January 2019

Utilizing Log Files For Treatment Planning And Delivery Qa In Radiotherapy

Carl W. Stanhope
Wayne State University, cstan969@gmail.com

UTILIZING LOG FILES FOR TREATMENT PLANNING AND DELIVERY QA IN RADIOTHERAPY

by

CARL STANHOPE

DISSERTATION

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

2019

MAJOR: MEDICAL PHYSICS

Approved By:

_____

Advisor Date

_____

_____

_____

_____

DEDICATION

To my parents and brother for being an abundant source of inspiration in life and for always being there - you've helped make this journey pretty easy, even when it maybe wasn't!

ACKNOWLEDGEMENTS

I would like to thank my PhD advisor Dr. Di Yan for allowing me this opportunity and guiding me during my PhD. I would also like to thank each of my co-authors for helping in each of their respective ways. Many thanks to Doug Drake for helping me out with various clinical medical physics concepts. Thank you to Dr. Markus Alber and Dr. Matthias Sohn for providing their Monte Carlo expertise. Your e-mail responses from Germany were always quick and helpful. Thanks to Dr. Jian Liang for helping me to assemble the software and providing various research insights. I would also like to thank Dave Gersten for providing software support and being my coffee buddy. To my dissertation committee – Dr. Yan, Dr. Burmeister, Dr. Alber, Dr. Snyder, and Dr. McDermott – thanks for the oversight. Lastly, I would like to thank my colleagues – Charbel Habib, David Solis, Afua Yorke, Nick Myziuk, Ina Sala, and Evan Porter - for keeping things entertaining. Cheers, and may our paths intersect again!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 INTRODUCTION & BACKGROUND

1.1 Basics of Radiotherapy

Radiation has been shown consistently capable of destroying cancer cells. For this reason it is often used for the treatment of tumors. Most often this treatment involves the firing of high energy photons (a particle of electromagnetic radiation) out of a linear accelerator (linac) at the target region of interest. A sample linear accelerator is displayed in Fig. 1. The amount of radiation delivered to a patient is defined as dose and measured in units of Gray (Gy). Because radiation is dangerous to not only cancer cells, but also healthy tissue, the radiation field is shaped to the target of interest to minimize the quantity of harmful radiation from reaching healthy tissues. This shaping is done using primary collimators as well as multi-leaf collimators (MLCs). Primary collimators, also called jaws, are large heavy metal blocks that create a rectangular or square radiation field of a certain size. MLCs are tungsten leaves such as those found in Fig. 2 which move back and forth to block unnecessary radiation from reaching healthy tissues. MLCs can be used to create irregularly shaped fields that accurately deliver dose to the tumor while minimizing dose to healthy tissue.

Figure 1 An Elekta Infinity linear accelerator

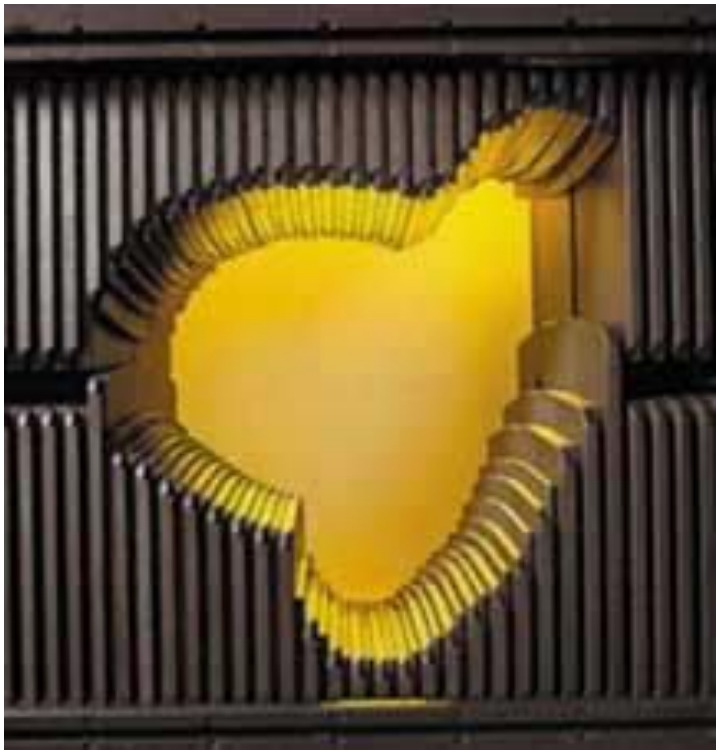

Figure 2. A multi-leaf collimator (MLC) shapes the radiation field

Two common methods (modalities) for radiotherapy are intensity modulated radiotherapy (IMRT) and volumetric modulated arc radiotherapy (VMAT). IMRT sub-modalities include step-and-shoot and sliding-window. For step-and-shoot IMRT, radiation is only turned on when the leaves are stationary. Leaves are moved into new orientations while radiation is off. For sliding-window IMRT, leaves are allowed to move while radiation is turned on. This allows for radiation to more precisely conform to the target. VMAT is similar to IMRT except that the treatment head rotates around the patient and dose rate is adjustable during delivery. This allows conformal target doses to be delivered in a relatively shorter period of time. Depending on the treatment site and prescribed dose, VMAT may also reduce toxicity probabilities.

Every patient's tumor shape and anatomy are different. Thus, every patient requires an individual treatment plan. A treatment plan is defined by the mechanical orientation of the MLCs and jaws as well as the amount of radiation being delivered (monitor units (MU)) and the orientation of the treatment head (gantry angle). The treatment plan is comprised of a sequence of control points containing each of these machines parameters. As the treatment is delivered, the MLC/collimators moves, the gantry rotates, and MU is outputted according to these control points.

1.2 Treatment Planning Optimization

Treatment planning optimization is an iterative process in which the aforementioned machine parameters are determined so as to achieve a desired dose distribution that uniformly covers the tumor with the prescribed dose while minimizing dose to healthy tissue. The treatment planning system (TPS) used in this study was Pinnacle3 by Philips. The optimization algorithm utilized herein as part of Pinnacle3 is a sequential quadratic programming (SQP)

optimization algorithm from the NPSOL library.[1] The dense SQP algorithm utilizes gradient descent to achieve progressively more optimal treatment plans.[2,3] During each iteration, the following steps are carried out: (1) a machine parameter is changed; (2) a 3D dose distribution is calculated for the new plan; and (3) the quality of the new plan is assessed. For step 2, dose is calculated using a dose calculation method. Two dose calculation methods are assessed as part of this study and will be introduced shortly. For step three, plan quality is assessed using a dose objective function. Put simply, a dose objective function is a series of dose-volume objectives that the plan should ideally meet. These dose-volume objectives are best explained in terms of dose-volume histograms (DVHs). Sample DVH curves are plotted in Fig. 3. Each curve displays how much dose is delivered to each structure. Specifically, DVHs display what percentage of each structure receives how much dose.

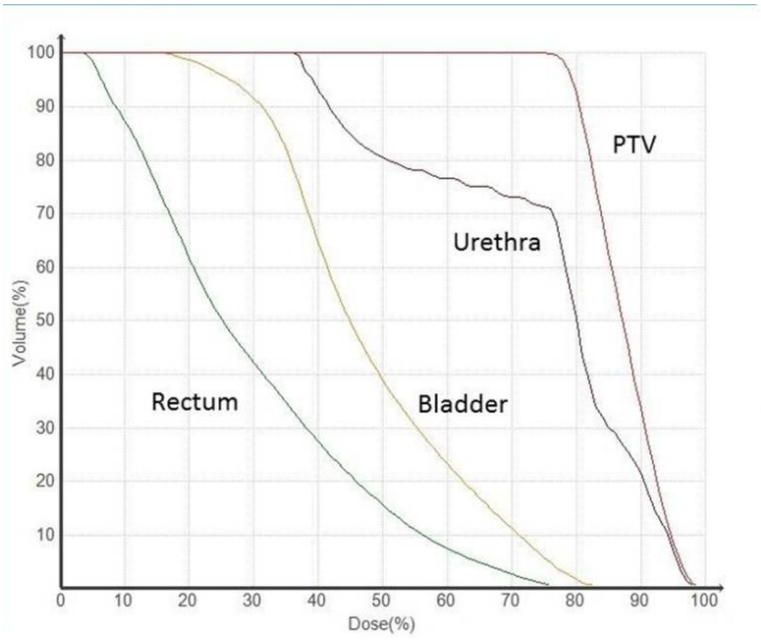

Figure 3 A sample dose-volume histogram (DVH) shows what percent volume of the regions of interest (ROIs) receive what percentage of the prescription dose.

A volume greater than the tumor itself is treated. The tumor itself is labeled the gross tumor volume (GTV). A margin of expansion is utilized to account for any nearby microscopic

cancer involvement as well as any regional involvement (e.g. lymph nodes). This expanded volume is labeled the clinical target volume (CTV). The internal target volume (ITV) is the CTV expanded by a margin that accounts for any intra-treatment tumor motion due to patient motion. Lastly, the planning target volume (PTV) is the ITV expanded by some margin to account for any potential setup errors. The prescribed dose is delivered to the PTV to assure that the CTV receives the prescribed dose.

1.3 Dose Calculation Methods

Dose calculation methods are responsible for accurately calculating the dose delivered to the patient. The two main parts of a dose calculation method are the dose calculation algorithm and the beam model. In general, there are three types of dose calculation algorithms: (1) convolution-superposition, (2) Monte Carlo, and (3) Boltzmann Solvers. Generally speaking, convolution-superposition (CS)-based algorithms are fast but less accurate while Monte Carlo (MC) and Boltzmann Solver-based algorithms are slower but more accurate. CS-based algorithms are most commonly used, however, as computers become increasingly faster, Monte Carlo and Boltzmann Solver are also becoming increasingly common. Two calculation algorithms were assessed as part of this study. First, Pinnacle3's collapsed-cone convolution-superposition algorithm was assessed. This is the algorithm that our institution is currently utilizing as part of treatment planning optimization. Second, ScientificRT's SciMoCa Monte Carlo algorithm was assessed. Unfortunately, Monte Carlo is still too slow to be routinely used in treatment planning and optimization, therefore, SciMoCa is utilized to verify the accuracy of CS dose calculation. Boltzmann Solver algorithms are introduced but are not an important part of this study.

Each brand of linac, each model of linac, and each individual linac is constructed differently. These differences mean that every linac must be modelled individually. Furthermore, every commissioned linac energy must also have its own beam model. This means every linear accelerator may have 2-3 beam models. Every treatment planning system (TPS) utilizes a slightly different beam modelling process, but largely speaking these parameters are consistent for each type of dose calculation algorithm. For instance, CS beam modeling parameters include values of MLC and collimator transmission, minimum leaf separation, effective source size, quantification of electron contamination, and many other factors. In order to set these beam modelling parameters, a wide variety of dose measurements are taken in a water tank. Beam modelling parameters are set so that the resultant calculated doses match the measured doses as accurately as possible. It is important to understand that dose calculation accuracy is not only dependent on the inherent dose calculation accuracy of an algorithm (e.g. CS or MC), but also dependent on a high quality beam model which is able to accurately calculate dose for a wide range of clinical cases. It follows that any differences in dose between CS and MC may be due to inherent differences in the calculation algorithms or may simply be due to poor beam modelling. Whereas the prior type of dose difference is unavoidable, the latter dose difference is fully avoidable. When evaluating MC vs. CS dose differences in this study, differences will be evaluated as avoidable or unavoidable. This step is critical from the perspective of treatment planning QA.

1.3.1. Monte Carlo

Task Group report of the AAPM 105 serves as a great introduction to the basics of Monte Carlo as well as current issues associated with the utilization of such algorithms.[4] Monte Carlo is a statistical simulation method for calculating dose that relies on physical principles of

radiation transport in matter. In Monte Carlo, a large (~10^8) number of particle histories are simulated and recorded. For example, let a photon be generated and fired at the patient of interest. Using physical interaction probabilities, the distance to interaction as well as interaction type can be determined. The particle is then transported to the interaction site and the interaction simulated. These steps are repeated until all secondary particles leave the dose grid or deposit all of their energy (E < pre-specified energy threshold). In Monte Carlo, quantities of interest are recorded over a vast numbers of particle histories. The statistical average of their cumulative effects is utilized to calculate the dose delivered to the patient.

The Monte Carlo algorithm utilized by this study was ScientificRT's SciMoCa algorithm. SciMoCa shares its fundamental concept with the voxel Monte Carlo (VMC) family of codes, e.g. VMC++ or XVMC.[5-7] SciMoCa is also based off a more recent series of papers by Sikora et al.[8-9] The beam model utilized is not clinically commissioned and is currently used for research purposes only. For all calculations, unless otherwise stated, particle histories are simulated to achieve 1% dosimetric variance at 70% max dose.

1.3.2. Convolution-Superposition

Convolution-superposition dose calculation was introduced by Thomas Mackie in 1984.[11-13] The algorithm allows for the calculation of 3-dimensional (3D) dose in heterogeneous patient anatomies for irregular field sizes. The basic process of CS is as follows. First, a Monte Carlo kernel is calculated. This kernel, displayed in Fig. 4 represents how radiation will deposit its energy in the patient or some medium relative to the interaction voxel (in bold). CS utilizes two of these Monte Carlo kernels. The first is the primary kernel which accounts for dose deposition for primary radiation – radiation originating from the target or source. The second is the scatter kernel which accounts for dose deposited by secondary/scattered radiation –

radiation originating from attenuated primary radiation. These kernels then undergo the process of convolution superposition. The easiest way to think of convolution-superposition is as follows. Every voxel in the patient or phantom is going to have some amount of primary and secondary radiation interaction occur in it – this quantity is the photon fluence. However, the dose from this interaction is not deposited entirely in the interaction voxel. Rather, the dose is deposited in surrounding voxels in accordance with the kernels. By considering every voxel in the patient as an interaction voxel, tracking the dose deposition to all other voxels using these kernels, and summing all of these doses, a 3D dose distribution is acquired.

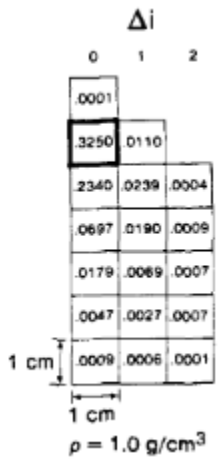

Figure 4 Monte Carlo calculated primary dose spread arrays. The numbers represent the dose deposited in the voxel normalized to the collision KERMA in the interaction voxel.

There are several sub-types of convolution-superposition algorithms. The CS algorithm utilized herein was Pinnacle3's collapsed-cone convolution-superposition algorithm. Collapsed-cone was introduced in 1988 by Anders Ahnesjo as a means of reducing the computational cost of standard convolution-superposition algorithms.[14] One of the key differences of collapsed-cone CS is that it utilizes an analytical Monte Carlo kernel calculated using Eq. 1, where A_θ, a_θ, B_θ, and b_θ are functions of the scattering angle θ and the beam energy acquired from pre-calculated lookup tables. The first term accounts for primary dose while the

second term accounts for secondary dose. It follows that dose can be calculated using Eq. 2 where D is the dose, f is a depth hardening correction factor, T is an approximate TERMA distribution, ρ is the medium density, and h is the poly-energetic kernel. The variables **r** and **s** correspond to the dose-deposition and interaction voxels respectively. The computational cost of this integral is $O(N^6)$, where N is the number of voxels along one side of the dose grid. However, by discretizing space into a spherical coordinate system, this computational cost can be greatly deduced. Using a spherical coordinate system (l,m,n) comprised of conical shell segments of solid angle Ω_{mn} around Θ_m, φ_n, and of thickness Δr, dose can instead be calculated using Eq. 3. Let η be a spatially varying function of the heterogeneous medium density. The radiant primary energy along the line through all segments on the line is calculated via recursion one voxel segment at a time as in Eq 4. Using this formalism, dose is instead calculated in $O(M \cdot N^3)$ where M is the number of cones. This is a huge speedup and the key advantage over collapsed-cone convolution-superposition over standard convolution-superposition algorithms.

$$\boldsymbol{h_w(r,\theta)} = \frac{\left(A_\theta e^{-a_\theta r} + B_\theta e^{-b_\theta r}\right)}{r^2} \tag{1}$$

$$\boldsymbol{D(r)} = \frac{f(r)}{\rho(r)} \iiint \boldsymbol{T(s)\rho(s)h(s,r)d^3s} \tag{2}$$

$$\boldsymbol{D(r)} = \frac{\eta(r)}{\rho(r)} \frac{1}{d^2 u} \sum_m \sum_n \left[a_m R^p_{mn}(r) + b_m R^s_{mn}(r)\right] \tag{3}$$

$$\boldsymbol{R^s_{mn}(r_i)} = R^s_{mn}(r_{i-1})(1 - b_m \eta_i \Delta r) + T_i \rho_i \Omega_{mn} d^2 u \frac{B_m}{b_m} \Delta r \tag{4}$$

1.3.3 MC vs CS

As stated recently, the primary difference between MC and CS is speed versus accuracy. The loss of accuracy in CS (with respect to MC) is primarily seen in two circumstances. The first inadequacy of CS is when the patient anatomy is particularly heterogeneous. Arnfield et al. showed differences in Monte Carlo (PEREGRINE) vs. collapsed-cone convolution-superposition as large as 10% in a heterogeneous lung phantom for a 4x4 cm^2 square field.[15] The effects of heterogeneity on Monte Carlo and collapsed-cone calculated doses will be further assessed as part of this study. The second inadequacy of CS is when the treatment plan is comprised of many small segments. Zhao et al. showed collapsed-cone CS consistently over-estimated PTV dose by 5% compared to MC.[16] Zhen et al. showed collapsed-cone was 3-4% higher for IMRT and VMAT plans than Accuros XB's (AXB) Boltzmann Solver algorithm.[17] For reference, AXB is typically considered similar in accuracy to Monte Carlo.[18] For each of these publications, collapsed-cone is shown to over-estimate target dose for complicated IMRT and VMAT fields. The effect of beam complexity on this discrepancy will be analyzed as part of this study.

Dose differences (MC vs. CS) due to heterogeneity are a result of inherent differences in the calculation algorithms. CS is known to do a poor job of calculating dose in heterogeneous patient anatomies.[19] This is because the Monte Carlo kernels utilized by CS are calculated in a homogeneous water phantom. When heterogeneous (non-water) voxels are present, CS accounts for patient heterogeneity using range scaling. In range scaling, the average mass density of voxels between the interaction and dose-deposition voxels is calculated and utilized when extracting values from the kernels. For single-scattered photons, dose deposition can only occur exactly on the line between the interaction and dose deposition voxels. Therefore, the range scaling approximation correction factor is perfectly accurate. However, for multiple-

scattered photons, they will likely have interacted in voxels not directly between the interaction and deposition voxels. As a result, the range scaling approximation factor will not be perfectly correct. It is important to note that for multiple-scatter photons, the mean density of the entire patient/phantom is utilized instead of the mean density between the interaction and deposition voxels. This is done for two reasons. First, photons do not necessarily interact only in voxels between the interaction and deposition voxels. Second, this approximation reduces the computational cost of the dose calculation. In comparison, MC physically models the interactions of particles (photons, electrons, positrons) in matter, thus allowing it to accurately model the interaction of these particles in heterogeneous anatomies.

For CS and collapsed-cone CS algorithms, dose is calculated per control point, where a control point is a planning parameter that defines the linac state (collimation positions, delivered MU, gantry angle) at some point in the planned treatment delivery. For example, a VMAT delivery might utilize a 4° control point spacing. This means that the collimator positions and MU output are defined every 4°. Dose is calculated at each of these discrete gantry angles. However, radiation is being outputted continuously along all gantry angles in the VMAT arc, not at discrete angles. In comparison, MC simulates particle histories continuously at all possible gantry angles. Thus, whereas the collapsed-cone algorithm investigated herein is susceptible to dose differences due to gantry discretization, MC is not.

1.4 Treatment Planning QA

There exist many sources of dosimetric error and uncertainty. It is the goal of radiotherapy QA to ensure that dosimetric accuracy falls within some maximum permissible level. If the difference between measured dose and expectation varies by more than this maximum permissible level, or action tolerance, then some action should be taken to rectify the

problem. The most commonly cited conclusion is that end-to-end dosimetric accuracy should be within ±5%.[20-22] However, ultimately what matters is not the delivered dose, but adequate control of the tumor as well as acceptable values of normal tissue complication probability (NTCP). Tumor control probability (TCP) is a formalism that defines the probability that the tumor will be controlled given the prescription dose. The TCP-formalism utilized in this study is the effective uniform dose (EUD)-based formalism developed by Niemierko et al.[23] It can be calculated via Eq. 5 and Eq. 6 where EUD is the absorbed dose to the patient that when homogeneously given to a tumor yields the same number of surviving clonogenic cells as the actual heterogeneous dose distribution. In Eq. 5, v_i is the partial fraction of the target volume that receives dose D_i. The parameter 'a' is a tissue-specific parameter that models how the tumor or normal tissue of interest responds to inhomogeneous irradiation. TCP is subsequently calculated from EUD where TCD_{50} is the homogeneous dose to the tumor that will result in a 50% control rate. $\gamma 50$ is a tissue-specific model parameter that quantifies the slope of the dose-response curve. The dose response slope determines the percent variation in tumor control probability (TCP) per percent difference in target dose. NTCP quantifies that probability that a dose of radiation to an organ at risk (OAR) will experience complications as a result of the prescribed dose. NTCP can be calculated per Eq. 7 and Eq. 8. where $TD_{50}(V)$ is the tolerance dose that results in a 50% complication probability due to the partial, uniform irradiation of volume V. The slope of the dose response curve is labeled m. Literature recommends that end-to-end dose accuracy should be within 5%[20-22], however, realistically, required dosimetric accuracy will vary from case to case. This is because the required accuracy will vary with the steepness/slope of the dose-response curve, γ_{50}, as well as what is clinically achievable. For instance, if γ_{50} is equal to 5.0, then TCP will drop by 5% per 1% decrease in target dose. If γ_{50}

is equal to 1.0, then TCP will only drop by 1% for a 1% decrease in target dose. It follows that a site with a low γ_{50} value will not require as stringent of dose accuracy, whereas a site with a high γ_{50} value will require more stringent dose accuracy – perhaps 2 to 3%. However, this level of accuracy is difficult to achieve. In the remainder of section 1.4, the limit of what is clinically achievable is addressed within the scope of the following dosimetric uncertainties: machine calibration, patient positioning accuracy, ROI segmentation and target deformation, tumor motion, electron density calibration, and dose calculation. All-together, the processes by which these dosimetric uncertainties are quantified and minimized is called treatment planning QA. Treatment planning QA is broadly described in TG-53.[24]

$$\boldsymbol{EUD} = \left(\sum_i (\boldsymbol{v_i D_i^a})\right)^{\mathbf{1/a}} \tag{5}$$

$$\boldsymbol{TCP} = \frac{\mathbf{1}}{\mathbf{1}+\left(\frac{\boldsymbol{TCD_{50}}}{\boldsymbol{EUD}}\right)^{\mathbf{4\gamma_{50}}}} \tag{6}$$

$$\mathbf{NTCP}\ \frac{\mathbf{1}}{\sqrt{\mathbf{2\pi}}}\int_{-\infty}^{t} \boldsymbol{e^{-t^2/2}\,dt} \tag{7}$$

$$\mathbf{t} = \frac{\boldsymbol{D - TD_{50}(V)}}{\boldsymbol{m\,TD_{50}(V)}} \tag{8}$$

The role of machine QA is to reduce systematic dosimetric uncertainties pertaining to the linac to within an acceptable limit. Without properly managing daily, monthly, and yearly machine QA, significant dosimetric uncertainties can find their way into clinical treatment. Task Group reports 40 and 142 lay out recommendations for machine QA of linear accelerators.[25,26] For IMRT cases, output constancy of the machine must be within 3% daily and 2% monthly. Lasers must be aligned within 1.5mm daily and 1.0mm monthly. In the monthly QA check, the radiation-light field coincidence, optical distance indicator (ODI), jaw positioning, and treatment couch positioning should all be within 1mm. The gantry read out

must also be accurate within 1 degree of actuality. Electron and photon beam profile constancy should be within 1% and electron beam energy constancy should be within 2%.

Correct patient positioning and immobilization ought to be assessed daily via image guidance (e.g. cone beam CT (CBCT)) to reduce any potential dosimetric uncertainties. Positioning uncertainty varies between treatment sites. For stereotactic brain radiotherapy utilizing a head ring, expected uncertainty is 1mm.[27] For breast and prostate radiotherapies utilizing a Vac-Lok, positioning uncertainty is expected to be 6mm and 2mm respectively.[28,29] Rotondo et al. shows that H&N positioning errors are typically a few millimeters.[30] Kaur et al. goes on to show that PTV dose volume constraints ($D_{2/50/95/98}$) vary by about 1% for H&N cases for setup errors of 1-2mm when using a 5mm PTV margin.[31] Rudat et al. showed that TCP can drop by 5% and 11% respectively for esophageal and prostate cancers as a result of clinically relevant setup errors.[32-33]

The management of target motion has become increasingly important with the introduction of image-guided radiotherapy. Intrafraction motions include respiratory, cardiac, and gastro-intestinal. Tumor motion causes the CT to be distorted; the tumor and OARs are blurred over the full range of their motions. Typically, the CTV is expanded by some margin to account for the full range of motion of the target, however, inadequate margins can result in under-dosing of the CTV. For lung tumor treatment, without the use of a more advanced respiratory management technique, Machalakos et al. showed GTV D_{95}, V_{95}, and TCP all changed by about $1.3 \pm 3.8\%$ due to intra-fractional breathing motion; a patient-specific 1-2cm CTV-PTV margin was utilized.[35] GTV dose difference for these metrics was as large as 10% for 4% of the treatment cases. Given the significance of this dosimetric error, TG-91 addresses the management of respiratory motions using more advanced techniques.[35] There exist multiple

more advanced methods for reducing the dosimetric uncertainties due to respiratory motion. The most common of these methods are respiratory gating, breath-hold, and abdominal compression. In respiratory gating, the tumor motion is tracked during the actual treatment. Radiation is delivered when the tumor position is within a certain range. Using a phase-based respiratory-gated 4D-CT technique, Lin et al. showed that dose differences to the PTV, bilateral lungs, and OARs were minimal.[36] For deep inspiration breath-hold (DIBH), the patient is coached to inhale deeply and hold their breath so as to produce a reproducibly deep inhale during pre-treatment CT and treatment. Using DIBH, Rosenzweig et al. showed PTV dose can be increased from 69Gy to 88Gy with no increase in NTCP values.[37] By compression the patient's abdomen, the patient's breathing is forcefully minimized in order to reduce tumor motion. This allows for reduced CTV-PTV margins and reduced NTCP rates. Sarkar et al. shows lung NTCP rates can be reduced by 2% for both conformal and VMAT lung cancer treatment plans.[38] In general, the dosimetric effects of respiratory motion is small when correctly accounted for using these techniques.

