June 2020

Analytical Setup Margin for Spinal SBRT Based on Measured Errors

Audrey Copeland
Louisiana State University and Agricultural and Mechanical College

ANALYTICAL SETUP MARGIN FOR SPINAL SBRT BASED ON MEASURED ERRORS

A Thesis

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Master of Science

in

The Department of Physics and Astronomy

by
Audrey Copeland
B.S., University of Texas at Austin, 2017
August 2020

Acknowledgements

I acknowledge the contributions of the other individuals who took part in the completion of this work. I thank my advisor, Dr. Fontenot, for his expert guidance on the direction of this work and its implications. I also thank my other committee members Dr. Matthews, Dan Neck, and Dr. Sprunger who offered invaluable support and advice on all aspects of this project. I thank Koren Smith for her help with my measurements using the Catalyst system and getting in touch with the C-RAD team, the dosimetry team at Mary Bird Perkins Cancer Center for their guidance with treatment planning and troubleshooting, and the therapists at MBPCC and Woman's hospital for their insight into my measurements. The tasks of scheduling committee meetings, advisor meetings, and handling any paperwork were all effortlessly facilitated by Susan Hammond, Katelynn Fontenot, and Yao Zeng who I also recognize.

My friends and classmates at LSU were also instrumental to the successful completion of this project. Stephanie Wang, Payton Bruckmeier, Andrew McGuffey, and Troy Jacobs provided a strong support system for me throughout my time at LSU. My officemates Phillip Wall and Cam Sprowls were particularly helpful with any questions I had about the more clinical aspects of my project.

Lastly, I recognize my parents Bill and Joni Copeland for always believing in me and supporting me completely as well as my partner Max Granat who listened to and critiqued countless hours of presentation practice, writing samples, and new ideas. No work of significance can be accomplished alone, and I certainly could not have persisted without the help of all those recognized and likely many more.

Table of Contents

List of Tables

List of Figures

Abstract

Purpose: No consensus currently exists in the radiotherapy community about the correct margin size to use for spinal SBRT. Margins have been proposed to account for various errors individually, but not with all errors combined to result in a single margin value. The purpose of this work was to determine a setup margin for spinal SBRT based on known and measurable errors during radiotherapy to achieve at least 90% coverage of the clinical target volume (CTV) with the prescription dose for at least 90% of patients and not exceed a 30 Gy point dose or 23 Gy to 10% of the spinal cord subvolume.

Methods: The random and systematic error components of intrafraction motion, residual setup error, and end-to-end system accuracy and the penumbral width of a spinal SBRT plan were measured. The patient's surface displacement was measured to quantify intrafraction motion, the residual setup error was quantified by re-registering accepted daily cone beam computed tomography (CBCT) setup images, and the measurement of the displacement between measured and planned dose profiles in a phantom quantified the end-to-end system accuracy. These errors and parameters were used to identify the minimum acceptable margin size. The margin recommendation was validated by assessing dose delivery across 140 simulated patient plans suffering from various random shifts representative of the measured errors.

Results: The errors were quantified in three dimensions and the analytical margin generated was 2.4 mm. With this margin applied in the superior/inferior direction only, at least 90% of the CTV was covered with the prescription dose for 96% of the 140 patients simulated. With this margin applied, there was minimal negative effect on the spinal cord dose levels.

Conclusions: The findings of this work support that a 2.4 mm margin applied in the superior/inferior direction can achieve at least 90% coverage of the CTV for at least 90% of

dual-arc volumetric modulated arc therapy (VMAT) spinal SBRT patients in the presence of errors when immobilized with vacuum bags and treated at Mary Bird Perkins Cancer Center (MBPCC).

Chapter 1. Introduction

1.1. Background and Significance

1.1.1. Spinal SBRT

Stereotactic body radiotherapy (SBRT) has been shown to improve outcomes for spinal metastases compared to conventional radiotherapy (Conti *et al.*, 2019). In a conventional treatment, typically about 2 Gy is delivered to the target during each fraction over 25 to 30 fractions. The dose distribution for this treatment may not necessarily be highly conformal; meaning the irradiated area may include a significant amount of normal tissue. However, since the dose per fraction is so low, the normal tissue within the irradiated area can recover from the damage and side effects are still minimized. Comparatively, SBRT utilizes higher doses per fraction, typically between 5 and 12 Gy, over just one to five fractions. Since the risk of normal tissue complications increases with dose and may even exceed tolerances when these higher doses are delivered at once, SBRT must be delivered with a highly conformal dose distribution to minimize the amount of normal tissue irradiated. Higher conformity results in steep dose gradients between high and low dose regions which may occur near critical structures in order to spare them. Therefore, any shift of the target relative to the planned dose distribution can result in a significant overdose of surrounding normal tissue, such as the spinal cord, or insufficient target coverage since the high dose region has shifted geometrically from its planned position.

1.1.2. Margins

Margins for conventional radiotherapy are added to the initial delineated target to balance the objectives of organ sparing and target coverage while accounting for uncertainties. The ICRU defines various volumes resulting from adding different margins to account for specific types of

uncertainties (2010). Figure 1 shows how these volumes are related to each other (Berthelsen *et al.*, 2007).

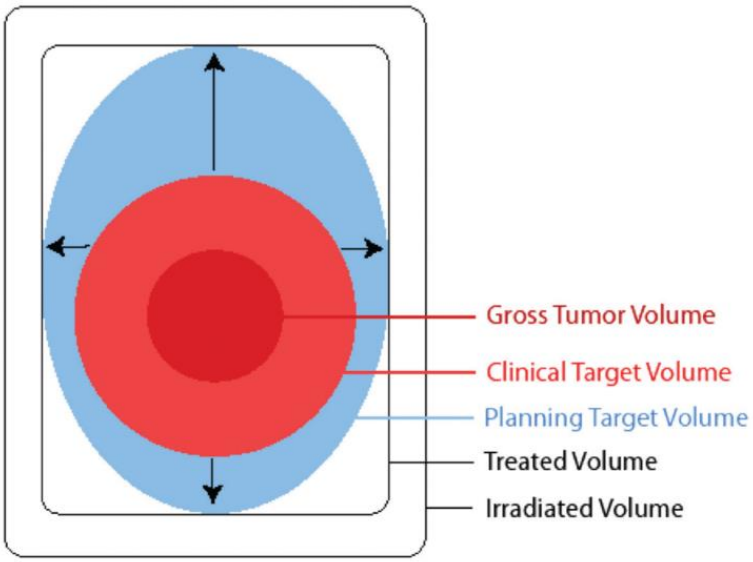

Figure 1. Representation of the ICRU-defined planning volumes created by adding margins (Berthelsen et al., 2007).

The gross tumor volume (GTV) is the initial delineated target and consists of the visible or palpable extent of the tumor, determined by the physician. The clinical tumor volume (CTV) is an expansion of the GTV to include the surrounding microscopic extent of the disease also requiring treatment. The CTV is also physician-defined and is typically drawn based on the type of tumor and known areas of disease extension determined from past treatment failures or visible disease growth (Berthelsen *et al.*, 2007). The internal target volume (ITV) is an expansion of the CTV to account for internal changes in the position and shape of the CTV within the body. When daily image guidance is used to locate the target and when the target is bony and not subject to much organ motion, the margin applied to create the ITV can be set to zero, as it will be in this work. A planning target volume (PTV) is created by adding a margin to the CTV to account for geometrical uncertainties due to interfraction and intrafraction uncertainties. Interfraction uncertainties occur when there are changes in the position or shape of the target between

fractions, or separate treatment sessions. Examples of interfraction uncertainties include differences in how the patient is set up on the table from day to day, target growth or shrinkage over the entire course of treatment, or unexpected changes in the treatment delivery system. Intrafraction uncertainties occur when there is a change to the target over the course of one fraction, such as patient motion on the table or breathing motion. The physician typically prescribes a certain dose to a minimum volume of the PTV and sometimes the CTV as well. For example, the physician may prescribe 5000 cGy to at least 95% of the PTV and 5500 cGy to at least 99% of the CTV and these objectives will guide the optimization of the treatment plan.

For conventional radiotherapy, there exist widely accepted analytical CTV to PTV setup margin calculation methods based on measured random and systematic geometrical uncertainties (Van Herk *et al.*, 2000). In the Van Herk formulation, random errors contribute to a blurring of the planned dose distribution which will combine with the existing penumbra of the planned distribution to affect how the dose falls off at the edges of the CTV. Random errors in this formulation may include uncertainties during the course of treatment which will have a random magnitude and direction on any given treatment day, such as the patient motion and daily setup error. The systematic errors contribute to a shift in the dose distribution from its planned position and can come from errors which cause the planned distribution to always be offset in the same way each treatment day such as imaging the target in a non-reproducible position for simulation or a shift in the radiation isocenter which persists throughout the patient's treatment course. The margin is then calculated to guarantee CTV coverage with a minimum dose for a given percentage of a population for an ideal, spherical CTV in an ideal, homogenous patient in the presence of these known errors.

However, assumptions made for calculating margins for conventional fractionation schemes are not necessarily applicable to hypofractionated treatments such as SBRT. For example, errors such as misalignment or patient motion present during each fraction are assumed to contribute only to the random error when there are many fractions and average out to zero over the whole course of a conventional treatment (Van Herk *et al.*, 2000). In hypofractionated treatments, these shifts may have a residual mean since there are so few fractions. A random shift during one fraction in one direction may not necessarily be compensated by another random shift in the opposite direction at a later fraction. Therefore, analytical margins commonly used for conventional treatments cannot be reasonably applied to a hypofractionated treatment (Van Herk *et al.*, 2000). Currently, for hypofractionated treatments like spinal SBRT, physicians use subjective isotropic PTV setup margins of up to two millimeters when feasible, but it is reduced where such a margin would overlap a critical structure such as the spinal cord (Redmond *et al.*, 2017). Setting the setup margin to zero has also been reported and there is not yet a consensus on a corresponding safe, objective margin recipe to use for SBRT treatment planning (Cox *et al.*, 2012). These nominal margins are not based on any known errors or measurements. With a generalized value for a margin, there is no guarantee that with inevitable and variable patient motion as well as known uncertainties in the delivery system that the coverage and dose limiting objectives can be achieved.

In the case of stereotactic radiosurgery (SRS) for cranial tumors, most often brain metastases, some potential margin recipes have been proposed (Zhang *et al.*, 2013; Ortega *et al.*, 2016). These are typically delivered in a single fraction, so the fractionation scheme is more similar to a spinal SBRT treatment than a conventional treatment. However, these SRS margins may not be directly applicable to spinal SBRT since these treatment sites have some important

differences. Brain metastases typically have well-defined borders and are spherical with a high degree of regularity. With a spinal treatment site, the target is usually highly irregular in shape. The spinal cord is also very close to the target, therefore requiring steep dose gradients. Non-spherical targets may have a different pattern of dose fall-off than a spherical target in a treatment field, so the dosimetric effects on CTV coverage and dose to nearby tissues with a given geometric shift may be different. These differences may cause varying dosimetric results between targets if the same margins were used.

Ortega performed an end-to-end system accuracy test of an SRS delivery system and included previously measured magnitudes of other error components to generate an analytical SRS margin (Ortega *et al.*, 2016). The margin only included systematic aspects of these errors for a single fraction cranial SRS treatment, excluding all random components. However, other works have shown that both systematic and random uncertainties should be accounted for in any margin calculation (Van Herk *et al.*, 2000). Especially for multi-fraction treatments consisting of a small number of fractions, both components will influence the dose distribution and should be considered. Zhang also proposed a cranial SRS margin recipe which relied on assumptions that are not applicable to spinal SBRT (Zhang *et al.*, 2013). In their work, a single-fraction 3D conformal technique was used, and it assumed that the shape of the CTV is highly regular and spherical. The treatment technique used affects the shape and width of the penumbra of the dose fall-off outside the target and will affect how the dose is missed in the presence of errors. These are all major differences between a cranial SRS treatment and a spinal SBRT treatment and the effects of using the margin for cranial SRS for spinal SBRT are unknown.

1.1.3. SDE2 Hypofractionated Margin Calculation Algorithm

A potential margin recipe for hypofractionated treatments based on both random and systematic components of measured uncertainties has been formulated by Herschtal et al. (Herschtal *et al.*, 2013). This recipe utilizes two different formulations and then a simulation step to verify the resulting margin and this method will be referred to as the "SDE2" algorithm. One formulation is based on the van Herk method (Van Herk *et al.*, 2000) for conventional treatments with a correction factor for a small and finite number of fractions; for the purposes of this work it will be referred to as the "Adjusted van Herk method" (AVH). The other formulation is an algorithm based on AVH which uses dose-population histogram (DPH) tables generated by calculating the dose to each point on a CTV with various added margins in the presence of random uncertainties or shifts, which for the purposes of this work shall be referred to as the "DPH method". The simulation step then uses the output of these formulations as the initial margin guess and measures the minimum dose to the CTV to determine if the goal was achieved for the desired percentage of cases.

The SDE2 algorithm requires as inputs the random and systematic components of the measured error standard deviations, the number of fractions, the minimum dose the CTV receives specified as a percentage of the prescription dose, the minimum proportion of the population receiving at least that dose level, and the penumbral width of the treatment plan. The standard deviation of the mean patient errors represents the systematic error and the root-mean-square (rms) sum of the individual patient standard deviations represents the random error (van Herk, 2004). The number of fractions is the amount of separate treatment sessions over which the total prescribed dose will be delivered. The minimum dose the CTV receives at any point is

given as a percentage of the total prescribed dose so that the calculation is independent of the absolute dose.

