival time. Inpatient status determines a patient’s probability of late arrival or absenteeism. We assume that an inpatient arrives on time and needs to be transported by

hospital staff. An outpatient is assumed to have a 5% chance of absenteeism. The arrival time of an outpatient is distributed within 18 minutes on either side of the patient's scheduled arrival time (Alexopoulos et al. 2008). Each patient is assigned a treatment plan that requires protons to be delivered from one, two, or three beam angles. Each treatment plan occurs with the same frequency. Finally, each patient has a scheduled arrival time, which may be based on the patient's inpatient status or treatment plan.

Upon arriving, a patient enters any available imaging room. If an imaging room is not available, the patient stays in the waiting room and waits for an imaging room to become available (Step 1 from Table 4-3). The priority of waiting-room patients to enter an imaging room varies between experiments. The default is first come, first served (FCFS). The center contains five imaging rooms which are attached to exactly one of the five gantry rooms. In the imaging room, the patient is loaded onto a patient transporter, immobilized, and receives a CT scan. Upon completion of the tasks in the imaging room (Step 2), if the gantry room associated with the imaging room is available, the patient is transported to the gantry room (Step 3). Otherwise the patient waits in the imaging room for the gantry room to become available.

When arriving in the gantry room, the patient transporter docks with the gantry arm in preparation for delivery of the proton beams (Step 4). After finishing the delivery of protons, there is a 60-second delay before the cyclotron is able to deliver protons to a new room. If no rooms require the cyclotron, the beam can be switched to a room in anticipation of the next patient to receive protons. If the cyclotron is busy or the beam needs to be switched to deliver the protons, a wait time is incurred in the gantry room. Between the delivery of protons from each prescribed angle (Steps 5, 7, 9), the gantry arm must

rotate (Steps 6, 8). One-beam plans omit Steps 6 through 9 and two-beam plans omit Steps 8 and 9. After receiving the protons, the patient dismounts the transporter and leaves the gantry room (Step 10). The room is prepared for the next patient. For each scheduling procedure, patient mix, and traffic mix, 100 days of arrivals were simulated.

Some experiments used a different distribution of late arrivals to simulate high congestion due to the urban location of the center. In 2012, Baltimore had the 17<sup>th</sup> worst congestion in the nation according to INRIX, a traffic data service (Thomson 2013). The average commuter spent 25.7 hours in traffic over the course of the year. The Maryland Proton Treatment Center is located near downtown and would be exposed to some of the worst traffic conditions in the city. During peak rush hour commutes (8am to 9am and 5pm to 6pm) the median commute time for highly traveled sections is between 2.4 and 2.7 times that of off-peak hours according to the 2012 Maryland State Highway Mobility Report (2012). Between 18% and 29% of highway miles traveled during peak hours are done in moderate congestion (taking greater than 30% more time than under normal conditions). During the worst 5% of days the commute can be between 5.6 and 6.8 times as long for some sections. The increased congestion and variability of traffic time leads to greater uncertainty in the arrival of outpatients scheduled during these hours. The patient arrival distribution is likely to be skewed late in addition to the greater variability of arrival time. Patients will be traveling from different locations and adjust for traffic differently in making their own decisions for commuting.

To model the effects of periods of high congestion, outpatients scheduled during peak traffic hours (8am to 9am and 5pm to 6pm) have a uniformly distributed variable between 0 and 10 minutes added to their arrival time to represent the effect of traffic under normal

weekday conditions. This uniform variable is in addition to the triangular late and early arrival distribution which allowed for patients to arrive during an eighteen minute window to either side of their scheduled time. Every twentieth day had an additional 10 minute delay added to all outpatients scheduled during peak traffic hours, these ten minutes represent an abnormally high traffic day (usually caused by an accident or construction which would affect all drivers on a stretch of road). The hours immediately preceding and following peak traffic hours had a uniformly distributed variable between 0 and 5 minutes added to their arrival time to better represent the traffic that leads up to and follows peak traffic.

## **4.5 Extended Simulation Results**

### **4.5.1 Mixtures of in-patients and out-patients**

In previous work (Price et al. 2013), we produced a large decrease in patient wait time by ordering patients based on the number of beam angles used in their treatment plans. Our work showed that the near maximum throughput for a five imaging room and five gantry room facility was 13.33 patients per hour. The patients were sequentially ordered, beginning with a one beam angle plan, followed by a two, then a three beam angle plan, repeating this ordering throughout the day. Since the cyclotron is the limiting resource of the facility, this schedule distributes patients who require more beam angles evenly. These patients arrived at the same uniform rate, but experienced only 70% the wait time of unordered patients. Further analysis of patient wait times showed that not all patients had the same experience. In Table 4-4, we see that patients with one beam treatment plan (who follow three beam treatment plans in this scheduling scheme) have longer treatment

times than two and three beam treatment plans. The longer wait time is due to the amount of time the cyclotron is occupied delivering protons to the prior patient.

Treatment Plan	Waiting Room (minutes)	Imaging Room (minutes)	Gantry Room (minutes)	Total (minutes)
1 Beam	0.04	0.52	3.25	3.81
2 Beam	0.05	0.47	2.38	2.90
3 Beam	0.04	0.41	1.72	2.17
All Patients	0.04	0.46	2.45	2.96

**Table 4-4:** Mean wait time of patients by number of beams used in treatment plan for sequential ordering schedule with evenly spaced arrivals at a rate of 13.33 patients per hour. All patients arrive at their scheduled time.

The different wait times based on the treatment plan suggests that the time between arrivals should be adjusted based on the treatment plan of the preceding patient. An adjusted spacing was used to more evenly distribute the wait times among patients being treated with each of three types of treatment plans. Adjusted spacing means a new patient will arrive more quickly after a patient who requires only a single beam angle and later after a patient requiring three beam angles. In Table 4-5, we show the results of schedules that use combinations of adjusting the arrival spacing and arrival ordering. While ordering patients by treatment plan and adjusting the spacing order have large effects on total wait times, taken together they show little improvement in waiting time over sequential ordering alone. Though using both adjusted spacing and sequential ordering does reduce the difference in wait times experienced between patients with different treatment plans.

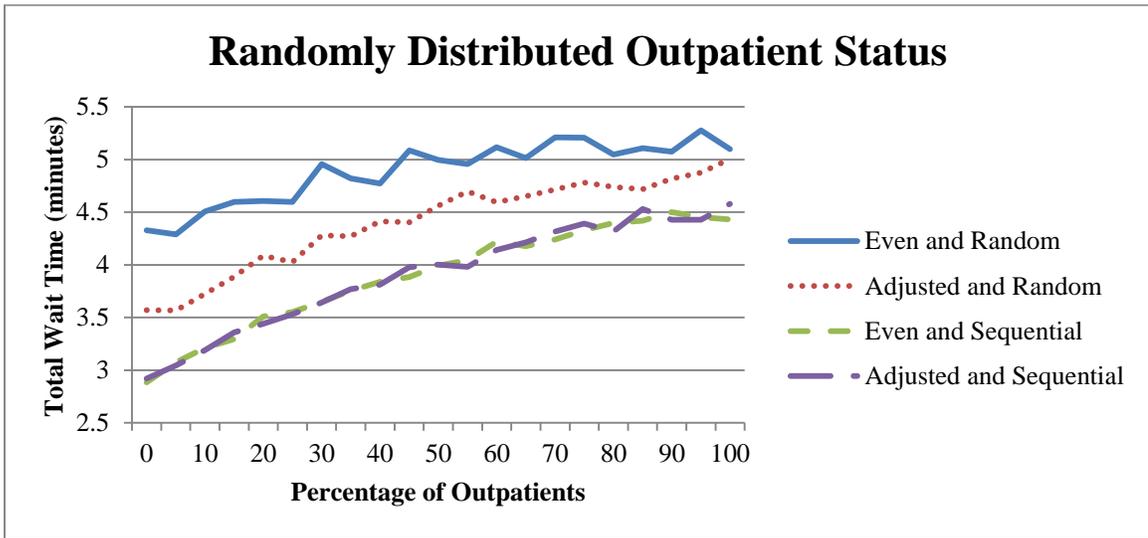
Despite all patients arriving at their scheduled appointment time, the stochastic nature of the treatment time causes about 38% of patients to complete treatment in a different order than they arrived. When early and late arrivals are included in the model the differences between the schedule and actual treatment order only increase.

	Waiting Room	Imaging Room	Gantry Room	Total	Standard Deviation
Even Spacing Random Order	0.10	0.94	3.20	4.24	4.11
Adjusted Spacing Random Order	0.56	0.62	2.46	3.65	3.35
Even Spacing Sequential Order	0.04	0.46	2.45	2.96	2.86
Adjusted Spacing Sequential Order	0.07	0.47	2.41	2.94	2.78

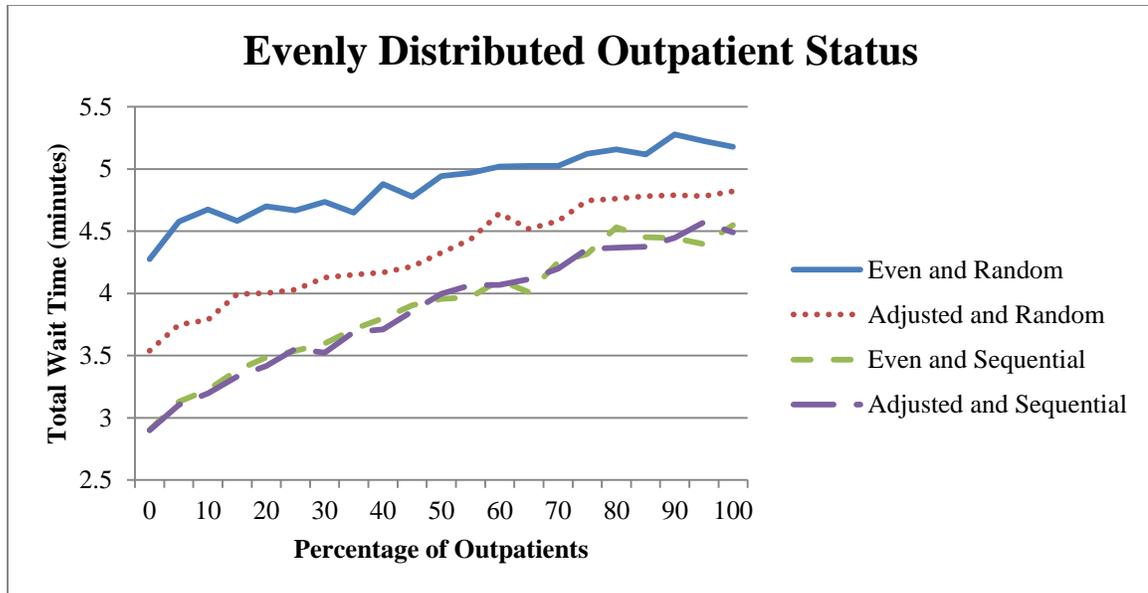
**Table 4-5:** Mean waiting time of patients scheduled without regard to treatment plan (random order) and scheduled sequentially according to treatment plan (sequential order) using evenly spaced arrivals (even spacing) and adjusted arrival times based on treatment plan of preceding patient (adjusted spacing). All schedules have an average arrival rate of 13.33 patients per hour.

Figure 4-8 shows how waiting times for the four schedules increase as the proportion of outpatients increases. Outpatients increase wait times by not arriving at their scheduled time, but can reduce waiting time too, due to absenteeism. The uncertainty in patient arrival time of outpatients increases the wait times more for sequential schedules (adding more than 90 seconds of waiting time), than for non-sequential schedules (less than 50

seconds for evenly spaced patients). By evenly distributing outpatients throughout the day, one might expect to improve performance by having a more uniform experience. Scheduling with respect to outpatient status does not, in general, reduce patient wait time very much. As can be seen in Figures 4-9, there is little difference in total wait times when compared to Figure 4-8. Only the “adjusted and random” curve was slightly improved by taking into account outpatient status.



**Figure 4-8:** Total patient wait time as a function of percentage of outpatients randomly distributed through schedule.



**Figure 4-9:** Total patient wait time as a function of percentage of outpatients evenly distributed through schedule.

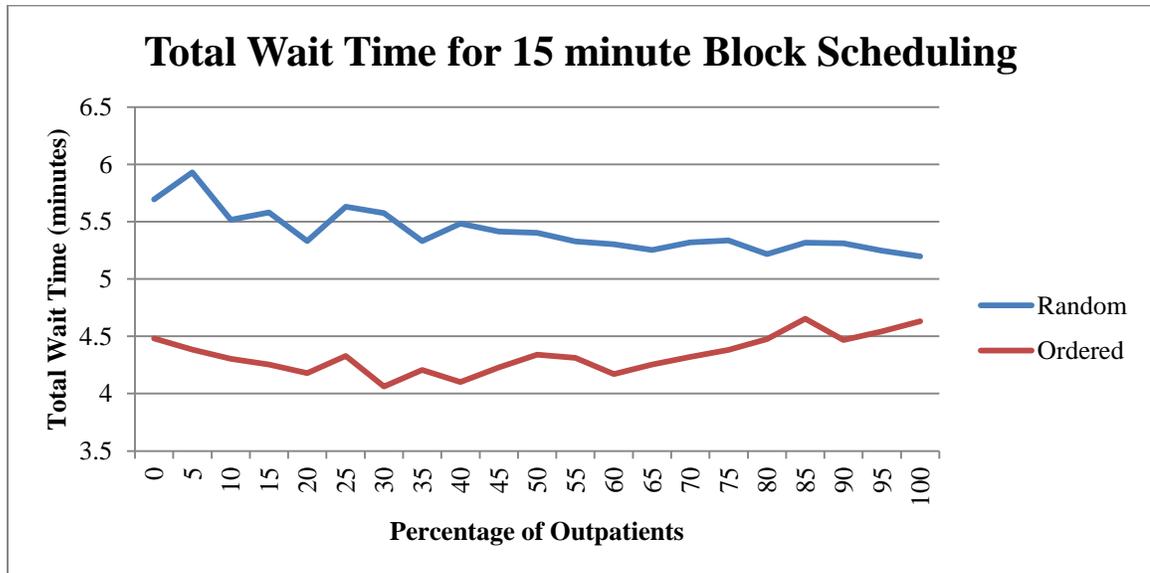
#### 4.5.2 Block Scheduling

We are now interested in determining the effects of block scheduling. By scheduling several patients to arrive simultaneously, the disruption to the system of any one of them arriving late is reduced. We schedule patients to arrive on the quarter hour in groups of three or four (maintaining an average hourly throughput of 13.33 patients). Figure 4-10 shows that the ordering patients by treatment plan is still important for decreasing total patient waiting time. This leads to more patients in the waiting room at the same time, especially when there is a high percentage of inpatients. In Table 4-6, we see that performance actually improves with the introduction of outpatients (and by extension early and late arrivals). Note that while exact scheduling outperforms block scheduling for every combination of inpatients and outpatients tested, exact scheduling sees a 66% increase in total wait time switching from all inpatients to all outpatients (Figure 4-9). Block sched-

uling has only a 12% difference between its shortest wait time and its longest. When treating 100% outpatients there is only a 2.5% difference between exact and block scheduling.

Percentage of Outpatients	Waiting Room	Imaging Room	Gantry Room	Total
0	1.15	0.71	2.78	4.63
25	0.83	0.75	2.89	4.48
50	0.77	0.81	3.01	4.59
75	0.79	0.86	3.07	4.71
100	0.91	0.94	3.14	4.99

**Table 4-6:** Average wait times for varying percentage of outpatients with 15 minute blocks, with an average arrival rate of 13.33 patients per hour.



**Figure 4-10:** Ordering patients by treatment plan results in between 10% improvement (for 100% outpatients) and a 20% improvement (for 0% outpatients) in total wait time experienced by patients under block scheduling.

We compared three decision algorithms to determine which patient is first to be admitted to an imaging room. First come, first served gives the baseline performance. In Table 4-7, we see that patients who arrive late have a shorter wait time than those who arrive on time. This is due to a shorter wait time in the waiting room, which occurs when multiple on-time arrivals do not find enough imaging rooms available.

