

12-2022

## Quantifying The Magnitude Of Total Dose Deviation Caused By Various Sources Of Error Among Iroc Phantom Irradiation Results

Sharbacha S. Edward

Follow this and additional works at: [https://digitalcommons.library.tmc.edu/utgsbs\\_dissertations](https://digitalcommons.library.tmc.edu/utgsbs_dissertations)



Part of the [Medical Biophysics Commons](#), [Oncology Commons](#), [Other Physics Commons](#), and the [Radiation Medicine Commons](#)

---

### Recommended Citation

Edward, Sharbacha S., "Quantifying The Magnitude Of Total Dose Deviation Caused By Various Sources Of Error Among Iroc Phantom Irradiation Results" (2022). *Dissertations and Theses (Open Access)*. 1226.  
[https://digitalcommons.library.tmc.edu/utgsbs\\_dissertations/1226](https://digitalcommons.library.tmc.edu/utgsbs_dissertations/1226)

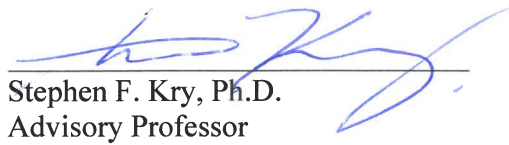
This Dissertation (PhD) is brought to you for free and open access by the MD Anderson UTHealth Houston Graduate School at DigitalCommons@TMC. It has been accepted for inclusion in Dissertations and Theses (Open Access) by an authorized administrator of DigitalCommons@TMC. For more information, please contact [digcommons@library.tmc.edu](mailto:digcommons@library.tmc.edu).

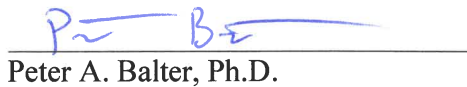
**Quantifying The Magnitude of Total Dose Deviation Caused by Various Sources of Error  
Among IROC Phantom Irradiation Results**

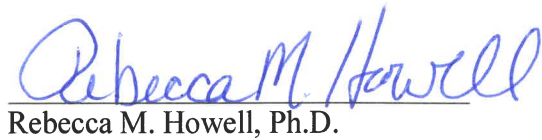
By

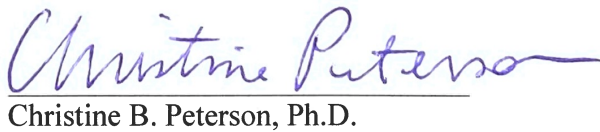
*Sharbacha Sherby Norma Edward, B.S.*

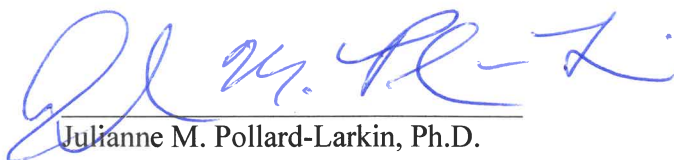
APPROVED:

  
Stephen F. Kry, Ph.D.  
Advisory Professor

  
Peter A. Balter, Ph.D.

  
Rebecca M. Howell, Ph.D.

  
Christine B. Peterson, Ph.D.

  
Julianne M. Pollard-Larkin, Ph.D.

APPROVED:

\_\_\_\_\_  
Dean, The University of Texas  
MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences

**QUANTIFYING THE MAGNITUDE OF TOTAL DOSE DEVIATION CAUSED BY  
VARIOUS SOURCES OF ERROR AMONG IROC PHANTOM IRRADIATION**

**RESULTS**

A

DISSERTATION

Presented to the Faculty of

The University of Texas

MD Anderson Cancer Center UTHealth Houston

Graduate School of Biomedical Sciences

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

by

Sharbacha Sherby Norma Edward, B.S.

Houston, Texas

December 2022

## **Dedication**

I dedicate this work and the past 5 years to my younger self.

Girl, I am living the dream that you dreamt for us.

You did it.

WE did it!

## **Acknowledgements**

What a journey this has been! I did not quite know what to expect when I started my PhD 5 years ago; I just knew that I loved Medical Physics and wanted to do more of it. Now, having survived a 1000-year hurricane and a global pandemic all while managing to keep going, I can say that I have evolved and grown in all the ways possible, and this journey has taught me more than just medical physics.

Thank you to UT Health MD Anderson Graduate School of Biomedical Sciences who believed in me enough to give me money to get a PhD. From the day that I interviewed I knew you guys were special. I have felt supported throughout my entire time here in Houston and I recommend this institution to everyone. If you can, get your degree from the GSBS, you won't regret it!

To the medical physics graduate program faculty and staff, thank you for giving of your time to educate us by teaching courses, administering labs, volunteering to sit on advisory and candidacy committees. Your volunteer efforts are worth it, and I am grateful.

To my advisor, Dr. Stephen F. Kry, I picked a good one. You were the perfect balance of laissez-faire and over achievement. You pushed me to aim for more without ever making me feel inadequate. That was vital to my success; thank you.

Thank you to my advisory committee: Drs. Balter, Howell, Pollard-Larkin and Peterson. Our committee meetings always felt like a few friends just hanging out, yet somehow, we always got things done. Thank you for offering your insight and expertise over the last few years.

To my IROC family, you are the best! Thank you for creating an inviting atmosphere that was both productive and relaxing.

- The phantom crew: Blake, Nadia, Hunter, Trang, Nick and Ben. Thanks for always helping me with equipment and questions and thanks for the fun chats along the way.
- Erika: Thank you for your immense help and support with everything and thanks for always doing it with a smile.
- Mallory: Thank you for holding my hand and leading me into the world of phantom research; you're 60% of the reason I joined this lab.

Thank you to Rob de los Reyes and Scott Hartzell from Raysearch Laboratories for teaching me how to plan and script in Raystation and answering my 10,000 questions long after class had concluded. Thank you to Dr. Aji Daniel and Dr. Peter Balter for providing me with the tools to get my scripting project up and running.