The AVH margin recipe is calculated by Equation 1, where *a* and *b* are the number of standard deviations of random and systematic error respectively corresponding to the desired levels of population percentage and prescription dose percentage and σ_{pen} is the penumbral width. The adjusted systematic error and random error are shown in Equation 2 and Equation 3 respectively, where Σ is the systematic error component, σ is the random error component, and *n* is the number of fractions.

$$m = a\Sigma' + b\sqrt{\sigma'^2 + \sigma_{pen}^2}$$

Equation 1

$$\Sigma' = \sqrt{\Sigma^2 + \frac{\sigma^2}{n}}$$

Equation 2

$$\sigma' = \sqrt{\frac{n-1}{n}}\sigma$$

Equation 3

The DPH method simulates shifts for a large number of patients based on combinations of errors and margins. The percentage of simulated patients with shifts sampled from the distribution of the measured errors who receive at least the minimum dose at any point on the CTV can be determined. The algorithm then looks up the value of the margin necessary to provide CTV coverage for the specified population percentage with the given error parameters. It outputs a multiplier on the random term in the AVH equation represented by *p* in Equation 4.

$$m = a\Sigma' + pb\sqrt{\sigma'^2 + \sigma_{pen}^2}$$

Equation 4

Both Chang et al. and Lyons et al. utilized this SDE2 recipe to determine SBRT margins for the spine and prostate respectively, but they did not include some known and measurable errors, such as the end-to-end system accuracy (Chang *et al.*, 2017; Lyons *et al.*, 2017). In these other works where spinal SBRT margins were determined, the spinal cord dose was not measured after expanding the target volume in the direction of the cord. Irradiating a larger volume that is now closer to a critical structure could impact whether the dose constraints for that structure are being exceeded.

1.1.4. Sources of Uncertainty

All known sources of uncertainty should be included when determining a safe margin for treatments requiring a high level of precision, including SBRT (Van Herk et al., 2000). Finnigan and Chang identified the components of error necessary for a margin recipe: residual setup error, end-to-end system accuracy, and intrafraction motion (Finnigan *et al.*, 2016; Chang *et al.*, 2017). Since the patient population, equipment, and procedures may differ between facilities and can affect the magnitudes of the error components, these quantities need to be measured on-site for the most accurate estimation (Zhang *et al.*, 2016; Parker *et al.*, 2002). The residual setup error is the remaining difference in position between the CTV just before treatment and the planned position of the CTV which was used to create the treatment plan. The end-to-end system accuracy consists of the measured geometric shift between the planned dose distribution and the deliverable dose distribution. Finnigan notes many individual mechanical sources of error, but these can be combined together with comprehensive end-to-end testing starting from simulation with a planning CT image and measuring the accuracy of the resulting dose delivery on the linear

accelerator (Finnigan *et al.*, 2016). Intrafraction motion consists of any patient or target motion during a fraction while radiation is being delivered. This error component is particularly dependent on factors such as imaging frequency, setup procedures, treatment time, and repositioning tolerances and should be measured at the same facility where the margin will be applied for the best estimate of its true value (Hoogeman *et al.*, 2008). The penumbral width of the beam is a measure of the dose fall-off or blurring of the planned dose distribution at the edges of the target and notably not a single open-field beam penumbra (Gordon and Siebers, 2007). This parameter should be quantified in a clinical plan representative of the treatments performed by the institution to account for dose delivered close to the CTV edge.

1.2. Motivation for Research

Currently, there is no consensus among physicians on how much margin to add to the CTV. Current practices range from adding no margin to adding 2 mm (Cox *et al.*, 2012). These values are not based on any measured errors or uncertainties so there is no guarantee of a certain probability of CTV coverage that an analytical margin based on those errors would provide.

Margins proposed for cranial SRS are not necessarily applicable for SBRT of the spine, where the vertebrae being treated are much denser than the surrounding soft tissue and the shapes can be highly irregular. These margins also do not include any effects of expanding the treated volume by the margin width on the nearby critical normal tissue, such as the spinal cord with spinal SBRT.

The motivation of this work was the lack of these necessary components of measured errors related to spinal SBRT included in an analytical margin recipe which can reliably be applied to a hypofractionated treatment of an irregular target. A verification of this margin with

simulated patient trials on a real treatment plan would verify that coverage and tolerance constraints are met, but this has not been done.

1.3. Hypothesis and Specific Aims

The hypothesis of this work is that a setup margin for spinal SBRT can be determined based on the SDE2 recipe (Herschtal *et al.*, 2013). This margin will provide at least 90% coverage of the CTV with the prescription dose and the spinal cord will receive no more than 23 Gy to 10% of its subvolume defined in the report of Task Group 101 for at least 90% of patients in the presence of geometrical uncertainties (Chang *et al.*, 2017; Gordon and Siebers, 2007; Hoogeman *et al.*, 2008). This margin will include both the systematic and random components of end-to-end system accuracy, setup uncertainty, and intrafraction motion. The dosimetric effects on the spinal cord of implementing such a margin will now be investigated.

1.3.1. Specific Aims

1. Quantify setup error, end-to-end system accuracy, intrafraction motion, and the penumbral width of a contemporary clinical SBRT treatment
2. Generate an analytical setup margin formula for spinal SBRT
3. Validate margin by investigating the dosimetric effects of simulated patient displacements in a spinal SBRT plan

Chapter 2. Methods and Materials

2.1. Aim 1

To achieve the first aim, the components of uncertainty present in the radiotherapy process for spinal SBRT were measured. These parameters were end-to-end system accuracy, residual setup error, intrafraction motion, and penumbral width and are facility-specific since they are sensitive to the specific workflows, treatment techniques, technologies, and other components used which can differ between institutions. These measurements would ideally be collected for spinal SBRT treatments. However, since a spinal SBRT program was not yet fully implemented at the time of this study, uncertainties were approximated from measurements of "similar" treatments (subsequently defined) as at Mary Bird Perkins Cancer Center (MBPCC).

"Similar" treatments were identified based on treatment site, treatment modality, immobilization device type, and treatment time. At Mary Bird Perkins Cancer Center, a spinal SBRT treatment would consist of a five-fraction dual-arc volumetric modulated arc therapy (VMAT) treatment with flattening filter-free beam (FFF). The patients would be immobilized with vacuum bags from the knees down and thermoplastic masks encompassing the shoulders and the head for lower and upper vertebral targets respectively. These "similar" treatments were for the spine or a closely adjacent structure to the spine and utilized the same immobilization devices to be used with spinal SBRT at MBPCC. Li showed that different immobilization devices affect the intrafraction motion magnitudes significantly (Li *et al.*, 2012). The treatment time should also be comparable to a dual-arc VMAT treatment because treatment time may also affect the magnitudes of intrafraction motion error (Hoogeman *et al.*, 2008). The median treatment time for a flattening-filter-free delivery of a dual-arc VMAT treatment reported was 6 minutes and 6 seconds (Jeon and Kim, 2018). These factors may affect measured errors, so it is

important that they are measured under the same or similar conditions. Measuring them at the same institution is also important because each component of the treatment simulation and delivery systems may have errors different enough from another at a different facility to affect the margin.

Every type of error contributing to the margin calculation will have a random and systematic component, where random errors serve to blur the ideal planned dose distribution, and systematic errors cause an overall shift of the distribution. In the conventional Van Herk margin formulation, the number of fractions was assumed to be large, a valid assumption for treatments of 25 to 30 fractions (Van Herk *et al.*, 2000). It was also assumed that each type of error was either completely random or systematic. This is because over time, the mean of the random errors would be zero with so many fractions. When considering a large number of random positional errors with random magnitudes and directions in three dimensions, the mean of all of these shifts will average out to zero as the number increases towards infinity. If the mean was nonzero, this would imply an intrinsic offset to the random error distribution further implying some systematic component to this random error type. However, when considering a smaller number of fractions, assuming a residual random error mean of zero may no longer be valid. For example, when considering the limiting case of a single fraction SBRT treatment, any patient motion that has a residual nonzero mean over the course of treatment will contribute to a shift in the delivered dose distribution, rather than contributing solely to an overall blurring of the dose distribution, indicating a systematic shift. Patient motion was initially considered a purely random error component in the conventional van Herk margin formulation. But with a hypofractionated treatment, this error has some systematic attributes if the mean is nonzero and must be factored into the margin calculation as such. This consideration must be applied to every

error source measured since each could have a random and systematic component (Oehler *et al.*, 2014).

2.1.1. Residual Setup Error Measurements

2.1.1.1. Residual Setup Error Measurement Methods

The residual setup error is the remaining difference in position between the CTV just before treatment and its planned position. It was estimated from cases similar to spinal SBRT treatments including spinal treatments delivered with conventional or palliative fractionation and other conventional treatments near the spine. For these non-spine treatment sites, the setup process closely resembled a spinal SBRT setup. A total of 195 setup CBCT images for 20 patients were collected and analyzed to estimate the mean value with 95% confidence to within 0.2 mm, assuming the standard deviation of the data is similar to other reported values (Chang *et al.*, 2017). Each case had cone beam CT (CBCT) setup images taken before delivery where the number of images for each patient ranged from 4 to 36. There were 18 spinal metastasis treatments with 10, 12, or 15 fractions with two to three static beams, one VMAT esophagus treatment with 28 fractions, and one VMAT larynx treatment with 36 fractions. The esophagus and larynx track along the spinal column, and the physicians for these patients required alignment to the spinal cord for setup image guidance.

The residual setup error was measured retrospectively after the CBCT setup images were taken as a routine part of the patient setup procedure. These images were taken after the patients were aligned to skin marks created at simulation to check that the internal target is aligned with the planning CT image. Then, for any patient, the setup CBCT image was registered with the planning CT image using automatic registration software (XVI) to measure the difference in the patient's current position from the planned position of the target. The therapist then manually

adjusted the registration if necessary, and the couch was automatically shifted by the calculated amount to match the target's current alignment most closely with its alignment at simulation. There were no verification images taken after the couch shifted the patient into the new position.

2.1.1.2. Residual Setup Error Data Analysis

To analyze the setup images and measure the residual setup error, the accepted daily registrations between the planning CT and the CBCT taken after patient setup for a fraction were re-registered with the accepted shifts already applied in the XVI software. The "Bone T+R" setting was used to automatically register the images and calculate a couch shift. The root-mean-square sum of the patient standard deviations represented the random error and the standard deviation of the patient averages represented the systematic error (van Herk, 2004).

Verification images were not taken after the couch corrections were automatically applied, so the error in the couch movement or any patient motion occurring between the initial CBCT image was taken and treatment is not included. This measurement method assumes the couch could move the patient into position perfectly and the patient did not move in the time between when the image acquisition started and when treatment began. In this way, the error measured is purely the setup error. The couch error and patient motion error can be measured purely using the respective methods described in this work and accounted for individually. The end-to-end system accuracy measurement includes the couch error since the couch is automatically adjusted in the procedure for this error component measurement. The patient motion occurring during beam delivery is also accounted for separately in the intrafraction motion error measurement.

2.1.1.3. Couch Error Measurement for Comparison

The measurement technique for the residual setup error for use in the margin calculation of this work assumed no couch error or patient motion. However, to compare the measurements in this work to previously reported values, the couch error was measured. Couch error is defined as the difference between the nominal distance of couch movement and the actual distance the couch was physically moved. A commercial image guidance phantom (QUASAR™ Penta-Guide, Modus QA, London, Ontario, Canada) was set up on the treatment couch daily as a part of routine morning QA. The therapist aligned marks on the phantom to the room lasers, then acquired a CBCT image. The couch was then shifted by a known amount which should result in a known displacement of the phantom and couch together. A verification CBCT image was then taken after the shift was performed by the couch. The displacement between the expected and measured positions of the phantom was calculated and recorded in three translational dimensions in the morning QA software. Measuring the couch error in this way, with a rigid phantom instead of a patient, removed any patient motion error so that the couch error could be measured alone.

2.1.2. Intrafraction Motion Measurements

2.1.2.1. Intrafraction Motion Measurement Methods

Intrafraction motion consists of any patient or target motion during a fraction while radiation is being delivered. As a surrogate for measuring the spine motion directly, patient surface monitoring data was collected with a commercial optical surface imaging system (Catalyst HD, C-RAD GmbH, Berlin, Germany). This system consists of three cameras which uses infrared light reflected off the patient surface to compare with a reference surface image and determine a displacement of the patient surface in six degrees. The single value representing the displacement is calculated by the average of the displacements of many surface points weighted

by their distance from where isocenter is projected to the surface. In order for the user to know how much the patient moves during treatment, a tolerance is set for each treatment type and the user may manually pause treatment if the patient surface moves outside of tolerance.