	Waiting Room	Imaging Room	Gantry Room	Total
Early	0.65	0.90	3.24	4.79
On Time	0.88	0.82	2.91	4.61
Late	0.68	0.78	3.06	4.52

**Table 4-7:** Wait time for patients who arrive early, on time, and late when given first come, first served priority in 15 minute blocks with average arrival rate of 13.33 patients per hour.

Taking into account that late arrivals should not be rewarded for their tardiness, we implement an ordering of admittances (Table 4-8) from the waiting room to an imaging room. In Table 4-9, we see that by applying the priority ordering the wait times are redistributed, thereby reducing the wait times of early and on time patients. The priority ordering ensures that late arrivals have the longest wait time of any group. We next tried to determine if it is possible to reduce the wait time of block scheduling by admitting patients in a smarter order. By admitting patients to an imaging room based on their treatment plan, a 4% improvement in wait times can be achieved over simple first come, first served. This is equivalent to the improvement for the early and on-time arrivals in the priority ordering from Table 4-8.

Order	Description
1. Early/On Time Arrivals from Previous Block	First priority is given to patients, who arrived early or on time, still waiting for treatment from previous blocks.
2. Early/On Time Arrivals from Current Block	Following the completion of early and on time patients from the previous block, patients from the current block arriving early or on time are treated FCFS.
3. Late Arrivals	After all patients arriving early or on time from all previous blocks and the current block are treated, late arrivals are seen FCFS.
4. Early Arrivals from Next Block	Finally, after all patients in the waiting room from previous blocks and the current block have been assigned to an imaging room, early arrivals from the next block can be seen.

**Table 4-8:** List of priority groups when admitting to an imaging room. Within each group, patients are given priority by arrival order.

	Waiting Room	Imaging Room	Gantry Room	Total
Early	0.59	0.83	3.18	4.60
On-Time	0.79	0.79	2.92	4.50
Late	0.80	0.81	3.07	4.68

**Table 4-9:** Wait time experienced by patients who arrive early, on-time, and late when served using priority listed in Table 4-6 with 15 minute blocks with average arrival rate of 13.33 patients per hour.

The major strength of block scheduling is its resilience to high uncertainty of arrival times. Due to the traffic conditions around the site of the treatment center we wished to test performance of block scheduling with greater variability of arrival time. Given the increased variability of outpatient arrival time during rush hour it is reasonable to want to

schedule inpatients during this time window to reduce the commute time of outpatients coming for treatment and to minimize any disruption caused by their potentially increased tardiness. As seen in Table 4-10, however, block scheduling is robust, having less than 4% longer wait times over the course of the day. It should be noted that while scheduling inpatients for treatment during high traffic hours may be more convenient for patients, it does not significantly affect their wait times.

	Waiting Room	Imaging Room	Gantry Room	Total
Normal Conditions	0.79	0.86	3.07	4.71
High Traffic (even mix)	0.84	0.92	3.13	4.89
High Traffic (inpatients)	0.93	0.88	3.10	4.90

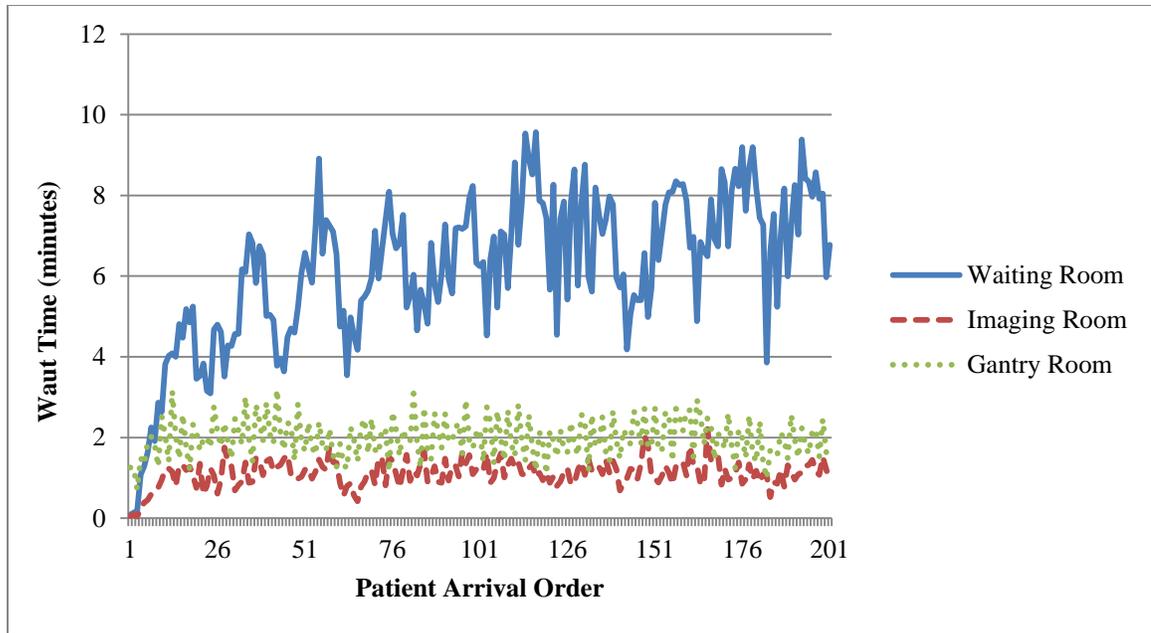
**Table 4-10:** Wait time experienced by patients with 15 minute blocks with average arrival rate of 13.33 patients per hour with a 75% outpatient mix. Simulated days include either a constant late arrival function (Normal Conditions) or include increased traffic for rush hour periods (High Traffic). Under high traffic conditions the patient mix either remained the same as the rest of the day (even mix) or switched to treating only inpatients (maintaining a daily average of 75% outpatients).

Thirty seven and one half percent of outpatients will be affected by the high traffic conditions when out patients remain uniformly mixed with inpatients throughout the day. These outpatients experience an increased probability of late arrival and a greater variance of arrival times. Despite these changes less than a 4% increase in average wait time is experienced. Similar to using block arrivals with only inpatients, we see an increase in

waiting room time when inpatients are scheduled to arrive during the hours of peak traffic.

#### **4.5.3 Equipment Failures**

The facility layout uses remote positioning of patients to help transfer patients from the imaging room to the gantry room. Each imaging room is associated with a single gantry room and vice versa. When an imaging room or gantry room is out of operation, the facility changes from a five imaging and five gantry room facility to a four imaging and four gantry room facility. If there is an equipment failure, it is important that patients still receive their scheduled treatment. Price et al. (2013) explored the differences in throughput of multiple facility configurations. They found that there is a decreasing return in throughput for each additional gantry room. Thus, closing a gantry or imaging room will not reduce throughput by 20%. In fact, a fifth gantry room allows for only 5.7% greater throughput. The patients seen in 15 hours of operation of the five gantry room system, would take four gantry rooms an additional 50 minutes. If patients are not rescheduled, this translates to an extra four minutes in the waiting room for each hour a gantry room is closed. Taking into account patient absenteeism, about one patient per 90 minutes needs to be rescheduled in order to maintain short wait times. Thus, an inpatient can be selected to be rescheduled in order to minimize the inconvenience to



**Figure 4-11:** Accumulation of wait time throughout day with one gantry room closed, 75% outpatient, and rescheduling one patient every 90 minutes.

commuting patients. Figure 4-11 shows the average wait time of 100 simulated days for a mixture of 75% outpatients.

#### 4.5.4 Special Needs

Pediatric patients have been shown to receive some of the greatest marginal benefits of proton therapy when compared to traditional IMRT (Mirabell et al. 2002, St Clair et al. 2004). The benefit of being able to deliver a radiation dose to a precise location with little incident radiation to surrounding tissue is contingent upon the radiation being delivered to the correct location. This is why an image is taken before treatment and why patients are immobilized. Pediatric patients require anesthesia in order to minimize movement during the imaging and treatment processes. The anesthesia is applied before patient immobilization in the imaging room. The patient remains anesthetized until he/she leaves

the gantry room to enter a recovery room. The anesthesiology team follows the patient from the imaging room, to the gantry room, to recovery, where a nurse stays with the patient until fully recovered. The bulk of the team's work is done in the imaging room, adding an extra five minutes to the preparation time. There is an extra two minutes of time in the gantry room, though beam delivery and gantry arm rotation times remain unchanged. As noted previously, a five gantry room facility has extra capacity. Despite the extra time required in the imaging room and the gantry room a single anesthesiology team does not affect throughput, while two teams will have a slowing effect of less than 2%. A single anesthesiology team with equipment in a single imaging room and gantry room is able to treat 8.5% of patients (or about 1.13 patients per hour). In order to minimize the amount of time a gantry room is unoccupied, the patient preceding pediatric patients should have long (three beam) treatment plans.

#### **4.6 Conclusions**

Due to the excess capacity of a five gantry room facility, adjusting for equipment failures requires the rescheduling of one out of 20 patients. This excess capacity, allows for patients with special needs to be treated without affecting throughput. We show that, while it is important to take into account all prior knowledge when scheduling a patient, waiting time improvements from doing so are reduced somewhat by the reality of patient tardiness and absenteeism. There is little extra wait time added (about 6%), in switching from individually scheduled arrivals to block scheduling in the presence of early and late patient arrivals.

## **5 THE IMPACT OF OPERATIONS RESEARCH ON PROSTATE CANCER RESEARCH**

### **ACKNOWLEDGEMENTS:**

This is research conducted with:

Brian Denton, University of Michigan  
Bruce Golden, University of Maryland  
Edward Wasil, American University

### **ABSTRACT**

Each year over 900,000 men worldwide are diagnosed with prostate cancer. In the United States alone, we spend almost \$12 billion in treatment. With so many men impacted by the disease and so many resources being spent on treatment operations research has an opportunity to make a big impact in many ways. Operations research techniques have been applied to screening with prostate specific antigen, detection of the disease through MRI, and treatment of the disease. We analyze the research done using operations research techniques and seek to quantify the impact it has had on the treatment of prostate cancer.

## 5.1 INTRODUCTION

Based on 2008-2010 data, approximately 15.3% of men will be diagnosed with prostate cancer during their lifetime. In 2014, about 233,000 men in the United States will be diagnosed with prostate cancer making it the most common form of cancer among men. In 2011, more than 2.7 million men in the United States were estimated to be living with prostate cancer. Prostate cancer is the second leading cause of cancer deaths among men with more than 29,000 deaths projected in the United States for 2014. All of the preceding numbers are taken from the Surveillance, Epidemiology, and End Results (SEER) program in the United States [17].

About \$11.85 billion is spent annually in the United States on the screening, detection, and treatment of prostate cancer (Mariotto et al. [13]). Clearly, improving the screening, detection, and treatment of prostate cancer has the potential to increase the lifespan and improve the quality of life for thousands of men.

Over the years, the battle against prostate cancer has been joined by researchers and practitioners who have used a wide variety of operations research (OR) models and methods to help screen, detect, and treat the disease. Specifically, a search on the Web of Knowledge using “prostate/prostatic and operations research method” revealed 402 published papers in 13 categories over the past 15 years or so (see Table 5-1). The citation count by year for these 402 papers is given in Figure 5-1. Clearly, this group of papers has been cited thousands of times since 2000.

In order to convey the importance and impact of OR methods and models in prostate cancer research, we annotate a representative set of 49 papers (selected from the 402 published papers). These papers were chosen because they deal with prostate specific

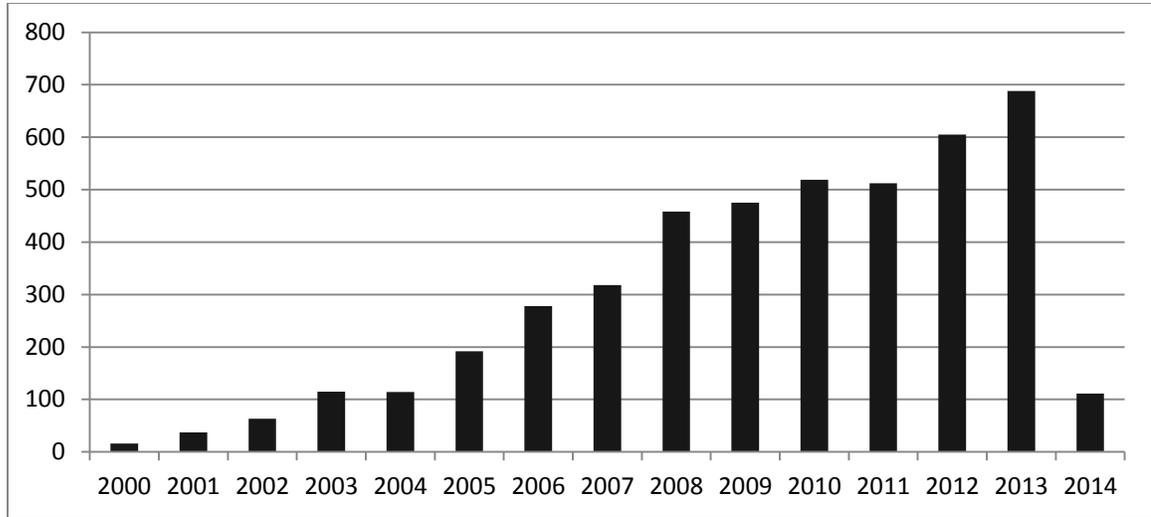
OR Model/Method	Number of Papers
Machine learning	125
Markov model	102
Markov chain	58
Integer programming	25
Linear programming	22
Optimal control	17
Microsimulation	17
Markov decision	16
Discrete event simulation	7
Dynamic programming	5
Analytic hierarchy process	5
Agent-based model	2
Multi-attribute utility model	1
Total	402

**Table 5-1:**Count of OR papers in prostate cancer research on the Web of Knowledge (2000-2014).

antigen screening (quantifying the effects of screening, predicting prostate cancer, screening policy), detecting tumors (improving biopsies, using magnetic resonance imaging to detect prostate cancer), and treatment (staging, active surveillance, hormone therapy and tumor size, brachytherapy, patient choice, recurrence, costs). They use almost all of the models and methods listed in Table 5-1 and some techniques that are not (e.g., nomograms). A categorization along these lines is given in Table 5-2. These papers are well cited (more than 1,900 citations in the Web of Knowledge as of April 24, 2014), interesting, and diverse in their applications.

In Sections 2, 3, and 4, we discuss OR models and methods for screening, detection, and treatment, respectively. Annotated papers listed in the bibliography are cited with ( ), while papers not appropriate for the bibliography are cited with [ ] and listed in the refer-

ences. In each section the annotations are given in reverse chronological order, starting with the most recent paper. We provide conclusions in Section 5.



**Figure 5-1:** Number of citations by year for 402 papers using OR methods and models on prostate cancer research from the Web of Knowledge search (April 24, 2014).