Thanks to Craig Martin, Dr. Dong Joo (DJ) Rhee and Dr. Julianne Pollard-Larkin for lending your time on many occasions so that I could complete parts of my project in the clinic.

Thank you to my classmates, the wonderful incoming class of 2017 (you know yourselves). I would not have made it through these courses, homework assignments and exams without our combined efforts. Thank you for your brilliance and comradery.

Thank you to the Medical Physics students of GSBS, some of whom I got to know very well. I have truly trained alongside the best and brightest in this field. You inspire me!

Thank you to my friends and family in the US, St. Lucia, Canada, South Africa and the UK. Thanks for checking on me, thanks for the fun times, thanks for keeping my life going outside of my work. You make life worth living .

Thank you to my parents for providing the type of life that every child deserves. You laid the foundation for this years ago in Vieux-Fort, St. Lucia. Thank you for your unwavering love and support. I love you both.

To my husband Michael: I left after one year of dating to move to Houston to pursue this degree. I honestly did not know if we would survive the next 5 years, but you made it so easy. You made it easy for me to focus on my work with no pressure and you make it easy to love you every day. Thank you for believing in me when I didn't believe in myself. This is OUR PhD. I love you ♥. I'm coming home!

# QUANTIFYING THE MAGNITUDE OF TOTAL DOSE DEVIATION CAUSED BY VARIOUS SOURCES OF ERROR AMONG IROC PHANTOM IRRADIATION RESULTS

Sharbacha Sherby Norma Edward, B.S.

Advisory Professor: Stephen F. Kry, Ph.D.

## **Abstract**

The Imaging and Radiation Oncology Core (IROC) phantoms are used as an end-to-end test of an institution's radiotherapy processes, and for clinical trial credentialing. Phantoms are treated like patients, and evaluation of the doses received by the thermoluminescent dosimeters (TLDs) inside the phantom, reflects the accuracy with which an institution can image, plan and irradiate a phantom or patient. Recent phantom results show that among the hundreds of various IROC phantoms irradiated annually, 8-17% of institutions fail this test. The purpose of this work was to investigate the various types of errors that may occur during the treatment process and quantify the magnitude of their contribution to planned treatment planning system (TPS) to measured TLD phantom dose deviation (TPS vs TLD dose deviation).

First, a preliminary study was conducted to identify the causes of failures among IROC phantoms. Categories of failure were established, and phantoms grouped accordingly. The results of this study lead to the investigation of three major error contributors: dose calculation error, delivery error and machine output error. Dose calculation error was assessed through independent recalculation of the phantom plans using a dose recalculation system (DRS). An acrylic output block containing TLDs was irradiated by each institution prior to phantom



irradiation, to measure machine output on that day. Machine output error was determined through an assessment of both the output block's measured TLD doses and the machine output dose reported by the institution using their in-house QA device or ion chamber. Delivery error was assessed by evaluating the machine log files associated with a plan delivery. Prior to collecting log files from institutions, a study was conducted to test the ability of the IROC phantoms to accurately capture log file (specifically MLC) errors. This study used the deliberate introduction of MLC errors into the plans, to assess how well they would translate to measured and log file dose deviations. Delivery log files from institutions irradiating the phantoms were collected and evaluated for MLC rms error and delivered dose error. All error types were assessed on an individual TLD basis. Results were categorized into two groups: TLDs with dose deviations greater than the threshold for TLD measurement uncertainty (3.2%) represented the poorer performing group of phantom TLDs, and those with dose deviations less than 3.2%, the better performing group of phantom TLDs.

The majority (60%) of spine and head and neck (H&N) phantom failures, which are static (no motion) and generally have more highly modulated plans, were caused by systematic dose errors. This was when the dose in the entire plan was either too high or too low throughout the entire plan, indicating errors in the institution's TPS dose calculations. The lung phantom, which moves to simulate patient breathing, failed primarily due to localization errors. Localization errors, which manifested as the correct amount of dose, but delivered to a location off-set from the PTV, represented 62% of lung phantom failures. Dose calculation errors were found in 47% of all spine phantom results and 42% of all lung results. However, among failing phantoms, this error was present in 93% of spine cases and only 35% of lung cases, indicating a greater impact of dose calculation error on the highly modulated spine treatment versus the lung. Machine output error showed positive correlations with increasing dose deviation for

spine ( $r = 0.55$ ,  $p < .001$ ), H&N ( $r = 0.63$ ,  $p < .001$ ) and lung phantoms ( $r = 0.45$ ,  $p < .001$ ), indicating that machine output accuracy has an impact on phantom performance. The IROC phantom was found to detect MLC errors that were comparable to clinical results. Random MLC errors produced average dose deviations in the PTV of up to -2.8% for H&N and 0.7% for spine plans. Whole bank MLC shifts resulted in average PTV dose deviations of up to 8% for H&N and 7.1% for spine plans. Analysis of delivery error among IROC phantom log file results showed that compared to the average phantom TPS vs TLD dose deviation of 2.1% (max = 6.1%) for the H&N phantoms, and 2.3% (max = 8.5%) for the lung phantoms, delivered dose error was relatively small. The amount of dose deviation due to delivery errors ranged between -0.3% to 0.5% for the H&N phantom and -0.8% to 0.2% for the lung phantom.

Overall, dose calculation error was found to be the greatest contributor of dose deviations among highly modulated static phantom irradiations (spine and H&N), output error contributed almost equally to all three phantoms and delivery error was minimal with no correlation to phantom performance. Lung phantoms are primarily plagued by motion management related dose deviations which are more difficult to quantify and assess via a remote phantom audit program such as IROC's. Therefore, among the errors evaluated, which were dosimetric in nature, we were able to quantify 56% of error among H&N phantoms, 68% among spine and only 19% of lung dose deviations.