Accurate contouring of the GTV, CTV, and PTV volumes is necessary in order to adequately account for microscopic cancer spread, patient motion, and setup uncertainty. Accurate OAR contouring is necessary to accurately predict normal tissue complication probabilities (NTCP). However, our ability to accurately delineate these structures suffers from inadequate training of dosimetrists, human error, and a lack of consensus in regard to how the planning volumes ought to be defined.[39] In a study by Fiorino et al., inter-observer variability in rectum contouring resulted in dose differences of 3-4% (standard deviation).[40] In a SRS study by Stanley et al., the median absolute difference in conformity index (CI) between reference contours of the metastases and alternative contours drawn by 8 other physicians was 0.35.[41]

Although DVH-based impacts were not assessed as part of this study, a mean absolute CI difference of 0.35 is large. CI values can range from 0 to 1 (ideal coverage), so a difference of 0.35 is very significant.

The patient's anatomy is subject to change over the course of radiation treatment relative to the planning CT. However, dose is only calculated on the pre-treatment planning CT. Ideally, a new CT would be taken every day, a new daily-plan generated, and a new daily-dose calculated. However, given current constraints of the clinical workflow this is not feasible. Instead, this new CT/plan/dose is only acquired when the change in the patient's anatomy is large enough to warrant pushing the patient's treatment back a day. The dosimetric impact of anatomical variation can be significant. For instance, van der Horst et al. showed that CTV D_{98} coverage can drop by 10% due to inter-fractional tumor position variations.[42] In the case of IMRT treatment of H&N cancer, Cheng et al. showed that reduction in the size of the target volume over the course of treatment often resulted in increased dose the spinal cord, brain stem, and parotid glands.[43] For prostate cases treated using VMAT, Chow et al. showed that reduction in the size of the patient over the course of treatment yielded approximately a 2-3% increase in prostate and normal tissue dose per centimeter in reduction of the patient's size (reduction measured in terms of depth of the tumor).[44] In order to mitigate these dose discrepancies, several techniques have been developed. These include cone-beam CT (CBCT), deformable image registration (DIR), and adaptive radiotherapy (ART). These processes are not covered as they fall outside of the scope of this thesis.

CT images are 3D arrays of values which quantify the attenuability of radiation in matter relative to water and air. This attenuability is quantified in Hounsfield Units (HU), also called CT number. However, in dose calculation algorithms, calculation is based on the electron

density of each voxel, not the HU. This is because the stopping power, or absorption rate, of electrons is dependent on the electron density of matter. Thus, it is necessary to convert HU into electron density (ED). This conversion is carried out using a CT-ED calibration curve. Any inaccuracies in this curve can potentially introduce dosimetric error into the treatment plan. Das et al. displays that CT-ED errors are more relevant for IMRT cases and for treatment sites with pronounced high-density bone present.[45] For IMRT cases, dosimetric differences due to inaccurate CT parameters could be as high as 5% in the presence of bony anatomies.

Dose calculation accuracy is one of the greatest factors in the overall end-to-end dosimetric accuracy of the radiotherapy workflow. Dose calculation accuracy depends partially on the inherent accuracies of the dose calculation algorithm, but also upon the quality of the measurement data that was used in the beam modeling as well as the accurate determination of beam model parameters in the treatment planning system. Dose calculation accuracy in heterogeneous media needs to be thoroughly established. The ability to accurately calculate dose for complex treatment fields also needs to be thoroughly vetted. In section 1.3.3 the effects of heterogeneity were shown to cause local calculation differences on the order of 10%. In the same section, collapsed-cone IMRT and VMAT fields were shown to consistently over-estimate dose by approximately 3% relative to Monte Carlo dose calculation.[17]

The linear accelerator needs to be able to accurately deliver the treatment plan precisely. This includes accurate reproduction of field sizes, gantry angle, machine output, MLC positioning, etc. A study by Moran et al. tabulates the accuracies of various delivery components as well as their corresponding dosimetric impacts.[46] Collimator output factors are typically accurate within 0.5-1.0%.[47] MLC positioning accuracy is typically less than 1mm, however the effect this error has on dose is highly dependent on the gap between adjacent

leaves.[48] For example, Nithiyanantham et al. shows dosimetric errors can be as high as 6% per millimeter of MLC error for plans with steeper dose gradients. The gross effect of leaf transmission errors can result in dose differences as high as 3% for poorer treatment models.[49] The table top must be accurately accounted for in the treatment planning system. Not accounting for the table top, or accounting for it incorrectly can result in local dose differences as high as 10-15% at certain gantry angles.[50]

Although it is fairly easy to isolate each of these dosimetric uncertainties and quantify them separately, it is more difficult to evaluate their combined effects. Furthermore, the effects of these dosimetric uncertainties will vary case to case. So then how do we know if the dosimetric accuracy of each patient's plan is within the ±5% limit that is required? The best way to verify your treatment planning QA is to plan, treat, and measure dose as delivered to an anthropomorphic (heterogeneous anatomically-realistic) phantom. The radiological physics center (RPC) at M.D. Anderson Cancer Center provides these anthropomorphic phantoms to institutions for the credentialing of linear accelerators. After your institution has delivered a radiation treatment plan to the phantom, RPC evaluates the dose that was delivered to implanted TLD detectors to determine that your institution was able to properly deliver the specified radiation dose to the target while sparing surrounding tissues. Molineu et al. analyzes the ability of 763 institutions to accurately deliver IMRT dose to an anthropomorphic head and neck (H&N) phantom.[99] This data, acquired from 2001-2011, provides an idea of what is a clinically achievable measure of dose accuracy. 31% of the institutions failed to deliver dose to the target within ±5%. A more recent study by et al. looks at data for the same H&N phantom but between November 2014 – October 2015.[100] Results show that 77% of institutions were able to pass a 5%/4mm gamma criteria. 50% of failures were due to systematically low delivered dose.

Although verifying your institution's end-to-end dosimetric accuracy using an anthropomorphic phantom is ideal, such a phantom may not always be available.

Short of using an anthropomorphic phantom with implanted dosimeters, TG-119 is an apt commissioning test for verifying accurate dose. TG-119 puts forth a variety of clinically relevant test cases which institutions can use to assess the accuracy of IMRT delivered dose.[49] It is the task of each institution's physicists to create a treatment plan for each test case, irradiate a solid water stack in accordance with said treatment plan, measure the resultant dose at two pre-specified high dose and low dose points using an ion chamber, and calculate the percent difference in calculated and measured doses. This commissioning test verifies that your treatment planning QA was effective. Specifically, TG-119 verifies that your machine QA, machine delivery, and dose calculation are correct/accurate. Because this test lacks an anthropomorphic phantom, these tests are incapable of verifying correct patient positioning, accurate accounting for patient heterogeneity, and an accurate CT-ED table. This test also does not assess the correct handling of inter- and intra-fraction target motion. TG-119 contains results for 10 institutions. Dosimetric accuracy is reported as a confidence limit (the mean calculation error (relative to measurement) plus 1.96 times the standard deviation).[51] On average, over all treatment sites and all institutions, confidence limits were 4.5% in high dose regions and 4.7% in low dose regions. This data suggests that dosimetric differences greater than 5% occur about 5% of the time. In comparison, RPC showed that dose difference was greater than 5% about 20% of the time. This difference is attributed to the fact that TG-119 utilizes a homogeneous solid water phantom whereas the RPC results utilize an anthropomorphic phantom. These results also agree with previous studies that a 5% dose accuracy is achievable and should be the minimum accuracy an institution aims to meet. In

conclusion, although TG-119 is not a true end-to-end test set, it is an easy way to evaluate the combined dosimetric effects of several important treatment planning uncertainties.

1.5 Patient-Specific Quality Assurance

1.5.1 Phantom-based QA

Linear accelerators are complicated devices comprised of many moving parts. It follows that there will be some non-zero delivery error when delivering a treatment plan. The purpose of patient-specific QA is to verify that radiation is being correctly delivered for each specific patient's treatment plan. Patient-specific QA is conventionally carried out using phantom-based systems. The phantom-based systems utilized herein are Sun Nuclear's MapCHECK and ArcCHECK systems, shown in Fig. 5 and Fig. 6. MapCHECK is a planar device that contains a planar array of diodes. A diode is a small ($0.019 mm^3$) device for measuring dose. The ArcCHECK phantom is a cylinder made of nearly tissue-equivalent material. Inside of the cylinder are 1386 diodes arranged helically throughout the cylinder. Whereas MapCHECK2 is used to QA radiation fields without gantry rotation, the ArcCHECK is utilized for fields where gantry rotation is present. These QAs are carried out as follows. First, dose is calculated on the phantom geometry. Next, the treatment plan is physically delivered onto the phantom. Dose is measured using the diodes and compared to that of calculation to make sure the treatment plan can be accurately delivered by the linac. MapCHECK2 and ArcCHECK QAs both utilize gamma analysis to determine whether the QA is deemed passing or failing. The equations underlying gamma analysis, shown in Eq.9 and Eq. 10, were developed by Low et al.[52] Gamma analysis is based on the concepts of dose difference (DD) and distance-to-agreement (DTA), where DD is simply the percent difference in dose between calculation and measurement at the spot of the diode of interest and DTA is the minimum distance it takes from the diode of interest

to reach a voxel in the calculated dose distribution with the case dose as the diode measurement. Let $r(r_m,r_c)$ and $\gamma(r_m,r_c)$ be the DTA measurement and DD measurement percent between the diode of interest r_m and some voxel in the calculated dose distribution r_c. Similarly, let Δd^2_m and ΔD^2_M be the DTA and DD criteria that we are looking to meet. Common gamma criteria are 2%/2mm – 3%/3mm. A 3%/2mm gamma criteria is utilized by our institution. Eq. 9 solves voxel-wise for Γ, a composite quantity evaluating the combined effects of DD and DTA in voxels proximal to the diode of interest. The per-diode gamma index, γ, is simply the minimum value of Γ surrounding the diode. If γ is less than 1, the measurement result of the diode is considered passing. If γ is greater than 1, the result for that specific diode is considered failing. The combined gamma passing rate tells us what percentage of the diodes are passing or failing. 95-100% is considered passing, 90-95% is passing but warrants investigation, and less than 90% is considered failing. In the case of failure, the QA should be re-done. If the plan still fails QA, the plan should be re-planned.

$$\boldsymbol{\Gamma(\mathrm{rm},\mathrm{rc})} = \sqrt{\frac{r^2(\mathrm{rm},\mathrm{rc})}{\Delta d_m^2} + \frac{\delta^2(\mathrm{rm},\mathrm{rc})}{\Delta D_M^2}} \tag{9}$$

$$\boldsymbol{\gamma(r_m)} = \mathbf{min}\{\boldsymbol{\Gamma(r_m, r_c)}\}\forall\,(\boldsymbol{r_c}) \tag{10}$$

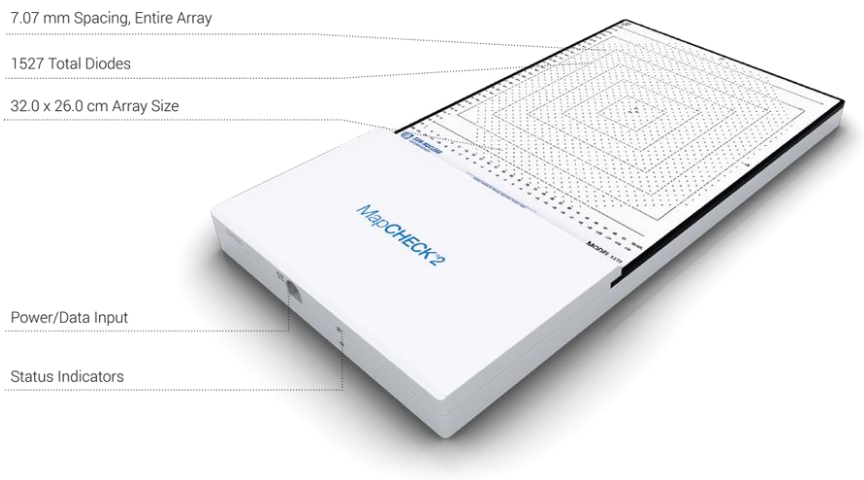

Figure 5. SunNuclear's MapCHECK2 planar diode array.

Figure 6. Sun Nuclear's cylindrical ArcCHECK phantom

MapCHECK2, ArcCHECK, and other phantom-based QA methods have been the standard of care for years. However, these measurement-based systems have come under fire in recent years. Several studies have shown diodes to be inaccurate measurement devices. Saini et al. shows that diodes typically under-respond on the order of 1-4% for lower instantaneous dose rates (~25 cGy/min) as opposed to typical clinical dose rate (> 200 cGy/min).[53] Letourneau et al. shows MapCHECK diodes under-respond by up to 2% at 50 MU/min compared to 600 MU/min.[54] Diode energy dependence has been explored by several

studies, albeit to a lesser extent.[55-57] Yin et al. shows that unshielded Scanditronix diodes at 15cm depth can over-respond by as much as 15% to lower energy patient-scattered radiation. However, for out-of-field diodes where the radiation spectrum can be significantly softer, one should expect an even greater relative over-response by the diodes. For example, Rickner et al. theorizes that diodes will over-respond to 100keV and 500keV photons by a relative 70% and 30% respectively.[58] Comparing ArcCHECK diode measurement in a solid water slab to an ion chamber, the ArcCHECK diode was shown to over-respond to radiation by up to 1% 5cm out of field.[101] Czarnecki et al. showed field size calibration factors to vary by up to 2.0% between 1x1 cm^2 and 10x10 cm^2 field sizes.[59] However, after applying field size correction factors, Chaswal et al. shows diode dose error as a result of field size dependence to be within 0.4% for 5x5 cm^2 to 20x20 cm^2 fields.[60] The accuracy of the ArcCHECK diodes is assessed as part of this study.

Debate-style manuscripts by Siochi et al. and Smith et al. state that measurement-based IMRT is both time-consuming and potentially inaccurate since the measurements are taken in a phantom and not the patient's anatomy.[61,62] Furthermore, the reliability of these systems is worrisome; 20% of institutions fail RPC's 7%/4mm gamma criteria despite presenting with acceptable gamma pass rates.[63] Multiple studies have shown IMRT QA passing rates to be weakly correlated with DVH-based metrics.[64-66] Given this concern, there has been a general movement towards DVH-based QA metrics. However, when it comes to DVH-based metrics, phantom-based systems tend to fall short. SunNuclear does offer an ArcCHECK-specific software package, 3DVH, for reconstructing 3D dose on the patient geometry from measurement, however, Tyagi et al. shows that log files are better able to reproduce delivered dose on the patient anatomy than this 3DVH software package.[67] As part of this study, log file

QA on the patient anatomy is compared to AC QA on the ArcCHECK geometry. The clinical relevancy of AC QA passing rates is assessed. Furthermore, the effectiveness of phantom-based QA as an anatomical surrogate is assessed.

1.5.2 Log File QA

Log files are records of treatment deliveries stored after each delivery. These files contain a recording of collimator (jaw and MLC), gantry positions, delivered dose rate, collimator angle, and MU output as a function of time. The Elekta log file utilized records log files samples at 25Hz (every 40ms). In addition, the log file contains error values for the collimator positions, gantry position, and collimator angle. Using these log files, a treatment plan is reconstructed based on machine parameters from the actual delivery and dose calculated. The log file reconstructed dose is then compared to the clinically planned dose. Log file dose can be calculated on either the phantom geometry or patient anatomy. In the scenario that log file dose is calculated on the phantom geometry, log file dose is compared to the clinically planned dose using gamma analysis. However, when calculated on the patient anatomy, log file dose can be compared to the clinically planned dose using the same DVH-based metrics that were used to optimize the initial treatment plan

Log file QA offers multiple significant advantages over conventional phantom-based patient-specific QA systems. Whereas phantom-based systems are time consuming, LF QA is a fully automatable software system. Utilizing phantom-based systems, per-fraction dose cannot be tracked. For example, ArcCHECK is setup before treatment to verify correct plan delivery but is not utilized during actual patient treatment. In comparison, LF QA can be carried out during every actual treatment fraction. By tracking dose to the patient per-fraction, a more accurate cumulative dose distribution to the patient can be determined. In the case of significant

dosimetric discrepancy relative to the clinically planned dose, the treatment plan can be adapted. In addition, phantom-based systems are largely incapable of root-cause analysis. Phantom-based QA is either passing of failing. For example, if ArcCHECK QA fails, it does not say why. In comparison, LF QA allows for a robust root cause analysis – e.g. was the source of error due to MLC error, gantry error, unstable MU output, or some other issue. Whereas phantom-based systems ignore patient-specific anatomical variation, LF QA is capable of tracking heterogeneous anatomic dose. In addition, LF QA provides the potential for online treatment tracking and/or real time treatment plan modification. By tracking the log file in real-time, dosimetric discrepancy could also be tracked in real time. If the discrepancy were significant enough, the treatment could be halted. Alternatively, the treatment plan could be adapted in real time to instantaneously account for delivery error. It is noted that the current log file system which has been provided to us by Elekta can only be used retrospectively; the log file is not acquired in real time.

Skepticism with regard to the reliability of measurement-based phantom-based QA has led researchers to investigate software-based log file QA as an alternative patient-specific QA method.[68-74] Using Varian's dynalog system, Stell et al. showed MU errors were greater for high dose rate deliveries.[72] Handsfield et al. showed log-file based QA utilizing a secondary Monte Carlo dose calculation method is an effective and efficient means of TomoTherapy QA.[70] Kumar et al. shows that log files can carry out IMRT QA with the same accuracy as film.[75] Log files are also a viable method of VMAT QA for both phantom and patient geometries.[67,76] Defoor et al. shows that log file gamma passing rates as calculated on the patient anatomy matched clinical gamma pass rates as calculated on the Delta^4 phantom within 1%.[76] Tyagi et

al. shows that log file dose as calculated on the patient anatomy consistently agreed with the planned dose within 2%.[67]

The primary concern with the use of log files is that they rely upon a reliable relationship between machine read-out and the actual delivery. For example, in a Varian machine, MLC leaf motor counts are converted into leaf position and output to the log file. As a result, a malfunctioning MLC motor, faulty t-nut (connects the motor to the MLC), or loss of encoder counts may result in incorrect log file reported MLC positions. Agnew et al. illustrates this exact problem - a loose t-nut resulted in incorrect log file MLC leaf positions.[77] However, for the Elekta machines at our institution, MLC positional information is determined based upon the optically measured position of reflectors on each leaf, thus eliminating the possibility of these catastrophic errors assuming that the optical reflectors have been calibrated correctly. At our institution, MLC calibration is carried out via monthly picket fence/phantom tests. These tests nominally ensure collimation components are within 1mm of post-calibration baseline positions, however they can be somewhat subjective and don't always directly isolate individual leaf position errors. In addition, calibration can degrade over time. For these reasons, it may be useful to perform Integrity's Calibration Workflows more often or to implement a more objective QA testing. In addition, log files do not monitor variations in output and beam tuning. Thus, these factors must be sufficiently monitor via machine QA.

A second concern is that LF QA requires an accurate dose calculation method. Dose calculation errors will go unnoticed by LF QA. It follows that having a robust dose calculation method (algorithm + beam model) is of the utmost importance. In general, dose calculation methods have been thoroughly vetted. For instance, using collapsed-cone CS, Butson et al. showed dose calculation can be accurate within 5% of delivered dose for an anthropomorphic

lung phantom.[78] However, this 5% includes more uncertainties than just dose calculation accuracy. It follows that the accuracy of the CS calculation method is likely in the 2-4% range. In a study by Haga et al., MC and CS calculated doses typically agreed within ±2-5% with ion chamber measurement in both solid water and cork phantoms.[79] Calculations consistently agreed within 1.5% of each other in target regions. These studies show that a dose calculation accuracy of 2-3% is achievable. However, it is the responsibility of each institution to validate the accuracy of their beam model(s). For instance, when it comes to LF QA, it is essential that the calculated leaf position precisely matches the actual leaf position to account for subtle discrepancies in leaf position as recorded in the log file. For CS, this is adjusted via the rounded leaf tip offset table in Pinnacle's beam modelling system.[80] For MC, SciMoCa utilizes a parameter that offsets the leaves from their nominal position as determined using a picket fence test. All MLC positioning systems (log file, Pinnacle, SciMoCa) are nominally accurate to 0.1mm. Given the importance of dose calculation accuracy, two dose calculation algorithms as well as two forms of LF QA were assessed as part of this study. The two dose calculation algorithms were Pinnacle's adaptive collapsed-cone convolution-superposition (CS) algorithm and a research version of ScientificRT's SciMoCa Monte Carlo (MC) algorithm.

The two forms of LF QA analyzed as part of this study were CS-based LF QA (LF-CS QA) and MC-based LF QA (LF-MC QA). In LF-CS QA, the log file reconstructed and planned doses both utilize the same CS dose calculation algorithm. Because the doses utilized the same dose calculation method, patient-specific dose calculation accuracy is unknown. However, for LF-MC QA, the log file reconstructed dose instead uses Monte Carlo for dose calculation. Thus, LF-MC QA not only assesses the ability of the machine to accurately deliver the treatment plan, but also acts as a secondary dose verification system. All three terms, LF QA, LF-CS QA, and

LF-MC QA will be used during this study. Where LF QA will be used when referring to log file quality assurance as a whole, LF-CS QA and LF-MC QA are more specific terms signifying that log file reconstructed dose is being calculated using either CS or MC respectively.

LF-MC QA also functions as a means of treatment planning QA. As part of treatment planning QA, the dosimetric impact of various beam modelling parameters and machine delivery parameters are analyzed. These include jaw positioning accuracy, MLC positioning accuracy, and linear MU output. The log file monitors each of these factors. Furthermore, LF QA can be used to calculate the dosimetric impact of these factors per-patient on their unique anatomy. It may be the case that current machine QA tolerances for these factors are too lax and need tightened, or perhaps the tolerances we have in place are not that necessary. Log files may also present with MLC/gantry errors that are greater than the tolerances set by machine QA. More specifically, machine QA requires MLC positioning to be within ±1% and gantry positioning to be within 1°, however, the log file may contain errors greater than these values. In this circumstance, the dosimetric impact of these errors ought to be investigated. If deemed significant the cause of this error can then be investigated and potentially corrected for. As an example, perhaps MLC errors greater than what are allowed are occurring when MLC velocities are greater than some value. Or perhaps gantry errors greater than the 1° machine QA tolerance are occurring when gantry accelerations exceed some value. MLC velocities and gantry acceleration values could then be constrained inside of the treatment planning system so that MLC and gantry errors are within tolerance. It is noted that the effect of MLC and gantry rotation on MLC and gantry error are not included in routine machine QA. By utilizing LF-MC QA, the effects of beam complexity and patient heterogeneity can also be investigated. Previously it was shown that MC and CS can disagree by 3% for IMRT treatments, despite

agreeing well for simpler conformal plans.[17] This suggests these calculation differences likely result from imperfect CS and/or MC beam modeling and/or treatment planning parameters.

1.6. Chapter Overview

The primary purpose of this study was to compare LF QA to ArcCHECK QA (AC QA). In this chapter, many of the primary advantages and disadvantages of the two systems were investigated. Chapter 2 is the first of two core chapters in this dissertation. Herein, diode sensitivity dependencies of AC QA diodes were quantified. In addition, LF QA was compared to AC QA when all doses were calculated on the ArcCHECK phantom geometry. In Chapter 3, LF-CS QA and LF-MC QA was carried out on the patient's anatomy, allowing for log file reconstructed dose to be directly compared to the clinically planned anatomical dose. Because LF-MC QA is not well established, a protocol for assessing log file reconstructed anatomical dose differences was developed. Using this protocol, anatomically-calculated LF-MC QA results were compared to clinical phantom-based AC QA results. Chapter 4, provides a brief overview of the LF QA software that was developed as part of this project. This software includes automation tools for both LF-CS and LF-MC QA. Finally, in Chapter 5, conclusions and the key clinical takeaways of this study are discussed.

CHAPTER 2 LF QA FOR TREATMENT PLANNING AND DELIVERY QA: PART 1 – ARCCHECK GEOMETRY

2.1 Introduction

This is largely a comparative study, seeking to determine whether log file QA can provide ArcCHECK equivalent accuracy. First, accuracy of the log file is investigated and the Monte Carlo beam model is validated. Next, the log file dose reconstruction process along with strategies for reducing the computational cost of said process was detailed. Third, the dosimetric uncertainty of ArcCHECK QA was re-investigated. Variation in diode sensitivity as a function of dose rate, energy, and field size were investigated. The effect of these parameters on measurement and subsequently ArcCHECK QA results was evaluated. Calculated vs. AC-measured dose differences were quantified as functions of diode dose rate and energy dependence. LF and AC QAs were evaluated for experimental plans of varying delivery complexity. Lastly, clinical LF QA and AC QA results were compared. In this Chapter, log file doses are all reconstructed on the ArcCHECK geometry.

2.2. Methods and Materials

For this study, an Elekta Infinity linac equipped with the Agility beam-limiting device was utilized. Plans were selected from previously treated patients that were planned and optimized using the Pinnacle3 treatment planning system. Eleven dual-arc VMAT patients were selected for this study (9 H&N, and 2 low dose rate brain). For the entirety of this thesis, all fields were 6MV photon fields. The head and neck cases were selected for their high degree of complexity and somewhat lower gamma pass rates. The two brain cases were selected to compare how log file and ArcCHECK QAs respond to low dose rate deliveries. All 22 arcs

were delivered to Sun Nuclear's ArcCHECK phantom and log files recorded. All beams were flattened, 6 MV.

To compare AC QA to LF QA, five dose distributions were determined for each plan: Plan-CS, Plan-MC, LF-CS, LF-MC, and AC. 'Plan' denotes the dose distribution that was calculated from the original clinical treatment plan file optimized in Pinnacle, whereas 'LF' doses were calculated from log files. CS and MC denote the two dose calculation methods. 'AC' denotes ArcCHECK measurement. To begin, Plan-CS is recalculated using Monte Carlo to get Plan-MC. Next, each plan was delivered on the ArcCHECK phantom, measurement taken, and log files recorded. For LF-CS dose calculation, in-house code was used to reconstruct the log file beams in Pinnacle with a reduced number of log file samples (see Section 2.2). For MC dose calculation, the DICOM toolkit (DCMTK) library by OFFIS was utilized to convert log files into DICOM RTPLANs containing every log file sample with non-zero MU. DCMTK is a collection of open-source C/C++/ANSI libraries and applications implementing the DICOM standard. Whereas LF-MC is calculated using every useful log file sample, LF-CS is calculated using a reduced number of log file samples.