**Table 5-2:** OR models and methods used in prostate cancer research.

<u>Area</u>	<u>Paper Number in Bibliography</u>	<u>Model/Method</u>
Data Mining	1	KNN
	4	GA, KNN, PCA, SVM
	6	CL
	7	SVM
	8	RPART, RT
	17	CART, RT
	20	QUEST, RT
	24	ICA, NMF, PCA, PLS, SVM
	26	ANN
	30	ANN
	33	ANN, SVM
	37	ANN, NOM
	38	ANN, FKNN
	39	ANN, FKNN
	40	ANN, NOM, CART, RT
	44	CART, RT

Table 5-2. (continued)

Decision Aiding/ Analysis	3 5 11 29 41	DT MAU MM AHP MAU
Mathematical Programming	15 16 27 28	GA GA MIP MIP
Simulation	9 10 12 13 14 18 19 21 36 42 46 47 48 49	MIC, MM MC MIC, MM MIC, MM MIC, MM MIC, MM MIC MC MC, MM TDS MX MIC, MM MDP MDP
Statistics	2 22 23 25 31 32 34 35 43 45	LR NOM NOM KF LR NOM TSLS LR NOM LR

---

AHP	Analytic Hierarchy Process	MIP	Mixed Integer Program
ANN	Artificial Neural Network	MM	Markov Model
CART	Classification and Regression Tree	MX	Mixture Model
CL	Clustering	NMF	Nonnegative Matrix Factorization
DT	Decision Tree	NOM	Nomogram
FKNN	Fuzzy K-Nearest Neighbor	PCA	Principal Components Analysis
GA	Genetic Algorithm	PLS	Partial Least Squares
ICA	Independent Component Analysis	QUEST	Quick, Unbiased, and Efficient Statistical Tree
KF	Kalman Filter	RPART	Recursive Partitioning and Regression Tree
KNN	K-Nearest Neighbor	RT	Regression Tree
LR	Logistic Regression	SVM	Support Vector Machine
MAU	Multi-Attribute Utility	TDS	Three Dimensional Simulation
MC	Monte Carlo Simulation	TSLS	Two Stage Least Squares
MDP	Markov Decision Process		
MIC	Microsimulation		

## 5.2 Prostate Specific Antigen Screening

Since the FDA approved the PSA test for prostate cancer in 1994, millions of men have been tested. In the 2010 census, there were over 67 million American men over the age of 40 (Howden and Meyer [10]) whose screening decisions might be affected by screening recommendations. In 1980, prior to PSA screening, there were 106.0 new cases of prostate cancer per 100,000 men (SEER [17]). By 2000, after PSA screening was introduced, the rate jumped to 183.1, an increase of 72% [17]. Screening accounts for 89% of new prostate cancer diagnoses, with screen detected tumors being more clinically localized than clinically detected tumors (tumors detected due to the onset of other symptoms) (Hoffman et al. [9]). In 2009, Medicare spent an estimated \$447 million on PSA screening (Ma et al. [12]). However, the benefits of screening are unclear. In 2012, the U.S. Preventive Service Task Force (Moyer [14]) recommended against PSA screening in men of average risk due to its limited benefit and potential harm from unnecessary treatment. There are three important questions that the field of operations research has helped address with respect to prostate cancer screening. What are the effects of prostate cancer screening? How can the use of the PSA test in detecting prostate cancer be improved? What should be the policy for PSA screening?

### 5.2.1 Quantifying the Effects of PSA Screening

The benefit of PSA screening is the early diagnosis of prostate cancer. The time between when the cancer is diagnosed by screening and when it would have presented clinically in the absence of screening is the lead time. The lead time is important because early treatment of the cancer can help improve outcomes.

Typically, prostate cancer is slow growing with a low mortality rate. Many men diagnosed with prostate cancer due to screening may never have been affected by prostate cancer within their lifetimes if the cancer had gone undetected. Men who are diagnosed due to screening and would not have presented the clinical symptoms within their lifetimes are referred to as having been *overdiagnosed*. Overdiagnosis can lead to overtreatment, i.e., treatment of a likely indolent cancer, resulting in some men unnecessarily living with the side effects of surgery or radiation therapy, with potential negative effects such as urinary and bowel dysfunction, erectile dysfunction, and loss of fertility. An important question is: Do the benefits from early diagnosis outweigh the risks of overdiagnosis?

There have been two major clinical trials focused on the effect of PSA screening on mortality rates from prostate cancer. The European Randomized Study of Screening for Prostate Cancer (Schröder et al. [16]) found a 20% decline in mortality due to prostate cancer as a result of screening. The U.S. Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (Andriole et al. [2]) found that screening had no statistically significant effect on reducing mortality. These conflicting results have contributed to the controversy over the benefits and harms of prostate cancer screening. These contradictory results have been attributed to a number of causes of bias, such as contamination of the control arms by men who received screening.

The failure of clinical trials to provide definitive answers about the merit of PSA screening provides motivation for the use of OR models to answer these questions. We summarize six papers that used statistics, Markov modeling, and simulation to estimate the effects of PSA screening on mortality, quality of life, and lead time.

---

Gulati et al. (19) used a simulated population to build a nomogram for estimating the probability that a screen-detected prostate cancer would not have been diagnosed within the patient's lifetime. Overdiagnosis rates in the United States are estimated to be between 23% and 42% for screen-detected prostate cancer. This high rate of overdiagnosis leads to overtreatment which can be detrimental to a patient's health and happiness. Overdiagnosis is not directly observable. After an individual has screen-detected prostate cancer and has been treated, we do not know if the cancer would have been diagnosed within the individual's lifetime without screening. Using a microsimulation model of the progression of the disease, a population of 10,000 simulated prostate cancers was developed to train a logistic regression model indicating if a patient was overdiagnosed. Patient age, cancer Gleason score, and PSA level were used to predict if the patient had been overdiagnosed. The probability of overdiagnosis increased with age, decreased with PSA level, and decreased if the Gleason score was greater than or equal to seven. The nomogram had an area under the ROC curve (AUC) of 0.75.

Estimates of lead time and overdiagnosis due to PSA screening varies widely in the literature. Draisma et al. (9) used three independently developed models to compare predicted results for lead time and overdiagnosis of prostate cancer. The lead time can be interpreted three ways: lead time for non-overdiagnosed cancers only, censored lead times for both non-overdiagnosed and overdiagnosed cancers with lead time for overdiagnosed patients stopping at death, and uncensored lead times (for overdiagnosed patients, these continue until the patients would have been clinically diagnosed). The au-

thors used two microsimulation models from the literature and a statistical mixed model (Tsodikov, Szabo, and Wegelin (46)) to compare the estimates of lead times and overdiagnosis. A microsimulation model simulates the health outcomes for each patient of the population. The overdiagnosis estimates varied from 23% to 42% of all screen-detected cancers, which is consistent with population-based trial estimates. The non-overdiagnosed lead times ranged from 5.4 to 6.9 years, which allows for treatment to begin at an earlier stage, thereby improving outcomes. The censored lead time ranged from 5.7 to 7.8 years (this includes the time that patients who would die from other causes are diagnosed with prostate cancer). The uncensored lead time ranged from 7.2 to 10 years (this includes the time that patients who die from other causes would gain before diagnosis without PSA levels).

Etzioni et al. (13) used the microsimulation model to measure the effect of PSA screening on advanced stage prostate cancer and prostate cancer mortality. In 1986, PSA screening started in the United States. From 1990 to 1999, there was a 21% decline in observed mortality from prostate cancer. The model compared outcomes with and without PSA screening (with some percentage of men receiving biopsies based on PSA level). PSA screening accounted for 80% of the observed drop in the incidence of advanced prostate cancer. PSA screening only accounted for two-thirds of the 21% drop in mortality from prostate cancer.

Schröder and Kattan (37) analyzed the results of 36 nomograms and ANNs from the literature to compare their efficacy against using PSA level alone. The authors reviewed 23 papers that used three or more variables for risk assessment in the form of a nomogram or ANN. Model accuracy was compared using the AUC when available; otherwise sensitivity and specificity were used. The studies drew from different populations, with some including referred patients and those participating in screening studies. The populations ranged in size from 151 to 8,851 patients. The models considered a variety of input variables, depending on available information. Age, family history, and PSA velocity often lacked significance and were not included in the final models. Model validation on external populations is an important step in arguing generalizability. Ten of the 36 models were externally validated on a total of 16 external populations. In 13 external validations, the AUC decreased, in two it increased, and in one it stayed the same. In general, the models improved the AUC over PSA alone by approximately 0.10. The authors raised important questions about the independence of the input variables, some of which, like age, prostate volume, and PSA level, have relationships that are well understood.

Tsodikov, Szabo, and Wegelin (46) constructed a statistical model to estimate lead time, overdiagnosis, and other relevant characteristics of prostate cancer screening. The authors used a three-stage model of the natural history of the disease (disease-free stage, pre-clinical stage, clinical stage). Overdiagnosis was defined as the fraction of screening detected cancers that would not be detected in the absence of screening. Lead time was defined only for non-overdiagnosed patients. The study was limited to men over 50, since the probability of prostate cancer is small for men under 50. Data on 350,000 cases of

prostate cancer were taken from the Survey and Surveillance, Epidemiology and End Results (SEER) database and population counts for the relevant areas of diagnosis. To simulate PSA testing schedules to be used by the statistical model, a simulator for PSA schedules from the National Cancer Institute's Statistical Research and Applications Branch based on data from the National Health Interview Survey and SEER was used. The statistical model estimated a 6-year mean lead time and 25% overdiagnosis among detected patients.

Etzioni et al. (14) applied a microsimulation of PSA screening and prostate cancer development to estimate the rate of overdiagnosis in a hypothetical cohort of men. A hypothetical cohort based on census data of two million men between 60 and 84 years old in 1988 was used. The simulation was run on the group with and without screening. Testing and detection rates for the model varied by year, age, and race. The model used three, five, or seven years mean lead time to estimate how long before the cancer would have been clinically detected without screening. Overdiagnosis was estimated at 28.8% for white men and 43.8% for black men using five-year and seven-year lead time estimates, respectively. The overdiagnosis rate was sensitive to the mean lead time provided by screening, but not to the relative frequency of screen-detection. When the lead time was reduced to three years, the overdiagnosis rate dropped to 17.7% for white men and 20.3% for black men.

---

### 5.2.2 Predicting Prostate Cancer from the PSA Level

When the FDA approved the use of PSA level to test for prostate cancer in 1994, a threshold of 4 ng/mL was used as the upper limit for a normal PSA level (FDA [7]). Using a PSA threshold of  $>4\text{ng/mL}$  to indicate a need for a biopsy, about 20% of prostate cancers would be detected and 30-40% of patients without prostate cancer would be above this threshold (Greene et al. [8]). In 2009, the American Urological Association (AUA) recommended against using a single threshold level, and advocated the use of additional information to improve detection (Greene et al. [8]). For example, PSA level changes with age.

Improving the predictive value of PSA screening means fewer unnecessary biopsies which carry the potential for infection, temporary erectile dysfunction, and lingering urinary problems. In the eight papers summarized, researchers applied machine learning and statistical methods to improve the prediction of prostate cancer by increasing sensitivity (probability of correctly identifying prostate cancer) and, in some cases, identifying only clinically significant prostate cancer (having Gleason Score greater than six).

---

Prostate biopsies can have one of five outcomes: benign, atypical small acinar proliferation (ASAP), high-grade prostatic intraepithelial neoplasia (PIN), non-significant prostate cancer, or clinically significant prostate cancer. Benign means that the biopsy was non-cancerous. ASAP means the biopsy result was indeterminate and could not be classified as benign or malignant. PIN is thought to be a precursor of a malignant tumor. Non-significant prostate cancer is considered not life threatening, while clinically significant prostate cancer typically requires treatment. Lawrentschuk et al. (26) constructed a polychotomous logistic regression model and an artificial neural network model to predict biopsy results. Age, PSA level, DRE, presence of hypoechoic lesion (a visible abnormality) during transrectal ultrasound (TRUS), and TRUS prostate volume for 3,025 men with PSA level less than 10 ng/mL who underwent a TRUS-guided biopsy were used to construct the models. The models were trained on two-thirds of the data and tested on the remaining one-third. The regression and neural network models correctly identified benign tumors 86% and 88% of the time, respectively and clinically significant prostate cancer 65% and 66% of the time, respectively. Neither model correctly identified any ASAP/PIN outcome. Both models predicted only 2% of non-significant prostate cancer correctly. Prostate volume and a positive TRUS lesion were the most significant in correctly identifying patients with clinically significant prostate cancer from benign biopsy results with odds ratios of .19 and 5.2, respectively. The authors concluded that additional predictors would be necessary to correctly distinguish among the five outcomes.

Many predictive models for prostate cancer biopsy results are developed using a population of patients referred for early cancer detection or urinary tract symptoms. Therefore,

these models may not generalize to all men being screened for prostate cancer. Sooriakumaran et al. (43) used a population of screened patients to construct two predictive models in order to reduce unnecessary biopsies. The authors began with 3,838 men from the Tyrol Prostate Cancer Screening Study. They removed patients missing total PSA (tPSA) level, DRE, prostate volume, or percent free PSA (fPSA), leaving 2,271 patients. The authors used a urologically referred population of 599 patients, from the Weill Cornell Medical College, for external validation. Multivariate logistic regression models were constructed where Model 1 used age, DRE, and tPSA, and Model 2 added percent fPSA. Nonlinear relationships between the features and the outcome were evaluated using a multiple fractional polynomial method. A bootstrap method with 200 bootstrap samples was used during internal validation. The AUC for the two models were .691 and .710, respectively. The authors then constructed nomograms based on the two models. If patients with a 10% or greater risk of cancer were biopsied, then 30 biopsies (1.3%) would be avoided and two cancers (.3%) would be missed by Model 1 applied to the Tyrol sample. Model 2 avoided 95 biopsies (4.2%) and missed six cancers (.9%). When applied to the urologically referred population for patients with 20% or greater risk of cancer, Model 1 avoided 126 biopsies (21.0%) and missed 18 cancers (9.8%). Model 2 avoided 169 biopsies (28.2%) and missed 19 cancers (10.3%). The authors concluded that their models had sufficient predictive power to aid in clinical decision making regarding the need to biopsy.

Gülkesen et al. (20) developed a decision tree to classify the risk level of prostate cancer for 750 patients with serum fPSA and PSA levels less than or equal to 10 ng/mL who underwent prostate biopsy at the Urology Department of Akdeniz University Hospital. Age, PSA level, free PSA level, percent-free PSA, DRE 1 (with three possible outcomes for prostate cancer: negative, suspicious, or positive), and DRE 2 (with two possible outcomes for prostate cancer: negative or not negative) were used to construct the model. The QUEST algorithm (quick, unbiased, and efficient statistical tree) produced a model using a training set of 562 patients. The algorithm classified the patients into five groups with the risk of cancer ranging from 0% to 25%. The AUC was .62 compared to .68 for logistic regression using the same features. The lowest risk group, with no prostate cancer in either the training or testing sets, identified by the decision tree had patients with a PSA level less than or equal to 5.98, a negative DRE, and an fPSA level greater than .81.

Nam et al. (32) constructed a nomogram to assess the risk of a patient having prostate cancer with a Gleason score greater than or equal to seven. In a data set of 3,010 men with PSA levels greater than 4.0 ng/mL or abnormal DREs referred to the Prostate Centers of the University of Toronto, 2,700 had PSA levels less than 50 ng/mL, were able to provide complete information including family history of prostate cancer, and consented to participate in the study. An additional 408 men with a PSA level less than or equal to 4.0 ng/mL agreed to undergo a biopsy and were added to the study. Age, ethnicity, family history, symptom score, PSA level, ratio of free PSA to total PSA, and DRE were used as predictor variables. A model to predict the probability of no cancer, low-grade cancer (Gleason score less than or equal to six), and high-grade cancer (Gleason score greater than or equal to seven) was constructed using ordinal logistic regression. The logistic re-

gression model used data from 2,108 patients and was tested on 1,000 patients. A nomogram was designed using the results of the logistic regression model. The AUC for predicting any prostate cancer was .74 based on the nomogram. The total AUC for predicting high-grade cancer was .77 based on the nomogram. The logistic regression model with the full set of predictor variables performed much better than the model based only on PSA levels and DRE results, which had an AUC of .59.

Aggressive prostate cancer, defined as having a Gleason score greater than or equal to seven, benefits from early detection and treatment. Spurgeon et al. (44) developed a classification and regression tree (CART) model to predict the presence of aggressive prostate cancer in patients. Age, PSA level, PSAD, DRE, race, family history, vasectomy, transrectal ultrasound (TRUS) findings, and prostate volume for 1,563 consecutively referred men from the Portland Veterans Administration Hospital were used in the training and testing of the CART model. The authors used 1,067 patients for training and 492 patients for testing. Their model identified 43 of the 47 cancer patients with aggressive prostate cancer (91.5% sensitivity) and 149 of the 445 patients with Gleason score less than seven (31.9% specificity). When their model predicted a Gleason score greater than or equal to seven (i.e., aggressive prostate cancer), it was correct only 12.7% of the time. When their model predicted a Gleason score below seven, it was correct 97.4% of the time.

Thompson et al. (45) used a logistic regression model based on data from the Prostate Cancer Prevention Trial to create an online risk calculator based on data for 5,519 men in the control arm of the trial. The statistically significant model risk factors were race, age, family history, PSA level, digital rectal exam result, and whether the patient had a prior biopsy or not. The AUC for out of sample data, based on 4-fold cross validation, was approximately 70%, indicating satisfactory discrimination. A more recent revision of the model in 2012, referred to as PCPTRC 2.0, adds the ability to estimate the likelihood of low- and high-grade cancers (Ankerst et al. (2)). Statistics about the risk calculator website indicate hundreds of thousands of visits in the last 48 months, suggesting the calculator is commonly used by physicians and/or patients.