Patients receiving surface guidance typically received treatment for a breast or pelvic site and the pelvis patients were determined to be a suitable surrogate for spine patients. The section of the anatomy that is tracked for each patient is less susceptible to respiratory motion than the breast patients and isocenter is aligned with the spine. The patient's surface was a suitable surrogate for measuring the spine directly because the spinal target is rigid and does not experience much internal motion or deformation. A total of 409 sets of surface imaging data were collected for 26 patients and analyzed to estimate the mean value with 95% confidence to within 0.2 mm. One set of surface imaging data consists of all the shifts calculated at many time points during a fraction for a patient. The number of sessions for each patient ranged between 4 fractions and 23 fractions. There were two spine patients who received 4 fractions each and the rest were pelvis patients. All patients either received dual-arc VMAT or 3D conformal therapy with conventional or palliative fractionation. These patients were immobilized with vacuum cushions (Vac-Lok™ Cushions, Civco Radiotherapy, Coralville, IA) from the knees down with a wingboard. This setup was identical to the setup for spine patients intended to be treated with SBRT. There can be significant differences between the magnitudes of patient intrafraction motion between different immobilization devices, so it was necessary to use the device intended for use with spinal SBRT patients (Li *et al.*, 2012). The time for treatment was comparable to the approximately 6 minute spine treatments and there can be a small time-dependence on the magnitudes of patient motion (Hoogeman *et al.*, 2008).

The intrafraction motion data was analyzed retrospectively after the patients finished treatment. To obtain the intrafraction motion measurements, the patient was first set up according to skin marks made during simulation while also using the surface imaging results being displayed in real time as additional guidance. Once the patient was aligned with their planned treatment position, a reference surface image was recorded and used as the reference for calculating the resulting intrafraction displacements. The radiation therapist then initiated treatment and the optical surface imaging software recorded the displacement of the patient's surface in real time by recording the averaged displacements collected at 200 fps over three seconds to an exportable .csv file. In this file, which was obtained after each treatment, the displacements in six degrees were reported in three second intervals and were correlated with the beam-on information.

2.1.2.2. Intrafraction Motion Data Analysis

The intrafraction error due to patient motion was analyzed using a program was written in R to analyze each patient file exported from the C-RAD Catalyst software. During one treatment session or fraction, the patient's displacement from a reference image taken just before treatment was recorded every three seconds. The displacements were calculated in three translational dimensions and the 3D vector displacement was shown as well. The average time of treatment defined as the time from first beam-on to the last beam-on was 5 minutes and 5 seconds.

To analyze the data, the first data point with the beam on was subtracted from every other point to show the true intrafraction motion during the radiation delivery as opposed to the displacement from some other arbitrary position. Then, only the points between the first and last beam on points are extracted to be further analyzed. Some treatments were dual-arc VMAT treatments or had other interruptions, so all the data during the "treatment" phase of the motion

measurements should be used. Occasionally, large false or entirely missing deviations were recorded when the gantry crossed in front of one of the C-RAD cameras for a VMAT treatment, but with the large amount of data collected, these would have little effect on the results. It may cause the results of average deviation and the standard deviations to be slightly larger, but this is in the conservative direction. Breathing motion effects are minimized by selecting the pelvic region of the patient's surface and by the averaging performed in the C-RAD software for the data recording.

The averages and standard deviations of displacement in all directions were calculated over each fraction for each patient. Then, the patients were binned together and patient averages and standard deviations over all fractions were calculated. The root-mean-square sum of the patient standard deviations represents the random error and the standard deviation of the patient averages represents the systematic error (van Herk, 2004).

To compare with reported values, the standard deviation over all fraction averages over all patients was quantified. All time points were included in the analysis of each fraction average. Other works used CBCT images taken immediately prior to and after treatment and used the displacement between the two images as the measure of intrafraction motion. Therefore to perform a comparison of this work with the studies which utilized the CBCT technique, the surface imaging data was averaged over the average CBCT acquisition time of a minute at the beginning and ending of each treatment. The beginning average displacement was subtracted from the final average displacement for each fraction and the standard deviation was taken over all intrafraction displacements for all patients.

2.1.3. End-to-end System Accuracy Measurements

2.1.3.1. End-to-end System Accuracy Measurement Methods

The end-to-end system accuracy consists of the measured geometric shift between the planned dose distribution and the deliverable dose distribution. The standard deviations of the end-to-end system accuracy were previously measured, and the data was analyzed in this work to obtain the random and systematic error components (Barron, 2018). The measurement procedure followed the process of a patient treatment. First, the commercial two-dimensional diode array (MapCHECK2 serial number: 76352038; Sun Nuclear Corporation, Melbourne, FL) underwent CT simulation (LightSpeed 16 slice RT, General Electric Company, Chicago, IL) for radiation therapy dose planning. A spinal SBRT plan was created for a real patient in the Pinnacle treatment planning system (TPS) (Philips Radiation Oncology Systems, Fitchburg, WI) in accordance with the RTOG 0631 clinical trial. Then, the planned radiation beams from this plan were copied onto the diode array in the TPS. The plan was created for the patient using the patient's CT data sets and based on their anatomy, and then those beam parameters were applied to the diode array CT data set in order to calculate what the dose would be to the diode array instead of the patient. The diode array was then set up on a motion stage, a platform which can be moved in more precise increments on the order of millimeters, on the treatment couch and aligned with the room lasers to be in the same position as it was planned in. To verify and localize its position, a CBCT setup image was taken. This image was then registered with the planning CT using the XVI volume imaging software (Elekta, Stockholm, Sweden) and shifts were automatically calculated and manually adjusted. These shifts were automatically applied to the couch. The plan was then delivered to the diode array and it was shifted by 1 mm using the motion stage afterwards. The plan was again delivered, and the process was repeated 10 times in

order to achieve 1 mm resolution of the dose in the direction it was shifted since the diodes were 1 cm apart along one lateral direction. The diode array was set up and taken down between each completed planar dose measurement consisting of delivering the plan 10 times with the 1 mm shifts in between plan deliveries. Then, the dose profiles measured by the diode array were compared to the profiles obtained from the dose distribution in the treatment planning system in order to measure the shift between them.

The measure of end-to-end system accuracy was the distance between specified isodose points in the planned and measured dose profiles. This measurement procedure includes all sources of mechanical error causing a shift of the dose distribution from any of the systems involved from simulation to beam delivery, including the isocenter accuracies of the CBCT system and the radiation beam, the localization uncertainty of the CBCT system and the couch, and the MLC uncertainty (Ortega *et al.*, 2016). A total of 14 dose planes were measured, where 10 planes were measured in the lateral direction, and 4 planes were measured in the longitudinal direction. 5 profiles were analyzed through each plane where the diodes were spaced 1 cm apart.

2.1.3.2. End-to-end System Accuracy Data Analysis

To represent the systematic error component, the standard deviation over the average planar shift values was calculated. The random component was calculated as the RMS sum of the individual intra-planar standard deviations (Ortega *et al.*, 2016). Because there were no significant differences in the standard deviations of the measurements of the two directions and the anterior/posterior direction was not measured, all measured planes were binned together. The standard deviations of the system accuracy were found individually for each linear accelerator used and there was not a significant difference between the machines, and so these were binned together as well.

2.1.4. Penumbral width measurements

Since the setup margin value depends inversely on the penumbral width and increases with decreasing penumbra, the minimum penumbral width was measured on an actual spinal SBRT plan used at MBPCC. The minimum penumbra, or sharpest penumbra, exists in the superior/inferior direction in a dual-arc VMAT plan as seen in Figure 2. The edges of the beam there are mostly defined by the MLCs and not by the attenuation of other parts of the arc beam through the patient and past the target. Applying the minimum penumbral width to the margin recipe will yield the most conservative margin in terms of CTV coverage because it will result in larger margin (Oehler *et al.*, 2014).

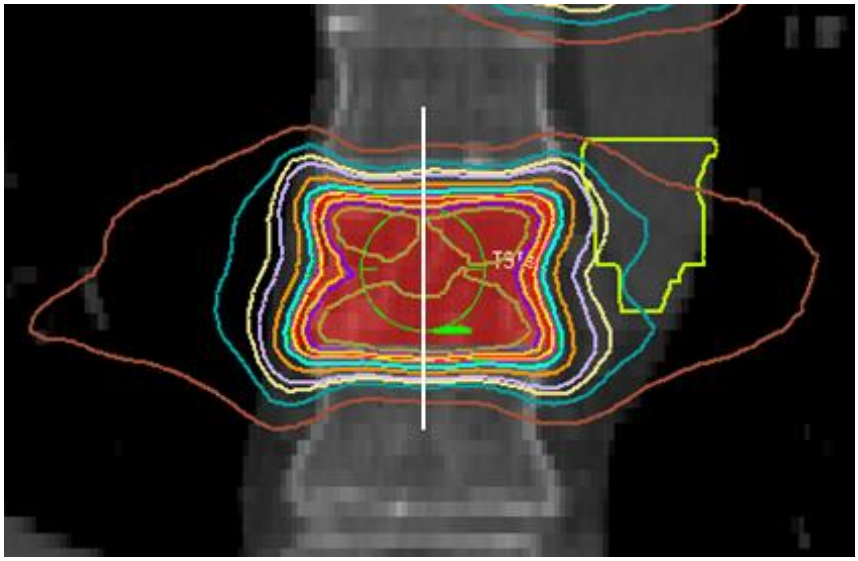

Figure 2. Coronal view of isodose lines resulting from the spinal SBRT plan used to measure the penumbral width. The white line shows the line along which the distance between isodose lines were measured.

To quantify the minimum penumbral width in the spinal SBRT VMAT plan, the distance between the 90% and 50% isodose lines was measured in the superior/inferior direction in a line through isocenter. This was done in the sagittal and coronal planes on the simulation CT images

in the Pinnacle TPS both above and below the target volume in the original plan for four total measurements. These four measurements were averaged to obtain the penumbral width value.

2.2. Aim 2

The second aim was to implement the SDE2 algorithm to determine a setup margin for an SBRT treatment which would meet the clinical constraints of CTV coverage and cord dose tolerances. This calculation included all known sources of uncertainty contributing to a geometrical shift from the planned dose distribution resulting in dosimetric changes to the target and surrounding normal tissue. The values used for the random and systematic error standard deviations and penumbral width were measured in the first aim. Figure 3 shows the flowchart of the inputs and outputs from each aim of this work. The magnitudes of the uncertainties used in the margin calculation were determined as the quadrature sum of all three types of uncertainties measured: intrafraction motion, residual setup error, and end-to-end system accuracy. The number of fractions was set to five because that is what spinal SBRT treatments will be treated with initially at MBPCC. The minimum acceptable dose to a point in the CTV was set at ninety percent of the prescription dose, in accordance with recommended dose limits in RTOG 0631. The minimum proportion of the population to receive the prescription dose to at least this

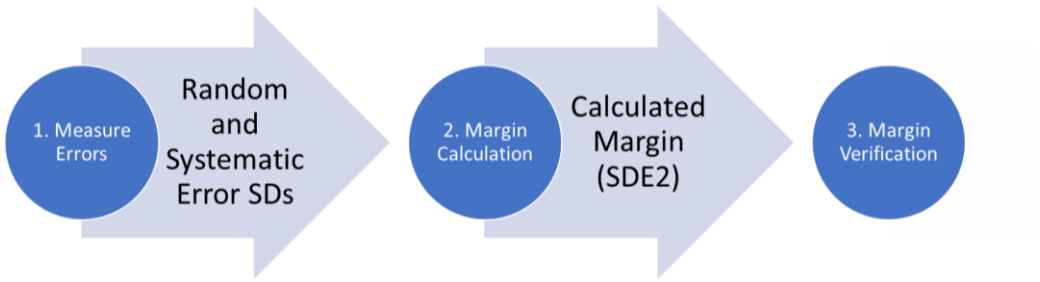

Figure 3. Flowchart of the three aims of this work showing how the result from the previous aim is the input to the following aim.

coverage level was also set to ninety percent as this has been used previously where margins have been calculated (Chang *et al.*, 2017; Gordon and Siebers, 2007; Hoogeman *et al.*, 2008).

The random and systematic errors were calculated as the combination of each error type: end-to-end system accuracy, intrafraction motion, and residual setup error as in Equation 5 and Equation 6. To achieve the most conservative margin which could be reasonably applied isotropically, the maximal directional errors were used for each error component. For example, for the random intrafraction error, the largest component was in the superior/inferior (Y) direction, so that value was used in Equation 5.

$$\sigma_{rand} = \sqrt{\sigma_{IF,rand}^2 + \sigma_{SE,rand}^2 + \sigma_{ETE,rand}^2}$$

Equation 5

$$\Sigma_{\mathrm{sys}} = \sqrt{\Sigma_{\mathrm{IF,sys}}^2 + \Sigma_{\mathrm{SE,sys}}^2 + \Sigma_{\mathrm{ETE,sys}}^2}$$

Equation 6

To investigate the sensitivity of the margin to the magnitudes of the error components and other parameters, the values of each component were varied in the model and the effects on the margin were measured. Only the intrafraction motion components of the systematic and random errors were varied about their measured values over the range of 0.1 to 1.1 mm because any error type will affect the result of the total random or systematic component in the same way. This type was chosen since it is the most likely component to change due to measuring the true spinal SBRT population or a change in immobilization device. Both the minimum percentage of the prescription dose inside the CTV and the percentage of the population receiving this dose were varied independently over a clinically relevant range of 80% to 99%. The number of fractions ranged from 1 to 5 since these are the number of fractions commonly used for spinal

SBRT (Redmond *et al.*, 2017). The penumbral width was also varied over a clinical range of values to determine sensitivity.

2.2.1. SDE2 Algorithm

The SDE2 algorithm is a simulation-based algorithm for calculating a margin for hypofractionated treatments, including SBRT treatments (Herschtal *et al.*, 2013). This algorithm consists of three separate steps in which first a dose-population histogram table is generated, then a margin is determined by interpolating between table entries, and then the margin value is validated with Monte-Carlo type simulations. The first two steps of the algorithm give an analytical result for the margin and the last part validates the result if the result is within a close enough range of the true value.