## Table of Contents

Approval Page .....	i
Title Page .....	iii
Dedication.....	iv
Acknowledgements.....	v
Abstract.....	viii
Table of Contents.....	xi
List of Figures.....	xvii
List of Tables .....	xxiii
Chapter 1: Introduction.....	1
1.1 Significance .....	1
1.2 Hypothesis and Specific Aims.....	3
1.3 Dissertation Organization .....	5
Chapter 2: Differences in the patterns of failure between IROC lung and spine phantom irradiations .....	6
2.1 Introduction.....	6
2.2 Methods .....	8
2.2.1 Phantom Design.....	8
2.2.2 Irradiation criteria .....	9
2.2.3 Data collection .....	9
2.2.4 Phantom classification .....	10

2.2.5 Data Analysis.....	10
2.3 Results.....	10
2.3.1 Category descriptions .....	11
2.3.2 Lung.....	15
2.3.3 Spine .....	17
2.3.4 Additional analysis .....	18
2.3 Discussion.....	20
2.4 Conclusion .....	23
Chapter 3: Quantifying dose calculation errors in IROC Phantoms.....	24
3.1 Introduction.....	24
3.2 Methods .....	26
3.2.1 Phantom .....	26
3.2.2 Phantom irradiation .....	27
3.2.3 Dose recalculation system .....	27
3.2.4 Data collection and analysis .....	28
3.3 Results.....	30
3.3.1 Spine .....	31
3.3.2 Lung.....	35
3.4 Discussion.....	39
3.5 Conclusions.....	42
Chapter 4: IROC phantoms accurately detect MLC delivery errors .....	44

4.1 Introduction.....	44
4.2 Materials and Methods .....	46
4.2.1 Phantoms & plans .....	46
4.2.2 Introduced delivery errors.....	47
4.2.3 Different prescription sizes .....	48
4.2.4 Delivery and Analysis.....	48
4.3 Results.....	49
4.3.1 Impact of leaf errors on phantom dose .....	49
4.3.2 Fraction size.....	52
4.4 Discussion.....	53
4.5 Conclusion .....	56
Chapter 5: Quantifying delivery errors in IROC phantoms.....	57
5.1 Introduction.....	57
5.2 Materials and Methods .....	59
5.2.1 Delivery Log files .....	59
5.2.2 Mobius FX .....	60
5.2.3 Plan analysis: delivery component .....	60
5.2.4 Predicted dose value .....	61
5.2.5 Sample analysis .....	62
5.3 Results.....	63
5.3.1 Phantom results.....	63

5.3.2 MLC error .....	65
5.4 Discussion.....	67
5.5 Conclusion .....	69
Chapter 6: Quantifying the magnitude of dose deviation caused by various major sources of error present in IROC phantom irradiations .....	70
6.1 Introduction.....	70
6.1.1 Dose calculation error.....	71
6.1.2 Delivery error.....	72
6.1.3 Machine output error .....	73
6.2 Methods and Materials .....	74
6.2.1 Dose calculation error.....	74
6.2.2 Delivery error.....	75
6.2.3 Output error .....	75
6.2.4 Error analysis .....	77
6.3 Results.....	78
6.3.1 All error types .....	78
6.3.1 (i) H&N.....	79
6.3.1 (ii) Spine .....	81
6.3.1 (iii) Lung:.....	83
6.3.2 Machine output .....	85
6.4 Discussion.....	86

6.5 Conclusion .....	88
Chapter 7: Discussion and Conclusions .....	89
7.1 Project summary .....	89
7.1.1 Preliminary study .....	89
7.1.2 Specific Aim 1 .....	91
7.1.3 Specific Aim 2 .....	92
7.1.4 Specific Aim 3 .....	94
7.2 Evaluation of the Hypothesis .....	95
7.2.1 H&N phantoms .....	95
7.2.2 Spine phantoms .....	95
7.2.3 Lung phantoms .....	96
7.3 Future Research and Applications .....	97
Appendix A: Supplement to Chapter 2 .....	99
Appendix B: Supplement to Chapter 3 .....	101
Appendix C: Supplement to Chapter 6 .....	105
Appendix D: Individual Cases & Additional Materials .....	107
Sample cases: IMRT H&N .....	107
Case 1 .....	107
Case 2 .....	110
Sample cases: Lung .....	112
Case 3 .....	112

Case 4.....	114
References.....	116
Vita .....	123



## List of Figures

- Figure 1.** Imaging and Radiation Oncology Core (IROC) thorax phantom with cork lung insert beside (left). Transverse axial computed tomography (CT) image of IROC thorax phantom with lung target insert showing the lung tumor (middle), and with spine target insert showing spine tumor, vertebral foramen, and the cord as the avoidance structure (right)..... 7
- Figure 2.** Lung phantom dose profiles showing superior-inferior (SI) localization error (top), a global error (middle), and a combination error (bottom). This combination error comprises an SI localization + internal target volume (ITV) exaggeration effect. The consequences of the ITV technique can be seen in the shoulder region of the profiles, where the planned dose (pink) is broader than the measured (blue) film profile. .... 13
- Figure 3.** Spine phantom dose profiles showing a systematic underdose (top), a dose fall-off region error (middle), and an organ at risk (OAR) overdose error (bottom). The OAR in this case is the spinal cord. .... 14
- Figure 4.** Thorax phantom used for both lung and spine irradiations with the cork lung insert (left). CT cross-sectional image of the phantom shows the lung tumor located inside the cork insert and the spine tumor located on the vertebral foramen, which houses the spinal cord (right). .... 26
- Figure 5.** Spine phantom results showing the number of phantoms in which the TPS (dark blue) dose calculation was more accurate than the DRS (light blue) dose calculation. .... 31
- Figure 6.** Average magnitude of dose difference values (D) present among subcategories of spine phantom plans. .... 32
- Figure 7.** Waterfall plot of spine target TLDs, distinguishing passing and failing individual phantom TLDs. The median for failing phantoms was much higher than that for all phantoms, showing an increase in dose calculation errors among failing spine phantoms. .... 34