Garzotto et al. (17) developed a decision tree using classification and regression tree analysis (CART) to improve the accuracy of PSA tests. The authors trained their model using 1,433 consecutive referred patients who underwent an initial prostate biopsy procedure. They used 5-fold cross validation with the cost of misclassifying cancer as normal three times more than the cost of classifying a normal prostate as potentially cancerous. They mirrored the medical decision process used by physicians by first creating a decision tree using only PSA and DRE data. Then they added demographic data and family history to identify patients at high risk of prostate cancer. Their model identified 278 of the 291 cancer patients (95.5% sensitivity) and 334 of the 882 noncancerous patients (37.9% specificity). If their model predicted a cancerous biopsy, it was correct 33.7% of the time. If their model predicted a negative biopsy, it was correct 96.3% of the time. Furthermore, of the 15 cancer patients missed by the CART model, 14 had Gleason scores less than or equal to six.

---

### 5.2.3 PSA Screening Policy

Researchers have tried to improve screening policies for men in the United States by analyzing when to start screening, how often to screen, and at what PSA level to receive a biopsy. In 2009, the AUA recommended that screening start as early as age 40 (Greene et al. [8]). In 2013, the AUA updated their guidelines and recommended against screening men at average risk of prostate cancer between the ages of 40 and 54 (Carter et al. [4]). In the six papers summarized, researchers used optimization and simulation to compare the effects of different screening policies on mortality, quality-adjusted life years (QALYs), overtreatment, and lead time.

### 5.3 Detecting Tumors

PSA screening is relatively inexpensive, costing between \$17 and \$62 (Ma et al. [12]). However, it is not a definitive test, because its sensitivity is too low to justify treatment for a patient. A biopsy is necessary for definitive diagnosis of prostate cancer. Typically, a biopsy involves taking 12 samples of the prostate using hollow core needles. The results are reviewed by a pathologist to provide an assessment which includes the Gleason grade of the cancer and other clinical factors associated with patient health outcomes.

In some cases, additional diagnostic tests may be used to provide information about the size, grade, and/or location of a tumor. These may include magnetic resonance imaging (MRI), or the use of additional biomarkers. There are a number of new biomarkers that have been discovered, and some have recently received FDA approval,

such as the urine based PROGENSA prostate cancer antigen 3 (PCA3) test. Similar to PSA, the PCA3 test provides a score that can be used to predict the presence and grade of cancer (Filella et al. [6]).

---

Gulati, Gore, and Etzioni (18) used a microsimulation model of prostate cancer to compare 35 screening policies based on lives saved, overdiagnoses, and mean time of life saved. A microsimulation model simulates at the unit of individual patients. This model simulated the PSA growth and disease progression of 100 million men. Thirty-two screening strategies that used starting age (40 or 50 years), stop age (69 or 74 years), screening intervals (annual or biennial), and biopsy referral conditions (PSA level of 4 ng/mL, PSA level of 2.4 ng/mL, PSA level of 4 ng/mL or PSA velocity of 0.35 ng/mL per year, or PSA level greater than 95<sup>th</sup> percentile for age) were compared. The authors evaluated screening policies recommended by the American Cancer Society (ACS), the National Comprehensive Cancer Network (NCCN), and the medical literature. They compared results to a base plan of annual screening from age 50 to 74 with biopsy threshold set at 4.0 ng/mL. The NCCN plan had the greatest improvement on reducing cancer deaths and increased the mean time of life saved from 0.86 month to 1 month. However, the NCCN plan doubled the number of false negatives and increased the probability of over diagnosis from 3.3% to 6%. The ACS plan performed the same as the base strategy for lives saved and overdiagnoses, but reduced the number of tests conducted by nearly a quarter. Using age-specific PSA thresholds for biopsy referral reduced false positives by 25%, overdiagnoses by 30%, and lives saved by 17%.

Underwood et al. (47) constructed a simulation, based on a non-stationary, finite horizon Markov process, to model various PSA-based screening policies. Each screening policy was represented as a set of PSA thresholds for biopsy referral by age. They optimized these policies using a genetic algorithm and compared their results to policies presented in the literature. The simulation used 51,294 PSA levels from 11,872 men from Olmsted County, MN and age-specific death and incidence data from a variety of other sources. The authors judged the policies in terms of quality-adjusted life years (QALYs), giving a one-time penalty for unnecessary biopsies. The genetic algorithm converged on a plan that had less than annual screening between the ages of 54 and 76, with the threshold for a biopsy based on a PSA level set lower than the current standard. The optimized policy slightly improved the results of the best policy given in the literature, and indicated the maximum number of QALYs that screening policies could provide. Underwood et al. suggested that screening over a shorter time period, with lower PSA thresholds, might do the most to improve QALYs.

Zhang et al. (48) constructed a partially observable Markov decision process to examine the benefits PSA screening to patients and society. Patients progress through health states (no cancer and cancer are not directly observable; treated, metastasis, and death are considered known) and observable PSA intervals. Each year a decision is made to perform a biopsy or defer the decision until the next time period. The objective was to maximize QALYs for a patient. Detriments to the quality of life were occurrence of biopsy, treatment upon detection of cancer, long-term complications resulting from treatment, and symptoms from metastasis and its treatment. A secondary objective, from the societal

perspective, is to maximize the expected monetary value of the QALYs minus the costs of screening, biopsies, and treatments. Prostate cancer probabilities conditional on PSA level were estimated using 11,872 patients from Olmsted County, Minnesota. A finite fixed-grid method was used to obtain the optimal policies for patients and society. The optimal policy from a patients perspective (maximizes QALYs) had screening stop at 76 and improved QALYs by 0.131 over no screening and 0.165 over current policy (note that current policy underperforms no screening. The optimal policy from a societal viewpoint stops screening at age 71 and improved QALYs by 0.110 over no screening and 0.161 over traditional guidelines.

Zhang et al. (49) used the partially observable Markov decision process model to examine prostate biopsy referral decisions. The optimal policy often took the form of a control-limit type policy, that is, a biopsy is performed only when the belief state for cancer exceeds a threshold value. There were three conditions for a control-limit type policy: annual probability of prostate cancer incidence is less than or equal to a multiple of the probability of treated patients developing metastasis, the annual probability of death from other causes is non-decreasing, and the annual probability of death from metastatic prostate cancer is non-increasing. There exist conditions when it is optimal to discontinue biopsy referral at a specific age. The optimal policy was estimated to improve the American Urological Association's age-adjusted guidelines by .115 QALYs (.306%). The authors noted that personalized utility assessment, consideration of comorbidity, and family history of prostate cancer might be added to the model in the future.

Prostate cancer screening strategies that use a PSA score vary in the age a patient begins screening, the threshold result from a PSA test to recommend a biopsy, and the time intervals to receive screening. It is impractical to run randomized screening trials for all possible screening strategies. Ross et al. (36) constructed a Monte-Carlo simulation that tested a range of screening policies based on the number of prostate cancer deaths prevented. They tracked the number of PSA tests and biopsies per 1,000 men taken over the entire run of the simulation. The simulation was based on a Markov model that transitioned between no prostate cancer, three levels of cancer, and mortality due to prostate cancer and other causes. They simulated populations of one million men starting at 40 years of age for 40 years. Eight policies (including a baseline of no screening) with different PSA thresholds, screening frequency, and age recommendations were tested. The medical standard at the time called for annual PSA testing from age 50 to 75 with a PSA score above 4 ng/mL requiring a biopsy. The authors constructed a policy of screening at ages 40, 45, and 50, and then every two years until 75 with the PSA level at 4 ng/mL. Their new plan reduced the lifetime biopsies per 1,000 men from 600 to 450 and the average number of PSA tests in a lifetime from 10.5 to 7.5. Prevented deaths increased to 3.3 per 1,000 men from 3.2.

Etzioni et al. (12) constructed a computer model to measure the effect of PSA screening strategies on years of life saved, PSA tests performed, false-positive PSA tests, and rate of over diagnosis. The authors constructed a simulation model of prostate cancer and its progression in the population using real-world data from the surveillance epidemiology and end result (SEER) registry. A Markov model simulated the disease progression of five-year birth cohorts. The mortality rates for all causes of death other than prostate can-

cer and deaths due to prostate cancer were computed for each cohort. Clinical histories were developed for the patients that included date of clinical diagnosis of prostate cancer (in the absence of screening) and date of death not due to prostate cancer. Not all men develop prostate cancer. If the clinical diagnosis of prostate cancer was made before death from other causes, a history of disease progression was needed. Using the simulated progression of prostate cancer, natural histories were constructed for patients that included dates for disease onset and progression of the cancer. PSA levels that resulted from screening were simulated and grew at different rates based on the health state of the prostate. Five screening strategies were tested: annual PSA screening with age-specific biopsy thresholds, biannual PSA screening with age-specific biopsy thresholds, annual PSA screening with  $>4\text{ng/mL}$  threshold, biannual screening with  $>4\text{ng/mL}$  threshold, and screening every five years with threshold of  $>4\text{ng/mL}$ . Biannual screening results in about half the total number of screenings and number of false-positive screenings, and captures 95% of the cases that would have been caught with annual screenings. Using age specific PSA levels reduces false positives by 13.7%, but reduces the percentage of patients that are clinically detected from 63% to 54.8%.

---

### 5.3.1 Improving Biopsies

Each year, approximately one million prostate biopsies are performed in the United States (Welch et al. [19]). Studies have shown that the standard sextant transrectal ultrasound-guided needle biopsy has a false negative diagnosis rate over 20% (Bankhead [3], Rabbani et al. [15]). Since the sensitivity of the standard biopsy is less than 80%, multiple biopsies may be needed to confirm prostate cancer. A biopsy is a painful, invasive procedure that has a risk of infection, with 4% of patients developing an infection. Within 30 days of a prostate biopsy, 6.9% of patients are admitted to the hospital (Loeb et al. [11]). To prevent the need for multiple biopsies, it is important to raise the accuracy of a biopsy. In the paper summarized, the authors tried to reduce the false negative rate by analyzing tumor locations and developing an optimal biopsy procedure.

---

Sofer, Zeng, and Mun (42) developed optimal biopsy procedures for fixed numbers of needles. The authors used 301 surgically removed prostates to generate a 3D statistical distribution of cancer occurrence in the prostate. The prostate was divided into 48 zones and 6,000 subzones. A biopsy protocol specified the number of samples to take and the zones where the samples are taken. There is variability as to the exact position and angle the physician aligns the needle for the biopsy. The authors optimized the biopsy protocols to maximize the probability of detecting cancer. Using their 3D model, they estimated the detection rate for the current sextant method at 67.3%. The optimized detection rate for a biopsy with six needles confined to the posterior of the prostate was 78.8%. By increasing the number of samples taken (to as high as 12), the authors showed that the detection rate could be as high as 85.5%. Biopsy procedures were developed for small and large

prostates that had better detection rates than the current sextant procedure. Further work will examine the possibility of alternate protocols based on age, race, and PSA level.

---

### **5.3.2 Using an MRI to Detect Prostate Cancer**

Using an MRI to detect prostate cancer has been shown to reduce the number of men needing a biopsy by 51%, lower the false negative rate, and identify low-risk tumors (Thompson et al. [18]). Each year, unnecessary biopsies are estimated to cost \$2 billion. An MRI could help prevent unnecessary biopsies, thereby reducing this cost (Ahmed et al. [1]). An MRI does not carry the same risks as a biopsy such as infection, bleeding at the biopsy site, and painful urination. Therefore, it seems reasonable to have an MRI and analyze the results before deciding to have a biopsy. In the four papers, the authors analyzed the MRI data of patients who had a tumor removed (the procedure is known as a prostatectomy) and used machine learning to automate the detection of tumors.

---

Proton magnetic resonance spectroscopic imaging (MRSI) allows for the detection and quantification of biochemical markers within the prostate and has been shown to be an improvement over conventional magnetic resonance imaging (MRI). A 3D grid divides the prostate into voxels (a voxel is a volumetric element that serves as the unit of analysis). Each voxel has its own spectra data from the MRSI. Matulewicz et al. (30) constructed an artificial neural network (ANN) model to automatically detect cancerous voxels in the prostate using MRSI data. MRSI data from 18 patients with positive biopsies for prostate cancer who later underwent radical prostatectomy was used to construct the model. A total of 5,308 voxels were used. One hundred forty eight voxels were labeled suspicious by an experienced spectroscopist, and 129 of the 148 voxels were confirmed to have cancer following the radical prostatectomy. Only voxels identified as cancerous by both histopathological maps and the spectroscopist were labeled as cancerous. The ANN model had input, hidden, and output layers, each with a single node. Two types of models were constructed. The first model had 256 variables based on the spectra. The second model added the percentage of the voxel that was in the peripheral zone, transition zone, periurethral region, and outside the prostate. Seventy percent of the data was used as a training set, 15% as a test set to prevent overtraining, and 15% as a validation set to measure performance. Since there is a random element in the generation of the ANN model, 100 models of each of the two types were applied to the same training and test sets. The model of each type that performed best on the test set was selected, and their performances on the validation set were compared. The AUC was .949 for model 1 and .968 for model 2. The sensitivity on the validation set for model 1 was 50% (identify-

ing 8 of 16 cancerous voxels) and 62.5% for model 2 (identifying 10 of 16 cancerous voxels). The specificity on the validation set for model 1 was 98.7% (identifying 770 of 780 noncancerous voxels) and 99.0% for model 2 (identifying 772 of 780 noncancerous voxels).

Anderson et al. (1) constructed logistic regression, nearest neighbor, and hybrid classifiers to predict the risk of prostate cancer using MRI data. Multi-parametric images from dynamic contrast MRI, diffusion weighted MRI, and MRSI were used to generate features (apparent diffusion coefficient, volume transfer constant, conventional average of T2 values, and spectroscopy scores) for 223 slices of prostates from 28 patients who had a radical prostatectomy (all of the prostate gland is removed). Gleason scores were assigned to each slice by an experienced radiologist. Leave-one-out cross-validation was used to separate the data into training and test sets. The models were trained to classify a slice as having a Gleason score between 0 and 4 or between 5 and 8. Logistic regression had an accuracy of 64.6% and an AUC of 0.66, and KNN had an accuracy of 74%. A hybrid approach was developed using linear regression with the number of cancerous neighbors from KNN. The hybrid approach had an accuracy of 77% and an AUC of 0.85. The hybrid approach identified the most aggressive cancers (Gleason score of 7 or 8) with an accuracy of 82% and an AUC of 0.86.

Parfait et al. (33) constructed a support vector machine (SVM) and a multilayer perceptron to classify magnetic resonance spectra (MRS) of prostates as healthy or pathological. MRS data from 22 patients with positive biopsies for prostate cancer were used. There were a total of 2,464 spectra, with 1,062 spectra localized in the peripheral zone where

the majority of prostate cancer lesions are typically found. The peripheral zone spectra were manually classified as undetermined (286), healthy (636), or pathological (140). A variety of preprocessing algorithms could be applied including phase correction, baseline correction, and normalization. Nine combinations of preprocessing algorithms and classification method were compared. The best average performance using 5-fold cross validation was SVM with preprocessing using phase correction and baseline correction. This combination correctly classified 82.2% of undetermined, 94.3% of healthy, and 75.0% of pathological spectra. When restricted to the healthy and pathological classifications, 4.5% are misclassified with 83.6% sensitivity and 98.1% specificity.

Kelm et al. (24) compared the use of quantitative and subspace feature extraction methods in linear and nonlinear machine learning classifiers for identifying prostate tumors with MRSI. A subspace method projects higher dimensional data onto a lower dimensional subspace. Two quantitative feature extraction methods from the MRS literature and a new quantitative method were compared to four subspace feature extraction methods— principal components analysis (PCA), partial least squares (PLS), independent component analysis (ICA), and nonnegative matrix factorization (NMF). These features were then used in either a linear classifier (logistic regression, generalized PLS, P-spline signal regression) or a nonlinear classifier (a random forest method that uses multiple decision trees, SVM, Gaussian processes). MRSI data from 24 patients were used to construct the models. Using published guidelines, 4,188 voxels were identified as healthy, undecided, or having a tumor. Cross-validation was performed by training on 23 patients and testing on the one remaining patient. None of the classifiers using quantification features outperformed a conventional ratio feature used by clinicians. Subspace methods improved the

performance of the classifiers, particularly the nonlinear classifiers. There was little difference among the performances of the four subspace methods.