The DPH table generation step creates a look-up table from which dose-population histograms can be created. Figure 4 shows an example DPH plot. This portion of the algorithm produces a table of minimum dose to the CTV for each combination of random error standard deviation, systematic error standard deviation, and margin width for a set number of fractions. In

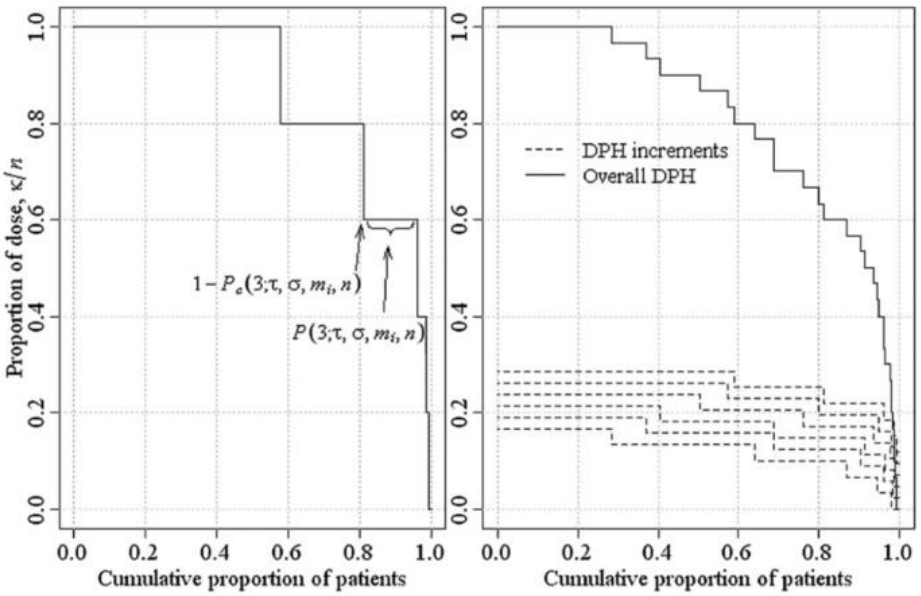

Figure 4. Example dose-population histogram (Herschtal *et al.*, 2013).

one row of the DPH table seen in Table 1, random shifts are sampled from the random and systematic error distributions given by the standard deviations of each. With the corresponding margin value added, the average minimum dose to a point on the CTV is recorded over a specified number of patients in the table. A separate table must be generated for each number of fractions per treatment. An example table is shown in Table 1 where the left three columns correspond to the set of parameters used for the row, and the numbered columns represent the average dose percent missed to a set of 100 simulated patients.

Table 1. Portion of a sample DPH look-up table generated from the first part of the SDE2 algorithm

Systematic	Random	Margin	1	2	3	4	5	6	7	8	9	10
0	0	1.6	5.48	5.48	5.48	5.48	5.48	5.48	5.48	5.48	5.48	5.48
0.02	0	1.6	5.84	5.84	5.84	5.84	5.84	5.84	5.84	5.84	5.84	5.84
0.04	0	1.6	6.23	6.23	6.23	6.23	6.23	6.23	6.23	6.23	6.23	6.23
0.06	0	1.6	6.63	6.63	6.63	6.63	6.63	6.63	6.63	6.63	6.63	6.63
0.08	0	1.6	7.05	7.05	7.05	7.05	7.05	7.05	7.05	7.05	7.05	7.05
0.1	0	1.6	7.49	7.49	7.49	7.49	7.49	7.49	7.49	7.49	7.49	7.49

The next part of the algorithm uses this table and interpolates between values of the error standard deviations, margins, and dose to determine the correct margin value for specified minimum dose and population values. Based on the user-specified systematic and random errors, it searches for the corresponding margin multiplication factor to meet the specified dose and population minimum goals within the table and generates a margin estimate. This factor is a multiplier for the random component of the van Herk margin recipe.

The final part of the algorithm, the simulation and verification step, simulates a large number of random shifts with a margin applied isotropically to a spherical CTV in three dimensions. It uses the calculated margin resulting from the second part of the algorithm as well as a range of margins about the calculated value. The dose is then calculated to the CTV for each of these margins and the process is summarized in Figure 5. First, the CTV is created as a set of

grid points on the surface of a sphere. Then, shifts are sampled from the measured error distributions for each fraction for each patient. For one example patient, the systematic error unique for that patient is set by generating a random number from a normal distribution with a mean of zero and a standard deviation equal to the measured systematic error. Then, the shifts for each fraction are set by generating random numbers from a normal distribution with a mean of the previously generated systematic error value and a standard deviation equal to the measured random error. The dose to the CTV resulting from these offsets generated in three dimensions is calculated by finding the cumulative normal percentile of each shifted CTV point from a distribution with a mean of zero and a standard deviation of the measured penumbral width. Then, the algorithm interpolates linearly between the margin multiplication factor values for the true value giving exactly the minimum population receiving the minimum dose, summarized in Figure 6. In the original version of the algorithm, if the true margin were not in the small range of margin values initially calculated from the initial input margin, it would not calculate an accurate margin estimate. To make this portion of the algorithm more user-friendly, error messages were added if the user-provided margin value was not in range of the values used for

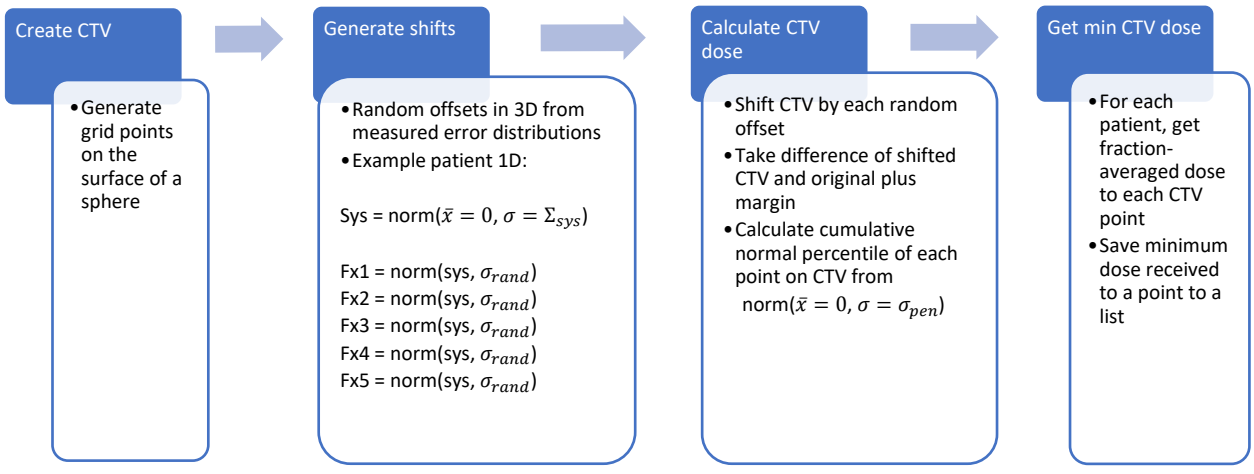

Figure 5. Dose calculation process in third module of SDE2 algorithm.

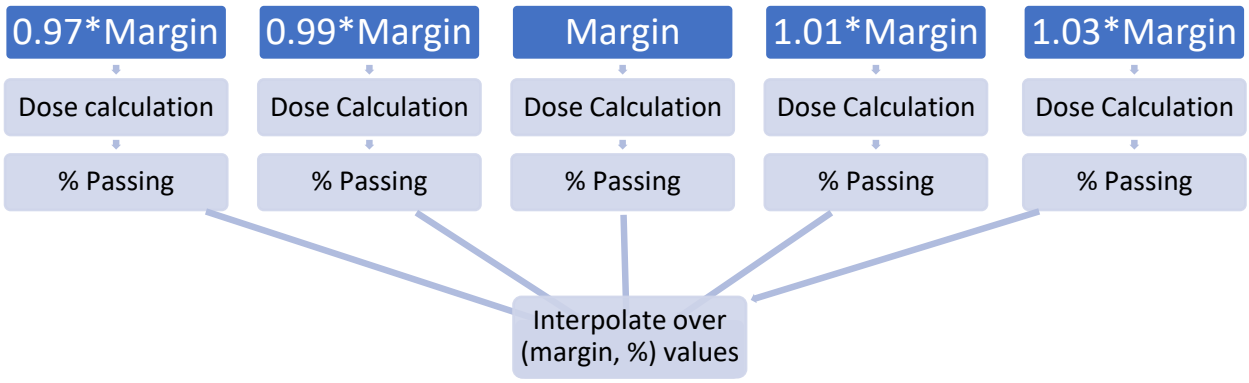

Figure 6. Margin verification and interpolation process in the third module of SDE2. Small variations of the margin are also calculated and simulated and then interpolated over.

interpolation. This would end the loop and exit the algorithm so that time is not wasted when the true margin value will not be given based on the initial input and a new guess is required.

The last verification portion of the SDE2 algorithm which simulates shifts sampled from the error distributions and interpolates to find the correct margin between the bounding values is computationally expensive and results in a value not clinically or significantly different than the average value of the margin resulting from the AVH method and the DPH method. It was determined to be more feasible to use the average values for sensitivity testing since the difference between the true margin calculated from the simulation code was not more than 0.01 mm different than the average between the two values generated from the margin calculation code. A potential clinical margin would be applied on the order of millimeters or tenths of millimeters, so this difference is not significant.

2.3. Aim 3

To achieve the third aim, the margin calculated in the second aim from measured parameters was validated by implementing random patient offsets from the measured error distributions and calculating the resulting dose to the CTV and the spinal cord in the TPS (Chuang *et al.*, 2007; Wang *et al.*, 2008). The margin formulations assume a spherical CTV,

homogenous patient, and uniform dose fall-off and conformity so it should be validated using a real plan. Four different margins were evaluated and compared. The results of adding the analytical margin both isotropically and in the superior/inferior direction only were compared with the standard 2 mm margin expansion typically used as well as with adding no margin to the CTV. Adding the analytical margin in the superior/inferior direction only was done because this is the direction with the minimum penumbral width which the margin calculation assumed. The other directions have much wider or larger penumbras and the calculated margin which used the minimum penumbral width is an overestimation of what is necessary to achieve the treatment objectives in the right/left and anterior/posterior directions.

First, new plans were created on the new PTVs resulting from each margin expansion all based on one clinical spinal SBRT plan. The original plan was a dual-arc, VMAT, spinal SBRT plan which was treated in five fractions to a prescription dose of 3000 cGy at MBPCC. The coverage of the CTV with the prescription dose was 97% in the original plan. The maximum cord dose was 2147 cGy. Each PTV was created from the expansion of the same CTV to make three different plans with expansions of (1) a 2 mm isotropic margin, (2) an isotropic margin calculated by the SDE2 algorithm, and (3) a superior/inferior margin calculated by the SDE2 algorithm.

The new plans were created by copying the original plan to a new trial and changing the inverse planning objectives to reflect the modified PTV. These plans were adjusted only to meet the dosimetric values of the original plan, but for the newly created PTVs. The goal was to create very similar plans to the original to get the most fair comparison between the different margins. It may bias the results if the new plans are created to be far superior to the original and then are compared with the original. To do this, the IMRT inverse planning objectives relating to the

original CTV were changed to apply to the newly created PTV in both cases and the plan was reoptimized with 25 iterations per cycle as many times as necessary until all the required constraints were met. The optimization objective weighting for the maximum dose to the PTV was increased from 75 to 85 to make the PTV maximum dose in the new plans more comparable to the original plan. These plans were then reviewed and accepted as suitable for treatment by a physician.

A script was written in the Pinnacle TPS to automatically perform the steps required for each margin validation. A separate program was used externally to generate the necessary random shifts for each fraction in each direction.

To generate the list of random displacements for multiple patients with five fractions per patient, a portion of code was used from the SDE2 algorithm. In the third step, during the dose calculations to the CTV, random displacements generated from the measured error distributions are calculated for each fraction for each patient. This portion of code is outlined in Figure 5 and does exactly what is required to generate the correct shifts used to verify the margin. These calculated shifts were subtracted from the original coordinates of the isocenter in the original plan to simulate a patient offset in that direction. The measured errors were measured with respect to the patient, so if the isocenter was to be changed to reflect a patient offset, it needed to be moved in the opposite direction. For example, assume the x-coordinate of the original isocenter was 15 cm, and the randomly sampled shift from the measured error distribution was +0.5 mm for that fraction. This would correspond to a shift of the patient in the positive x direction of 0.5 mm. In order to shift the dose distribution instead of the patient by that same shift, the position of the delivered dose determined by the isocenter coordinates should be shifted

in the negative x direction by 0.5 mm. This means the new isocenter should be 14.5 mm to reflect this patient offset.

The script in the Pinnacle TPS then performed the steps necessary to perform the margin validation. First, the isocenter of the original treatment plan was changed in the TPS to reflect each shift generated from the separate program described previously. Then, the dose was calculated to the patient and the spinal cord resulting from the beams being shifted in this way. The total number of monitor units delivered by the linear accelerator was kept constant from the original plan so that the dose was not renormalized to a different point with each isocenter shift. The dose-volume histogram (DVH) values were computed, and the desired values were output to a file to be analyzed. These values were the percentage of the CTV receiving 100% of the prescription dose (30 Gy), the maximum point dose in the spinal cord, and the volume of the spinal cord receiving 23 Gy. This process was repeated for every shift representing a new dose calculation for each fraction for each patient.