2.2.1 Preliminary Validations

First, the accuracy of our MLC calibration was assessed. The accuracy of the log file records is based on this calibration and thus is of the utmost importance. Elekta's Log File Convertor for Integrity R3.2 records linac delivery parameters (dose rate, gantry/collimator angle, leaf/collimator positions, MU) every 40ms. MU readings are determined via an ion chamber in the treatment head of the linac. Jaw positions are determined electronically via two redundant potentiometer voltages. MLC positions are determined from the video image positions of optical reflectors on each MLC leaf. The accuracy of jaw and leaf positioning, as

represented in the log file, hinges on machine calibration and QA. Jaw and MLC calibration is carried out annually as well as after any relevant field service. Calibration is performed using the internal Calibration Workflows contained in the Integrity R3.2 treatment control system (TCS), specifically the optical, diaphragm, and leaf workflows. For monthly QA, picket fence patterns are acquired with the iViewGT EPID and a BB array phantom aligned to the crosshairs and compared to baseline images acquired just after Calibration Workflows to verify that leaves and jaws are within 1mm of their baseline positions. If any of the collimation components appear to be out of tolerance, the appropriate Calibration Workflow is re-run and the appropriate picket fence test repeated.

The ArcCHECK phantom was supported by the iBeam evo H&N extension. To account for this added attenuation, a couch ROI was added to the phantom geometry in Pinnacle beneath the ArcCHECK. In order to create this couch ROI the following process was carried out. A solid water stack was set on top of the H&N extension and a standard farmer chamber was inserted into the solid water stack at 10cm depth. Seventeen 10x10 cm^2 fields were delivered to the solid water stack at 17 different gantry angles. These angles were 100° to 260° with 10° spacing between them. Measurement was compared to calculation with this added couch ROI. The ROI was optimized to maximum agreement between CS dose calculation and measurement for the greatest number of these 17 fields.

As discussed in Chapter 1, the accuracy of CS and MC is generally well established, however, beam modeling accuracy is linac specific and must be thoroughly validated by each institution. Thus, both CS and MC beam models were first thoroughly validated by comparing calculated dose to measurement data. Measurement data for the 6MV Elekta Agility linac was acquired using IBA's Blue Phantom^2. Inline/crossline profiles, PDDs, and output factors were

acquired for a variety of field sizes ranging from 1x1 cm^2 to 40x40 cm^2. These field sizes are displayed in Table 3 in the results section. Output factors and PDDs and percent depth dose data for field sizes less than 5x5cm^2 were acquired using a Sun Nuclear EDGE diode. All other measurements were acquired using an IBA CC13 ion chamber (0.13 cm^3 volume, 5.8 mm length, 3.0 mm radius). Crossplane profiles were offset half a leaf width (2.5 mm) in the inplane direction to minimize the effect of interleaf leakage and to scan close to the center of the Agility leaf tips which appear tapered in EPID images. All profiles were measured at 1.5/5/10/20 cm depths. Ideally, large field profiles should be acquired using a microdiamond or microchamber. For the ≥20x20cm profiles taken with a CC13 chamber, substantial penumbral blurring and thus poorer agreement between measurement and calculation is expected. The resulting beam model was validated by comparing calculated doses to each of the aforementioned measurements. All calculations utilized 1mm dose grid resolution in the direction of measurement. In regard to commissioning and QA of treatment planning dose calculation, Smilowitz et al. recommends low-gradient in-field regions match within 1.5%, penumbra agree within 3mm DTA, and out-of-field measurements agree within 3% of max dose.[81] Therefore, the fraction of calculations that meet each of these criteria was determined. Out-of-field doses were also compared using more strict 2% and 1% criteria. Low-gradient was defined as less than 1cGy/mm. Out-of-field was defined as less than 10% of CAX dose. Additional per-field calculations included percent difference in output factor, mean percent difference in PDD from 0-35cm, mean percent difference in beam width (50%-50%) over all eight profiles, and mean percent difference in penumbral width (80%-20%) over all eight profiles. Percent PDD differences were calculated relative to maximum dose. Due to insufficient scan length for the 40x40 field, out-of-field dose and penumbral width was left uncalculated and unanalyzed.

2.2.2 Strategies for Reducing the Computational Cost of LF-CS QA

Log files are reconstructed by converting *n* number of log file samples into *n* VMAT control points, where a control point is a planning parameter that defines the linac state (collimation positions, delivered MU, gantry angle) at some point in the planned treatment delivery. However, Pinnacle places an upper limit of 359 control points on each field, thus limiting the number of log file samples that can be reconstructed to 359, despite thousands of samples per log file. In addition to this software limit, utilizing every one of thousands of log file samples would result in extremely long CS calculation times. This is because CS dose calculation time scales linearly with the number of control points for the CS method. There are at least a few ways to reduce this cost: (1) utilize CPU and/or GPU parallelization, (2) utilize a dose calculation algorithm whose computational cost is independent of the number of control points (e.g. Monte Carlo), and/or (3) reduce the number of log file samples reconstructed. For this study, methods 2 and 3 were investigated. CS dose calculation utilized Oracle's Sun Server X4-2. MC calculation utilized an Intel i7-3770 and 8GB of RAM.

The effect of reducing the number of log file samples reconstructed was investigated by utilizing 1/2/3/4º control point spacing. For example, a VMAT arc starting at -179º, ending at 179º, and reconstructed with a 3º control point spacing would have the following control points: -179.xº, -176º, -173º, … 175º, 178º, 179.xº. Variable x stands for the exact gantry angle at which radiation started and stopped delivering radiation according to the log file. In addition to control point spacing, the effect of dose grid resolution (2/3/4 mm) on log file reconstructed dose accuracy was analyzed. Reconstructive effectiveness was quantified by comparing log file gamma pass rates (LF-CS vs. AC) with pass rates of the original plan (Plan-CS vs. AC).

This metric measures dosimetric improvement due to log files accounting for machine delivery errors. Because LF and AC both account for delivery errors, this metric should be ≥ 0.

2.2.3 Discrepancy in Calculated vs. Measured Diode Doses

The ArcCHECK phantom contains 1386 diodes whose job is to measure the dose delivered to the phantom. However, this measurement can be inaccurate due to various diode sensitivity dependencies. Diode sensitivity dependence is evaluated as a function of dose rate, energy, field size, and field complexity. By quantifying these dependencies the accuracy ArcCHECK dose measurement can be better understood.

2.2.3.1 In-Field Discrepancy: Dose Rate Effect

Calculated (Plan-CS) vs. measured (MapCHECK2 and ArcCHECK) doses were compared for a variety of dose rates (35, 70, 140, 280, and 570 MU/min). 150MU was delivered onto the MapCHECK2 diode array for $3x3cm^2$ and $20x20cm^2$ field sizes and each dose rate. Constant machine output was verified by placing 2cm solid water on top of the MapCHECK and inserting an ion chamber at CAX. In-field and out-of-field dose rate effects were investigated; penumbral diodes directly on the field edge were excluded from the analysis. MapCHECK2 was initially utilized in order to best isolate dose rate effects; the simpler device geometry minimizes any possible spectral effects. It is important to note that the same diodes are used in both MapCHECK2 and ArcCHECK systems. Systematic effects of dose rate on AC QA and LF QA were assessed by delivering a 25x25arc onto the phantom using 70 and 570 MU/min.

Dose rate is dynamically determined based upon gantry, jaw, and leaf speed limitations. Thus, for deliveries with fixed jaws and leaves, delivered dose rate should be calculable as a function of MU, gantry excursion (gantry degrees subtended), and maximum gantry speed

(6°/s). To prove this, 23 10x10 arcs (see Table 1) with varying gantry excursions and MU were delivered on the ArcCHECK phantom. 'Calculated' dose rate matched log file data for all cases. Next, percent difference in calculated (Plan-CS) and measured diode dose was tracked for each of the 23 fields and analyzed as a function of 'calculated' dose rate.

Table 1. 10x10 arcs of varying gantry excursions and MU

Gantry Start/Stop/Excursion	MU
-90 / -60 / 30	10, 20, 35, 50, 100
-90 / -30 / 60	20, 40, 70, 100, 200
-90 / 30 / 120	20, 40, 80, 140, 200, 400
-90 / 90 / 180	20, 40, 60, 120, 210, 300, 600

2.2.3.2 Out-of-Field Discrepancy: Energy Effect

Percent difference in calculated versus measured diode dose was analyzed as a function of off-axis distance to determine the effects of off-axis softening on AC QA. Because off-axis spectra vary with field size, static 2x2 cm^2, 10x10 cm^2 and 25x25 cm^2 arcs were assessed. 0% and 10% dose thresholds were utilized.

2.2.3.3. Delivery Complexity

The effect of beam segment complexity on calculated vs. measured dose difference was investigated for 42 fields delivered onto the ArcCHECK phantom. 100MU static fields were delivered at gantry 0° and -90° for 1x7 cm^2, 7x1 cm^2, 2x8 cm^2, 8x2 cm^2, 3x5 cm^2, 5x3 cm^2, 5x1 0 cm^2, 10x5 cm^2, 2x2 cm^2, 3x3 cm^2, 5x5 cm^2, 10x10 cm^2, 15x15 cm^2, and 20x20 cm^2 field sizes. 300MU arcs from -90° to 90° were delivered for the same field sizes. Two metrics, MPDD (mean percent dose difference) and MAPDD (mean absolute percent dose difference), were utilized. These metrics work by comparing any two of the five dose distributions described in Section 2.2 at each of 1386 ArcCHECK diode positions. Equation 1 shows how to calculate MAPDD for LF-MC vs. AC where i is the i-th diode (i = 1-1386) such that AC_i is greater than

some threshold percentage of max(AC). MPDD is simply MAPDD without taking the absolute value.

$$\boldsymbol{MAPDD = Mean\left(\left|\frac{LFMC_i - AC_i}{AC_i} * 100\right|\right)} \quad (11)$$

LF QA and AC QA should agree with one another even for complex fields. To verify this is the case, we divided a 12x12 cm^2 static field and 12x12 cm^2 arc into 1/4/9/16/36 sub-fields (i.e. 1 12x12 cm^2, 4 6x6 cm^2, 9 4x4 cm^2, 16 3x3 cm^2, and 36 2x2 cm^2 fields/arcs). Each static field was setup at gantry 0. Arcs were a full 358° CW or CCW. Plans were generated in Pinnacle on a 'prescription' geometry and then re-calculated and delivered on the ArcCHECK geometry. The prescription geometry, shown in Fig. 7, consisted of three structures. A 20x20x20 solid water cube centered on the isocenter serves as the patient volume. A 10x10x10 'Target' ROI centered on isocenter was utilized for DVH matching purposes; all static fields and arcs had matching DVHs. Lastly, an isocenter-centered 10x10x1 (1cm perpendicular to gantry 0° CAX) 'Prescription' ROI was created. Sub-field beam weights were optimized to achieve a flat beam profile at 10cm depth for the full field; dose was optimized to the Prescription ROI using a 1Gy minimum DVH objective and 1.1Gy maximum DVH objective of equal priority. Static field beam weights were copied to their corresponding arc fields. Each set of sub-fields was delivered onto the ArcCHECK phantom. Sub-field measurements were combined such that ten combined-field measurements (5 static field and 5 arc) were created. Differences in AC, Plan-CS, Plan-MC, and LF-MC doses were assessed as a function of the size and number of segments delivered.

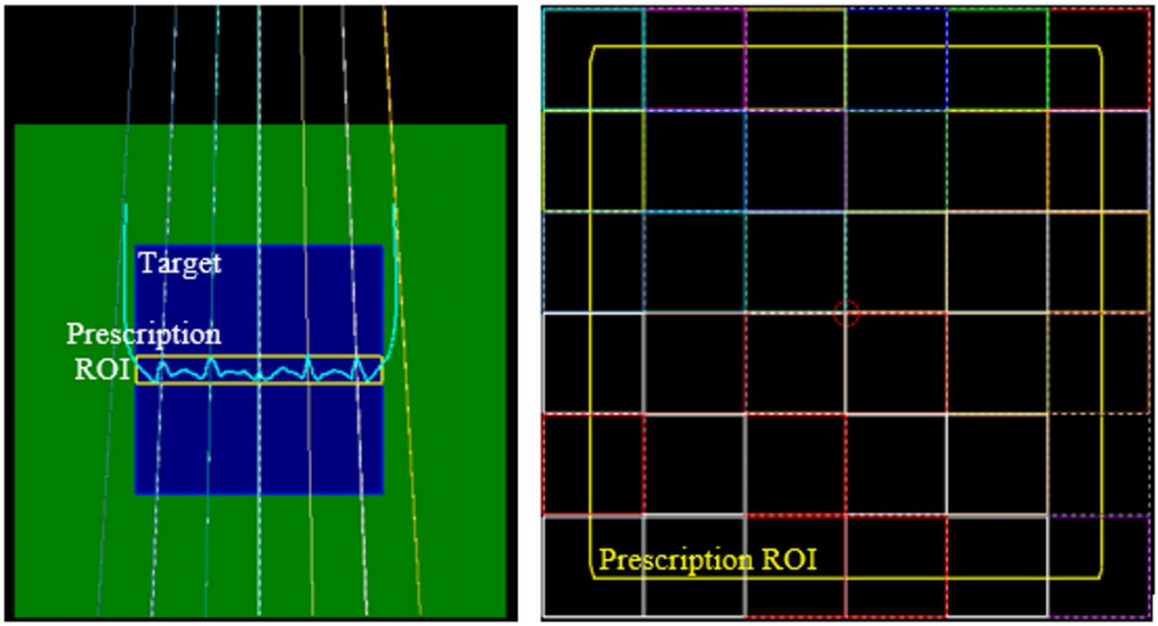

Figure 7. (left) The 'prescription' geometry. The teal isodose line displays the flatness of the combined 12x12 field when (right) 36 abutting 2x2 fields are delivered

2.2.4 Comparing LF QA & AC QA

Plan, LF, and AC dose distributions are subject to varying sources of dose uncertainty and share various correlated factors. AC QA compares 'Plan' and AC doses; discrepancies stem from phantom-dependent inaccuracies (e.g. setup error, diode miscalibration/drift, diode sensitivity dependencies, and physical diode limitations), dose calculation inaccuracy, delivery error, and variation in beam consistency (drift in beam output/symmetry). LF and AC doses both account for delivery error. It follows that LF and AC doses should agree better than Plan and AC doses. For LF QA, LF and Plan doses are compared. Conventionally this means comparing LF-CS to Plan-CS to isolate delivery error. However, dose calculation accuracy should also be verified. By comparing LF-MC to Plan-CS, both delivery error and dose calculation accuracy are investigated. It is important to note that whereas AC accounts for variations in beam consistency, neither LF nor Plan doses do. This means that LF QA ignores variation in beam output.

AC and LF QAs were carried out for 11 clinical dual-arc VMAT patients (9 H&N, 2 low dose rate brain). Dose was determined at each of ArcCHECK's 1386 diode locations for each arc and each of the five aforementioned dose distributions. Various pairs of these five plans were compared. Plan pair comparisons included (1) Plan-CS vs. AC, (2) Plan-MC vs. AC, (3) LF-CS vs. AC, (4) LF-MC vs. AC, (5) Plan-MC vs. Plan-CS, (6) LF-CS vs. Plan-CS, (7) LF-MC vs. Plan-MC, and (8) LF-MC vs. Plan-CS. Three comparisons metrics were utilized. First, 2%/2mm and 1%/1mm global gamma pass rates were calculated using 10% (conventional) and 85% (in-field/target approximation) dose thresholds. Second, global percent dose difference was calculated at each diode position and averaged over all diodes (1386 diodes/arcs * 22 arcs). 'Global' indicates that percent differences have been scaled relative to the maximum dose value being compared to, e.g. for X vs. Y where X and Y are two 1386 element dose matrices. Global percent diode dose difference (GPDDD) is calculated according to Eq. 12. This second metric was calculated in order to isolate and quantify systematic and statistical dose uncertainties due to delivery error, dose calculation difference, and the ArcCHECK phantom (e.g. diode sensitivity dependencies) as well as to confirm the gamma analysis results. The dosimetric comparisons studies as well as their significances are displayed in Table 2.

$$\boldsymbol{GPDDD} = \boldsymbol{mean}\left(\frac{\boldsymbol{X}-\boldsymbol{Y}}{\boldsymbol{\max(Y)}} * \mathbf{100}\right) \tag{12}$$

Table 2. Dosimetric comparisons along with their significance. When two comparisons are listed it is because those two comparisons are being compared.

#	Comparison 1	Comparison 2	Significance
1	Plan-CS vs. AC	Plan-MC vs. AC	Which dose calculation method matches AC better?
2	Plan-MC vs. Plan-CS	--	Dose difference due to calculation method is isolated.
3	LF-CS vs. AC	Plan-CS vs. AC	Log files should match AC better.
4	LF-MC vs. AC	Plan-MC vs. AC	Is this the case?
5	LF-MC vs. Plan-MC	--	Machine delivery error is isolated.
6	LF-CS vs. Plan-CS	--	

7	LF-MC vs. Plan-CS	Plan-CS vs. AC Plan-MC vs. AC	How do LF QA and AC QA compare?

2.3 Results and Discussion

2.3.1 Preliminary Validations

For the Elekta Infinity/Agility linac used in this study, Calibration Workflows resulted in leaf, leaf bank, and jaw positions agreeing with nominal positions within approximately 0.3mm each (0.1mm RMS). Despite the initial accuracy of these calibration processes, agreement with nominal positions can degrade over time. Monthly picket fence/phantom tests nominally ensure collimation components are within 1mm of post-calibration baseline positions, however they can be somewhat subjective and don't always directly isolate individual leaf position errors. For these reasons, it may be useful to perform Calibration Workflows more often if utilizing LF QA clinically, or to implement more objective collimator component QA testing. The necessary increased frequency of calibration or exploration of other QA methods were not investigated as part of this study.

The ArcCHECK phantom is supported by a H&N extension. A couch ROI was added to the phantom geometry in Pinnacle to account for this added attenuation. Calculation was compare to measurement for 17 10x10 cm^2 fields delivered at 17 different gantry angles passing through the H&N extension. With the added H&N ROI calculation agreed with measurement within 0.25% for all seventeen fields. On average, calculation agreed with measurement within 0.10% for these fields.

Verification data comparing MC calculation to measurement is shown in Table 3. The SciMoCa model utilized herein met each of the aforementioned AAPM recommended criteria 100% of the time. Furthermore, 96.5% of out-of-field measurement data matched calculation

within 2%. For field sizes ≤ 20x20 cm^2, 97% of out-of-field measurement data matched calculation with 1%. For the larger 30x30 cm^2 field size, 44% of measurement data matched within 1%. On average, over all field sizes, calculated output factors deviated from measurement by 0.20 ± 0.17% and PDDs agreed within 0.65±0.19%. On average, beam width varied by 0.54 ± 0.38%. On average, over all field sizes, and all profiles (both in-plane and cross-plane at each depth), penumbral width varied by 1.1 ± 0.5% of the field size. The absolute differences in penumbral width are shown per-field in Table 3. These values are shown in parentheses. For field sizes less than or equal to 10x10cm this difference was within 0.5mm, however penumbral width varied by almost 3mm for the 30x30 cm^2 field; these absolute differences are displayed in parentheses in Table 3. As hypothesized, deviation in penumbral width was larger for the profiles taken with the CC13 chamber (20x20 cm^2 and 30x30 cm^2) than the EDGE diode (< 20x20 cm^2). This larger deviation is expected to result from penumbral blurring and not poor beam modeling.

Table 3. 6MV Elekta Agility Monte Carlo beam model validation results are shown. Absolute percent differences in output factor (OF), percent depth dose (PDD), penumbral width (PW), and beam width (BW) are shown for each field size. Due to insufficient measured profile length, differences in 40x40cm penumbra were not calculated.

Field (cm)	OF |%Diff|	PDD Mean(|%diff|)	PW Mean(|%diff|)	BW Mean(|%diff|)
1x1	0.3%	--	--	--
2x2	0.4%	0.64 ± 0.23%	1.9 ± 1.0% (0.4mm)	0.9 ± 0.5% (0.2mm)
3x3	0.5%	0.95 ± 0.20%	1.5 ± 1.1% (0.4mm)	1.0 ± 0.5% (0.3mm)
4x4	0.0%	--	--	--
5x5	0.3%	0.22 ± 0.12%	1.0 ± 0.3% (0.50mm)	0.8 ± 0.5% (0.4mm)
6x6	0.4%	--	--	--
8x8	0.3%	--	--	--
10x10	0.0%	0.51 ± 0.11%	0.5 ± 0.3% (0.50 mm)	0.6 ± 0.3% (0.6mm)
12x12	0.3%	--	--	--
15x15	0.1%	0.61 ± 0.11%	--	--
20x20	0.0%	0.75 ± 0.10%	0.7 ± 0.4% (1.5 mm)	0.1 ± 0.1% (0.2mm)
25x25	0.2%	0.75 ± 0.12%	--	--
30x30	0.1%	0.72 ± 0.17%	1.0 ± 0.3% (2.9mm)	0.14 ± 0.11% (0.4mm)
35x35	0.1%	0.73 ± 0.17%	--	--

40x40	0.0%	0.64 ± 0.25%	--	0.25 ± 0.11% (1.0mm)
All	0.20 ± 0.17%	0.65 ± 0.19%	1.1 ± 0.5%	0.54 ± 0.38%

2.3.2 Strategies for Reducing the Computational Cost of LF-CS QA

Fig. 8 displays change in pass rate ((LF-CS vs. AC) minus (Plan-CS vs. AC)) as a function of control point spacing and dose grid resolution for 10 (8 H&N, 2 Brain) of the 22 clinical VMAT arcs used in this study. To make the plot clearer, error bars were plotted only in one direction for each curve. For the dose grid resolution curve, 2° control point spacing is utilized. 2mm resolution offers a 1.2±0.5% increase in pass rate compared to 3mm. For all subsequent studies, a 2mm dose grid resolution was utilized. Increased control point spacing resulted in lower LF-CS pass rates. Yielding a negative change in pass rate, control point spacing $\geq$ 2° was deemed unacceptable. 1° LF-CS spacing was used throughout the rest of this study. Similarly, Barbeiro et al. found log files reconstructed with ~3 times as many control points as the TPS agreed better with film than those reconstructed with the TPS' discretization.[82]

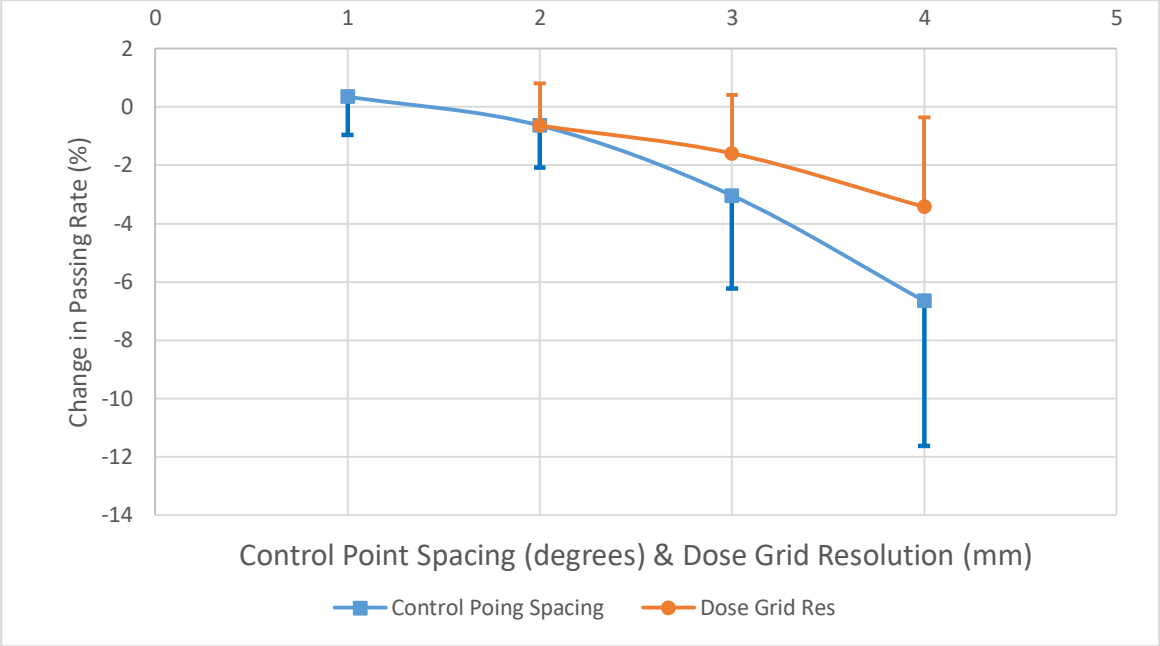

Figure 8. Difference in 2%/2mm pass rate ((LF-CS vs. AC) minus (Plan-CS vs. AC)) is plotted as a function of varying dose grid resolution (fixed 2° control point spacing) and control point spacing (fixed 2mm dose grid resolution). For visualization purposes, error bars (1σ) are plotted only in one direction.

Computational cost for CS was typically 12 seconds per control point. This corresponds to 18 minutes for a 358° arc planned with 4° control point spacing or 72 minutes for a log file plan reconstructed using 1° control point spacing. In comparison, MC dose calculation took ~5 minutes to calculate log file dose using every control point. This discrepancy in computational cost stems primarily from CS's computational cost being dependent on the number of control points, but also from the aforementioned differences in hardware.