---

## 5.4 Prostate Cancer Treatment

There have been great advances in the treatment of prostate cancer in the past 30 years. In 1985, the five-year survival rate was 75%; in 2006, it was 99.6% (SEER [17]). This increase is due to better treatments and earlier detection due to PSA screening. Surgery (prostatectomy) has long been the primary form of treatment. Surgery carries a risk of pain, bleeding, infection, and death. There is also a chance of incontinence and erectile dysfunction.

Other treatment options include external beam radiation therapy, brachytherapy, and hormone therapy. Often treatments are used in tandem. For example, hormone therapy is used to shrink a tumor before radiation treatment. Because prostate cancer can be slow growing, a patient with a low risk cancer (e.g., Gleason Score  $\leq 6$ , PSA  $< 10$ ) may forgo immediate treatment and elect to monitor the progression of the cancer (known as active surveillance) through regular clinical exams, PSA testing, and surveillance biopsies.

Operations research models have been used to improve the delivery of treatment, compare treatment options, and help patients make treatment selections.

### 5.4.1 Staging

Staging describes the classification of a cancer. This is important information when making a treatment decision. Prostate cancer can be slow growing, posing little risk to the life of some patients who have low risk based on clinical factors. Unnecessary treatment can result in worse health outcomes. Identifying a low-risk tumor can prevent unneces-

sary treatment, while identifying a high-risk tumor can encourage a more aggressive treatment strategy. We summarize four papers that used data mining to identify the type or severity of the prostate cancer.

---

Chandana, Leung, and Trpkov (4) constructed models with different combinations of automatic feature selection, sampling, and classifier to predict the stage of prostate cancer. Age, primary Gleason grade, secondary Gleason grade, PSA level, PSAD, DRE, TRUS, gland volume, positive biopsy core, total percent of cores involvement, and total cancer length in mm for 1,054 patients were used in the models. The prostate cancer was organ-confined in 934 patients. There was an extra prostatic extension in 120 patients. The authors generated 18 models with combinations of automatic feature selection (rough set features, PCA, GA-based continuous feature selection), sampling (under-sampling, synthetic minority over-sampling technique (SMOTE), combined under-sampling and SMOTE), and classifier (SVM, KNN). SMOTE generated synthetic examples of extra prostatic extension patients, based on those in the data set, that were used in training. Dempster-Shafer fusion produced classification probabilities based on the classification probabilities from multiple classifiers and generated new classifiers from combinations of the 18 models. A GA identified the best set of classifiers for fusion. SVM with rough set for feature selection and under-sampling had the best performance of the 18 models with an AUC of .8376. A Dempster-Schafer fusion of four models had a total accuracy of 90.1% and AUC of .8626.

Kattan et al. (22) developed a statistical model to predict the probability that prostate cancer is indolent or in need of treatment. Kattan et al. used data from 1,022 patients

treated with retropubic radical prostatectomy. Patients with a PSA level greater than 20, Gleason grade four or five, greater than 50% biopsy cores positive, total cancer in biopsy greater than 20 mm, or total benign tissue in all cores less than 40 mm were excluded as being unlikely to have indolent cancer. This left 409 patients. The authors constructed three statistical models. The base model used PSA level, clinical stage, and primary and secondary biopsy Gleason grades. The medium model added percent of cores that were positive, and ultrasound prostate volume to the features of the base model. The full model replaced percent of cores that were positive with millimeters of cancerous and noncancerous tissue. Bootstrapping was used to build the model and leave-one-out analysis was used to examine the predictive probabilities. The three models had AUC values of 0.64, 0.74, and 0.79, respectively. The full model predicted too many indolent cancers (17% too high). The authors then translated their statistical models into nomograms. A nomogram is a diagram that is used to calculate the probability of indolent cancer predicted by a specific model.

Bone scan and CT imaging are commonly used to detect metastatic cancer by identifying bone metastases and enlarged lymph nodes, respectively. This is important in determining whether treatment with curative intent, such as surgery, is appropriate. Merdan et al. (31) and Risko et al. (35) developed logistic regression models based on data from more than 80% of community urology practices in Michigan to estimate the probability of a positive bone scan and a positive CT scan, respectively. The models were found to provide very good discrimination, based on estimates of AUC greater than 80%. Furthermore, the models were used to estimate efficiency and effectiveness of proposed guidelines for determining which patients should receive imaging tests. Criteria of Gleason

Score > 7 or PSA > 20 (for bone scan guidelines) and Gleason Score > 20, PSA > 20, Clinical State > T2a (for CT guidelines) were estimated to result in fewer than 1% of patients having missed positive results. Implementation of these guidelines was estimated to substantially reduce the number of negative imaging results and the total number of imaging studies overall. The latter results are particularly important given the high cost of imaging.

---

#### 5.4.2 Active Surveillance

Active surveillance has become a viable alternative to treatment. Active surveillance involves close observation of a patient with frequent visits to a doctor for a DRE, a PSA test, and possibly a biopsy. If there is a change in the tumor, a patient may need to decide on a treatment. Active surveillance allows a patient to delay treatment, thereby avoiding any side effects. The two papers summarized compared active surveillance to different treatment options.

---

Eldefrawy et al. (11) compared the cost of active surveillance to common treatment options (radical retropubic prostatectomy, robotic assisted radical prostatectomy, external beam radiation therapy, brachytherapy) for low-risk prostate cancer. Screening for prostate cancer has led to an increase in diagnosis, particularly in early stage prostate cancer. Though 17% to 20% of men will be diagnosed with prostate cancer within their lifetime, only 3% will die from prostate cancer. Overtreatment can have a negative impact on patient quality of life due to treatment side effects such as erectile dysfunction and incontinence. Active surveillance delays treatment, while having the patient undergo surveil-

lance involving DRE, PSA testing, and biopsies. Procedure cost, professional fees, and inpatient costs were estimated based on Medicare reimbursement levels in the Miami area and the costs for a Miami area hospital for 2010. The authors estimated costs for treating low-risk prostate cancer over a 10-year period using a Markov model. The model considered the probabilities of various complications and recurrence from treatment options and the resulting costs. For example, each TRUS guided biopsy had a 1% chance of causing sepsis. Robotic assisted radical prostatectomy, the most commonly performed procedure for prostate cancer in the United States, had the second highest costs with a one-year cost of \$17,824 and a 10-year cost of \$22,762. Active surveillance had the lowest costs with a one-year cost of \$1,154 and a 10-year cost of \$13,116.

Hayes et al. (21) used a simulation model to compare the QALYs for men with low-risk prostate cancer associated with active surveillance and initial treatment (brachytherapy, IMRT, radical prostatectomy). Active surveillance involves regular PSA testing, DRE, and biopsies. Active surveillance delays treatment until the prostate cancer progresses or the patient chooses to begin treatment. In the United States, 16% to 40% of newly diagnosed prostate cancer patients meet the criteria for active surveillance. The authors constructed a state transition model that they analyzed using Monte Carlo simulation. The model considered probabilities of side effects associated with brachytherapy, IMRT, and radical prostatectomy and with the disease in general. Probabilities were estimated based on a systematic literature review. Patient utilities were estimated by asking men not diagnosed with prostate cancer using the time-trade-off method. A hypothetical cohort of 65-year-old men with newly diagnosed, clinically localized, low-risk prostate cancer was used as the patient population. Patients undergoing active surveillance had the best

performance with 11.07 QALYs, compared to 10.57 QALYs for brachytherapy, 10.51 QALYs for IMRT, and 10.23 QALYs for radical prostatectomy. Active surveillance had a higher risk of death from prostate cancer (11% compared to 9% for initial treatment). For those men undergoing active surveillance, 61% received treatment (brachytherapy, IMRT, radical prostatectomy) having been under active surveillance for a median of 8.5 years.

---

#### **5.4.3** Hormone Therapy

Hormone therapy can be used to shrink a cancerous tumor in the prostate before it is treated with radiation. Hormone therapy shrinks a tumor to a minimum size, after which the therapy loses effectiveness and the tumor starts growing. Less radiation needs to be delivered when a tumor is at its smallest, thereby reducing the side effects from the treatment. The paper summarized tried to identify when a tumor has reached its minimum size.

---

Lavieri et al. (25) used a dynamic Kalman filter model to predict when to begin radiation therapy. A Kalman filter is a recursive procedure that computes the optimal estimator of the state vector at each time period based on a series of noisy measurements. Since it is believed that the PSA level is lowest at the same time that the tumor's size is at a minimum, PSA level is used in lieu of tumor size. PSA level over time is modeled using a log quadratic curve, which has an average  $R^2$  of 0.9 for the 163 patients in the study. The authors estimated the prior distribution of the curve parameters, then the distribution of the estimated time of the nadir. A Kalman filter is used to update the estimates of the curve parameters. Finally, clustering was used to identify subgroups of patients with similar responses to hormone therapy. Protocol in British Columbia required that radiation therapy begin if the PSA level started to rise, if after four months the PSA level was below 0.05 ng/mL, or eight months after beginning the therapy. Four new policies were proposed that used the predictive model to determine if the nadir had a threshold probability of being reached within a fixed time period of PSA tests. Two of these policies outperformed the current policy; beginning treatment if the nadir is predicted to occur before the next PSA test or beginning treatment if the nadir is likely to occur within one month of the next PSA test. The mean absolute difference between the time of the nadir to the beginning of radiation therapy is 45 days for the current policy, but only 29 or 36 days for the two new policies.

---

#### 5.4.4 Brachytherapy

Brachytherapy places radioactive seeds inside the prostate. The placement of the seeds affects the amount of radiation received by the tumor and other parts of the prostate. When developing a treatment plan, the objective is to deliver an appropriate amount of radiation to the tumor and to the tissue identified as at risk of cancer, while sparing healthy tissue. Brachytherapy carries risks including urinary problems and erectile dysfunction. We summarize five papers that used mathematical models to optimize the placement of the seeds and needles and analyzed the risks of underestimating the Gleason score when treating with brachytherapy.

---

Ferrari et al. (15) proposed a mathematical model that allowed for the simultaneous optimization of seeds and needles in brachytherapy planning. The model took into account the prostate, urethra, rectum, and bladder. The volume was divided into discrete points where the radioactive seeds could be placed. Several quality metrics were used including the percentage of each tissue type receiving given dose levels, dose non-uniformity ratio, and percent of treatment volume achieving the prescribed dose. The dose for each point was calculated based on the sum of the radiation received from the seeds (inversely related to the distance from the seed). Each seed is delivered by a needle and the needle can only deliver a finite number of seeds in a straight line that meet certain spacing criteria. The model minimized the number of needles and seeds, while maximizing the percentage of treatment volume above the prescribed dose and healthy tissue below a certain dose threshold. The authors developed a genetic algorithm to find good solutions to the mathematical model. The GA used a two-dimensional chromosome of integers that corresponded to the seed placement and needles. The GA was used to develop plans for 11 test

patients. The authors established criteria for acceptable plan performance based on how well the plan matched the prescribed dose. The GA took an average of 20 minutes and found acceptable solutions in 10 of 11 cases.

Lee and Zaider (28) devised optimization models and computational techniques for real-time intraoperative 3D treatment planning in brachytherapy. Traditionally, treatment plans had been constructed several days or weeks prior to the implantation of the radioactive seeds, following an ultrasound or computerized tomography scan of the prostate. The manual construction of a treatment plan is a lengthy process and can take several iterations, and the images used to construct the treatment plan are often different from the images obtained when it is time to insert the seeds. This is in part because the prostate volume measured at the first imaging is often different from the volume observed in the operating room. Discrepancies between the original imaging and the imaging during implantation can increase the chance of undesirable side effects. Lee and Zaider formulated a mixed integer program (MIP) to optimize the placement of the seeds throughout the prostate. Constraints included dosimetric constraints for the tumor and critical structures. Constraints desired by the physicians, such as limiting the number of seeds, can also be included. In order to solve the MIP in real time, various techniques, including matrix reduction and a branch-and-cut environment, were used. The resulting plans delivered the prescribed dose to 98% of the prostate, reduced the number of seeds implanted by 20-30%, and reduced urethra dose by 23-28% compared to other computerized techniques.

Cambio et al. (3) used the probability of each outcome and the cost of each decision to assess the cost-benefit and outcome of recommending brachytherapy to men with a pros-

tate biopsy Gleason score of six. The Gleason score from a sextant biopsy agrees with the Gleason score from a prostatectomy 46% to 63% of the time. An extended core biopsy (10 or more cores) agrees 76% of the time. A Gleason score of six is considered low risk of biochemical recurrence of prostate cancer, while seven is considered intermediate risk. Brachytherapy is considered an appropriate standalone treatment for a patient with a Gleason score six or lower, though it is not recommended by itself for a patient with a Gleason score of seven or higher. The authors estimated the treatment cost and quality of life changes of brachytherapy from previous research. A group of 60-year old patients with a prostate biopsy Gleason score of six were used for the analysis. Patients either had a true Gleason score of six or they had been undergraded and had a Gleason score of either seven or eight to 10. Depending on the true state of the cancer, there were different probabilities that brachytherapy would successfully prevent biochemical failure (a rise in PSA levels after treatment considered to signal the recurrence of prostate cancer) within five years. A variety of biopsy accuracies were evaluated. For example, raising the agreement of the biopsy and prostatectomy Gleason score from 60% to 80% reduced average costs of treatment and recurrence from \$63,780 to \$62,929 per patient.

Fu, Yu, and Liu (16) developed a genetic algorithm for planning of prostate brachytherapy prior to the beginning of surgery, and replanning after some of the radioactive seeds have been placed. During brachytherapy surgery, needles can be deflected, changing the placement of radioactive seeds. Dynamic intraoperative treatment planning adjusts the placement of the remaining seeds based on the actual placement of the seeds already inserted into the prostate. The authors developed a genetic algorithm that first optimized the needle pattern (assuming one of a few seed spacing rules) and then optimized the seed

spacing. An original plan was constructed for a patient. Needles were randomly deflected, and seeds delivered as planned along the deflected needle. After one-quarter of the needles were delivered, the actual positions were known through imaging, and a new plan was constructed. After one-third of the needles in the second plan were delivered, a third plan was constructed. After one-half of the needles in the third plan were delivered, a fourth plan was constructed. After all the needles in the fourth plan were delivered, a fifth plan was constructed to ensure the tumor had received the necessary radiation. The quality of the reoptimized plans was compared to the original plan based on the number of seeds used, maximum urethral dose, and maximum rectal dose. The performance was based on three simulated runs for each of 10 patients using two different types of radioactive seeds, iodine-125 with a prescription dose of 145GY and palladium-103 with a prescription dose of 115Gy. Reoptimization led to a 10% to 20% increase in the number of seeds used, depending on the type of seed used. The deflections led to a 17% to 28% increase in maximum urethral dose and 16% to 42% increase in maximum rectal dose, depending on the type of seed.

Lee and Zaider (27) formulated a mixed integer program (MIP) to optimize the placement of the seeds throughout the prostate. A grid is placed over the potential seed locations in the prostate and binary variables were used to indicate whether a seed is placed at each location. The seed locations contribute to the radiation dosage at every point, and constraints include upper and lower bounds of radiation dosage at each point. Since it is not always possible to simultaneously meet all dose constraints two models are constructed that either maximize the number of points meeting these constraints or minimize the

weighted sum of the deviations from the dose bounds. Constraints included dosimetric constraints for the tumor and critical structures. Constraints desired by the physicians, such as limiting the number of seeds, can also be included. Computational strategies included matrix reduction, perturbation, and a penalty-based adaptive primal heuristic procedure. The MIP solver was able to solve sample problems within 15 CPU minutes and improve over the computer aided planning methods available. The MIP solver improved by using fewer seeds and needles, and providing better coverage and conformity.

---

#### 5.4.5 Patient Choice

Men who receive an early prostate cancer diagnosis and are aggressively treated have a higher survival rate than those diagnosed with later stages of the disease. However, there is concern that screening can lead to unnecessary treatment, particularly for men with low-risk disease. When deciding on the best treatment for prostate cancer, a patient needs to consider two key factors: (1) the risk of prostate cancer progression and mortality and (2) the potential side effects of the various treatment options. We summarize three papers that developed models to help a patient select a course of treatment.