Since a loop could not be created in Pinnacle to repeat this process, a script was written in R to repeatedly write the same set of Pinnacle commands to a file for each shift of isocenter, with the new isocenter values for each shift changed each time. This resulting file can be directly run by Pinnacle as a set of continuous commands which results in an output file containing the new isocenters which were set, and the corresponding results of the volume of the CTV receiving at least the prescription dose, the volume of cord receiving 23 Gy, and the maximum dose to the cord.

A total of 140 different 5-fraction simulated treatments were performed as described above for each of the four margins to estimate the percentage of treatments receiving at least 90% of the prescription dose to the CTV to within 10 percent.

Chapter 3. Results

3.1. Aim 1

The standard deviations of random and systematic components of residual setup error, intrafraction motion, and end-to-end system accuracy are tabulated in Table 2 and graphically represented in Figure 7 and Figure 8.

Table 2. Measurements of all the error components with uncertainties based on standard errors of standard deviations.

		σ_x (mm)	σ_y (mm)	σ_z (mm)	Vector (mm)
Intrafraction Motion	Systematic	0.319 ± 0.045	0.520 ± 0.074	0.347 ± 0.049	0.434 ± 0.061
	Random	0.760 ± 0.039	1.19 ± 0.066	0.736 ± 0.033	1.10 ± 0.078
Residual Setup Error	Systematic	0.054 ± 0.009	0.034 ± 0.006	0.074 ± 0.012	0.075 ± 0.012
	Random	0.097 ± 0.007	0.087 ± 0.007	0.082 ± 0.005	0.111 ± 0.007
End-to-end System Accuracy	Systematic	0.312 ± 0.061	0.312 ± 0.061	0.312 ± 0.061	0.312 ± 0.061
	Random	0.539 ± 0.033	0.539 ± 0.033	0.539 ± 0.033	0.539 ± 0.033

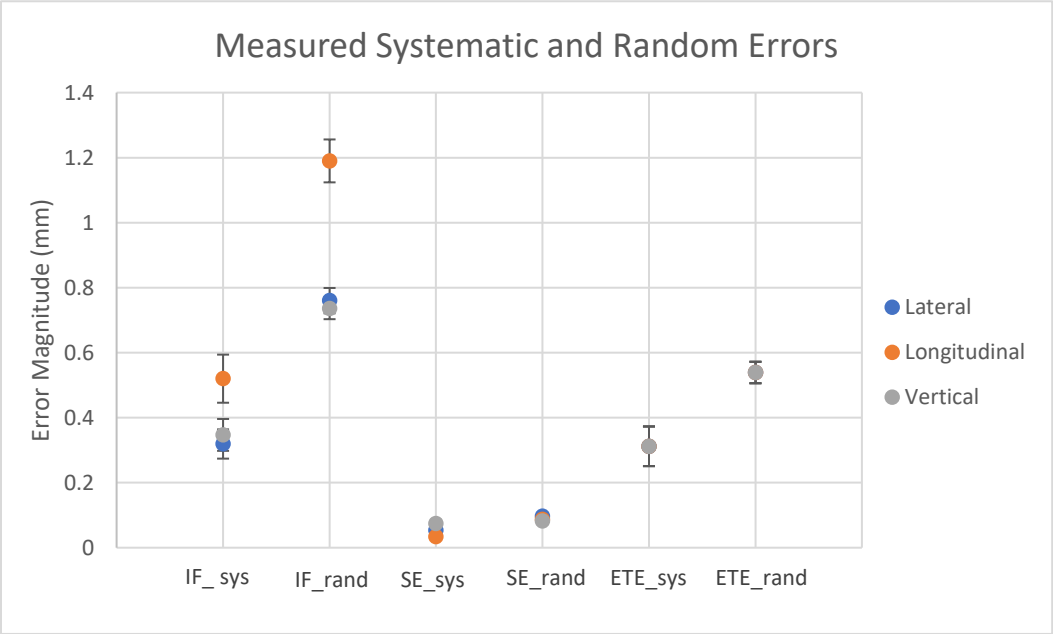

Figure 7. Standard deviations of random and systematic error components of all measured errors.

Figure 8. Standard deviations of random and systematic error components of all measured errors.

3.1.1. Residual Setup Error Results

For each patient, the average and standard deviation of the residual setup error was calculated over all 195 images. The standard deviation of the patient averages was used as the systematic error, and the RMS sum of the patient standard deviations was used as the random error in each direction (van Herk, 2004). For comparison between the values measured in this work and values from the literature, the standard deviation over all individual residual setup error measurements were quantified for each translational direction and is quantified in Table 5 in the Discussion chapter.

A total of 195 images for 20 different patients were analyzed in XVI. The results for systematic and random error in three dimensions as well as the vector magnitude are tabulated in Table 2. The largest random error occurred in the lateral or right/left direction, and the largest systematic error occurred in the anterior/posterior direction.

3.1.2. Intrafraction Motion Error Results

The results for intrafraction motion systematic and random error in three dimensions as well as the vector magnitude are tabulated in Table 2. The largest systematic and random errors were in the longitudinal or superior/inferior direction. This direction was significantly different from the other two directions, while the other two directions were not significantly different from each other.

3.1.3. End-to-end System Accuracy Results

The standard deviations of the of the end-to-end system accuracy as described in the first aim was previously measured at MBPCC and the data was analyzed to obtain the systematic and random components of error (Barron 2018). Figure 9 shows an example profile through one planar dose measurement. The results for the systematic and random standard deviations of end-to-end system accuracy measurements are tabulated in Table 2.

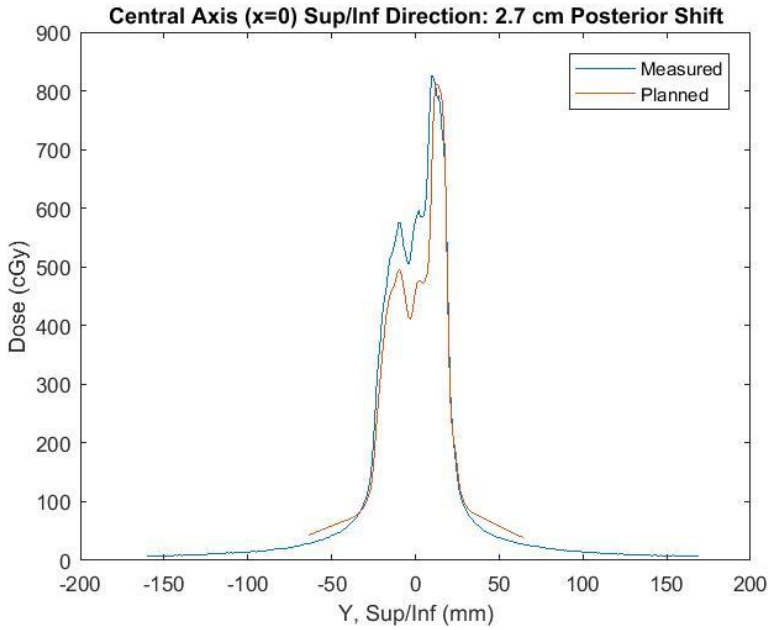

Figure 9. Example dose profile through a measured plane on the diode array to measure end-to-end system accuracy

3.1.4. Penumbral Width Results

The minimum penumbral widths over the four superior and inferior directions from the sagittal and coronal plane views were averaged and divided by 1.28 to achieve a 3.57 ± 0.03 mm penumbral width measurement for use in the margin recipe. Figure 10 shows the coronal view of the planned dose distribution used for the penumbral width measurement.

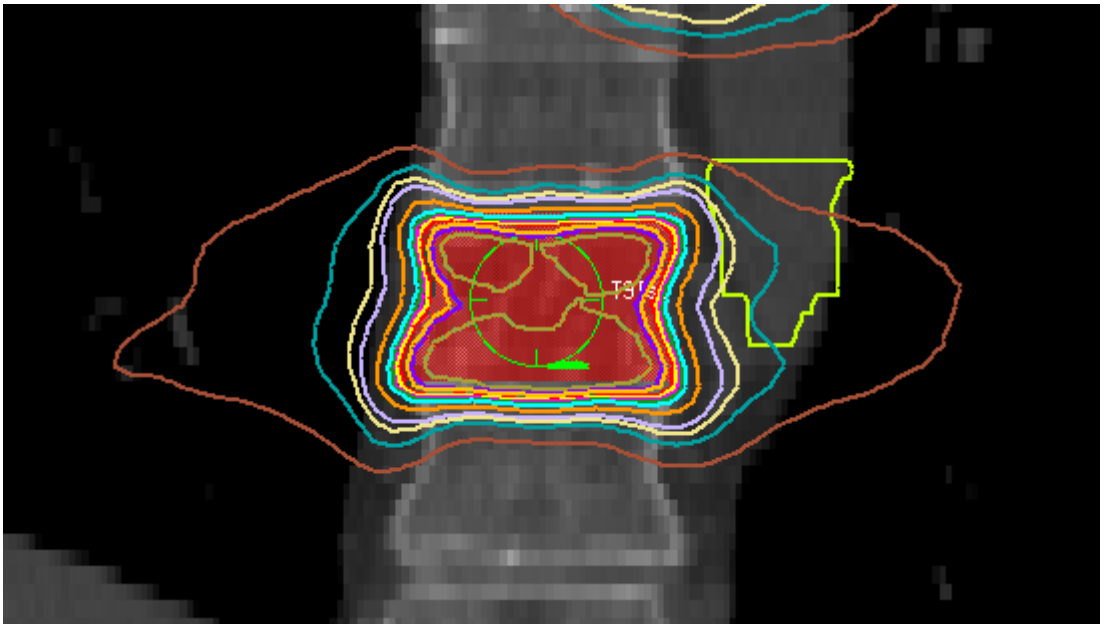

Figure 10. Coronal view of the dose distribution of a clinical spinal SBRT dual-arc VMAT plan.

3.2. Aim 2

3.2.2. Margin results

The margin calculated with the SDE2 algorithm for the clinical parameters chosen and tabulated in Table 3 was 2.42 mm. The margin values reported in Figure 11 through Figure 16 show the sensitivity of the margin to the input parameters to the margin calculation: fraction number, desired minimum population percentage, desired minimum dose level, penumbral width, and the random and systematic components of error standard deviation. The margin

Table 3. Measured parameters and values set based on recommendations and intended practices. These values were input into the margin recipe and will be used as the standard values with which to compare for sensitivity testing.

Penumbral Width	3.57 mm
Systematic Error	0.611 mm
Random Error	1.31 mm
Fractions	5
Population Percentage	90 %
Dose Percentage	90 %

values shown for fractions 2 through 5 are the average of the AVH method and the DPH method. The 1 fraction margin values reported were calculated with the AVH margin recipe alone and were verified with the third step of the SDE2 algorithm since the DPH method could not give a result for a one fraction treatment. The "current standard" values in these figures corresponds to the 2.42 mm margin from the parameters in Table 3. Figure 11 shows the margin dependence on the number of fractions, which increases over 1 mm when decreasing the number of fractions from 5 to 1. Figure 17 shows the directional margins calculated from the corresponding directional error values, but with the same minimum penumbral width of 3.57 mm, 5 fractions, and dose and population objectives.

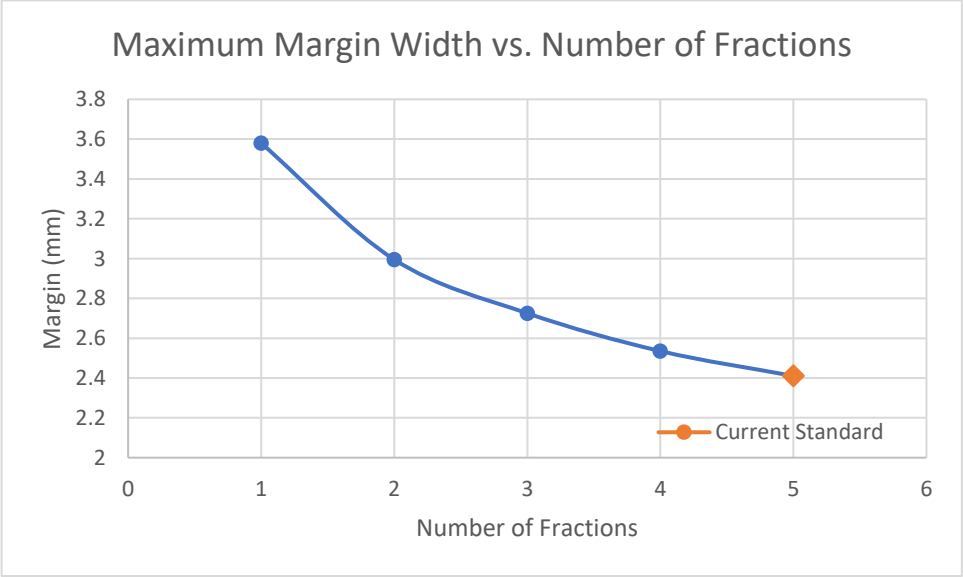

Figure 11. Results of sensitivity testing for margin dependence on number of fractions

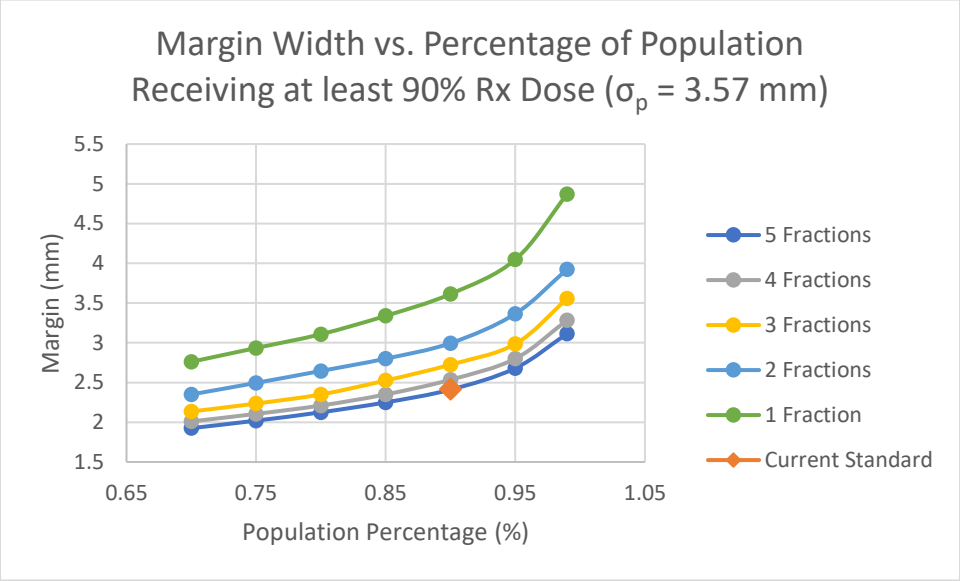

Figure 12. Results of sensitivity testing for percentage of population receiving at least 90% of prescription dose, with all other parameters set to standard.