2.3.3 Discrepancy in Calculated vs. Measured Diode Dose

2.3.3.1 In-Field Discrepancy: Dose Rate Effect

Fig. 9 plots in-field reduction in diode measurement with decreasing dose rate with respect to 570 MU/min for 3x3cm and 20x20cm MapCHECK2 fields, a 25x25cm ArcCHECK field, and the 23 10x10cm ArcCHECK fields in Table 1. For the MapCHECK2 measurements, low dose rate (35 MU/min) measurements were -1.2±0.4% and -1.6±0.4% lower than the high

dose rate (570 MU/min) measurements for 3x3cm and 20x20cm fields respectively. The trend shown in Fig. 9 is similar to Letourneau et al's, although slightly less in magnitude; Letourneau measured an under-response of -2.0% at 50MU/min.[54] Variance in dose rate effect with field size is believed to result from the reduced signal observed with smaller field sizes. Jursinic's 'hypothesized reaction scheme' would suggest smaller signals are less susceptible to the dose rate effect.[83] Eq. 13 quantifies the ratio of diode sensitivity when traps are open to that when traps are closed. The terms τ_d, τ_{RG1}, and τ_{trap1} are the charge lifetimes for the detection circuit, recombination-generation (R-G) centers, and traps respectively. Charge lifetime is the average time it takes for charge to enter and exit a system. Traps are locations in a solid that restricts the movement of charge. These traps results as a function of chemical impurities or irregular atomic spacings. R-G centers result from atomic impurities in semiconductors that introduce additional allowed electron levels in the band gap, slowing charge collection in the semiconductor diode. This would also explain why out-of-field diodes registered identical doses for low and high dose rates. Delivering a static 25x25cm arc so that all diodes were in-field, ArcCHECK measurement behaved similarly to MapCHECK2; measurement decreased with decreasing dose rate. Lastly, delivered dose rate was calculated from MU and gantry excursion for the 23 10x10cm arcs shown in Table 1. This 'calculated' dose rate matched the actual log file determined dose rate for all cases. Subsequently, as displayed by the solid line in Fig. 9, variation in in-field diode measurement as a function of dose rate agreed well with previous results.

$$\frac{S_{traps\ open}}{S_{traps\ closed}} \propto \sqrt{\frac{\frac{1}{\tau_d}+\frac{1}{\tau_{RG1}}}{\frac{1}{\tau_d}+\frac{1}{\tau_{RG1}}+\frac{1}{\tau_{trap1}}}} \qquad \textbf{(13)}$$

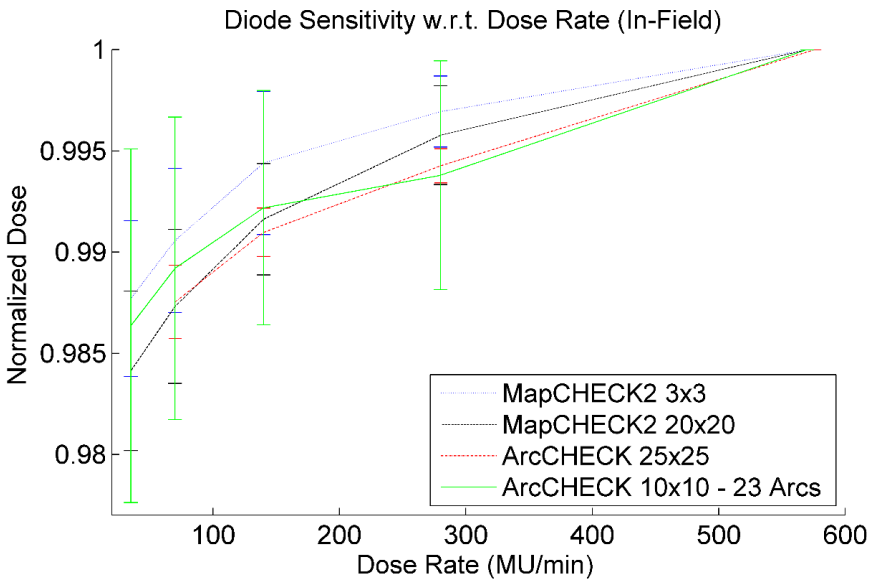

Figure 9. Reduction in in-field diode measurement with decreasing delivered dose rate is plotted for (1) 3x3 and 20x20cm MapCHECK2 fields, (2) a 25x25 ArcCHECK arc, and (3) 23 10x10 arcs. Reduction is normalized to 570 MU/min.

To better understand the clinical effect of dose rate dependence on AC QA, MC and CS calculated doses were compared to AC measurement for each diode in an open 25x25cm arc for both 70 and 570 MU/min dose rates. Plan-CS vs. AC percent dose differences were -2.1 ± 0.5% and -0.7 ± 0.5% at 570 and 70MU/min dose rates respectively. In comparison, Plan-MC vs. AC percent dose differences were 0.3 ± 0.8% and 1.7 ± 0.8%. Interestingly, because MC doses were generally higher than CS doses and higher dose rates resulted in greater measurement, MC dose differences were smaller for high dose rates, while CS dose differences were smaller for low dose rates. In comparison to AC QA, LF QA should be invariant with respect to dose rate. To verify whether this is indeed the case, percent difference in 70MU/min vs. 570MU/min LF-MC dose was calculated for the same 25x25cm arc; per-diode dose difference was 0 ± 0.85% and 0 ± 0.21% when calculated with 1% and 0.25% MC dose uncertainties, respectively. This shows that whereas AC QA is sensitive to dose rate and could yield false positives/negatives, LF QA does not have this issue.

Table 4. 10x10 arcs of varying gantry excursions and MU

Gantry Start/Stop/Excursion	MU
-90 / -60 / 30	10, 20, 35, 50, 100
-90 / -30 / 60	20, 40, 70, 100, 200
-90 / 30 / 120	20, 40, 80, 140, 200, 400
-90 / 90 / 180	20, 40, 60, 120, 210, 300, 600

2.3.3.2 Out-of-Field Discrepancy: Energy Effect

Fig. 10 displays percent difference in calculated (Plan-CS and LF-MC) versus AC measured diode dose as a function of off-axis distance for a 10x10cm 358° arc. For the in-field region, measurement and calculation agreed well. In the penumbral region, significant dose difference is seen. Penumbral differences are attributed to imperfect geometry (e.g. imperfect linac isocentricity, imperfect leaf calibration, and setup error) and physical diode limitations (e.g. finite size), as well as the two issues discussed in the next paragraph.

Out-of-field, calculation is much lower than measurement. This disagreement is comprised of two major components. First, the calculation model could be underestimating collimator scatter and head leakage at greater off axis distances. For CS, this is typical of many institutions and is known to be the case at our institution as well.[84-88] Second, diodes could be over-responding to the low energy patient-scattered photons. Rickner et. al theorizes 30% and 70% increases in diode sensitivity to 0.5MeV and 0.1MeV photons.[58] The typical energy range of patient-scattered photons is right in this range.[89] Thus, one can expect an approximate 50% over-response to low energy photons, and an approximate 30% net over-response after multiplying by the low energy component of the fluence. Figure 10 illustrates the percent difference in calculated dose (Plan-CS and LF-MC) and the ArcCHECK measured dose as a function of off-axis distance for a 10x10 cm^2 358° arc. The MC results of are in good agreement with Rickner et al.; a relative 10-30% over-response by out-of-field diodes is concluded for this

10x10 cm^2 field. It is important to note this is for a 0% dose threshold. A 10% dose threshold is explored below to determine clinically relevant effects only.

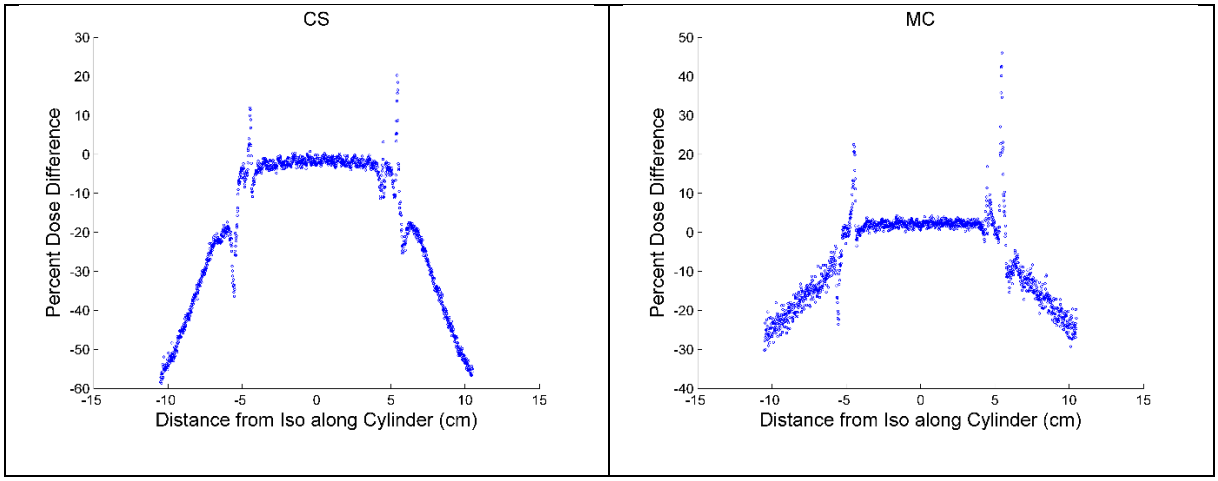

Figure 10. Calculated vs. measured dose differences are plotted as a function of axial distance from isocenter for a 10x10 arc using a) Plan-CS and b) LF-MC calculated doses.

Field size plays a critical role in determining the energy spectrum of photons at each diode. Here, the effect of field size on energy spectrum, and subsequently diode sensitivity was studied. Three static 200MU arcs (2x2 cm^2, 10x10 cm^2, 25x25 cm^2) were delivered on the ArcCHECK phantom. Fig. 11a and Fig. 11c plot histograms of percent difference in calculated versus measured diode dose for each field size and for CS and MC respectively. Fig. 11b and Fig. 11d plot the same histograms with a 10% dose threshold applied. Histograms consisted of a tight primary peak and prolonged 'spectral' out-of-field distribution for both CS and MC.

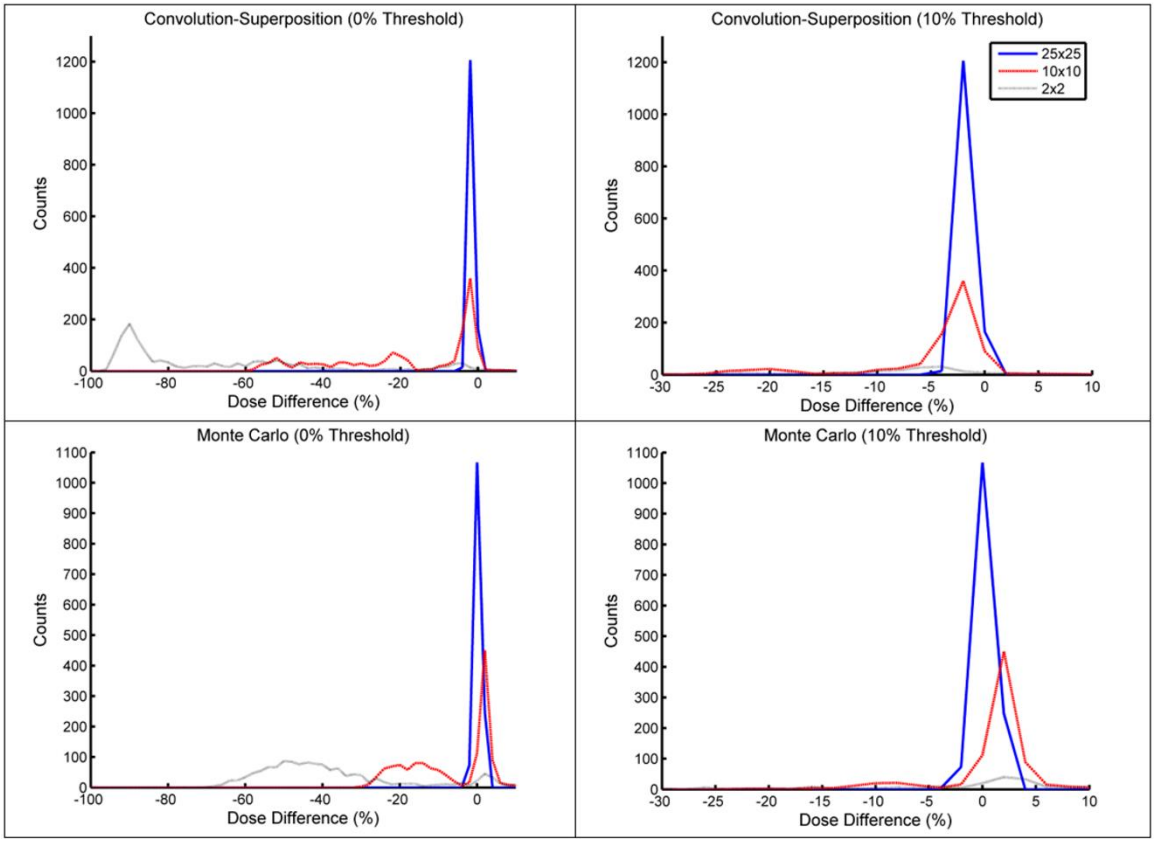

Figure 11. Histograms of percent diode dose difference (calculated vs. measured) for CS and MC with respect to ArcCHECK measurement. The left plots (a, c) utilize no threshold, while the right plots (b, d) utilize a 10% threshold based off the maximum measured dose.

As discussed previously, this 'spectral' distribution results from inaccurately calculated collimator and head leakage as well as diode over-response to patient-scattered radiation. Dose differences increased with decreasing field size. By shrinking the field, the distance from the field edge to distal diodes increases. This results in more medium to scatter off and thus a softer photon spectrum and greater diode over-response. As illustrated by the relative positions of their out-of-field distributions as well as the integral area of these distributions, MC yielded better agreement with out-of-field measurement than CS. For the 2x2 cm^2 field size, a good portion of CS's spectral distribution is centered around -90%. This corresponds to a 10 times greater measured versus calculated dose. This massive difference is in part due to an

underestimation of out-of-field dose by the CS method, but also largely due to the extremely low dose – a low absolute dose will yield a larger relative dose difference. For the 2x2cm MC field, measurement is 70% higher than calculation. Given the diode over-response values mentioned in 3.3.2, 70% over-response is reasonable. It is important to note that some portion of this discrepancy will be due to calculation inadequacies.

Despite the very large discrepancies seen in Fig. 11a/c, when a standard 10% dose threshold is applied (Fig. 11b/d), much of this out-of-field component is eliminated. For instance, for the 2x2cm field where the absolute quantity of patient scatter is minimal, there is a miniscule out-of-field distribution. Thus, diode energy dependence is going to result in very few failing diodes. However, for the 10x10cm field where the absolute magnitude of patient scatter is greater, the number of failing out-of-field diodes that are above the measurement threshold is more noticeable - energy effects are seen on the order of 3-5% of the maximum measured dose. Although systematic uncertainties due to energy and dose rate will cancel out to some degree, when compared to standard 2-3% gamma criteria, 3-5% over-response to lower energy radiation remains worrisome.

2.3.3.3 Delivery Complexity

Static fields/arcs with a wide variety of field sizes were delivered onto the ArcCHECK phantom and log files reconstructed using Monte Carlo. LF-MC calculated dose is compared to AC diode measurement using MPDD and MAPDD along with an 85% dose threshold. Fig. 12a and 12b display MPDD and MAPDD respectively as a function of field size ($FS_X \cdot FS_Y$). Smaller fields yielded greater MAPDD values, indicating greater LF-MC vs. AC disagreement for smaller field sizes. Negative and positive MPDD values were seen because an 85% dose threshold will include some field penumbra. Consequently, any small shift in diode position

due to setup error or MLC position due to leaf calibration error can result in large differences between calculation and measurement. Depending on the position of the diode with respect to CAX (e.g. ±x̂/ŷ), direction of setup error, and direction of leaf calibration error, measured diode dose may increase or decrease with respect to calculation.

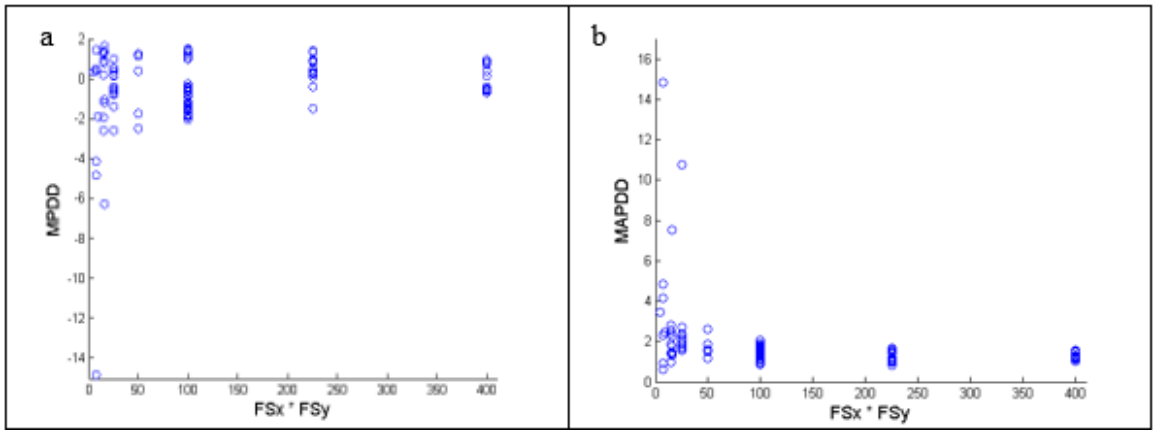

Figure 12. LF-MC and AC doses are compared for varying field sizes using an 85% dose threshold. a) Mean Percent Dose Difference (MPDD) and b) Mean Absolute Percent Dose Difference (MAPDD) are plotted against FSx*FSy for all the aforementioned fields in Section 2.2.3.

Delivering various combinations of abutting fields onto the ArcCHECK phantom, calculated (LF-MC, Plan-MC, Plan-CS) vs. AC measured dose difference was evaluated as a function of segment size and quantity. Table 5 displays MPDD values for each static and arc field arrangement using both 10% and 85% dose thresholds. As the number of segments increased, Plan-CS increasingly deviated from measurement (max 13%). This Plan-CS dose difference likely occurs due to the out-of-field dose mismatch seen between CS and AC in section 2.3.3.2. This discrepancy is seen not only for the 10%, but also the 85% dose threshold. This is because what is considered out-of-field for one segment may be considered in-field for the combined 12x12cm field. Plan-MC and LF-MC matched measurement within a few percent for all cases. LF-MC and Plan-MC deviated from measurement by 1.42 ± 1.2% and 1.56 ± 1.2%

on average respectively. This shows improvement by accounting for delivery errors was only 0.14%.

Table 5. MPDD (mean percent dose difference w.r.t AC) values are displayed for various field complexities, static and arc fields, and for 10% and 85% dose thresholds.

Field(s) → / **Threshold ↓**		12x12cm	4 6x6cm	9 4x4cm	16 3x3cm	36 2x2cm
				Static Fields		
LF-MC	10%	1.0 ± 11.0	-0.4 ± 11.1	0.2 ± 10.5	0.4 ± 9.4	4.5 ± 17.4
	85%	-1.1 ± 1.4	-2.7 ± 2.2	-1.2 ± 2.4	-0.8 ± 3.3	1.8 ± 4.8
Plan-MC	10%	3.4 ± 10.2	2.3 ± 12.2	0.9 ± 9.3	-0.5 ± 8.5	2.3 ± 16.6
	85%	1.4 ± 1.4	-0.1 ± 1.6	-0.5 ± 1.6	-1.9 ± 3.0	-1.0 ± 3.6
Plan-CS	10%	-0.2 ± 11.6	-2.1 ± 13.6	-3.6 ± 10.7	-6.7 ± 9.2	-11.4 ± 13.5
	85%	-1.5 ± 1.6	-4.3 ± 3.5	-4.6 ± 3.5	-6.9 ± 3.0	-7.6 ± 2.3
				Arcs		
LF-MC	10%	0.3 ± 6.2	-2.4 ± 6.6	-2.7 ± 8.2	-3.8 ± 10.3	-0.5 ± 21.5
	85%	0.2 ± 1.3	-1.8 ± 1.6	-0.9 ± 2.0	-0.1 ± 1.9	-1.6 ± 6.2
Plan-MC	10%	2.7 ± 5.9	-0.2 ± 6.6	-2.1 ± 8.0	-3.7 ± 10.2	-0.7 ± 21.6
	85%	2.8 ± 1.3	0.6 ± 1.6	-0.1 ± 1.9	-0.3 ± 1.9	-3.6 ± 3.2
Plan-CS	10%	-0.2 ± 11.6	-2.1 ± 13.6	-3.6 ± 10.6	-6.6 ± 9.2	-13.0 ± 24.8
	85%	-1.4 ± 1.6	-4.2 ± 3.5	-4.6 ± 1.9	-6.9 ± 3.0	-11.7 ± 2.9

2.3.4 Comparing LF QA & AC QA

Gamma pass rates were calculated for each of 22 clinical beams and each dose-pair shown in Table 6 using a 10% dose threshold. Comparing #3 to #1, we saw lower LF-CS than Plan-CS pass rates, suggesting even 1° control point spacing was insufficient. Comparing #4 to #2, reconstructive accuracy was evaluated for the full 25Hz log file; pass rates were identical or improved for 17/18 (2%/2mm) and 14/18 arcs (1%/1mm). High pass rates for #5 and #6 indicate both dose calculation differences and machine delivery errors were small. LF QA results (#7) were similar to AC QA results (#1/2).

Table 6. 2%/2mm and 1%/1mm gamma pass rates are shown for each dose-pair.

#	**Comparison**	**2%/2mm**	**1%/1mm**
1	Plan-CS vs. AC	97.5 ± 1.6%	81.6 ± 6.5%
2	Plan-MC vs. AC	96.5 ± 4.2%	78.1 ± 9.9%
3	LF-CS vs. AC	97.0 ± 2.4%	79.2 ± 8.1%
4	LF-MC vs. AC	97.3 ± 3.2%	78.9 ± 10.3%

5	Plan-MC vs. Plan-CS	99.0 ± 2.4%	85.0 ± 8.2%
6	LF-MC vs. Plan-MC	99.6 ± 0.4%	91.0 ± 2.7%
7	LF-MC vs. Plan-CS	96.8 ± 5.1%	76.3 ± 11.6%

Values are calculated for the same 22 beams for each plan pair in Table 7 Comparison #5 isolates differences in dose calculation method; MC and CS agreed within 1-2%. Comparisons #6/7 isolate delivery errors for CS and MC respectively; the effect of delivery error was ~0±1%. Comparing #3 to #1 and #4 to #2, the effects of delivery error are again seen to be very small for this Agility system. #2/4/5 show MC dose was systematically 1.2% higher than CS dose when utilizing an 85% dose threshold. This is believed to result from slight MC vs. CS beam model differences stemming from how CS dose is normalized to a 10x10cm field at 10cm depth and MC is not. Rather, the MC beam model was simply designed to best match calculation to measurement for the greatest amount of PDDs and dose profiles. Accounting for both delivery error and calculation difference, comparison #8 is this study's standard LF QA comparison. Summing systematic differences linearly and statistical differences in quadrature, #8 equals #5 plus #7 for both dose thresholds. Eliminating phantom error sources resulted in substantially reduced uncertainty; values for #5-8 were ~0.8 ± 1.2% lower than #1-4. We conclude phantom-based uncertainty is ~0.8 ± 1.2%. It is important to note that phantom-dependent uncertainty includes not only measurement and setup uncertainties, but also any potential beam inconsistencies due to imperfect machine QA that the log file does not track. Gamma and statistical analyses both led to the same conclusions.

Table 7. Global percent diode dose differences (mean ± σ) are displayed for various plan pair comparisons and for both 10% (standard) and 85% (target) dose thresholds.

	Comparison	10% Threshold	85% Threshold
#1	Plan-CS vs. AC	0.8 ± 1.8%	0.8 ± 2.6%
#2	Plan-MC vs. AC	1.0 ± 1.8%	2.1 ± 2.5%
#3	LF-CS vs. AC	1.0 ± 1.8%	0.8 ± 2.6%

#4	LF-MC vs. AC	1.0 ± 1.7%	2.1 ± 2.5%
#5	Plan-MC vs. Plan-CS	0.2 ± 1.1%	1.2 ± 1.7%
#6	LF-CS vs. Plan-CS	0.2 ± 1.0%	0.4 ± 1.4%
#7	LF-MC vs. Plan-MC	0.1 ± 0.9%	0.0 ± 1.3%
#8	LF-MC vs. Plan-CS	0.2 ± 1.3%	1.2 ± 2.1%

2.4 Summary

As discussed briefly in the introduction, there exist distinct advantages and disadvantages of the two QA systems. The key advantage of ArcCHECK is that it directly measures dose; physicists can trust measurement within the ranges of uncertainty set forth in this study. However, it is difficult to determine the cause of failing pre-treatment QA. In comparison to AC QA, LF QA is unaffected by phantom-dependent uncertainties and capable of isolating the cause of dose difference. The downside of LF QA is that it relies heavily on machine QA and accurate dose calculation. Log file reconstructed dose would be more accurate if beam limiting device (BLD) calibration was carried out more often. Reconstruction accuracy could be further improved by better maintaining beam output and symmetry whether through more frequent QA or tighter tolerances. Furthermore, LF QA is unable to account for dose calculation differences when utilizing the same dose calculation method as the TPS. Thus, LF QA should ideally utilize an independent dose calculation method that has been thoroughly vetted as detailed in TG-106.[90] In the case of dose deviance between the two methods, ArcCHECK can be utilized as a second check. Assuming a non-MC TPS dose calculation engine, Monte Carlo is well suited for LF QA. Conventionally, MC is deemed accurate, but slow. However, reconstructing every log file sample with no increase in cost, LF-MC QA is not only accurate, but exceptionally fast. Once properly setup, LF-MC QA is a fully automatable patient-specific QA technique that can quickly and accurately calculate per-

fraction delivered dose. However, given the increased reliance on machine QA and dose calculation accuracy, LF QA should at least initially be used only as a supplement to current QA systems.

2.5 Conclusion

In Chapter 2, LF QA is assessed as a potential supplement or potential replacement of ArcCHECK QA. First, preliminary validations were carried out to assure there are no flaws in the work that we are carrying out here. The beam models and MLC calibration were both verified to be accurate. For LF-CS QA, varying control point spacings were investigated in order to determine the optimal number of log file samples that should be utilized. Utilizing control point spacing greater than 1° resulted in substantially degraded dosimetric accuracy. LF-CS QA utilizing 1° control point spacing and LF-MC QA utilizing the full 25Hz log file both yielded equivalent results to AC QA. Next, the ability of diodes to accurately measure dose was investigated. Calculation and ArcCHECK measurement differed by up to 1.5% in-field due to variation in dose rate and up to 5% out-of-field. For highly segmented experimental plans, despite CS calculation deviating by as much as 13% from AC measurement, Plan-MC and LF-MC doses generally matched AC measurement within 3% (mean 1.56% and 1.42% respectively). Finally LF-CS QA and LF-MC QA as carried out on the phantom geometry was carried out and compared to conventional AC QA. Carrying out AC QA and LF QA for 22 clinical VMAT arcs, phantom-dependent, calculation method-dependent, and delivery error-dependent dose uncertainties were found to be 0.8 ± 1.2%, 0.2 ± 1.1%, and 0.1 ± 0.9% respectively. It follows that by eliminating phantom-dependent uncertainty, LF QA is theoretically more accurate than AC QA. However, none of the doses utilized (AC, MC, CS)

can truly be considered a gold standard. Without a gold standard, we merely conclude that LF QA and AC QA offer similar dosimetric accuracy.