<b>Figure 8.</b> A combination plot showing the fraction of the TPS vs TLD dose deviation (area under purple line) that can be attributed to dose calculation error (blue shaded area). A direct correlation can be seen between the TPS vs. TLD dose difference and the magnitude of the dose difference value (D).....	34
<b>Figure 9.</b> Lung phantom results showing the number of phantoms in which the TPS (dark blue) dose calculation was more accurate than the DRS (light blue) dose calculation. ....	36
<b>Figure 10.</b> Average magnitude of dose difference values (D) present among subcategories of lung phantom plans.....	37
<b>Figure 11.</b> Waterfall plot of lung target TLDs, distinguishing passing and failing individual phantom TLDs. The median for failing phantoms was similar to all phantoms, showing no increase in dose calculation errors among failing lung phantoms. ....	38
<b>Figure 12.</b> A combination plot showing the fraction of the TPS vs TLD dose deviation (area under purple line) that can be attributed to dose calculation error (blue shaded area). A very low correlation can be seen between the TPS vs. TLD dose difference and the magnitude of the dose difference value (D) for individual lung phantom TLDs. ....	39
<b>Figure 13.</b> Phantoms with ion chamber inserts and A1SL chambers inside: Thoracic (spine) (left), IMRT head and neck (right) .....	46
<b>Figure 14.</b> Dose deviations of phantom plans due to random and systematic MLC error in the TPS. The solid boxes represent the random shifts, and the hollow ones represent the systematic (whole bank) MLC shifts.....	50
<b>Figure 15.</b> Dose comparisons between TPS, measured (ion chamber) and delivered (log file) doses for the H&N phantom, showing differences in recorded dose deviation across all 3 modes. Dose deviations represent difference from the original unperturbed plan, for the 3 plans	

that were irradiated: 50%, 100% and 2mm MLC shifts. Each datapoint represents an ion chamber point on the phantom; 3 plans each for VMAT and sMLC, with 4 ICs each. .... 51

**Figure 16.** Dose comparisons between TPS, measured (ion chamber) and delivered (log file) doses for the spine phantom, showing differences in recorded dose deviation across all 3 modes. Dose deviations represent difference from the original unperturbed plan, for the 3 plans that were irradiated: 50%, 100% and 2mm MLC shifts. Each datapoint represents an ion chamber point on the phantom; 3 plans each for VMAT and sMLC, with 2 ICs each. .... 51

**Figure 17.** Distribution of all MLC RMS errors for the two VMAT plan pairs for the H&N (top) and Spine (bottom) phantoms, irradiated on the Varian 2100EX machine. .... 52

**Figure 18.** Distribution of delivery error compared with TPS vs TLD dose deviation (phantom performance), showing that one does not increase with the other. Delivery error remained relatively unchanged as TPS vs TLD dose deviation increased (phantom performance worsened), indicating minimal contribution from this error type..... 64

**Figure 19.** Distribution of MLC rms error for the two Varian machine classes (Truebeam and Clinac) in relation to phantom performance, represented by absolute TPS vs TLD dose deviation (top), and delivery error (bottom). MLC rms error was found to be machine dependent rather than based on phantom type, magnitude of dose deviation or delivery error. 66

**Figure 20.** IROC acrylic output block containing three double loaded TLD's (top). Output block setup at the LINAC with pointer measuring distance to the machine's output specification point (bottom) ..... 76

**Figure 21.** (a) Illustration and (b) distribution of the total dose deviation quantified for the two subgroups of H&N phantom TLD results; 56% of dose deviation was quantified in TLDs with >3.2% dose deviation. This is shown in the bar graph where the black empty column represents dose deviation, and the colored bars account for this deviation with various error types (c)

Comparison of absolute TPS vs TLD dose deviation (gray area) to each error type. As TLD dose deviation increases (from left to right), phantom performance gets worse. Dose calculation error was the most prevalent and had the strongest correlation (0.59,  $p < .01$ ) with phantom performance, as can be seen from the large number of positive values for this error type towards the right of the graph. Output error was less positively correlated (0.43,  $p < .01$ ) with TLD performance and delivery error remained constant (-0.02,  $p = 0.8$ ) throughout..... 80

**Figure 22.** (a) Illustration and (b) distribution of the total dose deviation quantified for the two subgroups of spine phantom TLD results; 68% of dose deviation was quantified in TLDs with  $>3.2\%$  dose deviation. This is shown in the bar graph where the black empty column represents dose deviation, and the colored bars account for this deviation with various error types. (c) Comparison of absolute TPS vs TLD dose deviation (gray area) to each error type. As TLD dose deviation increases (from left to right), phantom performance gets worse. Dose calculation error was the most prevalent and had the strongest correlation (0.65,  $p < .01$ ) with phantom performance, as can be seen from the large number of positive values for this error type towards the right of the graph. Output error was less positively correlated with TLD performance (0.47,  $p < .01$ ). ..... 82

**Figure 23.** (a) Illustration and (b) distribution of the total dose deviation quantified for the two subgroups of lung phantom TLD results; 19% of dose deviation was quantified in TLDs with  $>3.2\%$  dose deviation. This is shown in the bar graph where the black empty column represents dose deviation, and the colored bars account for this deviation with various error types. (c) Comparison of absolute TPS vs TLD dose deviation (gray area) to each error type. As TLD dose deviation increases (from left to right), phantom performance gets worse. Dose calculation error had the strongest correlation (0.48,  $p < .01$ ) with phantom performance, and output had a correlation of 0.27( $p < .05$ ) with delivery error remaining relatively unchanged throughout.... 84