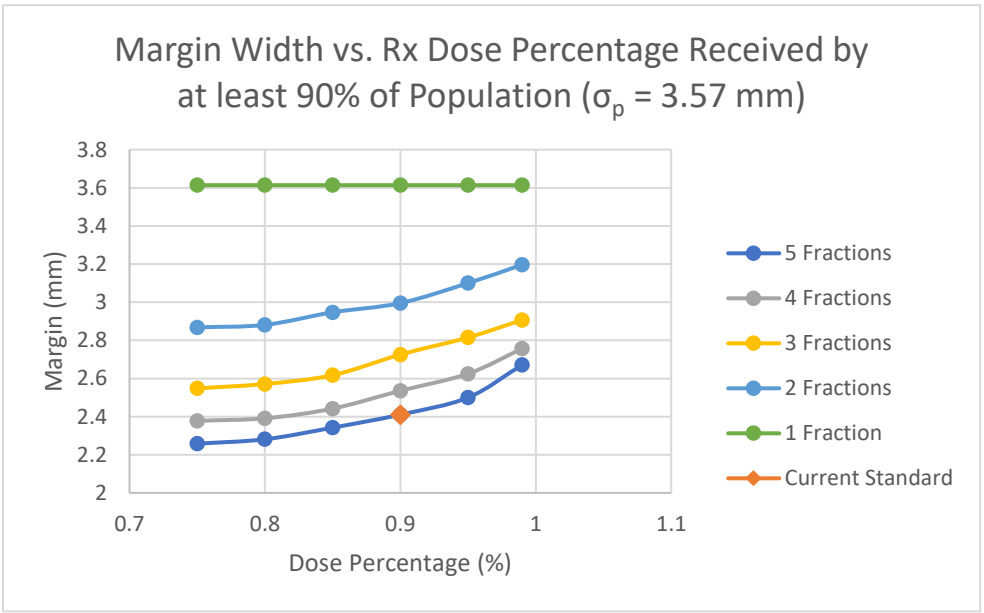

Figure 13. Results of sensitivity testing for percentage of dose received by at least 90% of patient population, with all other parameters set to standard.

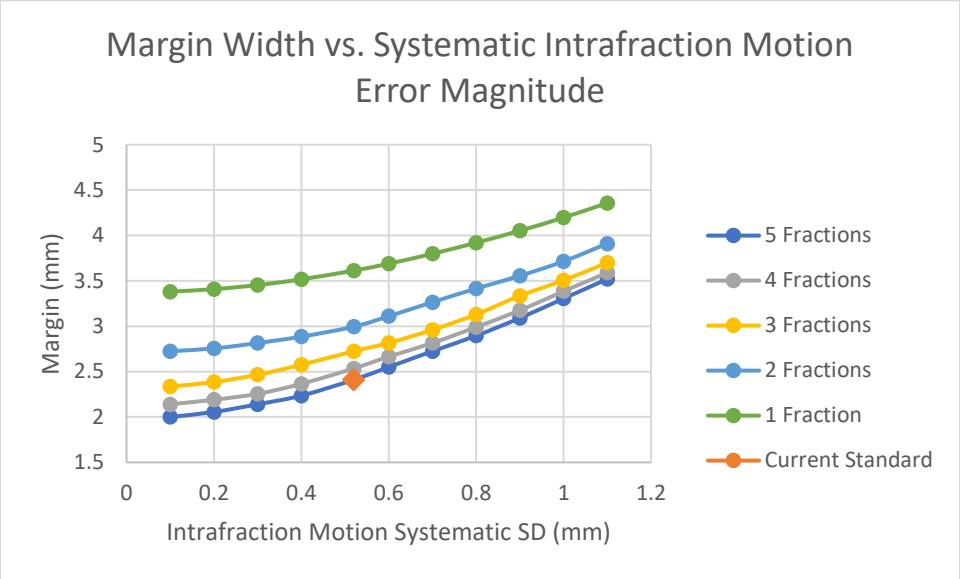

Figure 14. Results of sensitivity testing for margin dependence on a systematic error magnitude over various fraction numbers

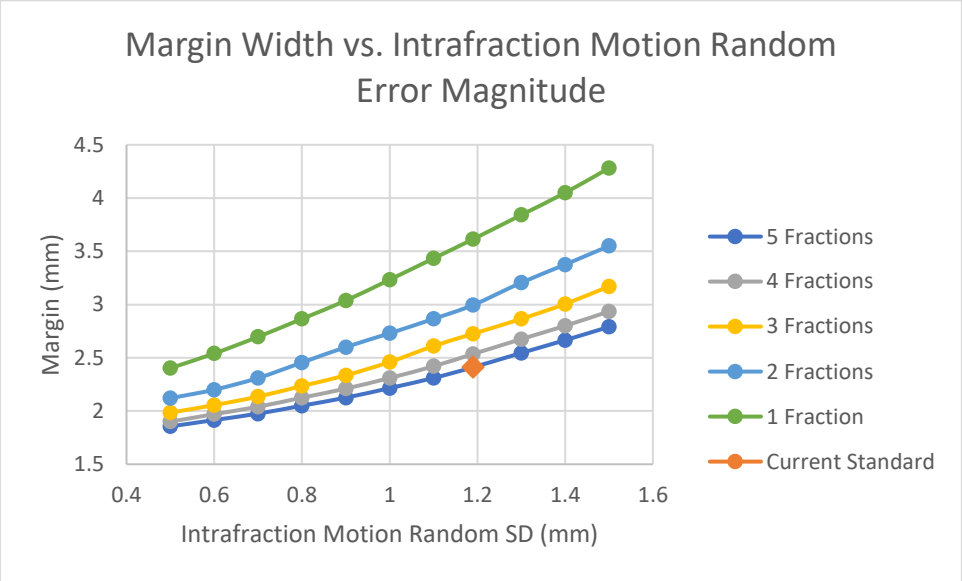

Figure 15. Results of sensitivity testing for margin dependence on a random error magnitude over various numbers of fractions

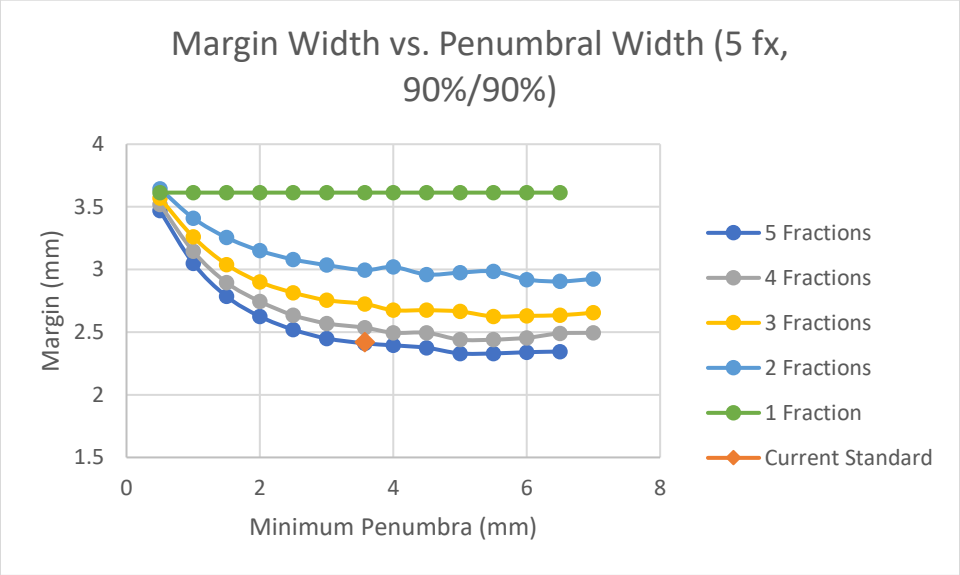

Figure 16. Dependence of margin width on penumbral width for varying fraction numbers.

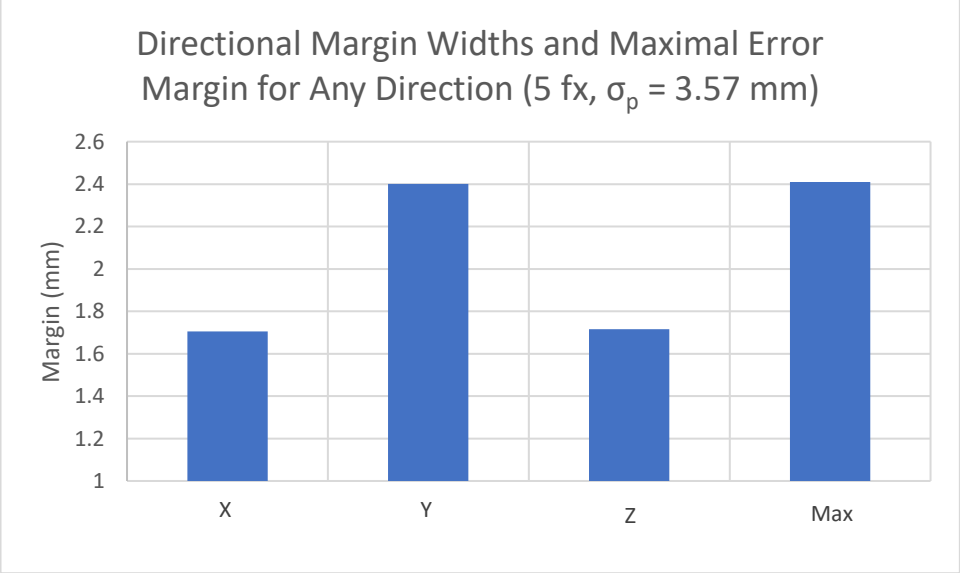

Figure 17. Differences in margin widths for maximum error values and measured directional values.

3.3. Aim 3

The file output by Pinnacle as a result of computing dose and volume constraints for each isocenter shift was analyzed by another R script. First, the data was grouped by constraint type, such as the maximum dose to the spinal cord. Then the patient average of each constraint was

computed over all fractions for each patient. Then the percentage of all patients passing each dose or volume criteria was computed and output into the results file which tabulates all relevant values. The total percentage of patients which passed all criteria was also computed and written to the output file. The results are tabulated in Table 4.

Table 4. Results of margin validation. Percentages of 140 patient population passing the required DVH criteria for a spinal SBRT plan with ± 5% error

	No Margin	Current Margin 2mm isotropic	Analytical Margin 2.4 mm isotropic	Analytical Margin 2.4 mm SI only
> 30 Gy to > 90% of CTV	81 %	100 %	100 %	96 %
< 30 Gy Max Dose to Cord	100 %	100 %	100 %	100 %
< 23 Gy to < 10% of Cord	99 %	99 %	99 %	100 %
Passing All Criteria	81 %	99 %	99 %	96 %

With no margin added, only about 80% of patients will receive at least the full prescription dose to at least 90% of the CTV. Also, with no margin added, it is expected that the organ at risk constraints to the cord would not be significantly affected, and only one patient did not meet the constraint of less than 23 Gy to 10% of the cord. With the analytical margin added in all directions, again only 1 patient did not meet the cord criteria, but 100% of the 140 simulated patients received greater than 90% coverage of the CTV with the prescription dose. Even with the 5% uncertainty, approximately 95% of patients still would have received enough dose to enough of the CTV. When the analytical margin is added only in the direction where the penumbral width matched most closely to the value used in the margin calculation, the proportion of the population came closer to the predicted value of 90%. Figure 18 shows the DPHs resulting from the addition of the 3 different margins and the original plan with no margin added. The DPH for the plan created with the analytical margin applied in the superior/inferior direction only shows that it came the closest to achieving greater than 90% CTV coverage for 90% of patients. These treatment goals summarize the DPH goal shown by the “Desired 90/90”

blue point in Figure 18. The DPH for the original plan shows that it failed to achieve the DPH goal and did not achieve coverage for enough patients. Both plans with the isotropic margins added far exceeded the DPH goal, meaning more normal tissue would be irradiated than is necessary to meet the stated treatment objectives. All plans shown achieved at least 85% coverage of the CTV for all patients.