CHAPTER 3 LF QA FOR TREATMENT PLANNING AND DELIVERY QA: PART 2 – PATIENT ANATOMY

3.1. Introduction

In the previous chapter, log file doses were reconstructed on the ArcCHECK geometry. Ultimately, LF-MC QA and LF-CS QA were deemed to be equivalent to ArcCHECK QA given standard TG-40 and TG-142 machine QA protocols are carried out. In this chapter, LF QA is instead carried out on the patient specific anatomy. DVH-based metrics as well as the aforementioned TCP formulation are used to evaluate the differences between the calculated treatment planning dose and the dose reconstructed using LF-MC QA. In addition, LF-MC QA is utilized as a treatment planning QA tool.

There are three key objectives of this study: (1) to establish the accuracy of LF-MC QA, (2) to compare the performance of patient-specific LF-MC QA with phantom-based QA, and (3) to quantify the causes of discrepancy determined by LF-MC QA. To achieve objective 1, the accuracy of MC dose calculation is assessed versus commissioning data as well as phantom measurement setups. To fulfill objective 2, LF-MC QA results on the patient anatomy are compared to clinical ArcCHECK results. For objective 3, dose discrepancy is broken down into delivery error, gantry discretization difference, and dose calculation difference. Dosimetric discrepancy due to calculation method was further broken down as a function of beam complexity and patient heterogeneity. Although LF-MC QA on the patient anatomy is the main focus of this study, Monte Carlo may not be available for every institution. Thus, delivery error was also assessed using LF-CS QA.

3.2. Materials & Methods

3.2.1. Dose Calculation Validation

Both dose calculation methods were first thoroughly validated with respect to measurement. Calculation was first investigated for square fields as delivered on a the homogeneous IBA Blue Phantom water tank. It is mandatory that calculation methods agree with measurement well for these simple square fields in a homogeneous medium. If they don't agree well for square fields and a homogeneous geometry then they will not agree well for more complicated plans delivered on heterogeneous patient anatomy. Next, more complicated clinical fields were delivered onto a homogeneous solid water stack. Measurement was compared to calculation to assure that we are able to accurately deliver complicated fields. However, neither of these tests assess the ability of CS and MC to accurately calculate dose in a heterogeneous environment, thus CS and MC calculation was also compared to measurement in a heterogeneous cork phantom for square fields.

3.2.1.1. Water Tank Measurement Data – Square Fields

It is important to note that the MC beam model was redone using a newer set of measurement data (M2) since Chapter 2. Thus, for this paper, CS and MC are commissioned using two different, but similar, sets of measurement data – M1 and M2 respectively. To investigate the quality of the beam models the following comparisons were made. CS and MC were compared to M1 and M2 respectively. M2 was compared to M1 to make sure the measurements were accurate within approximately 1%. Lastly, MC and CS were compared.

M2 was acquired in the same way as M1 except for two key differences. First, output factors less than $5x5cm^2$ and PDDs less than $2x2cm^2$ were acquired using a PTW60019 microDiamond. Second, M1 was affected by a software bug that M2 was not affected by. This bug in IBA's OmniPro-Accept software v7.3-7.4b resulted in water scans in which background may not have been subtracted, compressing and offsetting the vertical scale of the data.

Unfortunately, it was impossible to be sure which profile measurements were affected by this bug. It follows that some M1 cross-beam profiles had inflated tails. Ideally, the CS beam model would be redone using M2, however such an undertaking is outside of the scope of this dissertation. Furthermore, the CS beam model parameters were not affected by this discrepancy.

Per-field comparison metrics included (1) absolute percent difference in output factor, (2) mean absolute percent global difference in PDD from 0-35cm, (3) mean absolute percent local difference in in-field dose, (4) mean absolute percent global difference in out-of-field (<10%Dmax) dose, (5) mean absolute distance-to-agreement in the penumbra (80%-20%), (6) mean absolute difference in penumbra width, and (7) mean absolute difference in field width (50%-50%). Due to insufficient scan length, penumbra and out-of-field data for the 40x40cm field were not analyzed.

3.2.1.2 Solid Water Phantom Measurement Data – Clinical Fields

To determine the accuracies of the two calculation methods, point doses measurements were taken and compared to both CS and MC calculation. First, TG-119 measurements were carried out.[49] TG-119 cases included: prostate, H&N, C-shape (hard), H&N with integrated boost (HNSIB), and prostate with lymph nodes (PLN). It is noted that the latter two cases are additional cases provided to us by Henry Ford Health System that are not part of the TG-119 publication. HNSIB and PLN were included because they are more difficult than their H&N and prostate counterparts and thus push the QA harder. It is important to QA more difficult treatment cases. Although these two cases are not part of TG-119 for the rest of this chapter they will be included as if they are. Two measurement positions were utilized for each test, one at the center of the target ROI and one in an out-of-field organ-at-risk (OAR). In total, ten measurements were compared to ten MC calculations and ten CS calculations. All

measurements were taken using a 0.125cc Wellhofer IC-10 ion chamber inserted into a 30x30x15cm stack of solid water. As recommended in TG-119, dose to charge ratio was determined by delivering 10x10 AP/PA fields onto the solid water stack and comparing measurement to dose calculation. This dose per charge metric was calculated for both dose calculation methods; the difference in the two dose per charge factors was 0.3%. Finally, results were compared to those found in TG-119.

In addition to the standard TG-119 cases, 11 clinical VMAT plans (22 arcs) were delivered onto the same solid water stack. These plans were selected to cover a range of MC – CS dose variations. Measurement points were selected per plan to be in a high dose, low dose gradient regions. The measurement point was confined to the central 18x18x3cm of the slab to ensure 6cm of medium to laterally scatter off of at all gantry angles. The same dose per charge factors were utilized as in the TG-119 measurements. For Monte Carlo, the solid water was overridden to a density of 1.035. This density was found to maximize the agreement between measurement and Monte Carlo for the specific solid water material composition utilized. A dose grid resolution of 4mm was utilized to match the diameter of the ion chamber utilized. Monte Carlo statistical uncertainty was set to 0.25%. A variety of density overrides were analyzed for Convolution Superposition, however, ultimately, utilizing no density override was found to yield the greatest agreement between CS and measured doses.

3.2.1.3 Cork Phantom Measurement Data – Square Fields

Heterogeneous dose calculation accuracy was assessed for both CS and MC calculation with respect to ion chamber measurement in a "lung phantom." The lung phantom utilized in this study consisted of 10cm of composite cork sandwiched between 6cm (entrance) and 6cm (exit) of solid water. Composite cork is known to be dosimetrically equivalent to ICRU-44

lung tissue.[91] The solid water slabs on either side of the cork act like the front and back of the patient respectively. The phantom was imaged while the ion chamber was inserted into the cork. In the resulting CT images, the cork phantom had a mean CT density of -897HU. To account for various ion chamber positions, CT image slices were virtually shifted and re-imported into Pinnacle3. The ion chamber was inserted at 12 depth positions along the central axis after accounting for a 1.3mm shift in dose reference point relative to the chamber surface. [1.4, 6.5, 7.45, 8.4, 9.35, 10.3, 11.25, 12.2, 13.15, 14.1, 16, and 17cm depths]. The ion chamber utilized was a TN-34045 parallel plate ion chamber. Calculation and measurement were compared for both 4x4 cm^2 and 10x10 cm^2 square fields. Due to the large active volume of the parallel plate chamber, smaller field sizes were not investigated.

3.2.2 LF-MC and LF-CS QA on the Patient Anatomy

Log files are acquired using Elekta's Log File Convertor for Integrity R3.2. The log files record linac delivery parameters (dose rate, gantry/collimator angle, leaf/collimator positions, MU) every 40ms. For this study, an Elekta Infinity linac equipped with the Agility beam-limiting device was utilized. Plans were selected from previously treated patients that were planned and optimized using the Pinnacle3 treatment planning system.

LF QA was assessed on the patient anatomy using both MC and CS dose calculation algorithms. Four dose distributions were utilized in this study. Plan-CS is the clinical treatment plan dose that was planned and optimized in Pinnacle3 v14.0 using convolution-superposition. The clinical Plan-CS doses were planned using a 4° control spacing. Plan-MC is the corresponding Monte Carlo dose calculated using the exact same plan as Plan-CS. LF-CS dose is calculated using the linac log file and Pinnacle's convolution-superposition. Lastly, LF-MC dose is calculated using the linac log file and Monte Carlo. The log file dose reconstruction

methodologies match those explained in chapter 2. For LF-CS, log files are selected to achieve as close to 1° control point spacing as possible. This is done by converting log file samples into control points in a dynamic arc field. For LF-MC, every log file sample with non-zero MU is converted into a control point in a dynamic arc field. MC doses were calculated using 1% statistical dose uncertainty. CT-to-density table and ROI override information was the same for both Pinnacle convolution-superposition (CS) and SciMoCa (MC) dose calculation algorithms. All doses were calculated using the same 3mm dose grid resolution that is used to calculate Plan-CS doses clinically at our institution.

The patient-specific LF-MC QA tool presented in this study is designed to verify the accurate calculation and delivery of dose on each patient's unique heterogeneous anatomy. First, accurate dose calculation was verified. LF-MC QA utilizes three key dose comparisons. Comparing Plan-MC to Plan-CS, dose discrepancy resulting from dose calculation method differences was isolated. When comparing dose calculation methods, all comparisons were MC with respect to CS. Comparing LF-MC to Plan-MC, dose discrepancy due to delivery error was isolated. Lastly, LF-MC vs. Plan-CS was analyzed in order to account for the combined effects of delivery error and dose calculation. However, because delivery error was found to be near-zero, the combined effects were similar to the effect from the dose calculation discrepancy and used for the rest study.

The LF-CS QA tool presented in this study is simpler than the LF-MC QA tool in that it is only capable of catching delivery error; calculation accuracy is not verified. However, it is not as simple as comparing LF-CS to Plan-CS. In Chapter 2, it was shown that LF-CS QA ought to utilize a control point spacing no larger than 2° (1° is preferred). However, the clinical treatment dose, Plan-CS, was calculated with a 4° control point spacing. Thus, when comparing

LF-CS doses to Plan-CS doses, delivery error is not truly being isolated. In order to assess this addition "gantry discretization difference" Plan-CS(4°) control points were linearly interpolated in Pinnacle and Plan-CS(2°) dose calculated. Pinnacle does not easily allow for further interpolation, thus Plan-CS(1°) was not calculated. In total, three doses were compared in order to quantify the effects of gantry discretization difference: LF-CS(1°), Plan-CS(2°), and Plan-CS(4°). Two comparisons were carried out. First, LF-CS(1°) dose was compared Plan-CS(4°) dose. This dose difference is due to delivery error as well as gantry discretization differences. Second, LF-CS(1°) was compared to Plan-CS(2°). This dose difference is the same as before except with reduced gantry discretization difference. The goal of studying the effects of gantry discretization is to determine if Plan-CS(4°) needs to be interpolated to Plan-CS(2°) for some or all LF-CS QA cases. Whereas clinical Plan-CS dose on the patient anatomy is calculated using 4° control point spacing, in Chapter 2 on the ArcCHECK geometry, all Plan-CS doses were clinically calculated using 2° control point spacing. It is important to note that MC does not suffer from discretization error. MC simulates millions of particle histories at all possible continuous gantry angles for both Plan-MC and LF-MC. Thus, LF-MC vs. Plan-MC dose difference is an accurate quantification of delivery error.

42 patients (85 treatment fields) were selected for this study based on the following criteria. First, plans must have been delivered on the Elekta Agility machine so that linac log files could be acquired. Second, all investigated fields were dynamic arc deliveries. In total, 12 head and neck, 6 lung, 6 breast, 5 brain, 4 esophageal, 3 mediastinum, 2 prostate, 2 spine, 1 stomache, and 1 axillary node patients were analyzed.

LF-MC QA utilizes the following metrics to evaluate the effects of planned patient dose discrepancy. DVH-based metrics included PTV D_{mean}, PTV $D_{99\%}$, PTV $D_{95\%}$, and PTV $D_{1\%}$.

These metrics were also calculated for GTV ROIs. Site-specific organ at risk (OAR) constraints were utilized as gathered from TG-101 and site-specific RTOG protocols.[7] DVH-based metrics were evaluated per-plan. In addition to DVH-based metrics, 3%/2mm gamma pass rates were calculated on the patient anatomy using both 10% (standard) and 85% (target approximation) dose thresholds. Lastly, tumor control probabilities (TCP) were calculated per plan as presented by Niemierko et al.[23] TCP was calculated using GTV when available and CTV when not. A study by Okunieff et al. shows that γ_{50} values can vary significantly between treatment site.[92]

Treatment site-specific TCD50 and γ_{50} values were acquired from a variety of publications.[92-95] The following treatment site-specific γ_{50} values were utilized: H&N 1.03; brain 0.75; lung 0.57; esophagus 1.03; breast 1.03; mediastinum 0.41; spine 0.41; stomache 0.71; prostate 0.74. TCD_{50} values were selected based upon the staging of each patient's individual treatment. All LF-MC QA metrics were calculated per-plan.

3.2.3 Dose Discrepancy (MC vs CS)

In Chapter 1, the main sources of calculation difference between Monte Carlo and collapsed-cone convolution-superposition were stated as being complexity and heterogeneity. Dosimetric differences in the two calculation methods is assessed as a function of beam complexity and patient heterogeneity.

3.2.3.1 Beam Complexity

Monitor units (MU) per prescription dose (Rx) is a standard metric for quantifying beam complexity.[96] Variation in dose due to calculation difference was evaluated as a function of MU/Rx on the patient anatomy for both target and external (outer contour around the patient skin) ROIs.

3.2.3.2 Patient Heterogeneity

Monte Carlo is known to more accurately model anatomical heterogeneities than convolution-superposition algorithms.[14] To quantify this effect and potentially explain the dosimetric variation seen between the two algorithms, 10 patient plans (5 lung and 5 H&N) were recalculated with the entire patient volume being overridden to unit mass density. Dose variation was analyzed as a function of tumor position relative to air gaps as well as bony anatomies.

3.3 Results

3.3.1 Dose Calculation Validation

3.3.1.1 Water Tank Measurement Data

Table 8 displays validation results for the two dose calculation methods and two sets of measurement data. From the first two data columns we see that both calculation methods yield clinically acceptable results.[90] MC consistently agreed within 1% and 1mm of the M2 measurement data. Similarly, CS consistently agreed within 1% and 1mm of the M1 measurement data. Furthermore, the two sets of measurement data were also in good agreement with each other. The largest difference between the two sets of measurement data was seen out of field. This difference is attributed to a since-fixed software bug in IBA's OmniPro-Accept software where background was not being subtracted despite the software saying it had been. M1 contained this bug while M2 did not, resulting in higher tails. Lastly, MC and CS generally agreed within 1% / 1mm of each other. In comparing the two calculation methods, the greatest difference was found to be in the in-field region with a variation of 1%. Thus, for square-fields and a water tank, all measured and calculated data generally agreed within 0.5% or 0.5mm and almost always agreed within 1.0% or 1mm.

Table 8. Beam model validation results are summarized. *CS output factors are exactly determined from measurement.

	MC vs M2	CS vs M1	M2 vs M1	MC vs CS
Output Factor	0.15 ± 0.13%	0*	0.25 ± 0.31%	0.25 ± 0.31%
PDD	0.41 ± 0.17%	0.52 ± 0.13%	0.38 ± 0.19%	0.64 ± 0.28%
Penumbra DTA	0.23 ± 0.11mm	0.34 ± 0.16mm	0.50 ± 0.89mm	0.45 ± 0.17mm
Beam Width	0.64 ± 0.62mm	0.60 ± 0.27mm	0.63 ± 0.56mm	0.67 ± 0.32mm
In-Field	0.58 ± 0.19%	0.60 ± 0.15%	0.24 ± 0.17%	1.06 ± 0.48%
Out-of-field	0.20 ± 0.16%	0.33 ± 0.41%	0.80 ± 0.09%	0.38 ± 0.25%

3.3.1.2 Solid Water Phantom Measurement Data

Per-plan percent differences in calculated vs. measured doses for the ten TG-119 measurement points displayed in Table 9. Percent differences are normalized relative to the prescribed doses. In general, MC agreement with measurement was superior to that of CS. CS agreed with measurement within 3% for 7/10 cases. Disagreement as large as -4.6% was seen for the in-field C-Shape(Hard) case. In comparison, MC agreed with measurement within 3% for all cases and within 2% for 8/10 cases.

Table 9. TG-119 results are tabulated per treatment site for both high dose and low dose measurement regions as well as both CS and MC dose calculation methods.

	Percent Difference in Calculated vs. Measured Dose			
	Target Measurement Point		OAR Measurement Point	
	CS	MC	CS	MC
H&N	2.4%	0.9%	4.5%	0.4%
C-Shape(Hard)	-4.6%	-1.2%	3.1%	-2.7%
Prostate	0.3%	1.1%	1.7%	2.2%
Prostate + LN	-0.3%	0.9%	1.1%	-0.6%
H&N SIB	1.0%	-1.2%	-1.3%	-1.7%

As displayed in Table 10, TG-119 results found herein are compared to results found in the actual TG-119 task group report. High dose measurements are addressed first. Whereas in TG-119, a confidence level of 4.5% is reported, CS and MC confidence levels were 5.3% and 2.4% respectively. Next, low dose TG-119 results were addressed. Whereas in TG-119, a confidence level of 4.7% is reported, confidence limits for CS and MC were 6.1% and 4.2%

respectively. It is important to reiterate that the TG-119 literature results include some simpler treatment sites and IMRT deliveries instead of the VMAT deliveries utilized herein.

Table 10. TG-119 results from this study are compared to literature for both high dose and low dose regions.

	TG-119	CS	MC
	High Dose Region		
Mean Difference	-0.2 ± 2.2%	-0.2 ± 2.6%	0.1 ± 1.2%
Mean Absolute Difference	0.9%	1.7%	1.0 %
Confidence Limit	4.5%	5.3%	2.4%
	Low Dose Region		
Mean Difference	0.3 ± 2.2%	1.8 ± 2.2%	-0.5 ± 1.9%
Mean Absolute Difference	1.1%	2.3%	1.5%
Confidence Limit	4.7%	6.1%	4.2%

Per-plan percent difference in calculated vs. measured dose was calculated for eleven clinical cases. On average, MC agreed within -0.04 ± 1.61% of measurement. Absolute agreement was 1.20 ± 1.02%. In comparison, CS agreed with measurement within 1.65 ± 1.86%. Absolute agreement was 1.81 ± 1.69%. For all but one case (#4), MC agreed with measurement within 2%. CS agreed with measurement within 3% for all but the same case (#4). In general, MC outperformed CS; the MC confidence limit was 3.2% as opposed to the 5.3% of CS.

3.3.1.3 Cork Phantom Measurement Data

CS and MC calculation were compared to Markus chamber measurement at various cork depths in the phantom. Calculated MC and CS PDDs are plotted against ion chamber measurement for 4x4 and 10x10 fields in Fig 13. On average, CS agreed with measurement within 4.3 ± 2.6% and 2.1 ± 1.7% for 4x4 and 10x10 square fields at the measurement points shown. MC agreement was 1.3 ± 1.3% and 0.5 ± 0.4% for the same fields. Monte Carlo

agreement with measurement was better than that of CS especially for measurement points deeper into the cork. These results are as expected. As discussed in Chapter 1, Monte Carlo is expected to better calculate heterogeneous dose.

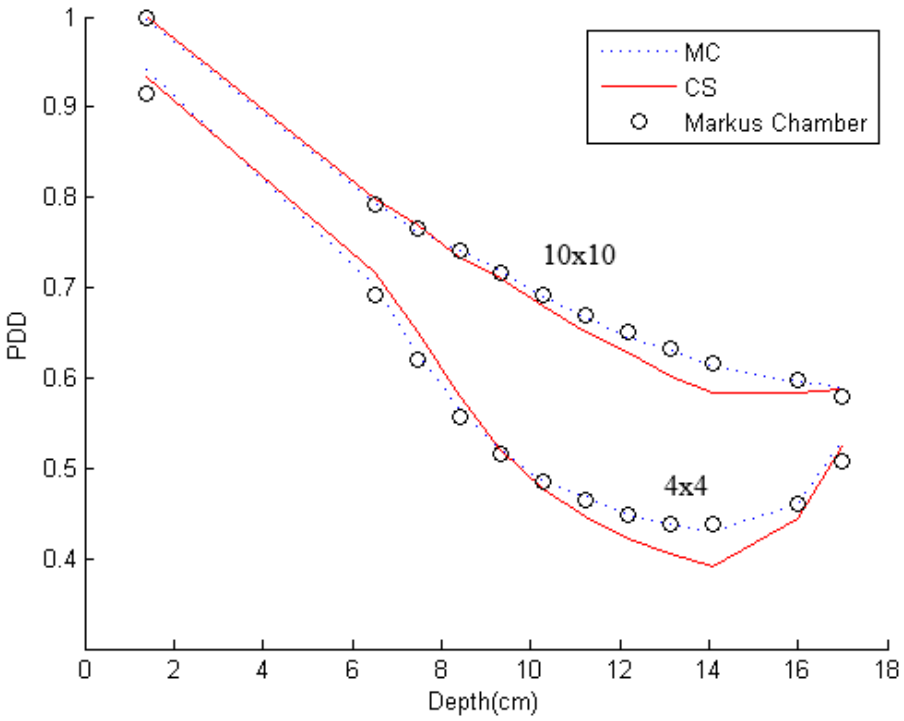

Figure 13. Calculated CS and MC PDDs are calculated and compared to ion chamber measurement in a water-cork-water phantom for 10x10 and 4x4 square fields.

In this section, the dosimetric accuracy of CS and MC was evaluated with respect to measurement for square fields as well as clinical VMAT fields as calculated on both a water tank and solid water phantom. The accuracy of these dose calculation methods in a heterogeneous environment was also assessed via the use of a heterogeneous cork phantom. For square fields in a homogeneous water tank, both CS and MC beam models were both accurate within 1% where these values are systematic difference with respect to measurement plus one standard deviation. For square fields in the cork phantom, CS and MC were accurate within 2.5% and 1.0% respectively. For clinical fields delivered on a solid water stack, CS

and MC were generally accurate within 3% and 2% respectively. It is important to note that these values are all within the goal of having a 3% dose calculation accuracy. This demonstrates that both dose calculation methods are doing a good job. Furthermore, it demonstrates that the beam models are high quality.

3.3.2 LF-MC and LF-CS QA on the Patient Anatomy

Fig. 14 displays histogram plots of variation in PTV Dose. Subfigure 14a illustrates percent variation in calculated dose for the PTV. Variation in dose due to calculation difference (i.e. dose calculation method) is broken down by treatment site and displayed in Table 11. Subfigure 14b displays the same data except for delivery error instead of calculation difference. Dose discrepancy due to dose calculation method was substantial. Percent differences in calculated dose for PTV D_{mean}, D_{99}, D_{95}, and D_1 values were -3.4 ± 1.9%, -4.6 ± 2.8%, -4.5 ± 2.1%, and -1.2 ± 2.8% (negative differences mean MC dose is lower than CS dose). Variation in GTV/CTV dose was generally a relative 5-10% lower than PTV dose variation. In comparison, dose variation due to delivery error was minimal. Dose variation was consistently within 1% for each of the PTV and CTV metrics. Percent differences in dose due to delivery error for PTV D_{mean}, D_{99}, D_{95}, and D_1 were -0.08 ± 0.08%, 0.02 ± 0.2%, 0.01 ± 0.12%, and -0.15 ± 0.17%.

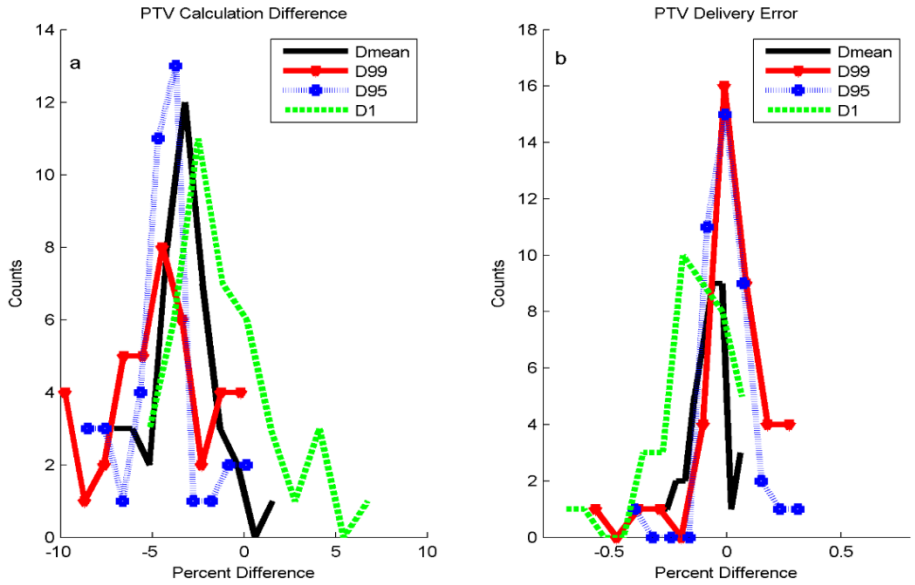

Figure 14. Histogram plots of percent difference in key PTV DVH metrics as a result of a) calculation difference, b) delivery error.

Table 11. Target and OAR dose variation due to calculation difference (MC – CS) is quantified per treatment site. Absolute dose difference is in units of Gy. The number of patients per treatment-site are shown in parentheses.