<b>Figure 24.</b> Output results for 548 institutions showing that institutions on average reported a value for measured machine output that was higher than the true output value.....	85
<b>Figure 25.</b> Lung phantom local dose failure .....	99
<b>Figure 26.</b> Lung phantom systematic overdose .....	99
<b>Figure 27.</b> Major categories of failure for lung phantoms .....	100
<b>Figure 28.</b> Major categories of failure for spine phantoms.....	100
<b>Figure 29.</b> Magnitude of dose calculation error vs Absolute TPS vs TLD dose deviation for the spine phantom results. ....	101
<b>Figure 30.</b> Magnitude of dose calculation error vs Absolute TPS vs TLD dose deviation for the lung phantom results.....	102
<b>Figure 31.</b> Phantom results categorized by TPS algorithm .....	102
<b>Figure 32.</b> Phantom results categorized by treatment type .....	103
<b>Figure 33.</b> Phantom results categorized by machine type .....	103
<b>Figure 34.</b> Lung phantoms categorized by respiratory management method .....	104
<b>Figure 35.</b> Magnitude of output error vs Absolute TPS vs TLD dose deviation for the spine phantom results. ....	105
<b>Figure 36.</b> Magnitude of output error vs Absolute TPS vs TLD dose deviation for the H&N phantom results.....	105
<b>Figure 37.</b> Magnitude of output error vs Absolute TPS vs TLD dose deviation for the lung phantom results.....	106
<b>Figure 38.</b> Dose profiles of IMRT H&N Case 1 showing systematic underdosing of the PTVs, indicating dosimetric inaccuracies.....	109
<b>Figure 39.</b> Dose profiles of IMRT H&N Case 2 showing local dose errors in various parts of the SI and LR profiles.....	111

<b>Figure 40.</b> Dose profiles of lung Case 3 showing systematic underdosing of the PTVs, indicating dosimetric inaccuracies.....	113
<b>Figure 41.</b> Dose profiles of lung Case 4 indicating ITV errors in the SI profile, where the planned dose (pink) appears broader than the film (blue) dose in the shoulder regions of the profile.....	115

## List of Tables

<b>Table 1.</b> Categories of failure for IROC lung and SBRT spine phantom irradiations from January 2012 to December 2018, along with head and neck phantom failure data from a previous IROC study <sup>6</sup> (Nov 2014 – Oct 2015).....	16
<b>Table 2.</b> Main lung phantom error types grouped by respiratory motion management technique. ....	17
<b>Table 3.</b> Demographics of the sample set .....	19
<b>Table 4.</b> Demographics of the phantom sample set .....	29
<b>Table 5.</b> MLC characteristics for Varian Truebeam and 21EX machines .....	49
<b>Table 6.</b> Comparisons of plan delivery accuracy between the two plan prescriptions, showing similar plan delivery accuracy despite varying prescription size.. ....	53
<b>Table 7.</b> Results for an example H&N phantom showing how delivery error is calculated using each individual TLD before finding the average for the phantom. ....	63
<b>Table 8.</b> Distribution of dose deviation results for phantom TLDs .....	78
<b>Table 9.</b> Raw TLD data for Case 1 .....	107
<b>Table 10.</b> Analyzed TLD data for Case 1 .....	107
<b>Table 11.</b> Raw TLD data for Case 2 .....	110
<b>Table 12.</b> Analyzed TLD data for Case 2 .....	110
<b>Table 13.</b> Raw TLD data for Case 3 .....	112
<b>Table 14.</b> Analyzed TLD data for Case 3 .....	112
<b>Table 15.</b> Raw TLD data for Case 4 .....	114
<b>Table 16.</b> Analyzed TLD data for Case 4 .....	114

## Chapter 1: Introduction

### 1.1 Significance

Every year, nearly 1.7 million new cases of cancer are diagnosed in the United States<sup>1</sup>. Over half of these patients will receive some sort of radiation therapy (RT) throughout the course of their treatment<sup>2</sup>. According to the International Commission on Radiation Units and Measurements (ICRU), radiation doses should be delivered to within 5% of the planned dose. Doses that are much higher than this threshold may result in further complications such as additional cancers, and doses that are much lower increase the risk of cancer recurrence<sup>3</sup>.

The RT process which comprises several steps, involves numerous healthcare professionals who work together to achieve the set treatment goals. In order to ensure a highly accurate and consistent standard of care in the RT community, institutions irradiate phantoms as a test of their procedures. Phantoms are physical representations of various human anatomy, comprised of materials with similar densities to corresponding tissue, bones and tumors. Such phantoms have been developed and used by the Imaging and Radiation Oncology Core (IROC) for decades, as a tool for clinical trial credentialing and quality assurance purposes<sup>4,5</sup>. Institutions are advised to treat the phantom as they would a patient, adhering to their regular procedures, which includes imaging, treatment planning, quality assurance and treatment delivery. Thousands of IROC phantoms have been irradiated to date, and the recent results have shown an average success rate of 85%, indicating an institutional phantom failure in 15% of cases. A failing irradiation occurs when an institution's measured TLD dose does not meet the established IROC criteria of being within  $\pm 7\%$  of the planned dose, or 85% of pixels failing to meet the gamma criteria which ranges from 5%/3mm to 7%/5mm among phantoms.



When an institution has an unsuccessful phantom irradiation, they receive a report with the results and analyses of their phantom measurement, and it is their responsibility to look into the possible missteps that may have caused this failure. Usually this is followed by an institution simply re-irradiating the phantom until they are successful. While there has been a slight improvement in phantom irradiation success over the years, there is still much unknown about the specific reasons why these failures still occur. The irradiation criteria for IROC phantoms is, of note, much looser than those used in clinical plans<sup>5,6</sup>, and so irradiations that meet the IROC criteria oftentimes still contain errors that could pose problems in the clinical setting for patient plans.

The overall objective of this work is to quantify the magnitude of the total dose deviation caused by various identified major sources of error among poorer performing IROC phantom irradiations. A poorer performing phantom TLD result is defined as one with a dose deviation greater than 3.2%. The sources of errors investigated are dose calculation errors and treatment delivery errors which comprise machine log file recorded dose errors and machine output & calibration errors. These error types were identified as possible areas for inaccuracies during the radiation therapy process and more importantly, errors that we can remotely assess and measure at IROC. Each error was quantified using data from the various IROC phantom TLD measurements.