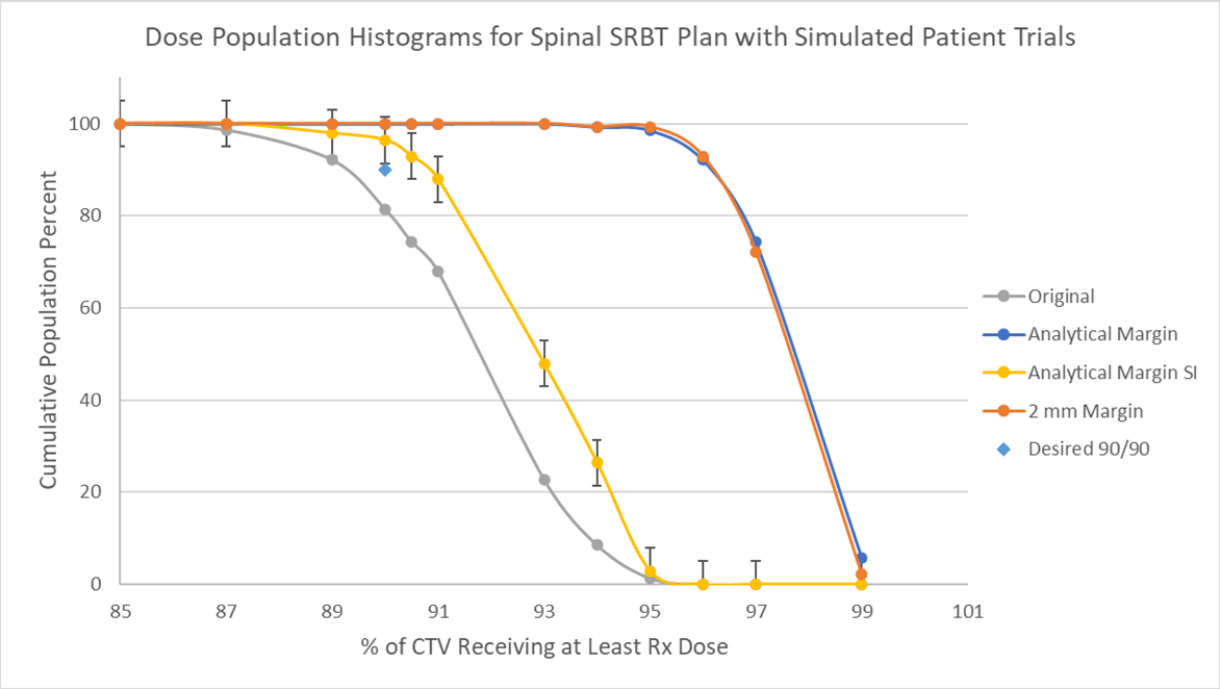

Figure 18. Dose population histograms resulting from the addition of different margins.

Chapter 4. Discussion

4.1. Aim 1

Residual setup error for spinal SBRT has been quantified previously at other institutions. In these other works, additional verification images were acquired after the automatic couch shifts were applied following the registration of the initial setup image with the planning CT. These verification images were taken just before treatment begins, so the displacements from the planning images were recorded as the measure of residual setup error. At MBPCC, it was not standard practice to obtain these verification images after the automatic couch shifts for all treatments; treatment began immediately after the couch is shifted assuming the shifts calculated were less than 1 cm. Obtaining an IRB to perform the additional imaging on the low number of spine patients at MBPCC would have been out of the time scope for this project and would not have allowed for a reasonable level of uncertainty in the results.

When comparing the residual setup error results to other works, this work reported significantly smaller values for overall standard deviation as tabulated in Table 5. This was most likely due to the measurement method in this work and its resulting exclusion of patient motion and couch shift error. Hyde reported a 10-minute average interval between verification and initial CBCT images, which included any displacement due to patient motion in that time. The intrafraction motion results reported in the same work are over a similar time interval and have a similar magnitude (Hyde *et al.*, 2012). Also, when obtaining the verification image after the couch has shifted automatically, the error in the couch movement and ability to reach the calculated displacement is included in the measured difference between the images.

Table 5. Measurements of overall standard deviation of residual setup error for comparison with literature values. Measured values significantly different from Chang and Hyde at p = 0.05.

	σ_x (mm)	σ_y (mm)	σ_z (mm)
Measured	0.102	0.104	0.152
Chang	0.4	0.3	0.25
Hyde	0.5	0.4	0.3

When the measured couch error was combined with the measured residual setup error, there was still a discrepancy between this work and others, but this was likely due to the exclusion of patient motion in the time interval in this work. However, the pure residual setup error was needed for the margin recipe, so this is what was used in the calculation, without adding in the couch or patient motion errors because those are accounted for separately. The couch error is tabulated in Table 6.

Table 6. Couch error overall standard deviation in each direction as well as the combined couch and residual setup error. Standard deviations were combined in quadrature.

	σ_x (mm)	σ_y (mm)	σ_z (mm)
Couch Error Alone	0.0549	0.0521	0.0528
Couch and Setup Error	0.116	0.116	0.161

The measured pure residual setup error may still be a conservative estimate due to the presence of manual adjustments in the accepted registrations. Since the registration software calculates shifts in all six degrees of freedom including rotational directions, there was usually a small discrepancy between the accepted shifts and the shifts calculated from the re-registration. Rotational shifts were not applied with the couches available. Since the therapists applied a manual shift after automatically registering the images, sometimes there were small deviations from the automatic result that were intentional. For this reason, the resulting mean and standard deviations should have overestimated the setup error since some of the deviation may have been intentional for some cases.

The intrafraction motion error has been quantified previously for spinal SBRT at other institutions. The method of measurement typically used was performing CBCT scans prior to treatment, after the initial setup was verified, and comparing them to the CBCT scans taken just after treatment ended. A displacement vector between the images was calculated and averaged over many patients to yield a measure of intrafraction motion (Kim *et al.*, 2009; Li *et al.*, 2012). This method was considered for use in this work, but with the time limits for the project and the requirement of an Institutional Review Board approval along with patient consent for a procedure which would add time and extra dose to a patient's treatment, it was not considered feasible. Additionally, the number of spine patients treated at MBPCC was limited since spinal SBRT was not yet fully implemented. The surface imaging technique has been shown to be comparable to the CBCT technique, so this was an acceptable substitute (Stieler *et al.*, 2013).

When comparing the results of the intrafraction motion errors in this work to other similar works, all were in close agreement as shown in Table 7. When data from all time points from this work was included in the analysis, the results were very close to previously reported values, with no significant differences. However, when comparing the "CBCT acquisition" results, there was a larger error compared to what was previously reported. Since Kim and Li utilized IMRT and had longer average imaging intervals during treatment of 40 minutes and 12 minutes respectively, this could have caused a discrepancy. This may also have been because the surface was measured in this work instead of the position of the rigid spine determined with

Table 7. Comparing overall standard deviations of measured intrafraction motion with literature values. *Significant difference from "Measured, CBCT Acquisition"

	σ_x (mm)	σ_y (mm)	σ_z (mm)
Measured, CBCT Acquisition	1.22	1.60	1.07
Measured, All	0.770	1.05	0.721
Kim	0.5*	0.9	0.6
Li	1.1	1*	0.7*

CBCTs, creating a larger variation in displacements over multiple patients and fractions. But when analyzing all time points measured, the outliers had less of an effect and the results aligned more closely with those reported previously.

Comparing the overall standard deviations over all fractions between measured directions, all were significantly different from each other. These measurements were performed on mostly conventional pelvis patients and not spinal SBRT patients as in the other works, but they had the same important treatment parameters and were an acceptable substitute.

Reporting the largest errors in the superior/inferior direction was consistent with other results (Tseng *et al.*, 2015; Kim *et al.*, 2009; Li *et al.*, 2012). Tseng used MRI imaging, but still reported the largest displacements occurring in the superior/inferior direction for bulk patient motion contributing to a shift of the spinal cord. No explanation was given for this and there was no information on whether this difference between the other two translational directions was statistically significant.

Ortega performed a very similar end-to-end system accuracy test as was performed for use with this work and obtained very similar results to those in this work for systematic and random error as shown in Table 8 (Ortega *et al.*, 2016). The results of this work are valid for MBPCC with all the same systems and equipment as when the measurements were taken. If a piece of equipment is replaced, this error may need to be measured again to check for any changes.

Table 8. Comparison of previously reported systematic and random errors to the measured results which were applied to all directions. No significant differences.

	Ortega			Measured
	x (mm)	y (mm)	z (mm)	Any (mm)
Random	0.6	0.4	0.4	0.539 ±0.033
Systematic	0.3	0.2	0.2	0.312 ± 0.061

The measured penumbral width of 3.57 mm was similar to the value used in another work where a margin was calculated with the SDE2 algorithm of 3 mm (Chang *et al.*, 2017). This value was measured only in one plan for one patient, but the minimum penumbral width occurs in the direction perpendicular to the gantry rotation. This will not change much for any other coplanar dual-arc VMAT plan since it will have a similar patient geometry relative to the gantry rotation.

4.2. Aim 2

The margins calculated in this work using SDE2 are conservatively appropriate for use at the Mary Bird Perkins Cancer Center with the proposed spinal SBRT treatment protocol. This involves using vacuum bag immobilization for a dual-arc FFF VMAT treatment, using the same workflows and protocols for patient setup using CBCT images, and the same equipment and QA procedures to ensure the end-to-end system accuracy is still representative of the treatment delivery system. However, a user at another institution may measure these parameters and utilize the SDE2 calculation in the same way to generate an appropriate margin.

The margin formulation shows different sensitivities to the various input parameters, so uncertainties in these parameters could influence the accuracy of the margin calculated. The penumbral width is not expected to change much from patient to patient if they also receive a VMAT treatment due to the geometry of the treatment site relative to the gantry rotation. However, even if there is some variation, Figure 16 shows that the margin does not change more than approximately 0.2 mm over 1 mm of penumbral width change. Increasing the desired population percentage has a more dramatic effect on the margin width than increasing the dose coverage that population receives, so the physician may adjust these parameters based on preference. The random and systematic intrafraction error standard deviations are the largest

error components and the most likely to change because in this work, actual spinal SBRT patients weren't being treated yet and could not be measured, so the true values may differ slightly. Also, the immobilization device used could be changed more easily than other treatment components like the linear accelerator. The margin is sensitive to the intrafraction motion error components with a nearly linear relationship for a 5-fraction treatment in the region about the measured errors.

Comparing the margin results generated by the AVH method alone and the SDE2 algorithm from the sensitivity testing of all margin dependencies over clinically relevant ranges, there were only a few large differences as seen in Figure 19. The outliers larger than 0.4 mm were due to very sharp penumbral widths included in the margin recipe, which are not likely to be seen clinically with the same treatment conditions. The SDE2 recipe was more accurate with these extreme conditions, but it is a complicated algorithm which takes a relatively long time to

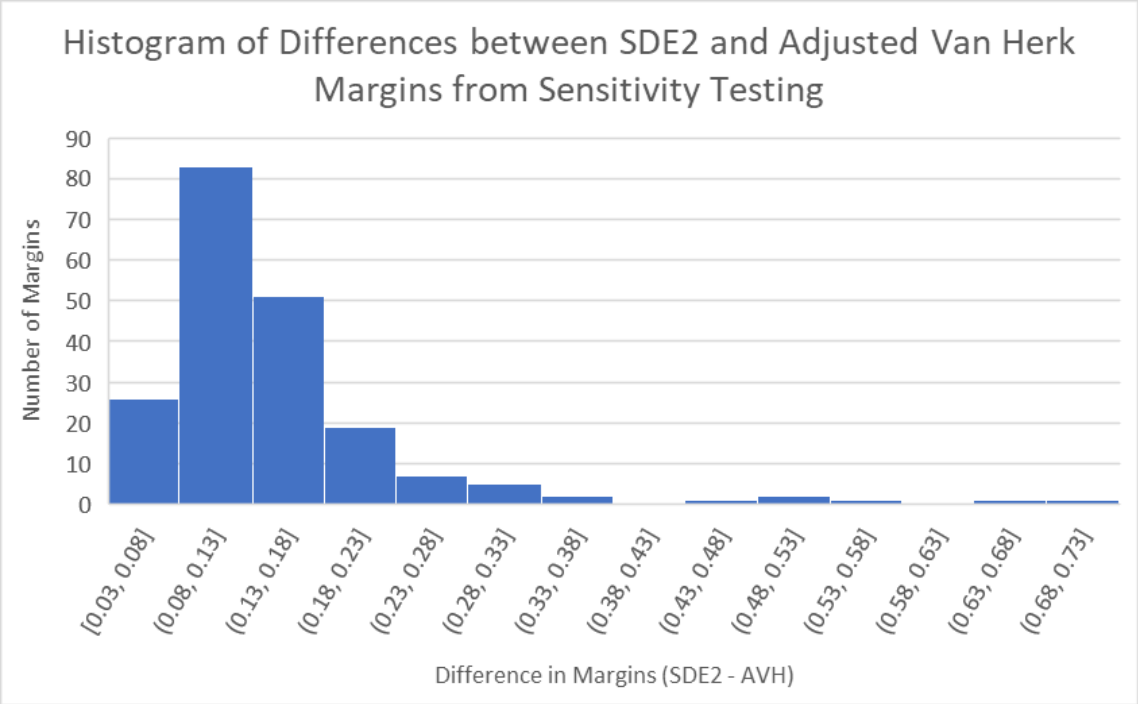

Figure 19. For varying combinations of parameters input into the margin recipe, the differences in margin between the SDE2 and AVH margins were calculated for a histogram.

generate results. The AVH recipe is a simple formula so results can be generated instantly, and it gives results very close to the SDE2 recipe within more realistic ranges of parameters.