H&N (11)	PTV D_{mean}	PTV D_{99}	PTV D_1	Parotid D_{mean}	Cord D_1	Stem D_1		
Absolute	-2.0 ± 0.9	-2.7 ± 1.8	0.6 ± 1.7	-1.2 ± 0.8	-1.6 ± 1.1	-1.6 ± 1.5		
Relative	-3.0 ± 1.2	-4.7 ± 2.6	1.1 ± 2.2	-5.6 ± 3.2	-5.8 ± 2.8	-3.8 ± 9.7		
Breast (6)	PTV D_{mean}	PTV D_{99}	PTV D_1	Lung D_{mean}	Heart D_{33}	Eso D_{33}	Cord	
Absolute	-3.2 ± 0.5	-3.3 ± 0.5	-2.3 ± 0.8	-1.0 ± 0.5	0.0 ± 0.2	-0.2 ± 0.1	-0.4 ± 0.6	
Relative	-7.3 ± 1.3	-8.9 ± 1.6	-4.7 ± 0.6	-8.9 ± 4.3	-0.1 ± 6.5	-8.6 ± 2.5	-6.7 ± 1.1	
Lung (6)	PTV D_{mean}	PTV D_{99}	PTV D_1	Lung D_{mean}	Cord D_1	Heart D_{33}	Eso D_{33}	
Absolute	-1.6 ± 0.7	-1.7 ± 1 0.	-1.4 ± 0.6	-0.5 ± 0.6	-0.6 ± 0.5	-0.4 ± 0.5	-1.2 ± 1.0	
Relative	-2.6 ± 1.2	-3.2 ± 1.9	-2.1 ± 0.9	-3.0 ± 2.5	-2.2 ± 1.6	-2.6 ± 3.0	-4.3 ± 1.6	
Eso. (4)	PTV D_{mean}	PTV D_{99}	PTV D_1	Lung D_{mean}	Cord D_1	Heart D_{33}	Liver	Kidney
Absolute	1.7 ± 0.4	-1.4 ± 0.8	-1.5 ± 0.6	-0.2 ± 0.2	-0.8 ± 0.4	-0.6 ± 0.4	-0.1 ± 0.3	-0.1 ± 0.1
Relative	-2.9 ± 1.5	-3.5 ± 1.3	-0.9 ± 0.6	-0.4 ± 0.4	-1.8 ± 1.0	-1.2 ± 1.0	-1.0 ± 1.4	-1.5 ± 1.7
Brain (5)	PTV D_{mean}	PTV D_{99}	PTV D_1	Cochlea D_1	Lens D_1	Nerve D_1	Stem D_1	Cord D_1
Absolute	-0.2 ± 1.3	-0.3 ± 0.8	1.1 ± 2.3	1.3 ± 1.4	-0.2 ± 0.3	-1.3 ± 1.1	-0.3 ± 0.8	-0.6 ± 1.4
Relative	-1.2 ± 3.4	-3.7 ± 6.8	1.5 ± 4.2	9.9 ± 51.5	-3.3 ± 6.8	-5.3 ± 3.5	-1.2 ± 2.2	-2.5 ± 3.6
Media. (3)	PTV D_{mean}	PTV D_{99}	PTV D_1	Lung D_{mean}	Cord D_1	Heart D_{33}	Eso D_{33}	
Absolute	-1.9 ± 0.9	-2.2 ± 0.9	-1.5 ± 0.8	-0.4 ± 0.2	-0.4 ± 0.2	-0.5 ± 0.44	-0.9 ± 0.6	
Relative	-4.8 ± 1.3	-6.4 ± 2.3	-3.5 ± 0.7	-7.4 ± 3.3	-2.6 ± 0.8	-6.8 ± 3.3	-6.2 ± 2.0	
Prostate (2)	PTV D_{mean}	PTV D_{99}	PTV D_1	Rectum D_{15}	Bladder D_{15}	Bulb D_{mean}	FemHead D_1	
Absolute	-2.2 ± 0.2	-1.7 ± 1.7	-2.0 ± 0.2	-1.0 ± 1.5	-1.0 ± 1.3	0.5 ± 0.8	-0.5 ± 0.5	
Relative	-2.5 ± 0.5	-2.4 ± 0.6	-2.6 ± 0.9	-1.7 ± 2.4	-2.3 ± 0.5	1.9 ± 2.7	-1.9 ± 1.9	

Per-plan 3%/2mm gamma pass rates were calculated on the patient anatomy using both 10% (standard) and 85% (target region) dose thresholds. Isolating delivery error, every pass rate was greater than 99%. Isolating calculation difference, the 42 per-plan gamma pass rates

were 87 ± 12% and 55 ± 27% for 10% and 85% thresholds respectively. The distribution of gamma results are shown in Fig. 15. Given the poor agreement in PTV D_{mean} dose between CS and MC in Fig. 14a, these poor gamma passing rates are expected.

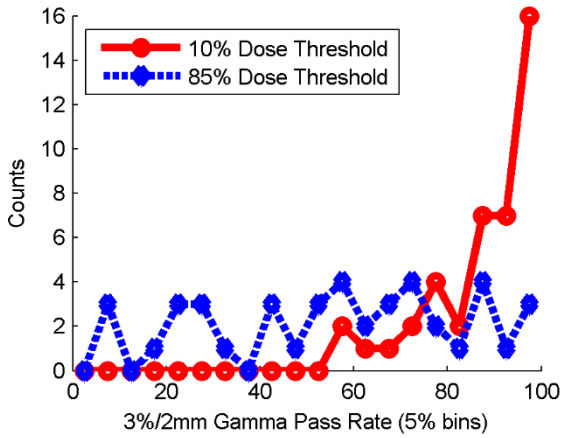

Figure 15. Histogram plot of 3%/2mm gamma pass rates (Plan-MC vs. Plan-CS) using both 10% and 85% dose thresholds.

Clinical per-beam 3%/2mm ArcCHECK gamma passing rates were compared to a variety of per-beam LF-MC QA metrics as calculated on the patient anatomy. These metrics included PTV D_{mean}, PTV D_{99}, PTV D_{95}, and PTV D_1 as well as 3%/2mm gamma calculated using 10% and 85% thresholds. Using a linear fit, R^2 correlations were drawn both when delivery error and calculation differences were isolated. When isolating delivery error (LF-MC vs. Plan-MC), R^2 values were all near zero. AC gamma pass rates had no correlation with delivery-error-only LF QA metrics as calculated on the patient's anatomy. Isolating calculation difference (Plan-MC vs. Plan-CS), R^2 values were small, but non-negligible. PTV D_{mean} yielded an R^2 correlation of 0.24 with respect to AC passing rate. The other PTV metrics had correlations of 0.12-0.14 while the gamma metrics were 0.05-0.08. AC gamma pass rates were largely uncorrelated with both gamma-based and DVH-based LF QA metrics on the patient's anatomy. It is the goal of radiotherapy to delivery accurate dose on the patient's anatomy. This

disconnect between AC QA results on the phantom geometry and LF QA results on the patient geometry suggest that AC QA is simply not capable of detecting clinically relevant dosimetric errors as calculated on the patient anatomy.

Difference in TCP ($TCP_{Plan\text{-}MC}$ - $TCP_{Plan\text{-}CS}$) was on average -3.1 ± 1.8%. Out of 42 plans, 29/18/13/6/3 plans had drops in TCP greater than 2/3/4/5/6%. A histogram plot of the per-plan decrease in TCP as measured using LF-MC QA is displayed in Fig. 16. The significance of this drop in TCP is discussed later, however, suffice it to say for now that these are fairly significant drops in TCP.

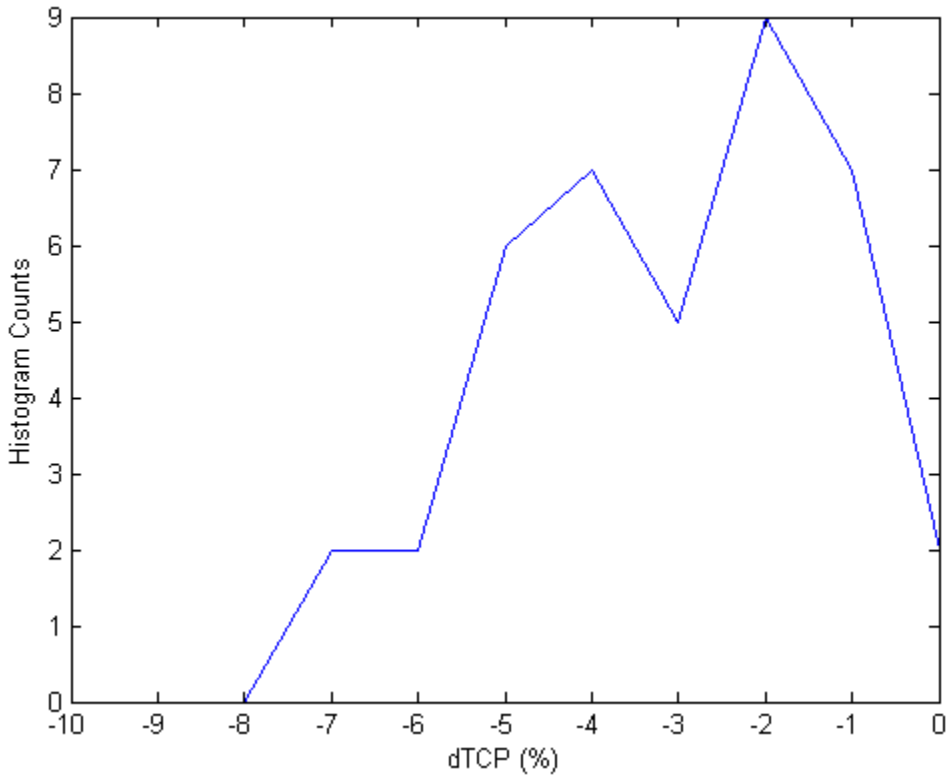

Figure 16. A histogram plot displaying reduction in TCP (Plan-MC vs. Plan-CS).

Gantry discretization difference, illustrated in Fig. 17, was assessed for LF-CS QA. Subfigure 17a displays the effects of gantry discretization difference (LF-CS(1°) vs. Plan-

CS(4°)) on various PTV DVH metrics. Subfigure 17b displays the same values except for reduced gantry discretization difference (LF-CS(1°) vs. Plan-CS(2°)). Dose variations for LF-CS(1°) vs Plan-CS(4°) were 0.45 ± 1.95%, 0.14 ± 0.96%, and 2.63 ± 4.07% for PTV D_{mean}, D_{99}, and D_1 respectively. Dose variations for LF-CS(1°) vs Plan-CS(2°) were 0.0 ± 1.99%, 0.20 ± 1.16%, and 0.67 ± 1.12%. Utilizing Plan-CS(2°) instead of Plan-CS(4°) resulted in reduced dose differences. Mean PTV D_{mean} dose difference decreased from 0.45% to 0.0%. Mean PTV D_1 dose difference decreased from 2.63% to 0.67%. These results show that the dosimetric difference due to gantry discretization is typically on the order of 0-2%, however PTV D_1 differences as large as 4% are possible for more severe cases. Although a 0.5% difference in PTV D_{mean} is not hugely significant, this dose difference is large with respect to dose difference due to delivery error. Thus, it is the recommendation of this study that Plan-CS(4°) be interpolated to Plan-CS(2°) when carrying out LF-CS QA.

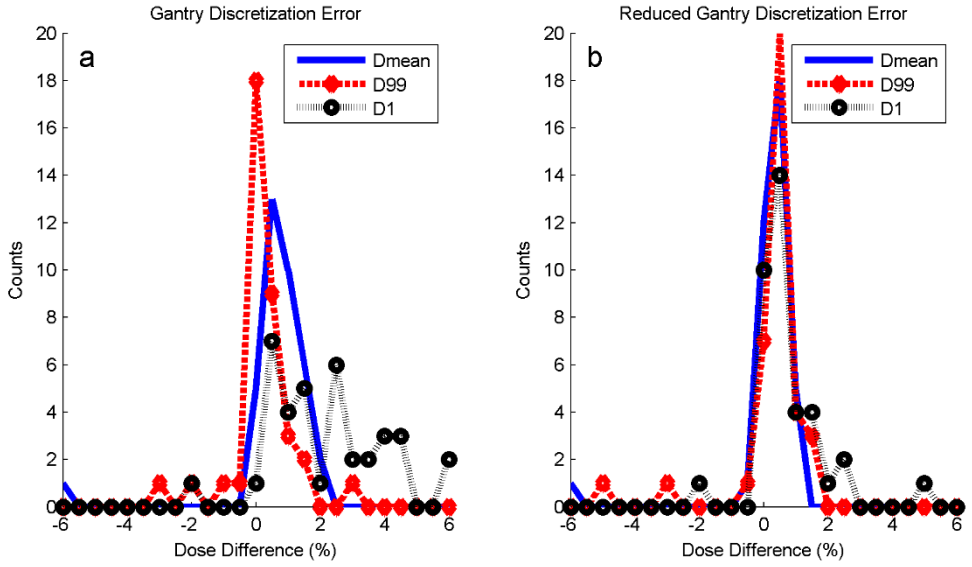

Figure 17. Histogram plots of percent dose difference in key PTV DVH metrics as a result of a) gantry discretization difference and b) reduced gantry discretization difference.

3.3.3 Dose Discrepancy (MC vs CS)

In section 3.1, both dose calculation algorithms were shown to be accurate within 3%, however, when dose is calculated and compared on the patient anatomy in section 3.2, significant dosimetric differences greater than this 3% were seen. Thus, the cause of this difference was investigated within the scopes of beam complexity and patient heterogeneity.

3.3.3.1 Beam Complexity

Variation in PTV D_{mean}, PTV D_{99}, GTV D_{mean}, and GTV D_{99} with respect to MU/Rx yielded R^2 correlations of 0.35-0.41. R^2 correlation for the external ROI was 0.67. These correlations suggest a significant portion of LF-MC QA dose discrepancy as calculated on the patient anatomy is due to beam complexity. GTV D_{99}, the target ROI with the greatest correlation value, is plotted against MU/Rx in Fig. 18. As displayed in Table 12, variation in GTV D_{99} was investigated for simple (< 3.5 MU/Rx) plans and complex (> 3.5 MU/Rx) plans. For simple plans, ΔGTV D_{99} was -3.89 ± 1.92%. Dosimetric variation was greater for the complicated plans; ΔGTV D_{99} was -6.64 ± 1.97%.

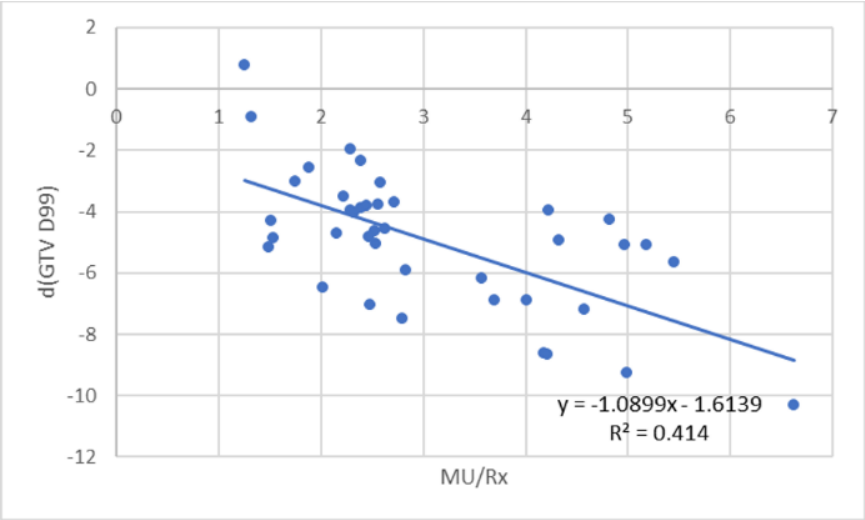

Figure 18. Per-plan Percent variation in GTV D99 (MC – CS) is plotted as a function of the beam complexity MU/Rx.

Table 12. Dose discrepancy is evaluated for simple (MU / Rx < 3.5) and complicated (MU / Rx > 3.5) patient cases.

Metric (MC – CS)	27 Simple Cases	15 Complex Cases
ΔGTV D99	-3.9 ± 1.9	-6.6 ± 2.0
ΔPTV D99	-3.7 ± 2.2%	-6.4 ± 2.6%
ΔTCP	-2.7 ± 1.6%	-3.9 ± 1.9

Variation in TCP with respect to MU/Rx yielded an R^2 correlation of 0.15. For the 27 simple plans, ΔTCP was -2.57 ± 1.70%. For the 15 complicated plans, ΔTCP was even greater at -3.94 ± 1.89%. The low correlation between TCP and MU/Rx is to be expected. Above, GTV D_{99} was shown to have a 0.42 R^2 correlation with MU/Rx. All other DVH metrics had lower R^2 values. It follows that TCP, which is one step further removed from MU/Rx, will thus have an R^2 values lower than 0.4. In addition, we know that ΔTCP is proportional to Δ(dose difference) multiplied by γ_{50} in the central non-shoulder part of the dose response curve. However, the 42 cases studied all have varying values of γ_{50}. This will result in even further reduced values of correlation between TCP and MU/Rx.

3.3.3.2 Patient Heterogeneity

Displayed in Fig. 19, per-plan percent difference in PTV D_{mean} was calculated for both the heterogeneous patient anatomy, as well as when the anatomy was fully overridden to unit density. Heterogeneity was evaluated per-site for two treatment sites with greater heterogeneity - lung and H&N. Five lung and five H&N plans having a range of dose calculation agreement were selected. For the five H&N patients, MC – CS agreement was actually poorer on the homogeneous anatomy. Dose agreement decreased by 10 ± 23% when overriding the patient to water. This indicates heterogeneity effects were largely insignificant for the H&N cases. For the five lung patients, MC – CS agreement was better for four of them on the homogeneous anatomy. On average the dose variation was 18 ± 21%. This indicates ~20% of the lung dose

variation due to calculation differences was due to heterogeneity effects. CS dose variation with respect to MC was mostly seen at the outer surface of higher density tumor voxels proximal to low density lung tissue. This over estimation of target dose by CS is likely due to poor estimations when it comes to loss of lateral scatter equilibrium in low density media.[16]

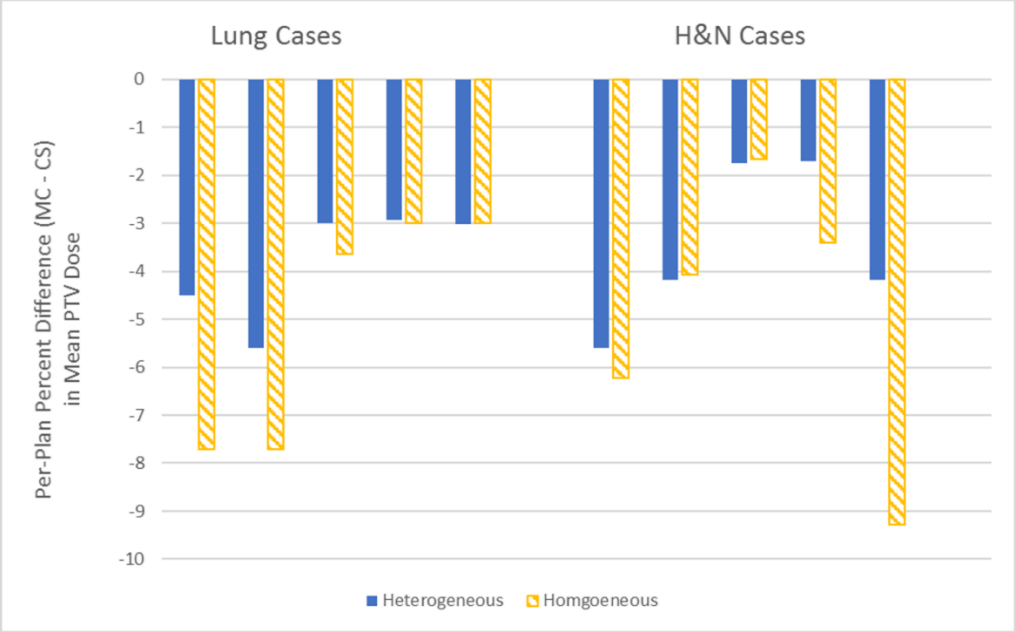

Figure 19. Per-plan percent difference in PTV D_{mean} due to dose calculation difference is displayed.

3.4. Discussion

LF-MC QA yielded PTV D_{mean} and PTV D_{99} dose discrepancies greater than 5% for 14% and 40% of patients respectively. The clinical impact of this discrepancy is best evaluated within the scope of TCP as no other metric can as succinctly quantify the effectiveness of a treatment plan. Decrease in TCP, calculated using GTV, was on average -3.1 ± 1.8%. Out of 42 plans, 29/18/13/6/3 plans had drops in TCP greater than 2/3/4/5/6%. The effect of these drops in TCP can best be evaluated in terms of recurrence rates. By definition, TCP quantifies the probability that 0 clonogenic cells survive the radiotherapy treatment. However, it is not necessary to kill every clonogenic cell. In reality, only enough cells need to be killed that the

re-development of a symptomatic tumor is prevented for the remaining duration of the patient's life. This relapse time, or time at which a symptomatic tumor is expected to return, can be calculated via Eq. 13, where G(t) is the probability of recurrence at time t, P_m is the probability

$$\boldsymbol{G(t) = \sum_{m=0}^{\infty} P_m [G_1(t)]^m} \quad \textbf{(14)}$$

that m number of clonogenic cells remain after radiotherapy, and $G_1(t)$ is the probability that a clonogenic cell does not become a symptomatic tumor at time t.[15] For example, given a decrease in TCP from $TCP_{Plan\text{-}CS}$ to $TCP_{LF\text{-}MC}$, a new probability distribution function detailing the probability of m number of clonogenic cells surviving can be calculated. Subsequently a new, higher, probability of recurrence at time t can be calculated. Following this train of thought, allowable decrease in TCP is really a patient-specific problem.

Despite substantial dose disagreement as determined using LF-MC QA on the patient anatomy, ArcCHECK QA results were consistently passing. LF-MC QA dose discrepancy stemmed primarily from dose calculation method. Values of PTV D_{mean} and PTV D_{99} were -3.4 ± 1.9% and -4.6 ± 2.8% lower for MC than CS for the 42 patient cases. 77/7/1 fields had pass rates of 95-100%/90-95%/87%. Furthermore, DVH-based LF-QA results were largely uncorrelated with clinical AC gamma pass rates. It follows that the coefficients of determination between the LF-MC QA results and gamma results were quite low. PTV D_{mean} displayed the greatest correlation with clinical 3%/2mm gamma passing rates with an R^2 of 0.24. All other LF QA metrics had R^2 values less than 0.14. Discrepancy in LF-MC QA and AC QA results can be attributed to several factors. Foremost is the mismatch in geometries utilized. A homogeneous, symmetric cylindrical phantom is not representative of the patient's anatomy. The phantom will not be able to account for patient-specific heterogeneities. In addition, when delivering IMRT planned doses which have been optimized on the patient's

anatomy onto a different geometry (e.g. ArcCHECK), the uniform target dose is blurred across a greater region. Simultaneously, the dose difference (e.g. MC vs. CS) is spread out over a greater volume, resulting in lower local percent variations in dose. This percent dose variation is further decreased when utilizing a global dose criteria (e.g. Van Dyke). Lower dose variance will yield higher gamma pass rates. Lastly, the diode detectors are placed at a fixed depth that likely does not correspond to the target location. If follows that the diode measurements are not as applicable as they maybe ought to be.

The failing AC QA result was addressed. The case was a four arc VMAT H&N case. Of the four arcs, one yielded an ArcCHECK pass rate (Plan-CS vs. AC) of 87.2%. Comparing Plan-MC to AC, a 3%/2mm gamma pass rate of 92.3%. Upon further investigation it was determined that this failure resulted from delivery error. By reconstructing LF-CS dose on the ArcCHECK geometry, an LF-CS vs. AC passing rate of 100% was calculated. Similarly, LF-MC vs. AC also yielded a 100% passing rate on the ArcCHECK geometry. Digging into the log file, this delivery error was determined to result from the non-linear delivery of MU/degree as well as gantry error up to 1.4 degrees. Importantly, this non-linearity in MU/degree coincided at the same gantry angle as where the gantry error occurred. When dose is calculated, MU output is assumed to be linearly delivered between control points, however, this is not always the case. The ArcCHECK failure pattern which was caused by this delivery error is circled in Fig. 20. Fig 21 and Fig. 22 display the non-linear MU/degree output and gantry error respectively. Despite this delivery error and low 3%/2mm gamma passing rate, dosimetric discrepancy due to delivery error on the patient anatomy was minimal; variations in PTV D_{mean} and PTV D_{99} were -0.7 and -2.6% respectively. This demonstrates that delivery errors as

measured on the ArcCHECK phantom do not necessarily translate into DVH-based errors on the patient anatomy.

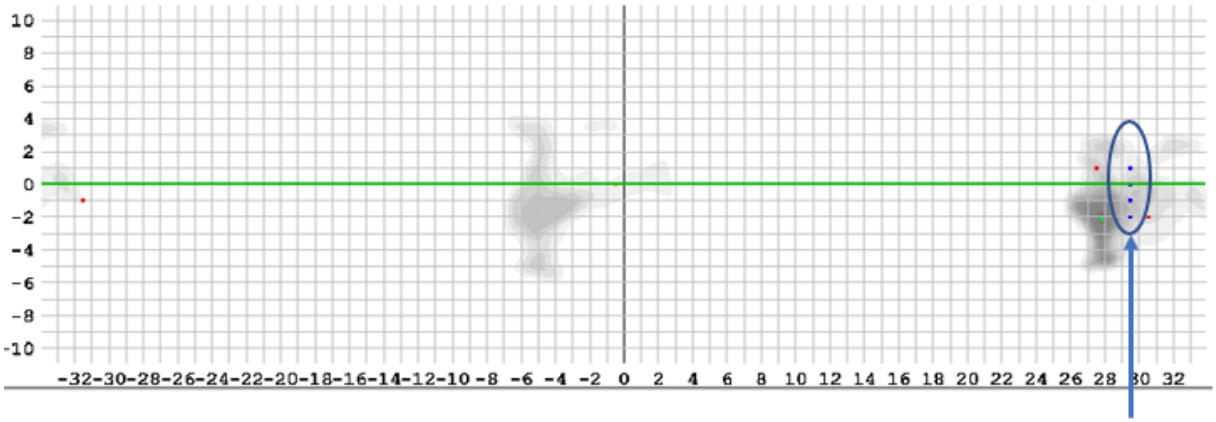

Figure 20. An ArcCHECK failure pattern for Plan-MC vs. AC resulting from gantry error and non-linear MU/° is shown.