By quantifying the magnitudes of these errors, we will be better equipped with the knowledge to inform the community on areas that are lacking in the RT process, to enable more targeted approaches for improvement in treatment accuracy. Given that IROC performs credentialing for hundreds of institutions across the US and abroad, the scope of this project is large and widespread. We have access to a wealth of data from various phantom irradiations

which allows us the opportunity to accurately represent the entire medical physics community on a scale that has not been previously done.

## **1.2 Hypothesis and Specific Aims**

We hypothesize that the sources of error evaluated will comprise, on average, at least 50% of the magnitude of total dose deviation that is currently seen among all poorer performing IROC phantom irradiations. This will reduce the total unknown dose discrepancy between the TPS and TLD from currently 6.5% on average, to less than 3.25%, which is below the ICRU recommended dose deviation limit for patient treatment.

First, a preliminary study was conducted to identify the major causes of failure among various IROC phantoms. This was done to identify the reasons why the phantoms failed the IROC irradiation and determine what types of failures were most prevalent among different phantom groups. This provided us with a general overview of the types of errors to expect and which phantom groups contained similar challenges with irradiation accuracy.

### **Specific Aim 1: Quantify the magnitude of dose deviation caused by dose calculation error in poor performing IROC phantom irradiations**

Lung and spine phantom plans were evaluated in this aim. Plans were recalculated using an independent dose recalculation system and results were analyzed to determine the magnitude of dose calculation errors contained in this phantom population. Similar analysis previously done on the IROC head and neck phantom showed evidence of TPS inaccuracies<sup>6,7</sup>, and so expanding this work to other phantom types was necessary to provide an indication of how the community performs in terms of radiation plan dose calculation for the different phantoms, and hence different patient treatments.

## **Specific Aim 2: Quantify the magnitude of dose deviation caused by delivery error in poor performing IROC phantom irradiations**

Three studies were performed under this aim. Firstly, the impact of various random and systematic MLC errors on the IROC H&N and spine phantoms was evaluated by introducing these errors into treatment plans, delivering them on the phantoms, and evaluating the delivery log files. We also examined the differences in delivery accuracy between the IROC phantom prescription size compared to typical clinical single fraction sizes, to determine whether the MLCs motion was different enough to cause clinically relevant dose discrepancies between two plans. Secondly, using the results of the first study, the IROC phantoms were evaluated to quantify the magnitude of delivery error that exists among phantom results. Lastly, the impacts of machine output and calibration error were examined to determine the contribution of inaccurate treatment machine dose output to dose deviations among IROC phantom irradiations.

## **Specific Aim 3: Combine all error contributions to determine the total magnitude of dose deviation accounted for among poor-performing IROC phantom irradiations**

To comprehensively understand all identified error contributions to phantom dose deviations, we applied the error analysis methods from the first two aims to a third group of phantoms, to investigate the combined effects of all errors on these phantom results. Measurement uncertainty, which is an inherent error that exists due to the use of TLDs, was used as a benchmark to determine the poorer performing group of phantoms. These findings will be compiled and assessed for three types of IROC phantoms: H&N, spine and lung. Our **working hypothesis** for this aim and the overall project was that the sources of error identified

would comprise at least 50% of the magnitude of total dose deviation among poor-performing IROC phantom irradiations.

### **1.3 Dissertation Organization**

Chapter 1 provides an introduction and details the specific aims of the project. Chapters 2 through 6 are self-contained studies each containing an introduction, methods, results, discussion, and conclusion section. Chapter 2 contains the preliminary study that was conducted to identify the various reasons for IROC phantom failures. Chapter 3 evaluates dose calculation errors among spine and lung phantoms. Chapter 4 contains details about the study conducted on the IROC H&N and spine phantoms to evaluate the IROC phantom's ability to detect delivery errors through measurement and log file analysis. Chapter 5 details the assessment and quantification of delivery errors found among the H&N and lung phantoms irradiated by various institutions. Chapter 6 covers machine output error and evaluates the phantoms overall to determine the distribution of each error type for each phantom. Chapter 7 evaluates the hypothesis and contains overall project conclusions and discussions. The appendices contain additional figures and tables as well as information about the location of raw data files on the IROC network drive, for future reference.

## **Chapter 2: Differences in the patterns of failure between IROC lung and spine phantom irradiations**

This chapter is based on the following publication:

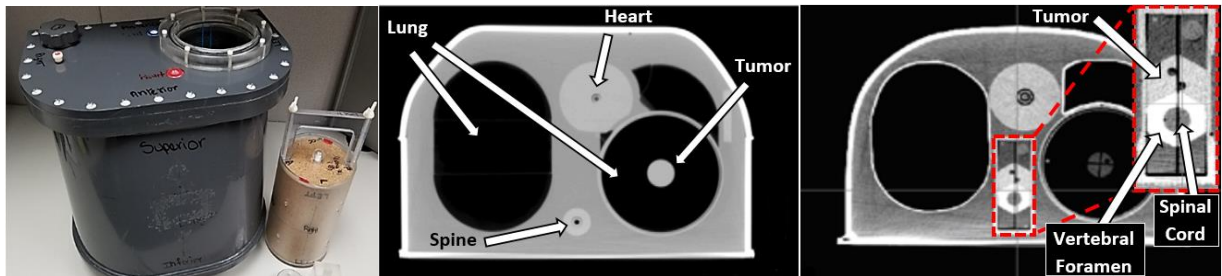
**Edward S.S.**, Alvarez P.E., Taylor P.A., Molineu H. A., Peterson C.B., Followill D.S., Kry S.F. “Differences in the patterns of failure between IROC lung and spine phantom irradiations”, *Pract Radiat Oncol.* May 2020. doi:10.1016/j.prro.2020.04.004

### **2.1 Introduction**

Radiation therapy is used to treat approximately 50% of cancer cases in the United States<sup>8</sup>. However, it has been repeatedly shown that in order for this treatment to be effective, the correct dose must be delivered to the correct treatment site, otherwise, overall patient survival decreases dramatically<sup>9,10</sup>. To maintain a high standard of quality and consistency in the radiotherapy community, and particularly among the National Cancer Institute’s (NCI’s) radiotherapy clinical trials, participating centers across the US and abroad irradiate patient surrogates (phantoms) from the Imaging and Radiation Oncology Core (IROC)<sup>5</sup>. These phantoms comprise tissue equivalent material, along with thermoluminescent dosimeters (TLDs) and radiochromic film, to measure dose delivered to targets and critical structures. These phantom irradiations evaluate an institution’s ability to deliver the planned dose correctly.