---

Simon (41) constructed a multiattribute utility model to help a patient consider five treatment alternatives: surgery (radical prostatectomy), external radiation, seed radiation, dual radiation, and no treatment. In the first part of the model, the life expectancy of a patient and the probability of death from prostate cancer were determined. The probabilities of three side effects (impotence, incontinence, toxicity) that depend on the type of

treatment and factors specific to a patient were developed. In the second part of the model, a patient's preferences were analyzed. For the three side effects, a patient needed to determine what percentage of remaining life he would be willing to forego to avoid a particular side effect (this is known as the emotional weight). Simon expressed the possibility of each side effect in terms of a reduction in life span. For a patient, a life score was developed that took into account the length and quality of life. It provided a patient with a weighted average utility for a specific treatment. For example, a life score of 95 for seed radiation would be preferred to a life score of 85 for surgery. Simon ran the model on different patient profiles and found some interesting results. Younger men who did not have an aversion to side effects had the highest life score for surgery. Older men had high life scores for treatments that were less aggressive than surgery. One result was controversial: external radiation was used too often. Simon implemented a web-based version of this model that had an average of 400 hits per week in 2007.

Liberatore et al. (29) developed a decision counseling method, based on the Analytic Hierarchy Process (AHP), to help men make informed decisions about whether to screen for prostate cancer. AHP is a technique for analyzing complex decisions that takes into account the factors affecting the final decision. As part of a randomized controlled trial, the authors scheduled sessions for 129 men with a trained health educator to discuss prostate cancer screening. They used AHP to process a participant's three top decision factors as identified by the participant during the session and to generate a preference score that reflected a participant's preference to screen. The score was computed during the session using a calculator and was then given to the participant prior to the scheduling decision. The preference score was a statistically significant predictor of a participant's decision to

schedule a screening exam. An analysis of the preference score showed affective decision factors (emotional factors), such as fear of getting cancer, were the most influential in a patient's decision to be screened. Liberatore et al. concluded that AHP was successful in eliciting decision factors. They supported its implementation as part of decision counseling in the future.

Chapman et al. (5) constructed a multi-attribute utility (MAU) model to examine the trade-offs between different treatment options for prostate cancer patients, with the goal of aiding patients in making treatment decisions. The authors considered five health attributes that can be affected by patient treatment: pain, mood, sexual function, bladder and bowel function, and fatigue and energy. Their study had 57 patients from two Chicago Veterans Administration health clinics with either localized or metastasized prostate cancer. The patients' preferences were measured using time trade off (TTO) judgments. The participants were asked how many years of full health were equivalent to 10 years in one of three levels for each health attribute. The patients then divided 100 points between the five health attributes to indicate their relative importance. Patients tended to place the most weight on pain followed by bladder and bowel function. The MAU model was compared to a global TTO preference assessment. The result of the comparison indicated that the MAU model may be useful in aiding patient decision making.

---

#### 5.4.6 Recurrence

There are many forms of treatment available for prostate cancer. Younger patients, who are healthy, are often encouraged to receive a radical prostatectomy. Older patients who are not good candidates for surgery are more often steered away from a radical prostatectomy towards other forms of treatment such as external beam radiation therapy or brachytherapy. Regardless of the treatment selected there is the potential for recurrence following treatment.

If a patient's PSA level rises above 0.1 ng per mL following a radical prostatectomy, the patient is considered to be at risk that his cancer persists. According to Catalona [5], the five-year probability of nonprogression after a radical retropubic prostatectomy for prostate cancer is 78%. We summarize seven papers that used data mining to identify patients at risk of prostate cancer recurrence.

---

Dancea et al. (7) developed a support vector machine (SVM) model to divide patients into risk classes prior to a radical prostatectomy to aid in surgical decisions. They used a data set with 14 medical attributes for 399 patients: classification of malignant tumors, Gleason score, presence of median endo-vesical lobe, prostate volume, pre-operative PSA level, international index of erectile function, quality of life, abort operation, surgery technique, nerve sparing, surgery time, postoperative hospital time, complications, and risk class. They trained their SVM model on 369 patients and tested on 30. The SVM model correctly classified 93% of the test set; it correctly classified all low-risk and high-risk patients.

Shariat et al. (40) reviewed the literature to compare nomograms, risk groupings, artificial neural networks, probability tables, and regression trees to determine their effectiveness in predicting the risk of cancer recurrence. The published studies evaluated models on predictive accuracy, performance characteristics according to risk level, generalizability, and level of complexity. The studies that were examined by the authors applied nomograms and another model to a common external data set with known patient outcomes. Nomograms had superior predictive accuracy to risk groupings and probability tables. Regression trees constructed with CART analyses were easy to use and offered greater model-fitting flexibility than nomograms. In a head-to-head study, predictive accuracy was 70% for CART models and 84% for nomograms. ANNs have a high level of complexity and are typically used as a black box method, but they can fit very complex patterns. In a study that applied nomograms and ANNs to an external data set, predictive accuracy was 70.6% and 67%, respectively.

Androgen deprivation therapy is a common treatment for recurrent prostate cancer, but androgen-independent prostate cancer will eventually develop. NOXA and PUMA are two proteins whose presence may affect prostate cancer recurrence. Diallo et al. (8) generated recursive partitioning and regression tree (RPART) models with NOXA and PUMA to predict biochemical recurrence in prostate cancer patients. RPART models use a two-stage procedure to produce binary decision trees. The study had 43 patients with healthy prostate tissues, 62 patients presenting primary prostate cancer tissues, and 30 patients with hormone-refractory prostate cancer. The authors constructed RPART models using NOXA and/or PUMA with combinations of preoperative PSA level, Gleason

score, pathologic stage, and resection margin status (the presence of cancer in the surgical margin that surrounded the designated tumor removed during the radical prostatectomy). The RPART model with the best performance was based on NOXA, PUMA, and the resection margin status. Nine of the top 10 models included NOXA to predict biochemical recurrence, while only three of the top 10 included PUMA.

Biochemical failure-free survival (bFFS) refers to the probability that a patient's PSA level does not rise for three consecutive tests following external beam radiotherapy for prostate cancer. If a patient's PSA level rises for three consecutive tests following treatment, this is a strong indicator of recurrent cancer. In order to determine a patient's risk of relapse, Churilov et al. (6) applied clustering techniques to a set of post-treatment patients who had received external beam radiation therapy. The data set had Gleason score, tumor stage, PSA level at diagnosis, and age for 258 patients treated at the William Buckland Radiotherapy Center (WBRC) in Melbourne, Australia. The PSA level of each patient was taken at three-month intervals for the first year following the start of treatment, six-month intervals the next year, and then annually. Churilov et al. applied an optimization algorithm for clustering based on earlier work. The clustering algorithm grouped the patients into 10 clusters. WBRC used a rule-based method with seven clusters. In each cluster, a patient was identified as low, intermediate, or high risk of having a biochemical failure in the five years following treatment. The WBRC method assigned 51.2% of all patients to an intermediate-risk category and 8.2% to a low-risk category. The clustering algorithm developed by Churilov et al. assigned 26% to low risk and 34% to intermediate risk, while maintaining a similar probability of bFFS for the categories as the WBRC

method. Those defined as low-risk by the clustering method had a 71.6% of five-year bFFS, compared to 76.2% for those classified by the WBRC method. Patient's classified as intermediate-risk had a 53.1% and a 53.4% five-year bFFS for the clustering method and WBRC method, respectively. Churilov et al. improved the classification of patient's risk, allowing doctors to make better decisions regarding post-treatment care.

Seker et al. (38) constructed a Fuzzy K-Nearest Neighbors (FK-NN) classifier to predict the outcome of prostate cancer treatments for patients. Data from 41 men with histologically proven prostate cancer was used to construct FK-NN, linear regression, and artificial neural network models. Four conventional indicators of prostate cancer (tumor stage, skeletal metastasis, Gleason score, serum PSA level) and two experimental indicators (p53 immunostaining, bcl-2 immunostaining) were used to predict a patient's response to hormonal treatment, radical surgery, or observation. Response to treatment was classified as either having no response to any type of treatment, complete response to treatment (no tumor progression for patients undergoing observation), or relapse following successful start (tumor progression for patients undergoing observation). Each model was trained using a leave-one-out method. Each model achieved a best predictive accuracy using a different subset of the indicators. Serum PSA and treatment type were the best indicators of patient outcome. The logistic regression model had a predictive accuracy of 41.46%. The artificial neural network model had a predictive accuracy of 53.66%. The FK-NN model had a predictive accuracy of 60.98% if the number of nearest neighbors was set to 1 and 63.42% if it was set to 2 or 3.

Seker et al. (39) extended their work by constructing a hybrid neuro-fuzzy rule-based system to predict the treatment outcome of a prostate cancer patient. A set of rules was generated to determine the classification of a patient. Each rule had a premise (a set of variables and conditions on which the rule was based) and a consequence (the classification based on the premise). A neural network computed the parameters of the premise for each rule. Singular value decomposition selected the most and least important rules generated by using a *k*-means clustering method. The hybrid approach achieved a predictive accuracy of 63.42%.

Kattan, Wheeler, and Scardino (23) developed a nomogram to predict the progression of prostate cancer in men treated with radical prostatectomy. Pretreatment PSA level, Gleason sum from the surgical specimen, prostatic capsular invasion (the presence of cancer cells in the capsule around the tumor), surgical margin status (the presence of cancer cells in the capsule around the tumor), seminal vesicle invasion (the presence of cancer cells in muscular wall of the seminal vesicle), and lymph node status (the presence of cancer cells in the lymph nodes) for 996 men who underwent radical prostatectomy were used in the development and testing of the nomogram. Progression of prostate cancer was defined as PSA levels rising to 0.4 ng/ml or higher, a second PSA test that was higher than the first by any amount, or treatment with radiation or hormone therapy within seven years of surgery. The nomogram had an AUC of .88, implying that for 88% of pairs of patients, the patient with the larger score will relapse first. If variables whose values were known only after the radical prostatectomy were removed, the AUC would be reduced to .74.

#### 5.4.7 Costs

With an aging population and a current expenditure of \$11.8 billion for prostate cancer treatment in the United States (SEER [17]), it is important to understand where money is being spent and what costs are likely to be in the future. Since many of those affected by prostate cancer are above the age of 65, Medicare pays for much of the screening and treatment. Accurately predicting future Medicare expenditures is important in long-term planning. The two papers that are summarized explored the financial costs of prostate cancer with one paper examining the costs prior to treatment and the other predicting future Medicare costs.

---

Ekwueme, Stroud, and Chen (10) conducted a systematic review of published cost data for prostate cancer treatment in the United States and other industrialized countries, and then performed a statistical analysis of the data. The authors identified 262 articles published between 1980 and 2003 whose title or abstract implied cost information on prostate cancer treatment. From these, 28 (15 from the United States and 13 from other industrialized countries) were available in full text, written in English, contained original resource cost data, and included screening, diagnosing, and staging costs. All costs were converted to 2003 U.S. dollars using the consumer price index (for the country of origin) and exchange rates. A weighted mean for the cost per man screened was taken, weighted by the number of men screened in each study. A Monte Carlo simulation method was used to determine uncertainty in the pooled resource costs. For the United States, the cost for PSA screening was \$37.23 (\$30.92 internationally) and DRE was \$31.77 (\$33.54 internationally). In the United States, costs have decreased for biopsies and PSA screening, while they have increased for clinical staging, pathologic or histologic analysis, TRUS,

urology consultation, and DRE. Internationally, resource costs associated with biopsies, PSA screening, pathologic or histologic analysis, TRUS, urology consultation, and DRE have decreased over time (over the course of one or more of the published studies).

Penberthy et al. (34) analyzed the factors that predict Medicare expenditures in patients with breast, colorectal, lung, or prostate cancer. The authors combined data from the Virginia Cancer Registry, Medicare Provider Analysis and Review files, Medicare Automated Data Retrieval System, Medicare Health Insurance Master file, Medicare Annual Demographic files, Area Resource File, and 1990 Census Data for Zip Code Level information. Expenditures were analyzed for one year following diagnosis, with cost being defined as the amount the Health Care Financing Administration reimbursed the health care provider. The data set had 1,952 breast, 2,563 colorectal, 3,331 lung, and 3,179 prostate cancer patients. Treatments for prostate cancer were categorized as definitive surgery (287 patients), nonsurgical treatment (1,827 patients), surgery plus nonsurgical therapy (70 patients), and no treatment (995 patients). A two-stage least squares analysis was performed to predict costs. The R squared value for prostate cancer was 0.38. The mean cost for prostate cancers was \$14,361 for Virginia (1985-1988). Treatments that involved surgery tended to cost more as did treatment for patients with any comorbidity. Predicted cost increased with income of the patient. Patients who did not survive one year following diagnosis were found to cost 80% more than those who survived their first year.

---

## 5.5 Conclusions

Over the past 15 years, OR models and methods have been applied to diverse problems in the screening, detection, and treatment of prostate cancer. The results of these efforts have been impactful and important.

We expect that the use of OR will grow significantly in the next decade, not only in prostate cancer research, but also in more general research on healthcare problems. We hope that our bibliography and the attendant annotations will serve as a useful guidepost to help those conducting research on new problems in the near future.

## 6 CONCLUSIONS

IMRT would not be possible as a treatment without the use of operations research techniques such as mixed integer programming and simulated annealing. IMRT has a three stage treatment planning process: identification of the tumor and organs at risk, selection of beam angles for delivery of radiation, and optimization of radiation delivery from those angles. Without methods to optimize the delivery of radiation to each of the hundreds of pixels from each angle, IMRT would not be able to reduce the amount of radiation being delivered to the healthy tissue of patients because of the size and complexity involved in planning. These patients would suffer more acute side effects from their treatment due to damage to their organs from radiation and they would have greater long term risk of developing new tumors. But the optimization of treatment planning for IMRT is incomplete, because most institutions still manually choose beam angles. By creating a decision support tool to aid in the selection of beam angle sets, we have found better solutions in a fifth of the time of traditional search. By using machine learning models trained on previously evaluated beam angle plans, we were able to sort potential beam angle plans, without the time intensive step of fully evaluating them we sort in seconds a list that would take almost a day to evaluate. Evaluating just the top three plans from a list sorted using k-nearest neighbors yielded better results than evaluating 15 plans from an unsorted list. All machine learning models had significantly better results than unsorted lists, allowing for treatment planners to find better plans faster. Decision support tools of this manner provide an intermediary step allowing treatment planners to still play a role in beam angle selection by suggesting potential beam angle lists. Such an in-

termediary step, while perhaps not strictly necessary, could help the transition to completely automated beam angle optimization procedures.

Next, we created a tool that searches all potential beam angles. This tool has the ability to find even better solutions than those proposed by the treatment planner. We used the models developed by the machine learning algorithms as the fitness functions for a genetic algorithm that searched the space of feasible beam angle sets. We incorporated patient tumor geometry into our machine learning models, both improving performance and increasing the applicability to new patients without necessitating any time intensive evaluation of multiple treatment plans. Our method was able to improve upon the best known solutions by more than 2% and yet still run in about six minutes. Our genetic algorithm works in a fraction of the time it takes to evaluate a single beam angle set. We are not only saving the time a radiation oncologist would spend planning, but we are improving the quality of the treatment plans being generated.

In addition to our work improving the quality of the treatment plans for radiation therapy, we used simulation to aid in the design of proton therapy treatment centers and decrease patient wait times. Given the large capital investments and the unique interactions between system resources, it is crucial that the facility is efficiently designed and patients are scheduled intelligently. Our simulation showed the decreasing returns of adding additional gantry rooms, from a 26% increase in throughput for a fourth gantry room, to a 14% increase for the fifth, a 6% increase for the sixth, and no increase for further gantry rooms. The cyclotron is fully utilized at six gantry rooms, identifying the cyclotron as the limiting resource of the facility. We generat-

ed schedules that by ordering patients using treatment plans, reduced the average time spent waiting by over 30%, reduced the variance of the total wait time by 30%, and reduced the number of patients waiting longer than 10 minutes by 80%. As an outpatient procedure, patient's arrival time is highly uncertain. Outpatients have high probabilities of both absenteeism and tardiness. Such uncertainty increases patient wait times of the most precise schedules by up to 60%. In the presence of such uncertainty, simpler schedules such as block scheduling become competitive. Even using block schedules, it is still important to schedule patients based on treatment plan. When using block schedules, the even distribution of treatment plans through the day reduces patient wait time by 20%. Furthermore, simulation can prepare planners for patients with special needs and for equipment failures.