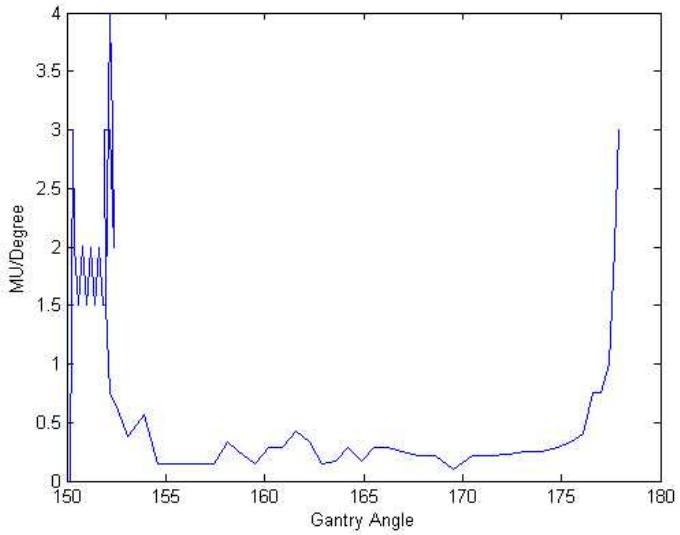

Figure 21. MU/° is plotted as a function of gantry angle.

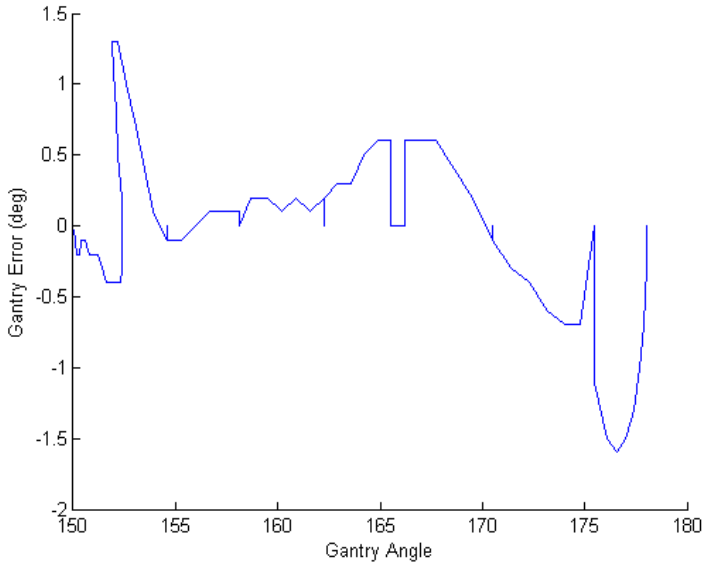

Figure 22. Gantry error is plotted as a function of gantry angle.

Dose discrepancy due to delivery error was extremely small for all cases studied. Delivery error peaked at 0.8% of the fractional prescribed dose and was typically within a few tenths of a percent. In comparison, despite excellent square-field water tank validation results, dose discrepancy due to dose calculation method was large. This disagreement was investigated within the scopes of tissue heterogeneity and beam complexity. The effect of tissue heterogeneity was greatest for lung tumors; 20% of dose discrepancy was explainable by heterogeneity effects. However, heterogeneity did not play a significant role for other treatment sites. In comparison, 40% of dose discrepancy was explainable by beam complexity over all treatment cases. Table 12 shows that dose discrepancies were much greater for the complicated (MU / Rx > 3.5) treatment cases. These results were investigated within the scope of treatment planning QA – "Given these results, is there some way this dose difference can be mitigated?". CS and MC inherently model tissue heterogeneities differently. CS is not designed to accurately account for inhomogeneities, thus little can be done to mitigate these calculation

differences. However, when it comes to beam complexity, action can be taken to reduce beam complexity. The most obvious method is to increase the minimum leaf pair gap that can utilized for VMAT treatment. This will result in smaller MU/Rx values, and thus should result in greater CS dose calculation accuracy. At out clinic a minimum leaf gap spacing of 6mm is used for VMAT treatment planning. However, several studies have recommended that minimum leaf gap spacings should be set to ≥2cm to minimize variation in calculated and measured dose.[102-104] Difference in MC and CS IMRT/VMAT dose is largely attributed to this difference.

Monte Carlo beam models may not always be available for every institution, machine, and/or energy. Thus, LF-CS QA was also evaluated. Specifically, CS gantry discretization difference was assessed. Assuming LF-MC vs Plan-MC to be the gold standard for estimating delivery error, dose variations for PTV D_{mean}, D_{99}, and D_1 were a very small 0.11 ± 0.10%, 0.12 ± 0.11%, and 0.20 ± 0.18%. However, comparing LF-CS to Plan-CS, dose variation could be large due to gantry angle discretization. Comparing Plan-CS(4°) to LF-CS(1°), PTV D_{mean} and D_1 dose variations were 0.5 ± 2.0% and 2.6 ± 4.1%. Linearly interpolating Plan-CS(4°) to Plan-CS(2°), dose variation with respect to LF-CS decreased to 0.0 ± 2.0% and 0.7 ± 1.1% respectively. Using 2 degree control point spacing, LF-CS QA is capable of verifying delivery error within fairly tight bounds; Plan-CS PTV metrics were generally accurate within 1% of LF-CS(1°) and almost always accurate within 2%.

MC agreement with measurement was superior to that of CS. In-field and out-of-field, MC confidence limits were 2.4% and 4.2% respectively. In comparison, CS confidence limits were 6.3% and 6.1%. Across all 10 institutions in TG-119, confidence limits were 4.5% in-field and 4.7% out-of-field.[49] However, whereas much of the data acquired in TG-119 is for IMRT cases and simpler treatment sites, the VMAT cases utilized in this TG-119 study were relatively more

complicated. This increased complexity is hypothesized to result in the higher CS confidence limit. However, given the ±5% dose goal, the 6% CS confidence limits are somewhat unacceptable. In comparison, the MC confidence limits, especially in-field, were very good.

Three methods are introduced for addressing the aforementioned decreases in dose and tumor control. First, it may be preferable, once the computational cost of Monte Carlo is sufficiently low, to simply utilize a TPS that utilizes MC in the plan optimization process. Second, the planned number of MU could be simply scaled by a scaling factor, SF = CS PTV D_{99} / MC PTV D_{99}. Third, the direct machine parameter optimization (DMPO) process could be constrained. For example, many of the worst cases in this study were highly complicated H&N or breast cases. An investigation into aperture shape revealed many segments with thin (< 1cm) snake-like apertures. Increasing the minimum dynamic leaf pair gap in the Pinnacle TPS should yield simpler beam segments and thus smaller ΔTCP values.

LF-MC QA is intended to fully replace conventional phantom-based QA systems. As such, new protocols are required to determine whether the results of LF-MC QA on the patient anatomy should be considered passing or failing. Initially, a $|\Delta TCP\ D_{99}| < 4\%$ action tolerance was investigated. This tolerance was selected given that dose calculation methods are typically expected to be accurate within 3-4%.[19] However, this resulted in a 41% failure rate. Given clinical workload constraints, the number of failing cases should not be more than 10%. Investigating various criteria, an action tolerance of $|\Delta PTV\ D_{99}| < 5\%$ and $|\Delta TCP| < 4\%$ was decided upon. Using this criteria, 8/42 cases failed. However, of these eight cases six were complicated fields (MU/Rx > 3.5). Given the dependence of dose discrepancy on beam complexity, it is the recommendation of this study that physicists/dosimetrists attempt to limit MU/Rx < 3.5 if using CS or an equivalent algorithm. With a total of 27 simple cases, the

remaining 2 failures represents a 7% failure rate. For these failures it is the recommendation that the prescription dose to the patient be multiplied by the aforementioned scaling factor, SF, and that the plan be re-optimized.

In general, OAR dose discrepancies were less significant than target dose discrepancies. This was largely due to MC calculation being systematically lower than CS. As negative OAR discrepancies will result in lower normal tissue complication probability (NTCP) values, negative discrepancies were not assessed. Variation in OAR dose due to calculation difference was assessed using 1Gy/3%, 2Gy/3%, and 3Gy/3% passing criteria. Over the 42 plans analyzed in this study, roughly 1000 dose-volume constraints were QAed. Of these 1000 constraints, 19 failed using a 1Gy/3% criteria. Of these failures, 12 failures were for Cochlea D1%; mean variation was 3.9 ± 1.8Gy. The remaining failures included brain stem, oral cavity, esophagus, bladder, penile bulb, brachial plexus, and mandible ROIs receiving 1.04, 1.21, 1.18, 1.05, 1.07, 2.18, and 3.83Gy more dose via MC than CS. Variation in OAR dose due to delivery error was assessed using a 0.5Gy/1% criteria. Of the 1000 dose-volume constraints, there was one failure; a cochlea D_1 varied by 0.7Gy/5%. Although a couple of these dose discrepancies were moderately large, the previously analyzed variations in target dose were more significant. Thus, given workload constraints, none of these variations were considered significant enough to label LF-MC QA as failing.

Although MC and CS calculation typically matched measurement within 2% and 3% respectively in a homogeneous media, calculation matched measurement poorly for a subset of the TG-119 & clinical VMAT cases. This trend is similar to what was shown by Bouchard et al.[98] Out of 1600 IMRT measurements in a phantom, 1300 agreed within 2%. Calculation disagreement ranged from -12.7% to 11.7% for the other 300 cases. It is important to note these

measurements were taken using the same size ion chamber as utilized herein. However, for this size ion chamber, correction factors for dynamic IMRT fields can be greater than 10%.[59] This suggests that any disagreement between calculation and measurement for IMRT/VMAT fields in this study could very well be due to measurement inaccuracy.

Comparing Plan-MC to Plan-CS, dose discrepancy was assessed as a function of heterogeneity, beam complexity, and gantry discretization. PTV D_{mean} dose discrepancy stemmed ~10% from heterogeneity, ~40% from beam complexity (MU/Rx), ~10% from gantry discretization, and ~40% from unknown sources. However, it is important to understand that these values are approximations. First, MU/Rx, although a simple and apt metric for indicating beam complexity, is simply an approximation of beam complexity. Second, overriding the patient anatomy to water, although a good experiment for assessing the effect of heterogeneity, is not simple to quantify. It follows that some additional fraction of the unexplained variance is likely due to beam complexity and/or heterogeneity.

3.5 Conclusion

Delivery errors, as tracked from linac log files, as well as calculation differences (ScientificRT's SciMoCa Monte Carlo vs. Pinnacle's Convolution-Superposition) were assessed on patient anatomies. Differences in DVH metrics due to delivery error were typically on the order of a couple tenths of a percent, indicating delivery errors for the Elekta Infinity linac equipped with an Agility treatment head were very minimal. Dosimetric difference due to gantry angle discretization error was moderate. When using LF-CS QA, Plan-CS should be interpolated to $\leq 2°$ control point spacing. Whereas phantom-based AC QA was consistently passing, LF-MC QA was able to catch clinically worrisome dose discrepancies as calculated on the patient's anatomy. Percent differences in PTV D_{mean}, PTV D_{99}, and TCP as large as 7%,

10%, and 7% on the patient anatomy. Reductions in TCP as large as 7% were also calculated. Dose discrepancy was primarily due to beam complexity. For instance, when using an action tolerance of [$|\Delta GTV\ D_{99}| < 5\%$ and $|\Delta TCP| < 4\%$] for patient anatomy-based planning QA, 7% of simple (MU/Rx < 3.5) cases and 53% of complicated (MU/Rx > 3.5) cases failed. Thus, it is recommended that treatment planning beam complexity be constrained within MU/Rx < 3.5 if dose is calculated using CS of an equivalent algorithm. It is concluded that, given proper machine QA and Monte Carlo commissioning, anatomical LF-MC QA is a fully automatable technique with the potential to fully replace phantom-based QA techniques.

CHAPTER 4 LF QA SOFTWARE

LF QA software was developed alongside the aforementioned research. Fig. 23 illustrates the general architecture of LF QA. There is a central LF QA server that is in charge of LF QA. The server receives log files from a linac or linacs. It queries Mosaiq to match a log file to the correct patient. It is connected to a linux server dedicated for the management and calculation of Monte Carlo doses. Lastly, it's setup so that one or multiple physicists' computers can easily initiate LF QA and analyze the results in Pinnacle. The LF QA software consists of three main components: (1) a system for automatically archiving log files in a manner where log files can easily be acquired when required; (2) LF-CS QA; and (3) LF-MC QA. Fig. 23 illustrates the key systems of LF QA.

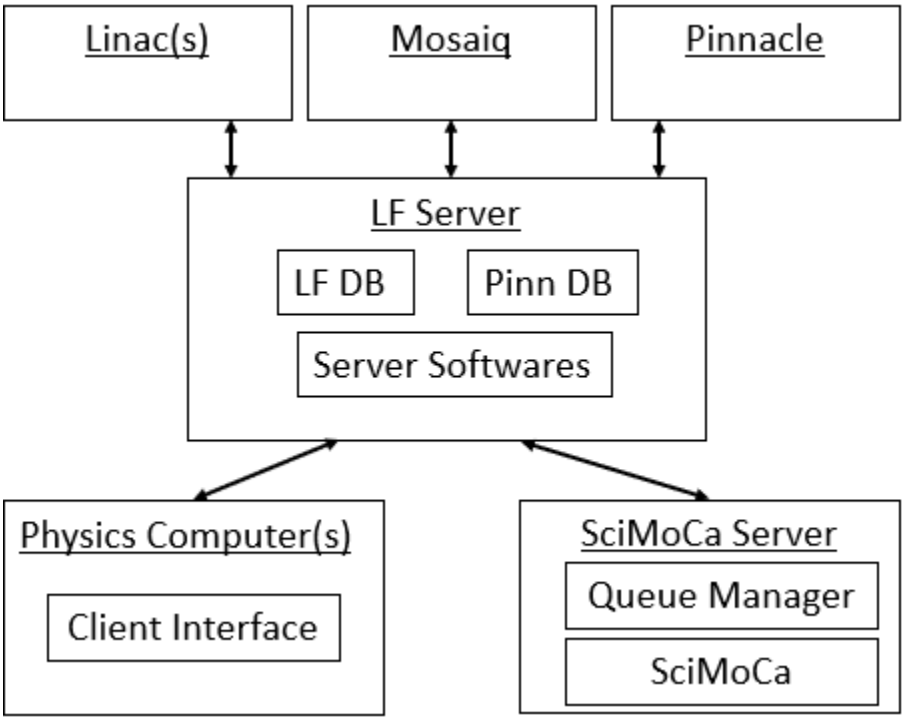

Figure 23. Log File QA Workflow. A central log file server acquires the log files from the linacs, matches the log files to a patient's plan using Mosaiq, calculates dose using SciMoCa, and pushes the resultant dose to Pinnacle for evaluation.

4.1 Log File Acquisition & Storage

First, acquisition of log files is discussed. Ideally, log files should be pushed to a log file server immediately upon the delivery of the field. This would allow the log file to be processed and log file dose to be calculated as soon as possible. Unfortunately, this method is not provided by Elekta. Thus, Elekta log files were acquired using the following two methods. First, log files can be manually exported by the physicist by putting the linac into service mode. This process takes roughly 10 minutes. The exported file is a zip file which contains per-field log files in binary .trf format. This export includes log files from the last 8 days. The benefit of this method is log files can be acquired and pushed to the log file server for processing in 10 minutes. The disadvantage of this method is that it requires the physicist to manually export log files. The second method used to acquire log files is routine and automatic, but not quick. The linac hard disk (specifically, the D: drive) is backed up to a set location on a weekly basis. Both manually exported and weekly automatically exported log files are sent to the same location. The log file server is setup to monitor this location and process files when they are acquired. For the weekly linac snap shots, the disk image is programmatically mounted as a virtual share drive. The log files (.trf) are then acquired from the share drive and the disk image is unmounted.

Binary trf files are then converted into ASCII .csv files. Unfortunately, the log file contains limited information as it pertains to patient and/or beam identification. However, we do know when the beam was delivered and on what machine. By querying Mosaiq this is enough information to determine what patient this log file is for, as well as which treatment plan and which treatment field, etc. Log files, both trf and csv, are then stored in a directory of

the following format: Database / MachineName / PatientID / TreatmentName / DeliveryDate / FieldID. A PostgreSQL database is setup to store information about all relevant log files.

4.2. LF-CS QA

LF-CS QA is carried out on the patient's anatomy using the methods described in previous chapters. The csv log file is converted into a Pinnacle script that in turn creates a plan in Pinnacle. From here, dose is calculated and analyzed in Pinnacle. This is a semi-automatic process. From the LF QA software, the physicist selects which fields they want to reconstruct log file dose for. This spawns a program which converts the log files of interest into two pinnacle scripts which are sent to the Pinnacle server. The physicist then opens Pinnacle, runs the first script (initialize beams and control points), runs the second script (sets beams to match the log file samples), and calculates dose. Pinnacle scripting is not particularly quick. Converting 358 log file samples into 358 control points for a 358° VMAT arc (1° control point spacing) can take close to an hour. Dosimetric analysis of the reconstructed log file(s) in Pinnacle is left to the discretion of the physicist.

4.3. LF-MC QA

LF-MC QA is initiated in the same way as LF-CS QA. The physicist selects which fields they want to reconstruct LF-MC doses for and clicks 'Run LF-MC QA' in the GUI. This spawns a program on the LF QA server. This program first queries a PostgreSQL database that contains path locations for each beam / trial in Pinnacle. The query returns a full path location on the Pinnacle server. From this location, a variety of Pinnacle files are acquired and converted into DICOM format. Monte Carlo calculation is carried out using these DICOM files (CTs, RTSTRUCT, RTPLAN). This process then converts the MC doses (Plan-MC, LF-MC) into binary Pinnacle format and creates a Pinnacle script to load the doses into Pinnacle. In total,

this process typically takes ~10 minutes. SciMoCa dose calculation usually takes ~2 minutes per VMAT arc using a 3mm dose grid and 1% statistical uncertainty. Fig. 24 shows the client GUI that physicists can use to automatically carry out Plan-MC and LF-MC doses as part of LF-MC QA. Fig. 25 shows a sub-GUI that is utilized to export doses to Pinnacle for further evaluation.

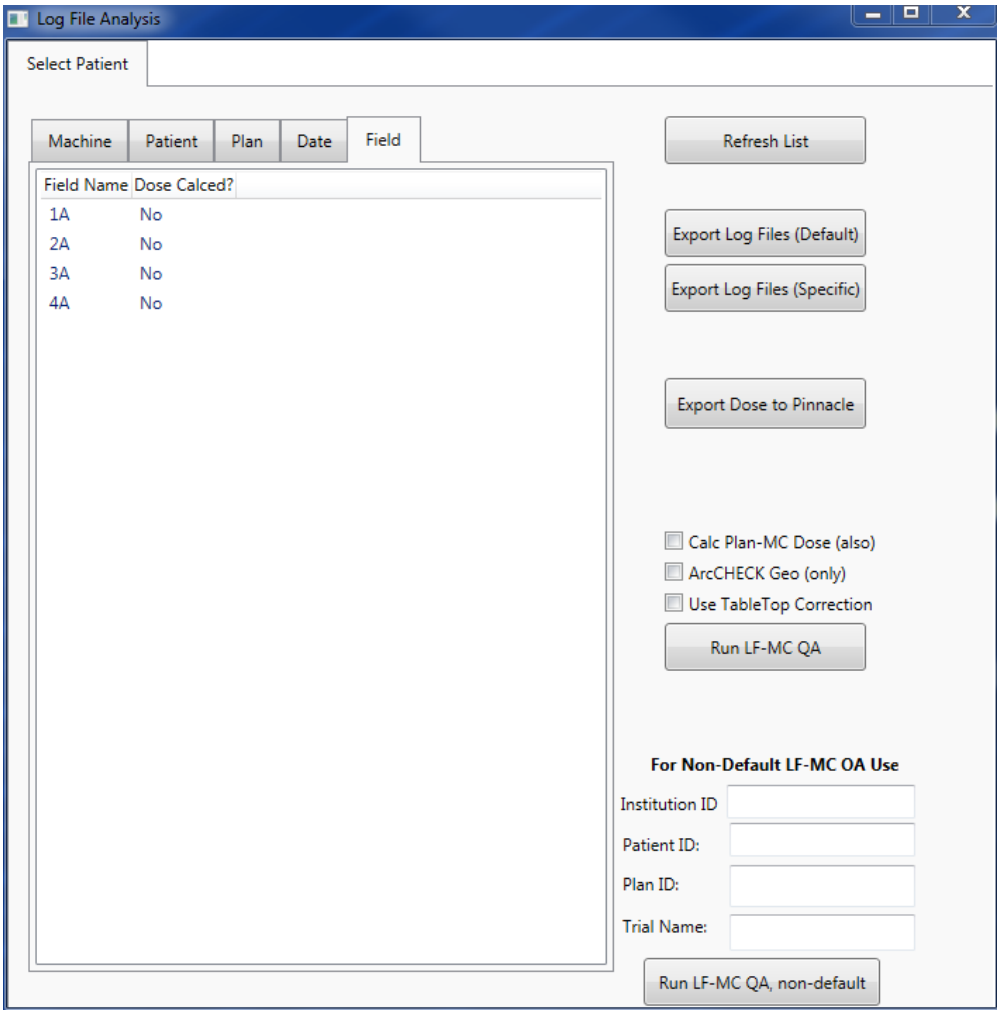

Figure 24. LF QA GUI for calculating Plan-MC and LF-MC doses. Doses can also be exported to a separate tool which is utilized for LF-CS QA.

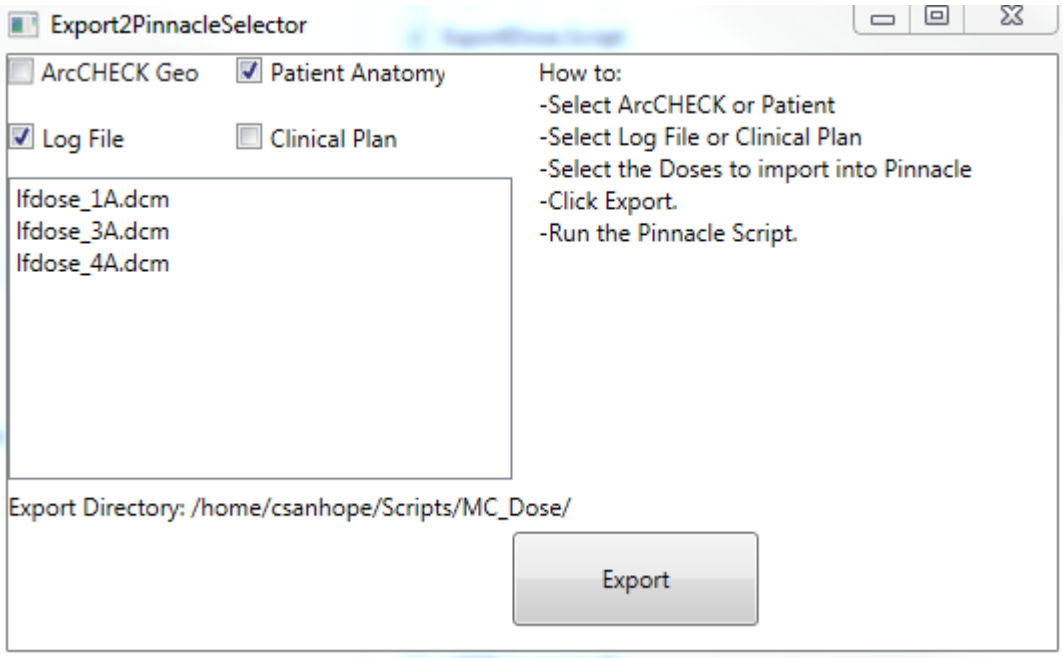

Figure 25. GUI for selecting doses to export to Pinnacle for evaluation.

CHAPTER 5 CONCLUSION

Accurately reconstructing dose on each patient's unique anatomy, LF-MC QA has shown great potential for patient-specific QA. Using Monte Carlo, dose calculation accuracy was ~1.2% (1σ) and generally within 2% for complex VMAT fields delivered on a homogeneous solid water stack. In addition, SciMoCa was shown to be able to accurately calculate dose in a heterogeneous cork phantom within 2%. In comparison, this study has shown conventional phantom-based QA to be generally lacking. In Chapter 2, diode sensitivity dependence on dose rate, beam complexity, and energy was investigated. Diode sensitivity varied up to 1.5% in-field with dose rate and up to 5% out-of-field with energy. Phantom-dependent uncertainties were on average 0.8 ± 1.2%. In Chapter 3, ArcCHECK gamma passing rates were consistently passing even though DVH differences ≥5% were calculated using LF-MC QA on the patient anatomy. In follows that AC pass rates were uncorrelated with both DVH-based and gamma-based LF-MC QA metrics as calculated on the patient anatomy ($R^2 < 0.24$). Finally, even when the ArcCHECK phantom registered a poor passing rate due to the combined effects of gantry error and non-linear MU, these delivery errors negligibly impacted anatomical DVHs. It is therefore the conclusion of this study that ArcCHECK, and perhaps phantom-based QA systems in general, is not an ideal method of patient-specific treatment plan QA. In comparison, LF-MC QA was able to not only accurately determine the delivered dose distribution on the patient, but also accurately determine the cause of potential dose differences. In addition, this QA technique also provide a possibility for online or real-time planning modification/treatment.

Despite the significant advantages of LF-MC QA over AC QA, there exist strict requirements that LF-MC QA must abide by. The two primary requirements of LF QA are that it relies on reliable machine QA and an accurate dose calculation method. The 6MV SciMoCa

Agility beam model utilized as part of this study was shown to accurately calculate dose within 2-3% for both heterogeneous anatomy and complicated MLC segments. However, it is essential that each institution thoroughly vet each of their MC beam models. LF-MC QA also relies on proper machine QA. Variations in output and beam tuning are not monitored via log files and thus must be sufficiently monitored via machine QA. In addition, MLC calibration must be maintained. However, these disadvantages are addressable – strictly maintain machine QA and/or expand upon current machine QA protocols. In conclusion, although LF-MC QA is not flawless, it is a superior form of patient-specific QA to AC QA.

LF QA on the phantom geometry yielded equivalent results as AC QA. It follows that LF QA on the phantom geometry can fully replace AC QA. Furthermore, LF QA on the phantom geometry allows for root cause of analysis of any potential delivery errors. However, given concerns with phantom-based QA in general, neither ArcCHECK QA nor LF QA as calculated on the phantom geometry are recommended forms of QA.

A benefit of this study and the benefit of LF-MC QA is that it is immediately clinically impactful. This study was immediately clinically impactful at our institution as well. This study showed us that machine delivery error is very small and generally clinically negligible for our Elekta Agility machine. Instead dose calculation was shown to be extremely important. Investigating dose calculation further, it was found that complexity was the primary concern of dose calculation accuracy at our institution. As a result, we changed our minimum leaf gap in our treatment planning system to be 1.5cm instead of 0.6cm. We also notified our dosimetrists that plans should ideally be within 3.5 MU / Rx. The greatest clinical takeaway from this study is that ArcCHECK QA and perhaps phantom-based QA is largely irrelevant. ArcCHECK QA was incapable of catching clinically relevant dosimetric

errors as determined via LF-MC QA on the patient anatomy. Thus, despite being the standard of care for decades, ArcCHECK absolutely should not be the primary form of patient-specific QA in clinics. In comparison, Elekta LF-MC QA has no real weaknesses. LF QA relies on accurate dose calculation, but Monte Carlo has been shown to be very accurate given proper beam modeling. LF QA also relies on accurate machine QA, but it is not the responsibility of patient-specific QA to reconfirm accurate machine QA.