The IROC head and neck phantom, which assesses intensity modulated radiation therapy (IMRT) performance, was previously evaluated on several criteria including the patterns of failure of the 10% of irradiations that did not meet IROC’s acceptability criteria<sup>6</sup>. It was found

that the majority of failing phantom irradiations were due to an incorrect dose (i.e., correct shape of the dose distribution in the correct location, but of the wrong magnitude)<sup>6</sup>, and that these were associated with inaccuracies in the institutional dose calculation<sup>7</sup>. Modern radiotherapy goes well beyond IMRT, and IROC phantoms have been established to test these other elements, including the moving lung and SBRT spine phantoms (Fig. 2-1).



**Figure 1.** Imaging and Radiation Oncology Core (IROC) thorax phantom with cork lung insert beside (left). Transverse axial computed tomography (CT) image of IROC thorax phantom with lung target insert showing the lung tumor (middle), and with spine target insert showing spine tumor, vertebral foramen, and the cord as the avoidance structure (right).

These two phantoms have recorded failure rates of 13% and 17% respectively between 2012 and 2018. These failures translate to a substantial number of institutions, and correspondingly patients, who may be receiving clinically suboptimal treatments. Also, institutions that successfully undergo the credentialing process, are generally better prepared for compliance with clinical trial protocol requirements<sup>11</sup>. To better understand the nature of these failing cases, and thereby to begin any rectification that may be appropriate at the corresponding institutions, we first need to better understand the nature of the failures. The purpose of this study was to evaluate the moving lung and SBRT spine phantom irradiation failures and classify them according to the nature of the failures. Results of this study are critical for the entire radiation oncology team in order to understand the risks and challenges of delivering high quality radiation therapy.

## 2.2 Methods

### 2.2.1 Phantom Design

The IROC anthropomorphic thorax phantom shell (Fig.2-1) is used to perform both lung and spine irradiations. This phantom is heterogeneous in order to simulate actual patient anatomy<sup>12</sup>, and specific lung and spine target and dosimetry inserts are used based on the anatomical site being irradiated (either a lung or spine insert). The target in the lung insert is an ovoid structure, measuring 5 cm in length and 3 cm in diameter, located in the center of low-density tissue equivalent material. The target in the spine insert mimics the shape of the vertebral body. Abutting the target is a structure representing the vertebral foramen, within which the organ at risk, the spinal cord, is contained, only 0.8 cm posterior to the edge of the target. The materials that make up this phantom include compressed cork for the lungs, nylon for the heart, polybutylene terephthalate-polyester for the spine, and polystyrene for the tumors<sup>13</sup>. The shell is also filled with water in order to represent soft tissue.

The lung phantom treatments are either static or include motion. Static lung treatments, despite not being representative of a typical patient lung treatment in terms of motion, test other aspects of the treatment process such as heterogeneity corrections and dose delivery. To simulate motion for gated and free-breathing/ internal target volume (ITV) treatments, the lung phantom is placed on a moving platform<sup>14</sup>. In the superior-inferior (SI) direction, the platform (and phantom) moves with a 2 cm amplitude. The breathing cycle contains 2 distinct breaths based on clinical patient breathing patterns, and a cycle of both breaths takes 11 seconds to complete. The phantom also moves a total of 0.5 cm in the anterior-posterior (AP) direction during motion. The AP direction was considered as the direction of minor motion for the purposes of this study, since the extent of this motion was within the distance-to-agreement gamma criterion ( $\pm 5$  mm).

Thermoluminescent dosimeter (TLD) capsules with double-loads of powder are placed inside the targets and organs at risk in order to record the dose to these structures<sup>15</sup>. The lung phantom contains 2 centrally located TLD capsules within the 35 cm<sup>3</sup> target. The spine phantom contains 4 TLD capsules within the 22 cm<sup>3</sup> target. Radiochromic film is also placed in the phantom in orthogonal planes in order to measure dose distributions<sup>15</sup>. Institutions are directed to deliver 6 Gy to the target and treat it as they would any radiotherapy patient in terms of imaging, treatment planning, setup, and delivery.

### **2.2.2 Irradiation criteria**

Successful irradiation of a lung phantom is achieved by producing TLD measurements in the target that are each within  $\pm 7\%$  of the planned dose (over the TLD contours). Additionally, a film pass rate of 80% is required for each of the axial, coronal and sagittal film planes, and a combined average of 85% for the three planes, with a film gamma index of 7%/5 mm. The criteria for successful irradiation of a spine phantom is that each measured TLD dose agrees within  $\pm 7\%$  of the planned dose, and a film pass rate of 85% each for the axial and sagittal film planes is achieved, with a film gamma index of 5%/3 mm.

### **2.2.3 Data collection**

Failing phantoms were identified through the IROC phantom records database. For this study, 116 failing lung and 42 failing spine phantom reports were analyzed individually. This includes all attainable data for failing irradiations for these two phantoms from January 2012 to December 2018. Reports were abstracted for irradiation result as well as demographic information (treatment delivery unit, planning system, etc.)