4.3. Aim 3

Based on the results of computing the dose received to the CTV and the spinal cord for simulated patients, a 2.4 mm margin applied only in the superior/inferior direction was sufficient to guarantee at least 90% of the prescription dose for at least 90% of simulated patients. The cord dose was less than the tolerance values for 100% of patients with this margin added. However, this formulation does not account for differences in importance between different areas of the CTV to receive the prescription dose. This formulation can only guarantee that some 90% of the volume will receive the prescription dose, but the particular volume may be variable. However, the results in Table 4 show that with the errors measured in this work, even the calculated margin of 2.42 mm applied in all directions did not cause the cord dose to exceed tolerances for 99% of cases. A margin providing the desired coverage level of the CTV which does not expand the treated volume towards the spinal cord is preferable, since moving the treated area closer to this critical structure increases the likelihood of exceeding the dose tolerances for the cord in the presence of uncertainties.

The measured percentage of the population meeting the CTV dose criteria was larger than what the margin calculation guaranteed likely due to some of the conservative assumptions made in the margin formulation. One conservative assumption was the perfect conformity of the prescription dose to the target (Van Herk *et al.*, 2000). The margin formulation assumes that with any shift, there is a loss of dose and coverage to the CTV. But in this work the CTV had an irregular shape, the spinal cord was in close proximity, and the minimum requirement for coverage of the PTV was 90%. When errors were introduced, the coverage did not necessarily

decrease as fast as the margin recipe assumed or even decrease at all. Another assumption was that the minimum penumbra used in the calculation described the Gaussian penumbra of the dose-fall off for the entire target. But if a shift occurred in a direction with a wider penumbra, less dose was missed to the CTV than if the direction of the shift were along the minimum penumbra direction. The irregular shape also means the dose fall-off may not follow a Gaussian shape and may not decrease as quickly. The algorithm is also designed to guarantee that the minimum dose point on the CTV does not fall below a desired level, which results in a higher dose to the entire volume. The coverage goals used in the clinic only require a certain volume to not fall below a desired dose level, so the constraint the margin was designed to meet was stricter than the clinical outcome desired.

Chapter 5. Summary and Conclusions

5.1. Limitations and Future Work

One limitation of this work is the lack of error measurements on spinal SBRT patients treated at MBPCC, especially for the intrafraction motion measurements. However, the pelvis patients measured had identical important treatment parameters. Once more patients are treated with spinal SBRT at MBPCC, an area of future work would be to measure the intrafraction motion of these patients and compare those results to those of this work for a potentially different margin result. Another limitation is that only one type of plan with one particular shape for the CTV was used for validating the margin. There are many different shapes the CTV can take based on the extent of the disease and its invasion to other areas of the vertebral bodies. This work used a plan with a T-spine lesion, which is the most common region for secondary metastasis (Conti *et al.*, 2019). Future work could include validating the margin for different plans with different CTV shapes, potentially where the CTV would include the spinal cord. Other types of treatment geometries may be used to treat spinal SBRT, such as using couch kicks and non-coplanar arcs, and the results from this work may not apply to those types of treatments. The minimum penumbral width may occur in a different direction other than superior/inferior which would impact how the margin works to ensure coverage of the CTV. It may also impact the spinal cord dose if the dose distribution has a significantly different shape in that region. Also, many conservative assumptions were made in this work, so the margin could potentially be reduced further to minimize the irradiation of normal tissue. However, this work shows that the margin needs only to be applied in the superior/inferior direction by 2.4 mm which does not include any critical structures. Another limitation was that different plans had to be created for each new volume created by the addition of a new margin. Efforts were made to ensure

similarity between these plans, but in order to achieve the most fair comparison, the plans should have nearly identical dose distributions inside the CTV and outside of it to avoid any bias of one plan having a more favorable dose distribution from the start.

5.2. Conclusion

This work showed that 96 ± 5% of patients would have sufficient target coverage with the prescription dose with the SI margin applied, but this does not imply that the remaining percentage of patients would necessarily fail treatment. The CTV is meant to account for microscopic disease surrounding the known tumor volume. However, the region of the CTV that did not receive the full dose may not actually have tumor cells. According to the DPH in Figure 18, 100 ± 5% of patients received greater than 87% coverage with the application of the 2.4 mm SI margin. This means that the patients who failed the 90% coverage criteria were not failing catastrophically, but only had a small loss in coverage relative to the desired level.

This work also establishes a framework for future margin calculations in case of changes to any of the treatment parameters or for use at a different institution. This margin calculation method could be applied elsewhere with measurements of the end-to-end system accuracy for the unique set of equipment used in the radiotherapy system, of intrafraction motion for spinal SBRT patients or of a similar group, and of the residual setup error representative of the institution's procedures. If the shape of the target were much different from the CTV from this work, the verification method of Aim 3 could be applied to the specific plan. This may be desired if the target includes multiple vertebrae, includes the spinal cord, or has a very different shape.

The results of this work support the hypothesis that a margin can be determined for spinal SBRT which meets the treatment goals of greater than 90% of the prescription dose to greater than 90% of the population without exceeding the cord dose tolerances of 30 Gy maximum point

dose and 23 Gy to 10% of the cord subvolume. The treatment goals were conservatively met with a 2.42 mm SI margin for a five-fraction dual-arc VMAT spinal SBRT treatment with vacuum bag immobilization at Mary Bird Perkins Cancer Center. All sources of known measurable error during radiotherapy were included in the margin formulation and the margin was verified on a clinical spinal SBRT treatment plan.

References

2010 The International Commission on Radiation Units and Measurements *Journal of the ICRU* **10** NP.2-NP

Berthelsen A K, Dobbs J, Kjellen E, Landberg T, Moller T R, Nilsson P, Specht L and Wambersie A 2007 What's new in target volume definition for radiologists in ICRU Report 71? How can the ICRU volume definitions be integrated in clinical practice? *Cancer Imaging* **7** 104-16

Chang J H, Sangha A, Hyde D, Soliman H, Myrehaug S, Ruschin M, Lee Y, Sahgal A and Korol R 2017 Positional Accuracy of Treating Multiple Versus Single Vertebral Metastases With Stereotactic Body Radiotherapy *Technol Cancer Res Treat* **16** 231-7

Chuang C, Sahgal A, Lee L, Larson D, Huang K, Petti P, Verhey L and Ma L 2007 Effects of residual target motion for image-tracked spine radiosurgery *Med Phys* **34** 4484-90

Conti A, Acker G, Kluge A, Loebel F, Kreimeier A, Budach V, Vajkoczy P, Ghetti I, Germano A F and Senger C 2019 Decision Making in Patients With Metastatic Spine. The Role of Minimally Invasive Treatment Modalities *Front Oncol* **9** 915

Cox B W, Spratt D E, Lovelock M, Bilsky M H, Lis E, Ryu S, Sheehan J, Gerszten P C, Chang E, Gibbs I, Soltys S, Sahgal A, Deasy J, Flickinger J, Quader M, Mindea S and Yamada Y 2012 International Spine Radiosurgery Consortium consensus guidelines for target volume definition in spinal stereotactic radiosurgery *Int J Radiat Oncol Biol Phys* **83** e597-605

Finnigan R, Lamprecht B, Barry T, Jones K, Boyd J, Pullar A, Burmeister B and Foote M 2016 Inter- and intra-fraction motion in stereotactic body radiotherapy for spinal and paraspinal tumours using cone-beam CT and positional correction in six degrees of freedom *J Med Imaging Radiat Oncol* **60** 112-8

Gordon J J and Siebers J V 2007 Convolution method and CTV-to-PTV margins for finite fractions and small systematic errors *Phys Med Biol* **52** 1967-90

Herschtal A, Foroudi F, Silva L, Gill S and Kron T 2013 Calculating geometrical margins for hypofractionated radiotherapy *Phys Med Biol* **58** 319-33

Hoogeman M S, Nuyttens J J, Levendag P C and Heijmen B J 2008 Time dependence of intrafraction patient motion assessed by repeat stereoscopic imaging *Int J Radiat Oncol Biol Phys* **70** 609-18

Hyde D, Lochray F, Korol R, Davidson M, Wong C S, Ma L and Sahgal A 2012 Spine Stereotactic Body Radiotherapy Utilizing Cone-Beam CT Image-Guidance With a Robotic Couch: Intrafraction Motion Analysis Accounting for all Six Degrees of Freedom *International Journal of Radiation Oncology*Biology*Physics* **82** e555-e62

Jeon S H and Kim J H 2018 Positional uncertainties of cervical and upper thoracic spine in stereotactic body radiotherapy with thermoplastic mask immobilization *Radiat Oncol J* **36** 122-8

Kim S, Jin H, Yang H and Amdur R J 2009 A study on target positioning error and its impact on dose variation in image-guided stereotactic body radiotherapy for the spine *Int J Radiat Oncol Biol Phys* **73** 1574-9

Li W, Sahgal A, Foote M, Millar B A, Jaffray D A and Letourneau D 2012 Impact of immobilization on intrafraction motion for spine stereotactic body radiotherapy using cone beam computed tomography *Int J Radiat Oncol Biol Phys* **84** 520-6

Lyons C A, King R B, Osman S O S, McMahon S J, O'Sullivan J M, Hounsell A R, Jain S and McGarry C K 2017 A novel CBCT-based method for derivation of CTV-PTV margins for prostate and pelvic lymph nodes treated with stereotactic ablative radiotherapy *Radiat Oncol* **12** 124

Oehler C, Lang S, Dimmerling P, Bolesch C, Kloeck S, Tini A, Glanzmann C, Najafi Y, Studer G and Zwahlen D R 2014 PTV margin definition in hypofractionated IGRT of localized prostate cancer using cone beam CT and orthogonal image pairs with fiducial markers *Radiation Oncology* **9**

Ortega C, Wunderink W, Delgado D, Moragues S, Pozo M and Casals J 2016 Evaluation of the setup margins for cone beam computed tomography-guided cranial radiosurgery: A phantom study *Med Dosim* **41** 199-204

Parker B C, Shiu A S, Maor M H, Lang F F, Liu H H, White R A and Antolak J A 2002 PTV margin determination in conformal SRT of intracranial lesions *J Appl Clin Med Phys* **3** 176-89

Redmond K J, Lo S S, Soltys S G, Yamada Y, Barani I J, Brown P D, Chang E L, Gerszten P C, Chao S T, Amdur R J, De Salles A A, Guckenberger M, Teh B S, Sheehan J, Kersh C R, Fehlings M G, Sohn M J, Chang U K, Ryu S, Gibbs I C and Sahgal A 2017 Consensus guidelines for postoperative stereotactic body radiation therapy for spinal metastases: results of an international survey *J Neurosurg Spine* **26** 299-306

Stieler F, Wenz F, Shi M and Lohr F 2013 A novel surface imaging system for patient positioning and surveillance during radiotherapy. A phantom study and clinical evaluation *Strahlenther Onkol* **189** 938-44

Tseng C L, Sussman M S, Atenafu E G, Letourneau D, Ma L, Soliman H, Thibault I, Cho B C, Simeonov A, Yu E, Fehlings M G and Sahgal A 2015 Magnetic resonance imaging assessment of spinal cord and cauda equina motion in supine patients with spinal metastases planned for spine stereotactic body radiation therapy *Int J Radiat Oncol Biol Phys* **91** 995-1002

van Herk M 2004 Errors and margins in radiotherapy *Semin Radiat Oncol* **14** 52-64

Van Herk M, Remeijer P, Rasch C and Lebesque J V 2000 THE PROBABILITY OF CORRECT TARGET DOSAGE: DOSE-POPULATION HISTOGRAMS FOR DERIVING TREATMENT MARGINS IN RADIOTHERAPY *Int. J. Radiation Oncology Biol. Phys.* **47** 1121-35

Wang H, Shiu A, Wang C, O'Daniel J, Mahajan A, Woo S, Liengsawangwong P, Mohan R and Chang E L 2008 Dosimetric effect of translational and rotational errors for patients undergoing image-guided stereotactic body radiotherapy for spinal metastases *Int J Radiat Oncol Biol Phys* **71** 1261-71

Zhang M, Zhang Q, Gan H, Li S and Zhou S M 2016 Setup uncertainties in linear accelerator based stereotactic radiosurgery and a derivation of the corresponding setup margin for treatment planning *Phys Med* **32** 379-85

Zhang Q, Chan M F, Burman C, Song Y and Zhang M 2013 Three independent one-dimensional margins for single-fraction frameless stereotactic radiosurgery brain cases using CBCT *Med Phys* **40** 121715

Vita

Audrey Copeland was born in Plano, Texas in 1995 and is the daughter of Joni and William Copeland. Audrey grew up in Garland, Texas and graduated from Garland High School in 2013. She enrolled at The University of Texas at Austin and received her Bachelor of Science in Radiation Physics with a certificate in Scientific Computation in 2017. She matriculated into the Master of Science in Medical Physics program at Louisiana State University and will graduate in the spring of 2020. In the summer of 2020, she will begin residency in Philadelphia at Thomas Jefferson University Hospital.