REFERENCES

1. PE Gill, W Murray, MA Saunders et al. User's Guide for NPSOL (Version 4.0): A Fortran package for nonlinear programming. Jan 1986; 60.

2. PE Gill, W Murray, MH Wright. Practical Optimization. New York: Academic Press. 1981.

3. DG Luenberger. Linear and Nonlinear Programming. Menlo Park: Addison-Wesley. 1989.

4. IJ Chetty, B Curran, JE Cygler et al. Report of the AAPM Task Group No. 105: Issues associated with clinical implementation of Monte Carlo-based photon and electron external beam treatment planning. *Med. Phys.* Dec. 2007;34(12):4818-53.

5. Kawrakow I, Fippel M, Friedrich K. 3D electron dose calculation using a voxel based Monte Carlo algorithm (VMC). *Med Phys.* 1996;23;445-457.

6. Fippel M. Fast Monte Carlo dose calculation for photon beams based on the VMC electron algorithm. *Med Phys.* 1999;26:1466-75.

7. Kawrawkow I, Fippel M. Investigation of variance reduction techniques for Monte Carlo photon dose calculation using XVMC. *Phys Med Biol.* 2000;45:2163-2183.

8. Sikora M, Dohm O, Alber M. A virtual photon source model of an Elekta linear accelerator with integrated mini MLC for Monte Carlo based IMRT dose calculation. *Phys Med Biol.* 2007;52:4449-4463.

9. Sikora M, Alber M. A virtual source model of electron contamination of a therapeutic photon beam. *Phys Med Biol.* 2009;54:7329-7344.

10. Sikora M. Virtual source modelling of photon beams for Monte Carlo based radiation therapy treatment planning. PhD thesis. http://bora.uib.no/handle/1956/5135?show=full.

11. Mackie TR, Scrimger JW. COMPUTING RADIATION DOSE FOR HIGH ENERGY X-RAYS USING A CONVOLUTION METHOD . 36-40.

12. Mackie TR, Scrimger JW, Battista JJ. A convolution method of calculating dose for 15-MV x rays. *Medical Physics*. 12: 188-96. PMID 4000075DOI: 10.1118/1.595774

13. Mackie TR, Bielajew AF, Rogers DW, Battista JJ. Generation of photon energy deposition kernels using the EGS Monte Carlo code. *Physics in Medicine and Biology*. 33: 1-20. PMID 3353444 DOI: 10.1088/0031-9155/33/1/001

14. Ahnesjo A, Collapsed cone convolution of radiant energy for photon dose calculation in heterogeneous media. *Med. Phys.* Aug 1989;16(4):577-92.

15. M Zaider, L Hanin. Tumor control probability in radiation treatment. *Med Phys.* Feb 2011;38(2):574-583.

16. Zhao Y, Qi G, Yin G, et al. A clinical study of lung cancer dose calculation accuracy with Monte Carlo simulation.

17. H Zhen, B Hrycushko, H Lee, et al. Dosimetric Comparison of Acuros XB with collapsed cone convolution/superposition and anisotropic analytic algorithm for stereotactic ablative radiotherapy of thoracic spinal metastases. *J Appl Clin Med Phys.* Jul 2015;16(4):181-192.

18. L Hoffman, M Alber, M Sohn, UV Elstrom. Validation of the Acuros XB dose calculation algorithm versus Monte Carlo for clinical treatment plans. *Med Phys.* Jun 16 2018.

19. M Zaider, L Hanin. Tumor control probability in radiation treatment. *Med Phys.* Feb 2011;38(2):574-583.

20. Determination of absorbed dose in a patient irradiated by beams of x- or gamma-rays in Radiotherapy procedures. International Commission on Radiation Units and Measurement Bethesda Report 24, 1976

21. Brahme, A. “Dosimetric precision requirements in radiation therapy.” Acta Radiol. Oncol. 23:379–391, 1984.

22. Brahme, A., J. Chavaudra, T. Landberg, E.C. McCullough, F. Nüsslin, J.A. Rawlinson, G. Svensson, H. Svensson. “Accuracy requirements and quality assurance of external beam therapy with photons and electrons.” Acta Oncol. Suppl. 1:5–76, 1988.

23. Niemierko A, Goitein M. Implementation of a model for estimating tumor control probability for an inhomogeneously irradiated tumor. *Radiother Oncol.* Nov 1993;29(2):140-147.

24. Fraass B, Doppke K, Hunt M et al. American Association of Physicists in Medicine Radiation Therapy Committee Task Group 53: Quality assurance for clinical radiotherapy treatment planning. *Med Phys.* Oct 1998;25(10):1773-1829.

25. GJ Kutcher, L Coia, M Gillin et al. Report of the AAPM Task Group No. 40: Comprehensive QA for radiation oncology. *Med. Phys.* April 1994;21(4): 581-618.

26. EE Klein, J Hanley, J Bayouth et al. Report of the AAPM Task Group No. 142: quality assurance of medical accelerators. *Med. Phys.* Sep 2009;36(9):4197-4212.

27. Schell, M.C., F.J. Bova, D.A. Larson, D.D. Leavitt, W.R. Lutz, E.B. Podgorsak, A. Wu. “AAPM Report No. 54. Stereotactic Radiosurgery.” Report of Task Group 42. Woodbury, NY: American Institute of Physics, 1995.

28. Malone, S., J. Szanto, G. Perry, L. Gerig, S. Manion, S. Dahrouge, J. Crook. “A prospective comparison of three systems of patient immobilization for prostate radiotherapy.” Int. J. Radiat. Oncol Biol. Phys. 48:657–665, 2000.

29. Nalder, C.A., A.M. Bidmead, C.D. Mubata, D. Tait, C. Beardmore. “Influence of a vac-fix immobilization device on the accuracy of patient positioning during routine breast radiotherapy.” Br. J. Radiol. 74:249–254, 2001.

30. Rotondo, R.L., K. Sultanem, I. Lavoie, J. Skelly, L. Raymond. “Comparison of repositioning accuracy of two commercially available immobilization systems for treatment of head-and-neck tumors using simulation computed tomography imaging.” Int. J. Radiat. Oncol. Biol. Phys. 70:1389–1396, 2008.

31. I Kaur, S Rawat, P Ahlawat et al. Dosimetric impact of setup errors in head and neck cancer patients treated by image-guided radiotherapy. *Med Phy.*Jun 2016;41(2):144-148.

32. Rudat, V., M. Flentje, D. Oetzel, M. Menke, W. Schlegel, M. Wannenmacher. “Influence of the positioning error on 3D conformal dose distributions during fractionated radiotherapy.” Radiother. Oncol. 33:56–63, 1994. 172.

33. Rudat, V., P. Schraube, D. Oetzel, D. Zierhut, M. Flentje, M. Wannenmacher. “Combined error of patient positioning variability and prostate motion uncertainty in 3D conformal radiotherapy of localized prostate cancer.” Int. J. Radiat. Oncol. Biol. Phys. 35:1027–1034, 1996

34. J Mechalakos, E Yorke, GS Mageras et al. Dosimetric effect of respiratory motion in external beam radiotherapy of the lung. *Radiother. Oncol.* May 2004;71(2):191-200.

35. Keall PJ, Mageras GS, Balter JM et al. AAPM Task Group Report 91: The management of respiratory motion in radiation oncology. *Med Phys.*

36. H lin, H Lu, L Shu et al. Dosimetric study of a respiratory gating technique based on four-dimensional computed tomography in non-small-cell lung cancer. *J Radiat. Res.* May 2014;55(3):583-588.

37. Rosenzweig, K. E., J. Hanley, D. Mah, G. Mageras, M. Hunt, S. Toner, C. Burman, C. C. Ling, B. Mychalczak, Z. Fuks, and S. A. Leibel. (2000). "The deep inspiration breath-hold technique in the treatment of inoperable non-small-cell lung cancer." Int J Radiat Oncol Biol Phys 48(1):81–87.

38. V Sarkar, L Huang, YJ H et al. Dosimetric evaluation of abdominal compression as a method to reduce the incidence of radiation-induced pneumonitis in lung SBRT treatment. *J Radiosurg. SBRT.* 2016;4(2):125-132.

39. CS Hamilton, MA Ebert. Volumetric uncertainty in radiotherapy. *Clin. Oncol.* Sep 2005;17(6):456-64.

40. C Fiorino, V Vavassori, G Sanguineti et al. Rectum contouring variability in patient treated for prostate cancer: impact on rectum dose-volume histograms and normal tissue complication probability.

41. J Stanley, P Dunscombe, H Lau et al. The effect of contouring variability on dosimetric parameters for brain metastases treated with stereotactic radiosurgery

42. A van der Horst, AC Houweling, G van Tienhoven, J Visser, A Bel. Dosimetric effects of anatomyical changes during fractionated photon radiation therapy in pancreatic cancer patients. *J Appl Clin Med Phys.* Nov 2017;18(6):142-151.

43. HCY Cheng, VWC Wu, RKC Ngan et al. A prospective study on volumetric and dosimetric changes during intensity-modulated radiotherapy for nasopharyngeal carcinoma patient. *Radiother Oncol.*. Sep 2012;104(3):317-23

44. JC Chow, R Jiang. Dosimetry estimation on variations of patient size in prostate volumetric-modulated arc therapy. *Med Dosim.* Spring 2013;38(1):42-7.

45. IJ Das, CW Cheng, M Cao, PAS Johnstone. Computed tomography imaging parameters for inhomogeneity correction in radiation treatment planning. *Med Phys.* Mar 2016;41(1):346.

Moran, J.M. and T. Ritter. "Limits of Precision and Accuracy of Radiation Delivery Systems." In Uncertainties in External Beam Radiation Therapy, J.R. Palta and T.R. Mackie, eds. Madison, WI: Medical Physics Publishing, 2011.

47. Zhu, T.C., A. Ahnesjo, K.L. Lam, X.A. Li, C.M. Ma, J.R. Palta, M.B. Sharpe, B. Thomadsen, R.C. Tailor. "Report of AAPM Therapy Physics Committee Task Group 74: in-air output ratio, Sc, for megavoltage photon beams." Med. Phys. 36:5261–5291, 2009.

48. LoSasso, T., C.S. Chui, C.C. Ling. "Physical and dosimetric aspects of a multileaf collimation system used in the dynamic mode for implementing intensity modulated radiotherapy." Med. Phys. 25:1919–1927, 1998.

49. Ezzell, G.A., J.W. Burmeister, N. Dogan, T.J. LoSasso, J.G. Mechalakos, D. Mihailidis, A. Molineu, J.R. Palta, C.R. Ramsey, B.J. Salter, J. Shi, P. Xia, N.J. Yue, Y. Xiao. "IMRT commissioning: multiple institution planning and dosimetry comparisons, a report from AAPM Task Group 119." Med. Phys. 36:5359–5373, 2009.

50. McCormack, S., J. Diffey, A. Morgan. "The effect of gantry angle on megavoltage photon beam attenuation by a carbon fiber couch insert." Med. Phys. 32:483–487, 2005

51. J Venselaar, H Welleweerd, B Mijnheer. Tolerances for the accuracy of photon beam dose calculations of treatment planning systems. *Radiother. Oncol.* 2001;60:191-210.

52. DA Low, WB Harms, S Mutic, JA Purdy. A technique for the quantitative evaluation of dose distributions. *Med Phys.* May 1998;25(5):656-61.

53. Saini AS, Zhu TC. Dose rate and SDD dependence of commercially available diode detectors. *Med Phys.* 2004 Apr;31(4):914-24.

54. Létourneau D, Gulam M, Yan D, Oldham M, Wong JW. Evaluation of a 2D diode array for IMRT quality assurance. *Radiother Oncol*. 2004 Feb;70(2):199-206.

55. Saini AS, Zhu TC. Energy dependence of commercially available diode detectors for in-vivo dosimetry. *Med Phys*. 2007 May;34(5):1704-11.

56. Griessbach I, Lapp M, Bohsung J, Gademann G, Harder D. Dosimetric Characteristics of a new unshielded silicon diode and its application in clinical photon and electron beams. *Med Phys.* 2005 Dec;32(12):3750-4. DOI: 10.1118/1.2124547

57. Yin Z, Hugtenburg RP, Beddoe AH. Response corrections for solid-state detectors in megavoltage photon dosimetry. *Phys. Med. Biol.* 2004 Aug;49(16):3691-3702.

58. Rikner G, Grusell E. General specifications for silicon semiconductors for use in radiation dosimetry. *Phys. Med. Biol.* 1987;32(9).

59. Czarnecki D, Zink K. Monte Carlo calculated correction factors for diodes and ion chambers in small photon fields. *Phys Med Biol.* 2013 Apr 21;58(8):2431-44. doi: 10.1088/0031-9155/58/8/2431

60. Chaswal V, Weldon M, Gupta N, Chakravarti A, Rong Y. Commissioning and comprehensive evaluation of the ArcCHECK cylindrical diode array for VMAT pretreatment delivery QA. *J Appl Clin Med Phys.* 2014 Jul 8;15(4):4832. doi: 10.1120/jacmp.v15i4.4832.

61. Siochi RA, Molineu A, Orton CG. Point/Counterpoint: Patient-specific QA for IMRT should be performed using software rather than hardware methods. Med Phys. 2013;40(7):070601. 3.

62. Smith JC, Dieterich S, Orton CG. Point/Counterpoint: It is STILL necessary to validate each individual IMRT treatment plan with dosimetric measurements before delivery. Med Phys. 2011;38(2):553–55.

63. Molineu A, Hernandez N, Nguyen T, Ibbott G, Followill D. Credentialing results from IMRT irradiations of an anthropomorphic head and neck phantom. Med Phys. 2013;40(2):022101.

64. Zhen H, Nelms BE, Tome WA. Moving from gamma passing rates to patient DVH-based QA metrics in pretreatment dose QA. *Med Phys.* 2011 Oct;38(10):5477-89. doi: 10.1118/1.3633904.

65. Kry SF, Molineu A, Kerns JR, et al. Institutional patient-specific IMRT QA does not predict unacceptable plan delivery. Int J Radiat Oncol Biol Phys. 2014;90(5):1195–201.

66. Nelms BE, Zhen H, Tomé WA. Per-beam, planar IMRT QA passing rates do not predict clinically relevant patient dose errors. *Med Phys.* 2011 Feb;38(2):1037-44. DOI: 10.1118/1.3544657

67. Tyagi N, Yang K, Yan D. Comparing measurement-derived (3DVH) and machine log file-derived dose reconstruction methods for VMAT QA in patient geometries. *J Appl Clin Med Phys*. 2014 Jul 8;15(4):4645. doi: 10.1120/jacmp.v15i4.4645.

68. Pawlicki T, Yoo S, Court LE, et al. Moving from IMRT QA measurements toward independent computer calculations using control charts. Radiother Oncol. 2008;89(3):330–37.

69. Agnew A, Agnew CE, Grattan MW, Hounsell AR, McGarry CK. Monitoring daily MLC positional errors using trajectory log files and EPID measurements for IMRT and VMAT deliveries. Phys Med Biol. 2014;59(9):N49–N63.

70. Handsfield LL, Jones R, Wilson DD, Siebers JV, Read PW, Chen Q. Phantom-less patient specific TomoTherapy QA via delivery performance monitoring and a secondary Monte Carlo dose calculation. Med Phys. 2014;41(10):101703.

71. Rangaraj D, Zhu M, Yang D, et al. Catching errors with patient-specific pretreatment machine log file analysis. Pract Radiat Oncol. 2013;3(2):80–90.

72. Stell AM, Li JG, Zeidan OA, Dempsey JF. An extensive log-file analysis of step-and-shoot intensity modulated radiation therapy segment delivery errors. Med Phys. 2004;31(6):1593–602.

73. Sun BZ, Rangaraj D, Boddu S, et al. Evaluation of the efficiency and effectiveness of independent dose calculation followed by machine log file analysis against conventional measurement based IMRT QA. J Appl Clin Med Phys. 2012;13(5):140–54.

74. Kruse JJ and Mayo CS. Comment on "Catching errors with patient-specific pretreatment machine log file analysis". Pract Radiat Oncol. 2013;3(2):91–92.

75. Dinesh Kumar M, Thirumavalavan N, Venugopal Krishna D, Babaiah M. QA of intensity-modulated beams using dynamic MLC log files. *J Med Phys* 2006;31(1):36-41. doi: 10.4103/0971-6203.25668

76. Defoor DL, Vazquez-Quino LA, Mavroidis P, Papanikolaou N, Stathakis S. Anatomy-based, patient-specific VMAT QA using EPID or MLC log files. *J Appl Clin Med Phys.* 2015 May 8;16(3):5283. doi: 10.1120/jacmp.v16i3.5283.

77. Agnew A, Agnew CE, Grattan MW, Hounsell AR, McGarry CK. Monitoring daily MLC positional errors using trajectory log files and EPID measurements for IMRT and VMAT deliveries. Phys Med Biol. 2014 May 7;59(9):N49-63. doi: 10.1088/0031-9155/59/9/N49.

78. Butson MJ, Elferink R, Cheung T, et al. Verification of lung dose in an anthropomorphic phantom calculated by the collapsed cone convolution method. *Phys. Med. Biol.* 2000;45(11).

79. Haga A, Magome T, Takenaka S, et al. Independent absorbed-dose calculation using the Monte Carlo algorithm in volumetric modulated arc therapy. *Radiat Oncol*. 2014 Mar 14;9:75. doi: 10.1186/1748-717X-9-75.

80. Cadman P, MCNutt T, Bzdusek K. Validation of physics improvement for IMRT with a commercial treatment-planning system. *J Appl Clin Med Phys.* Spring 2005;6(2):74-86.

81. Smilowitz JB, Das IJ, Feygelman V et al. AAPM Medical Physics Practice Guideline 5.a: Commissioning and QA of Treatment Planning Dose Calculations – Megavoltage Photon and Electon Beams. *J Appl Clin Med Phys.* 2015 September;16(5):14-34. DOI: 10.1120/jacmp.v16i5.5768

82. Barbeiro AR, Ureba A, Baeza JA et al. 3D VMAT Verification Based on Monte Carlo Log File Simulation with Experimental Feedback from Film Dosimetry. *PLoS ONE, 11*(11), e0166767. http://doi.org/10.1371/journal.pone.0166767

83. Jursinic PA. Dependence of diode sensitivity on the pulse rate of delivered radiation. *Med Phys.* 2013 Feb;40(2):021720. doi: 10.1118/1.4788763.

84. Huang JY, Followill DS, Wang XA, Kry SF. Accuracy and sources of error of out-of field dose calculations by a commercial treatment planning system for intensity-modulated radiation therapy treatments. *J Appl Clin Med Phys.* 2013 Mar 4;14(2):4139. doi: 10.1120/jacmp.v14i2.4139.

85. Howell RM, Scarboro SB, Kry SF, Yaldo DZ. Accuracy of out-of-field dose calculations by a commercial treatment planning system. *Phys. Med. Biol.* 2010;55(23):6999-7008. doi:10.1088/0031-9155/55/23/S03.

86. Joosten A, Matzinger O, Jeanneret-Sozzi W, Bochud F, Moeckli R. Evaluation of organ-specific peripheral doses after 2-dimensional, 3-dimensional and hybrid intensity modulated

radiation therapy for breast cancer based on Monte Carlo and convolution/superposition algorithms: Implications for secondary cancer risk assessment. *Radiother. Oncol.* 2013 Jan; 106(1):33-41. doi: 10.1016/j.radonc.2012.11.012

87. Van den Heuvel F, Defraene G, Crijns W, Bogaerts R. Out-of-field contributions for IMRT and volumetric modulated arc therapy measured using gafchromic films and compared to calculations using a superposition/convolution based treatment planning system. *Radiother Oncol*. 2012;105:127–132. https://doi.org/10.1016/j.radonc.2011.12.030.

88. Kry SF, Bednarz B, Howell RM et al. AAPM TG 158: Measurement and calculation of doses outside the treated volume from external-beam radiation therapy. Med Phys. 2017 Oct;44(10):e391-e429. doi: 10.1002/mp.12462.

89. Scarboro SB, Followill DS, Howell RM, Kry SF. Variations in photon energy spectra of a 6 MV beam and their impact on TLD response. *Med Phys.* 2011 May;38(5):2619-28. DOI: 10.1118/1.3575419

90. Das IJ, Cheng C, Watts RJ et al. Accelerator beam data commissioning equipment and procedures: Report of the TG-106 of the Therapy Physics Committee of the AAPM. Med. Phys., 35: 4186–4215. doi:10.1118/1.2969070

91. Chang KP, Hung SH, Chie YH, Shiau AC, Huang RJ. A comparison of physical and dosimetric properties of lung substitute materials. *Med Phys.* Apr 2012;39(4):2013-20.

92. Okunieff, P., D. Morgan, A. Niemierko, H.D. Suit. “Radiation dose-response of human tumors.” Int. J. Radiat. Oncol. Biol. Phys. 32:1227–1237, 1995

93. Brenner DJ. Dose, volume, and tumor-control predictions in radiotherapy. *Int J Radiat Oncol Biol Phys.* 1993;26(1):171-179.

94. Horiot JC, Le Fur R, N'Guyen T, et al. Hyperfractionation versus conventional fractionation in oropharyngeal carcinoma: final analysis of a randomised trial of the EORTC cooperative group of radiotherapy. Radiother Oncol. 1992;25(4):229–30. [PubMed]

95. The START Trialists' Group. The UK Standardisation of Breast Radiotherapy (START) Trial A of radiotherapy hypofractionation for treatment of early brast cancer: a randomized trial.

96. Stern R, Heaton R, Fraser MW et al. Verification of monitor unit calculations for non-IMRT clinical radiotherapy: Report of AAPM Task Group 114. *Med. Phys.* 2011 Jan;38(1). DOI: 10.1118/1.3521473.

97. Krieger T and Sauer OA. Monte Carlo- versus penfil-beam-/collapsed-cone-dose calculation in a heterogeneous multi-layer phantom. *Phys. Med. Biol.* 2005;50:859-868. DOI:10.1088/0031-9155/50/5/010.

98. Bouchard H and Seuntjens J. Ionization chamber-based reference dosimetry of intensity modulated radiation beams." Med Phys 2004; 31(9).

99. Molineau, A, Hernandez N, Nguyen T, Ibbott G, Followil D. Credentialing results from IMRT irradiations of an anthropomorphic head and neck phantom. *Med Phys.* Feb 2013;40(2):022101.

100. Carson M, Molineu A, Taylor P, Followill D, Kry S. Examining credentialing criteria and poor performance indicators for IROC Houston's anthropomorphic head and neck phantom. AAPM Annual Meeting 2016.

101. Li G, Zhang Y, Jiang X et al. Evaluation of the ArcCHECK QA system for IMRT and VMAT verification. *Physica Medica*. 2013;29:295-303.

102. Yao W, Farr JB. Determining the optimal dosimetric leaf gap setting for rounded leaf-end multileaf collimator systems by simple test fields. *J Appl Clin Med Phys.* Mar 2015;16(4):65-77.

103. Bedford JL, Thomas MDR, Smyth G. Beam modeling and VMAT performance with the agility 160-leaf multileaf collimator. *J Appl Clin Med Phys.* Mar 2013;14(2):172-185.

104. Childress N. Increasing safety by creating simpler treatment plans. Mobius Medical Systems.

ABSTRACT

UTILIZING LOG FILES FOR TREATMENT PLANNING AND DELIVERY QA IN RADIOTHERAPY

by

CARL STANHOPE

August 2019

Advisor: Dr. Di Yan

Major: Medical Physics

Degree: Doctor of Philosophy

Purpose: Monte Carlo-based log file quality assurance (LF-MC QA) is investigated as an alternative method to phantom-based patient-specific quality assurance in radiotherapy (e.g. ArcCHECK QA (AC QA)).

Methods: First, the shortcomings of AC QA were investigated. The sensitivity dependence of ArcCHECK diodes on dose rate (in-field) and energy (primarily out-of-field) was quantified. LF-MC QA was then analyzed on the phantom geometry. Planned ('Plan') and LF-reconstructed CS and MC doses were compared with each other and AC measurement via statistical (mean ± StdDev(σ)) and gamma analyses to isolate dosimetric uncertainties and quantify the relative accuracies of AC QA and LF-MC QA. LF-MC QA was then analyzed on the patient geometry. Calculation algorithm dependent (Plan-MC vs Plan-CS) and delivery error (LF-MC vs Plan-MC) dependent dosimetric discrepancies were isolated. Dose discrepancies were evaluated using PTV D_{mean}, D_{99}, and D_1 as well as tumor control probability (TCP). Dose discrepancy due to calculation algorithm was further assessed as a function of

heterogeneity and beam modulation complexity (MU/Rx). LF QA results were compared to clinical AC QA results. Various LF-MC QA pass/fail protocols were assessed.

Results: Calculation and ArcCHECK measurement differed up to 1.5% in-field due to variations in dose rate and up to 5% out-of-field due to energy effects. On the ArcCHECK geometry, phantom-dependent, calculation algorithm-dependent (MC vs. CS), and delivery error-dependent dose uncertainties were 0.8±1.2%, 0.2±1.1%, and 0.1±0.9% respectively. On the patient anatomy, percent differences in [PTV D_{mean}, D_{99}, D_1] were [-0.1±0.1%, 0.0±0.2%, -0.2±0.2%] for machine delivery error, [-3.4±1.9%, -4.6±2.8%, -1.2±2.8%] for dose calculation difference, and [0.5±2.0%, 0.2±1.2%, 2.6±4.1%] due to limited VMAT beam sampling. Drop in TCP due to calculation difference (MC-CS) was -3.1±1.8% [min -5.7%]. 41% of PTV D99 dose calculation difference was due to beam complexity. Heterogeneity effects were negligible for H&N. For lung, 18% of dose calculation difference on PTV Dmean was due to heterogeneity

Conclusions: ArcCHECK QA was consistently incapable of catching clinically relevant dose discrepancies as calculated on the patient anatomy using LF-MC QA.

AUTOBIOGRAPHICAL STATEMENT

CARL STANHOPE

Education:

2008-2012: B.S.E. Engineering Physics

University of Michigan: Ann Arbor

2013-2015: M.S. Medical Physics

Duke University

2015-2019: Ph.D. Medical Physics

Wayne State University