Finally, we have analyzed the impact of operations management on the treatment of prostate cancer. Operations management techniques have been applied to every stage of prostate cancer, from detection to postoperative monitoring. Machine learning methods have been applied to improve detection using both PSA and MRI. Large scale simulations have had an impact on policies surrounding PSA screening. Integer and linear programs have been used to improve the dose distribution of radiation during brachytherapy. Markov models have been used to aid in treatment choice decisions. This research has impacted both the recommendations of doctors and the design of the way they treat their patients. It has impacted policy and budgets.

Operations management has already had a significant impact on the delivery of healthcare in the United States. Its importance, in improving the efficacy and effi-

ciency of healthcare treatment, will continue to grow as demand for healthcare grows, with more people having access to healthcare, and advanced new treatments are developed. Along with the efficient delivery and utilization of medical resources, techniques that can harness the vast amounts of data being generated will play an increasingly important role. Using machine learning to find patterns in the treatments and outcomes of patients has the potential to quickly find adverse drug interactions, identify the best treatments for patients, identify risk factors in disease development, and to search for patterns in patient outcomes. As new treatments are developed for patients, simulation becomes important for determining throughput and designing workflow. Simulation can be used, prior to building new facilities, to aid in the intelligent design of schedules for patients, nurses, and doctors. Simulation can help insure the efficient use of equipment, which might otherwise become a bottleneck in the system. We have applied our models from system level decisions, about facility design and patient scheduling, to treatment level decisions, optimizing the radiation delivery to minimize damage to healthy tissue. We have seen how operations research models have affected the treatment of prostate cancer. The potential of operation research to improve healthcare delivery is only limited by healthcare practitioners' willingness to implement recommended changes. In order to have the greatest impact, researchers need to work to communicate not only the potential benefits but also the models and methods.

## 7 REFERENCES

American Cancer Society. Cancer Facts & Figures 2014. Atlanta, Ga. 2014

### 7.1 References for IMRT

Aha DW, Kibler D, Albert MK (1991) Instance-based learning algorithms. *Mach Learn* 6:37-66

Aleman DM, Romeijn HE, Dempsey JF (2008) Neighborhood search approaches to beam orientation optimization in intensity modulated radiation therapy treatment planning. *J Global Optim* 42:587–607

Aleman DM, Romeijn HE, Dempsey JF (2009) A response surface approach to beam orientation optimization in intensity modulated radiation therapy treatment planning. *INFORMS J Comput: Computat Biol Med Appl* 21:62–76

Bangert M, Oelfke U (2010) Spherical cluster analysis for beam angle optimization in intensity-modulated radiation therapy treatment planning. *Phys Med Biol* 55:6023-6037

Bertsimas D, Cacchiani V, Craft D, Nohadani O (2013) A hybrid approach to beam angle optimization in intensity-modulated radiation therapy. *Comput Oper Res* 40:2187-2197

Beyer K, Goldstein J, Ramakrishnan R, Shaft U (1999) When is “nearest neighbor” meaningful? *Database Theory—ICDT’99* (pp. 217-235). Springer Berlin Heidelberg

Bortfeld T (2006) IMRT: a review and preview. *Phys Med Biol* 51:363-379

Bortfeld T, Schlegel W (1993) Optimization of beam orientation in radiation therapy: some theoretical considerations. *Phys Med Biol* 38:291–304

Cao W, Lim GJ, Lee A, Li Y, Liu W, Zhu XR, Zhang X (2012) Uncertainty incorporated beam angle optimization for IMPT treatment planning. *Med Phys* 39:5248-5256

Cho BCJ, Roa HW, Robinson D, Murray B (1999) The development of target-eye-view maps for selection of coplanar or noncoplanar beams in conformal radiotherapy treatment planning. *Med Phys* 26:2367-2372

Craft D (2007) Local beam angle optimization with linear programming and gradient search. *Phys Med Biol* 52:127–135

- Dink D, Langer MP, Rardin RL, Pekny JF, Reklaitis GV, Saka B (2012) Intensity modulated radiation therapy with field rotation—a time-varying fractionation study. *Health Care Manag Sci* 15:138-154
- D'Souza WD, Meyer RR, Shi L (2004) Selection of beam orientations in intensity-modulated radiation therapy using single-beam indices and integer programming. *Phys Med Biol* 49:3465-3481
- D'Souza WD, Zhang HH, Nazareth DP, Shi L, Meyer RR (2008) A nested partitions framework for beam angle optimization in intensity modulated radiation therapy. *Phys Med Biol* 53:3293–3307
- Eclipse 11, Varian Medical Systems, 2013. Palo Alto, California, United States.
- Goitein M, Abrams M, Rowell D, Pollari H, Wiles J (1983) Multi-dimensional treatment planning: II. Beam's eye-view, back projection, and projection through CT sections. *Int J Radiat Oncol Biol Phys* 9:789-797
- Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten I (2009) The WEKA Data Mining Software: An Update. *SIGKDD Explor Newsl* 11:10-18
- Haykin, S (1994) *Neural networks: a comprehensive foundation*. Prentice Hall PTR.
- Hunt MA, Burman CM (2003) Treatment Planning Considerations Using IMRT. In: Fuks Z, Leibel SA, Ling CC (eds) *A Practical Guide to Intensity Modulated Radiation Therapy*. Medical Physics Publishing, Madison, WI, pp 105-121
- Lee EK, Fox T, Crocker I (2006) Simultaneous beam geometry and intensity map optimization in intensity-modulated radiation therapy. *Int J Radiat Oncol Biol Phys* 64:301–320
- Li Y, Yao J, Yao D (2004) Automatic beam angle selection in IMRT planning using genetic algorithm. *Phys Med Biol* 49:1915-1932
- Lim GJ, Choi J, Mohan R (2008) Iterative solution methods for beam angle and fluence map optimization in intensity modulated radiation therapy planning. *OR Spectr* 30:289–309
- Lim GJ, Cao W (2012) A two-phase method for selecting IMRT treatment beam angles: Branch-and-Prune and local neighborhood search. *Eur J Oper Res* 217:609-618
- Lim GJ, Kardar L, Cao W, Kelkar A (2012) Radiation therapy beam angle optimization methods: a comparison study. *IIE Annual Conference Proceedings*. Available at: <http://www.highbeam.com/doc/1P3-2813481701.html>

- Liu HH, Jauregui M, Zhang X, Wang X, Dong L, Mohan R (2006) Beam angle optimization and reduction for intensity-modulated radiation therapy of non-small-cell lung cancers. *Int J Radiat Oncol Biol Phys* 65:561–572
- MATLAB and Neural Network Toolbox Release 2012b, The MathWorks, Inc., Natick, Massachusetts, United.
- Misic VV, Aleman DM, Sharpe MB (2010) Neighborhood search approaches to non-coplanar beam orientation optimization for total marrow irradiation using IMRT. *Eur J Oper Res* 205:522-527
- Mohan R, Ling CC (1995) When becometh less more? *Int J Radiat Oncol Biol Phys* 33:235–237
- Myriantopoulos LC, Chen GT, Vijayakumar S, Halpern HJ, Spelbring DR, Pelizzari CA (1992) Beam's eye View volumetrics: an aid in rapid treatment plan development and evaluation. *Int J Radiat Oncol Biol Phys* 23:367–375
- Narayanan VKS, Vaitheeswaran R, Bhangle JR, Basu S, Maiya V, Zade B (2012) An experimental investigation on the effect of beam angle optimization on the reduction of beam numbers in IMRT of head and neck tumors. *J Appl Clin Med Phys*. Available at: <<http://jacmp.org/index.php/jacmp/article/view/3912/2562>>
- Philips International B.V. [www.healthcare.philips.com/main/products/ros/products/pinnacle3/](http://www.healthcare.philips.com/main/products/ros/products/pinnacle3/) Accessed 1 March 2013
- Potrebko, P.S., McCurdy, B.M.C., Butler JB, El-Gubtan AS, Nugent Z (2007) A simple geometric algorithm to predict optimal starting gantry angles using equiangular-spaced beams for intensity modulated radiation therapy of prostate cancer. *Med Phys* 34:3951–3961
- Pugachev A, Li JG, Boyer AL, Hancock SL, Le QT, Donaldson SS, Xing L (2001) Role of Beam Orientation Optimization in Intensity-Modulated Radiation Therapy. *Int J Radiation Oncology Biol Phys* 50:551-560
- Pugachev A, Xing L (2001a) Computer-assisted selection of coplanar beam orientations in intensity-modulated radiation therapy. *Phys Med Biol* 46:2467-2476
- Pugachev A, Xing L (2001b) Pseudo beam's-eye-view as applied to beam orientation selection in intensity-modulated radiation therapy. *Int J Radiat Oncol Biol Phys* 51:1361-1370
- Pugachev A, Xing L (2002) Incorporating prior knowledge into beam orientation optimization in IMRT. *Int J Radiat Oncol Biol Phys* 54:1565-1574

- R Core Team (2013) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. <http://www.r-project.org/>
- Romeijn HE, Dempsey JF (2008) Intensity modulated radiation therapy treatment plan optimization. *Top 16(2):215-243*
- Rowbottom CG, Webb S, Oldham M (1998) Improvement in prostate radiotherapy from customization of beam directions. *Med Phys 25:1171–1179*
- Rowbottom CG, Webb S, Oldham M (1999) Beam-orientation customization using an artificial neural network. *Phys Med Biol 44:2251–2262*
- Saher A, Sultan A (2010) Automatic Selection of Beam Orientations in Intensity- Modulated Radiation Therapy. *Electronic Notes in Discrete Mathematics 36:127-134*
- Schreibmann E, Lahanas M, Xing L, Baltas D (2004) Multiobjective evolutionary optimization of the number of beams, their orientations and weights for intensity-modulated radiation therapy. *Phys Med Biol 49:747–770*
- Söderström S, Brahme A (1995) Which is the most suitable number of photon beam portals in coplanar radiation therapy? *Int J Radiat Oncol Biol Phys 33:151–159*
- Stein J, Mohan R, Wang X, Bortfeld T, Wu Q, Preiser K, Ling CC, Schlegel W (1997) Number and orientations of beams in intensity-modulated radiation treatments. *Med Phys 24:149–160*
- Takamiya R, Missett B, Weinberg V, Akazawa C, Akazawa P, Zytkevich A, Bucci MK, Lee N, Quivey J, Xia P (2007) Simplifying intensity-modulated radiotherapy plans with fewer beam angles for the treatment of oropharyngeal carcinoma. *J Appl Clin Med Phys*. Available at: <http://www.jacmp.org/index.php/jacmp/article/view/2412/1256>
- Wang C, Dai J, Hu Y (2003) Optimization of beam orientations and beam weights for conformal radiotherapy using mixed integer programming. *Phys Med Biol 48:4065–4076*
- Webb S (2003) The physical basis of IMRT and inverse planning. *Brit J Radiol 76:678-689*
- Wu QJ, Yin F, McMahon R, Zhu X, Das SK (2010) Similarities between static and rotational intensity-modulated plans. *Phys Med Biol 55:33-43*

## 7.2 References for Proton Therapy Treatment Center

- Alexopoulos, C., D. Goldsman, J. Fontanesi, D. Kopald, and J. R. Wilson. 2008. “Modeling Patient Arrivals in Community Clinics.” *Omega* 36:33-43.

- Aitkenhead, A., D. Bugg, C. Rowbottom, E. Smith, and R. Mackay. 2012. "Modelling the Throughput Capacity of a Single-Accelerator Multitreatment Room Proton Therapy Centre." *The British Journal of Radiology* 85:1263-1272.
- Bolsi, A., A. J. Lomax, E. Pedroni, G. Goitein, and E. Hug. 2008. "Experiences at the Paul Scherrer Institute with a Remote Patient Positioning Procedure for High-throughput Proton Radiation Therapy." *International Journal of Radiation Oncology\* Biology\* Physics* 71:1581-1590.
- Cayirli, T. and E. Veral. 2003. "Outpatient Scheduling in Health Care: A Review of Literature." *Production and Operations Management* 12:519-549.
- Chakraborty, S., K. Muthuraman, and M. Lawley. 2013. "Sequential Clinical Scheduling with Patient No-Show: The Impact of Pre-Defined Slot Structures." *Socio-Economic Planning Sciences* 47:205-219.
- Fava, G., L. Widesott, F. Fellin, M. Amichetti, V. Viesi, A. J. Lomax, L. Lerderer, E. B. Hug, C. Fiorino, G. Salvadori, N. Di Muzio, and M. Schwarz. 2012. "In-gantry or Remote Patient Positioning? Monte Carlo Simulations for Proton Therapy Centers of Different Sizes." *Radiotherapy and Oncology* 103:18-24.
- Harper, P. and H. Gamlin. 2003. "Reduced Outpatient Waiting Times with Improved Appointment Scheduling: A Simulation Modelling Approach." *OR Spectrum* 25:207-222.
- Klassen, K., and R. Yoogalingam. 2009. "Improving Performance in Outpatient Appointment Services with a Simulation Optimization Approach." *Production and Operations Management* 18:447-458.
- Maryland Department of Transportation. 2012. "2012 Maryland State Highway Mobility Report."
- Miralbell, R., A. Lomax, L. Cella, and U. Schneider. 2002. "Potential reduction of the incidence of radiation-induced second cancers by using proton beams in the treatment of pediatric tumors." *International Journal of Radiation Oncology\* Biology\* Physics* 54: 824-829.
- National Association for Proton Therapy. 2013. Accessed March 8. <http://www.proton-therapy.org/>.
- Price, S., B. Golden, E. Wasil, and H. Zhang. 2013. "Optimizing Throughput of a Multi-Room Proton Therapy Treatment Center via Simulation." *Proceedings of the 2013 Winter Simulation Conference*  
R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, eds

Roylance, F. 2010. "University of Maryland to Build Proton Cancer Treatment Center." *The Baltimore Sun*. [http://articles.baltimoresun.com/2010-10-14/health/bs-hs-proton-cancer-20101013\\_1\\_proton-therapy-cancer-treatment-radiation-therapy](http://articles.baltimoresun.com/2010-10-14/health/bs-hs-proton-cancer-20101013_1_proton-therapy-cancer-treatment-radiation-therapy) [Accessed March 8, 2013].

St Clair, W. H., J. A. Adams, M. Bues, B. C. Fullerton, Sean La Shell, H. M. Kooy, J. S. Loeffler, and N. J. Tarbell. 2004. "Advantage of Protons Compared to Conventional X-ray or IMRT in the Treatment of a Pediatric Patient with Medulloblastoma." *International Journal of Radiation Oncology\* Biology\* Physics* 58: 727-734.

Thomson, C. 2013. "Baltimore Traffic 17<sup>th</sup> Worst in the Nation, Survey Shows." *The Baltimore Sun*. [http://articles.baltimoresun.com/2013-04-24/news/bs-md-traffic-rating-0425-20130424\\_1\\_inrix-baltimore-region-baltimore-traffic](http://articles.baltimoresun.com/2013-04-24/news/bs-md-traffic-rating-0425-20130424_1_inrix-baltimore-region-baltimore-traffic) [Accessed October 16, 2013].

Wilensky, U. 1999. NetLogo. Accessed January 10, 2013. <http://ccl.northwestern.edu/netlogo/>.

### 7.3 References for Prostate Cancer

[1] Howden, L. M. and Meyer, J. A. "Age and Sex Composition: 2010," *United States Census Bureau*, May (2011)

[2] Surveillance, Epidemiology, and End Results (SEER) Program Populations (1969-2012) ([www.seer.cancer.gov/popdata](http://www.seer.cancer.gov/popdata)), National Cancer Institute, DCCPS, Surveillance Research Program, Surveillance Systems Branch, released March 2014

[3] Hoffman, R. M., Stone, S. N., Espey, D., and Potosky, A. L. "Differences between men with screening-detected versus clinically diagnosed prostate cancers in the USA," *BMC Cancer*, 5(1), 27 (2005)

[4] Ma, X., Wang, R., Long, J. B., Ross, J. S., Soulos, P. R., Yu, J. B., Makarov, D. V., Gold, H. T., and Gross, C. P. "The cost implications of prostate cancer screening in the Medicare population," *Cancer*, 120(1), 96-102 (2014)

[5] Moyer, V. A. "Screening for prostate cancer: US Preventive Services Task Force recommendation statement," *Annals of Internal Medicine*, 157(2), 120-134 (2012)

[6] Schröder, F. H. et al., "Screening and prostate-cancer mortality in a randomized European study," *New England Journal of Medicine*, 360(13), 1320-1328 (2009)

[7] Andriole, Gerald L., et al., "Mortality results from a randomized prostate-cancer screening trial," *New England Journal of Medicine*, 360(13), 1310-1319 (2009)

[8] Food and Drug Administration, HHS., "Hybritech Inc.; Premarket Approval of Tandem-R, E, and ERA PSA Assays," Federal Register, 59(198), Friday, October 14, 1994, FR Doc No: 94-25516, From the Federal Register Online via the Government Printing Office, <http://www.gpo.gov/fdsys/pkg/FR-1994-10-14/html/94-25516.htm> (1994)

- [9] Greene, K. L., Albertsen, P. C., Babaian, R. J., Carter, H. B., Gann, P. H., Han, M., Kuban, D. A., Sartor, A. O., Stanford, J. L., Zietman, A., and Carroll, P., "Prostate specific antigen best practice statement: 2009 update," *The Journal of Urology*, 182(5), 2232-2241 (2009)
- [10] Carter, H. B., Albertsen, P. C., Barry, M. J., Etzioni, R., Freedland, S. J., Greene, K. L., Holmberg, L., Kantoff, P., Konety, B., Murad, M. H., Penson, D. F., and Zietman, A. L., "Early detection of prostate cancer: AUA Guideline," *The Journal of Urology*, 190(2), 419-426 (2013)
- [11] Ma, X., Wang, R., Long, J. B., Ross, J. S., Soulos, P. R., Yu, J. B., Makarov, D. V., Gold, H. T., and Gross, C. P. "The cost implications of prostate cancer screening in the Medicare population," *Cancer*, 120(1), 96-102 (2014)
- [12] Welch, H. G., Fisher, E. S., Gottlieb, D. J., and Barry, M. J. "Detection of prostate cancer via biopsy in the Medicare-SEER population during the PSA era," *Journal of National Cancer Institute*, 99(18), 1395-1400 (2007)
- [13] Bankhead, C. "Sextant biopsy helps in prognosis of PCa, but it's not foolproof," *Urology Times*, 25(8), (1997)
- [14] Farhang, R., Stroumbakis, N., Kava, B. R., Cookson, M. S., and Fair, W. R. "Incidence and clinical significance of false-negative sextant prostate biopsies," *The Journal of Urology*, 159(4), 1247-1250 (1998)
- [15] Loeb, S., Carter, H. B., Berndt, S. I., Ricker, W., and Schaeffer, E. M. "Complications after prostate biopsy: Data from SEER-Medicare," *The Journal of Urology*, 186(5), 1830-1834 (2011)
- [16] Thompson, J. E., Moses, D., Shnier, R., Brenner, P., Delprado, W., Ponsky, L., Pulbrook, M., Bohm, M., Haynes, A. M., Hayen, A., and Stricker, P. D. "Multiparametric magnetic resonance imaging guided diagnostic biopsy detects significant prostate cancer and could reduce unnecessary biopsies and over detection: A prospective study," forthcoming in *The Journal of Urology* (2014)
- [17] Ahmed, H. U., Kirkham, A., Arya, M., Illing, R., Freeman, A., Allen, C., and Emberton, M. "Is it time to consider a role for MRI before prostate biopsy?" *Nature Reviews Clinical Oncology*, 6(4), 197-206 (2009)
- [18] Lavieri, M. S., Puterman, M. L., Tyldesley, S., and Morris, W. J., "When to Treat Prostate Cancer Patients Based on Their PSA Dynamics," *IIE Transactions on Healthcare Systems Engineering*, 2(1), 62-77 (2012)

#### **7.4 Annotated Papers for Prostate Cancer analysis**

Anderson, D., Golden, B., Wasil, E., Zhang, H., "Predicting Prostate Cancer Risk Using Magnetic Resonance Imaging Data," forthcoming in *Information Systems and e-Business Management* (2014)

Cambio, A. J., Ellison, L. M., Chamie, K., de Vere White, R. W., and Evans, C. P., "Cost-Benefit and Outcome Analysis: Effect of Prostate Biopsy Undergrading," *Urology*, 69(6), 1152-1156 (2007)

Chandana, S., Leung, H., and Trpkov, K., "Staging of Prostate Cancer Using Automatic Feature Selection, Sampling and Dempster-Shafer Fusion," *Cancer Informatics*, 7, 57-73 (2009)

Chapman, G. B., Elstein, A. S., Kuzel, T. M., Nadler, R. B., Sharifi, R., and Bennett, C. L., "A Multi-Attribute Model of Prostate Cancer Patients' Preferences for Health States," *Quality of Life Research*, 8(3) 171-180 (1999)

Churilov, L., Bagirov, A. M., Schwartz, D., Smith, K., and Dally, M., "Improving Risk Grouping Rules for Prostate Cancer Patients with Optimization," In *Proceedings of the 37th Annual Hawaii International Conference on System Sciences, 2004.*, IEEE, 1-9 (2004)

Dancea, O., Gordan, M., Dragan, M., Stoian, I., and Nedevschi, S., "Postoperative Risk Classification of Prostate Cancer Patients Using Support Vector Machines," *IEEE International Conference on Automation, Quality and Testing, Robotics*, Vol. 3, 53-56 (2008)

Diallo, J. S., Aldejmah, A., Mouhim, A. F., Péant, B., Fahmy, M. A., Koumakpayi, I. H., Sircar, K., Bégin, L. R., Mes-Masson, A. M., and Saad, F., "NOXA and PUMA Expression Add to Clinical Markers in Predicting Biochemical Recurrence of Prostate Cancer Patients in a Survival Tree Model," *Clinical Cancer Research*, 13(23), 7044-7052 (2007)

Draisma, G., Etzioni, R., Tsodikov, A., Mariotto, A., Wever, E., Gulati, R., Feuer, E., and de Koning, H., "Lead Time and Overdiagnosis in Prostate-Specific Antigen Screening: Importance of Methods and Context," *Journal of the National Cancer Institute*, 101(6) 374-383 (2009)

Ekwueme, D. U., Stroud, L. A., and Chen, Y., "Cost Analysis of Screening for, Diagnosing, and Staging Prostate Cancer Based on a Systematic Review of Published Studies," *Preventing Chronic Disease*, 4(4), (2007)

Eldefrawy, A., Katkooori, D., Abramowitz, M., Soloway, M. S., and Manoharan, M., "Active Surveillance vs. Treatment for Low-Risk Prostate Cancer: A Cost Comparison," *Urologic Oncology: Seminars and Original Investigations*, 31(5), 576-580 (2011)

Etzioni, R., Cha, R., and Cowen, M. E., "Serial Prostate Specific Antigen Screening for Prostate Cancer: A Computer Model Evaluates Competing Strategies," *The Journal of Urology*, 162(3), 741-748 (1999)

- Etzioni, R., Gulati, R., Falcon, S., and Penson, D. F., "Impact of PSA Screening on the Incidence of Advanced Stage Prostate Cancer in the United States: A Surveillance Modeling Approach," *Medical Decision Making*, 28(3), 323-331 (2008)
- Fu, L., Yu, Y., and Liu, H., "Genetic Algorithm-Based Dynamic Intraoperative Treatment Planning for Prostate Brachytherapy," *Artificial Intelligence Applications and Innovations*, Eds. Li, D. and B. Wang, Proceedings of Second IFIP Conference on Artificial Intelligence Applications and Innovations, Springer US, 425-434 (2005)
- Garzotto, M., Beer, T. M., Hudson, R. G., Peters, L., Hsieh, Y. C., Barrera, E., Klein, T., and Mori, M., "Improved Detection of Prostate Cancer Using Classification and Regression Tree Analysis," *Journal of Clinical Oncology*, 23(19), 4322-4329 (2005)
- Gulati, R., Gore, J. L., and Etzioni, R., "Comparative Effectiveness of Alternative Prostate-Specific Antigen-Based Prostate Cancer Screening Strategies: Model Estimates of Potential Benefits and Harms," *Annals of Internal Medicine*, 158(3) 145-153 (2013)
- Gulati, R., Inoue, L. Y., Gore, J. L., Katcher, J., and Etzioni, R., "Individualized Estimates of Overdiagnosis in Screen-Detected Prostate Cancer," *Journal of the National Cancer Institute*, 106(2), (2014)
- Gülkesen, K. H., Köksal, İ. T., Özdem, S., and Saka, O., "Prediction of Prostate Cancer using Decision Tree Algorithm," *Turkish Journal of Medical Science*, 40(5), 681-686 (2010)
- Hayes, J. H., Ollendorf, D. A., Pearson, S. D., Barry, M. J., Kantoff, P. W., Stewart, S. T., Bhatnagar, V., Sweeney, C. J., Stahl, J. E., and McMahon, P. M., "Active Surveillance Compared with Initial Treatment for Men with Low-Risk Prostate Cancer: A Decision Analysis," *Journal of the American Medical Association*, 304(21), 2373-2380 (2010)
- Kattan, M. W., Wheeler, T. M., and Scardino, P. T., "Postoperative Nomogram for Disease Recurrence after Radical Prostatectomy for Prostate Cancer," *Journal of Clinical Oncology*, 17(5), 1499-1499 (1999)
- Kattan, M. W., Eastham, J. A., Wheeler, T. M., Maru, N., Scardino, P. T., Erbersdobler, A., Graefen, M., Huland, H., Koh, H., Shariat, S., Slawin, K., and Ohori, M., "Counseling Men with Prostate Cancer: A Nomogram for Predicting the Presence of Small, Moderately Differentiated, Confined Tumors," *The Journal of Urology*, 170(5), 1792-1797 (2003)
- Kelm, B. M., Menze, B. H., Zechmann, C. M., Baudendistel, K. T., and Hamprecht, F. A., "Automated Estimation of Tumor Probability in Prostate Magnetic Resonance Spectroscopic Imaging: Pattern Recognition vs Quantification," *Magnetic Resonance in Medicine*, 57(1), 150-159 (2007)

- Lavieri, M. S., Puterman, M. L., Tyldesley, S., and Morris, W. J., "When to Treat Prostate Cancer Patients Based on Their PSA Dynamics," *IIE Transactions on Healthcare Systems Engineering*, 2(1), 62-77 (2012)
- Lawrentschuk, N., Lockwood, G., Davies, P., Evans, A., Sweet, J., Toi, A., and Fleshner, N. E., "Predicting Prostate Biopsy Outcome: Artificial Neural Networks and Polychotomous Regression are Equivalent Models," *International Urology and Nephrology*, 43(1), 23-30 (2011)
- Lee, E. K., and Zaider, M., "Mixed Integer Programming Approaches to Treatment Planning for Brachytherapy—Application to Permanent Prostate Implants," *Annals of Operations Research*, 119(1-4), 147-163 (2003)
- Lee, E. K., and Zaider, M., "Operations Research Advances Cancer Therapeutics," *Interfaces*, 38(1), 5-25 (2008)
- Liberatore, M., Nydick, R., Daskalakis, C., Kunkel, E., Cocroft, J., and Myers, R., "Helping Men Decide About Scheduling a Prostate Cancer Screening Exam," *Interfaces*, 39(3), 209-217 (2009)
- Matulewicz, L., Jansen, J. F., Bokacheva, L., Vargas, H. A., Akin, O., Fine, S. W., Shukla-Dave, A., Eastham, J. A., Hricak, H., Koutcher, J. A., and Zakian, K. L. "Anatomic Segmentation Improves Prostate Cancer Detection with Artificial Neural Networks Analysis of 1H Magnetic Resonance Spectroscopic Imaging," *Journal of Magnetic Resonance Imaging*, Available at: <http://dx.doi.org/10.1002/jmri.24487> Accessed January 24, 2014
- Nam, R. K., Toi, A., Klotz, L. H., Trachtenberg, J., Jewett, M. A., Appu, S., Loblaw, D. A., Sugar, L., Narod, S. A., and Kattan, M. W., "Assessing Individual Risk for Prostate Cancer," *Journal of Clinical Oncology*, 25(24) 3582-3588 (2007)
- Parfait, S., Walker, P. M., Créhange, G., Tizon, X., and Mitéran, J., "Classification of Prostate Magnetic Resonance Spectra using Support Vector Machine," *Biomedical Signal Processing and Control*, 7(5), 499-508 (2012)
- Penberthy, L., Retchin, S. M., McDonald, M. K., McClish, D. K., Desch, C. E., Riley, G. F., Smith, T. J., Hillner, B. E., and Newschaffer, C. J., "Predictors of Medicare Costs in Elderly Beneficiaries with Breast, Colorectal, Lung, or Prostate Cancer," *Health Care Management Science*, 2(3) 149-160 (1999)
- Ross, K.S., Carter, H.B., Pearson, J.D., and Guess, H.A., "Comparative Efficiency of Prostate-Specific Antigen Screening Strategies for Prostate Cancer Detection," *JAMA: The Journal of the American Medical Association*, 284(11), 1399-1405 (2000)
- Schröder, F., and Kattan, M. W., "The Comparability of Models for Predicting the Risk of a Positive Prostate Biopsy with Prostate-Specific Antigen Alone: A Systematic Review," *European Urology*, 54(2), 274-290 (2008)

- Seker, H., Odetayo, M. O., Petrovic, D., and Naguib, R. N. G., "A Fuzzy Logic Based-Method for Prognostic Decision Making in Breast and Prostate Cancers," *IEEE Transactions on Information Technology in Biomedicine*, 7(2), 114-122 (2003a)
- Seker, H., Odetayo, M. O., Petrovic, D., Naguib, R. N. G., and Hamdy, F. C., "An Intelligent Hybrid Neuro-Fuzzy Rule-Based System for Prognostic Decision Making in Prostate Cancer Patients," in *Information Technology Applications in Biomedicine, 2003. 4th International IEEE EMBS Special Topic Conference on*. IEEE, 173-176 (2003b)
- Simon, J., "Decision Making with Prostate Cancer: A Multiple-objective Model with Uncertainty," *Interfaces*, 39(3), 218-227 (2009)
- Sofer, A., Zeng, J., and Mun, S. K., "Optimal Biopsy Protocols for Prostate Cancer," *Annals of Operations Research*, 119(1-4), 63-74 (2003)
- Sooriakumaran, P., John, M., Christos, P., Bektic, J., Bartsch, G., Leung, R., Herman, M., Scherr, D., and Tewari, A. "Models to Predict Positive Prostate Biopsies using the Tyrol Screening Study," *Urology*, 78(4) 924-929 (2011)
- Spurgeon, S. E., Hsieh, Y. C., Rivadinera, A., Beer, T. M., Mori, M., and Garzotto, M., "Classification and Regression Tree Analysis for the Prediction of Aggressive Prostate Cancer on Biopsy," *The Journal of Urology*, 175(3), 918-922 (2006)
- Tsodikov, A., Szabo, A., and Wegelin, J., "A Population Model of Prostate Cancer Incidence," *Statistics in Medicine*, 25(16), 2846-2866 (2006)
- Underwood, D.J., Zhang, J., Denton, B.T., Shah, N.D., and Inman, B.A., "Simulation Optimization of PSA-Threshold Based Prostate Cancer Screening Policies," *Health Care Management Science*, 15(4), 293-309 (2012)
- Zhang, J., Denton, B. T., Balasubramanian, H., Shah, N. D., and Inman, B. A., "Optimization of PSA Screening Policies: A Comparison of the Patient and Societal Perspectives," *Medical Decision Making*, 32(2), 337-349 (2012a)
- Zhang, J., Denton, B. T., Balasubramanian, H., Shah, N. D., and Inman, B. A., "Optimization of Prostate Biopsy Referral Decisions," *Manufacturing & Service Operations Management*, 14(4), 529-547 (2012b)