#### **2.2.4 Phantom classification**

We reviewed all phantom reports, including dose disagreements, profiles, and gamma analysis, in order to define categories and then categorize the patterns of failure. While most phantoms had a single clear cause of failure, some cases were found to contain multiple causes of failure (e.g., the dose was systematically low and the dose profile was shifted from target center). In such an instance, where the case would have failed given either one of the failures on its own, the case was counted twice: once in each of the failure categories represented. For a few phantoms, the failure was as the result of a culmination of causes, no single one of which on its own would have caused a failure. In these cases the failure mode was described as a “combination” and placed into the combination category.

#### **2.2.5 Data Analysis**

Failure mode totals were calculated for each category. Due to the double counting of some phantoms, the failure-mode total was greater than the number of individual phantoms evaluated. A 95% confidence interval was calculated for the rate of failures due to each main failure category using the Wilson interval method. This method was used to assess the likelihood of phantom failures falling under each category, given the total number of phantoms in the study. Other criteria that were analyzed for patterns of failure include beam energy, machine model, treatment planning system (TPS) algorithm and treatment technique, with the addition of motion management technique for the lung phantom. The association between failure and this demographic data was analyzed using the Chi-square and Fisher’s exact tests.

### **2.3 Results**

From January 2012 to December 2018, the lung phantom was irradiated 1052 times and recorded a failure rate of 13%. The SBRT spine phantom was irradiated 263 times and recorded

a higher failure rate of 17%. For this study, all available failing phantom records, totaling 116 lung (82% of failures) and 42 spine (91% of failures) were evaluated and categorized based on failure type. The 116 failing lung phantom cases were from 106 different institutions, with 7 institutions repeating the phantom 2 or more times. The 42 failing spine cases were from 33 different institutions, also with 7 institutions repeating the phantom 2 or more times. Four institutions recorded failures in both phantoms.

While some phantom irradiation cases contained multiple error types, all cases were categorized based on their most egregious error type that ultimately caused that irradiation result to fall outside of the established IROC criteria.

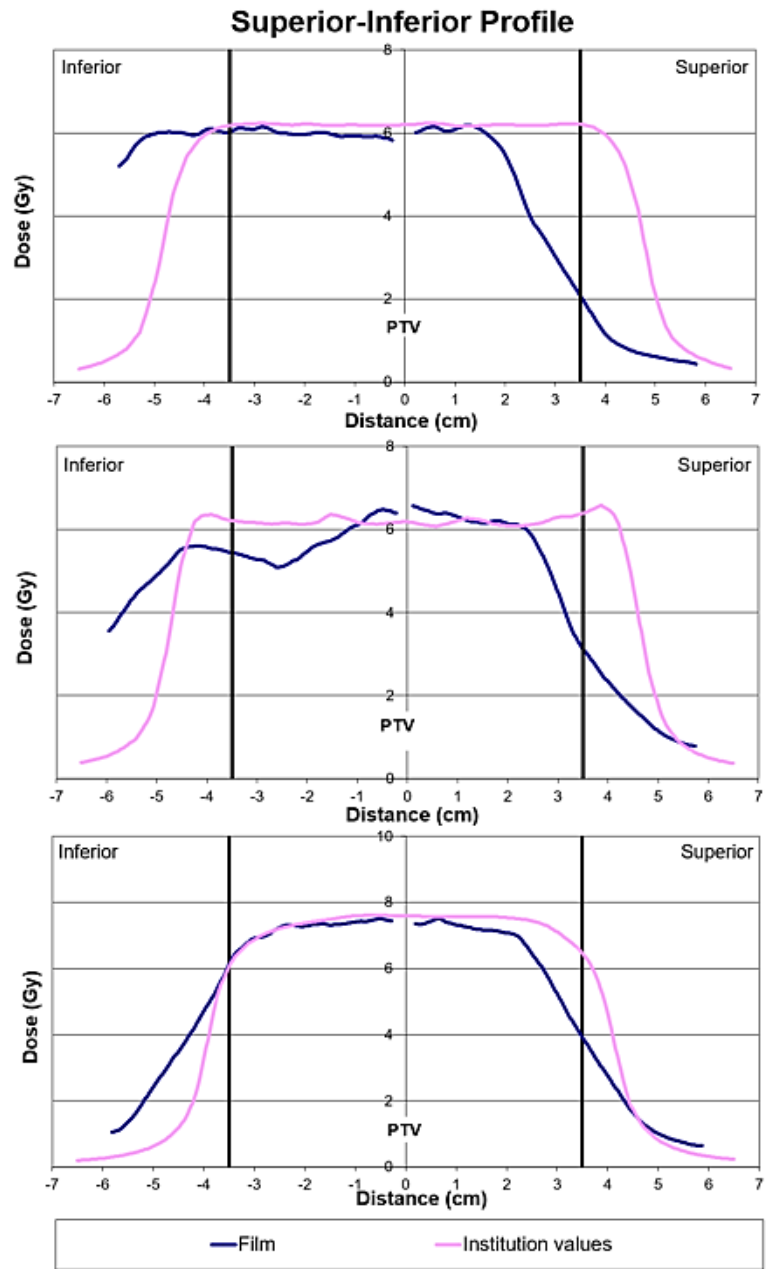
### **2.3.1 Category descriptions**

Seven lung and four spine categories were formed. These categories are described as follows:

#### **Lung**

1. Systematic dose: uniform overdosing or underdosing of the PTV
2. Local dose: dose error in an isolated area of the plan
3. Localization – major motion: dose distribution improperly aligned with target in the superior –inferior direction. The phantom moves 2.0 cm in this direction during the breathing cycle. This error is illustrated in Fig 2-2 (top), where the measured dose profile is shifted almost 2 cm inferiorly to the institution’s planned dose profile.
4. Localization – minor motion: dose distribution improperly aligned with target in the anterior-posterior direction. The phantom moves 0.5 cm in this direction during the breathing cycle.

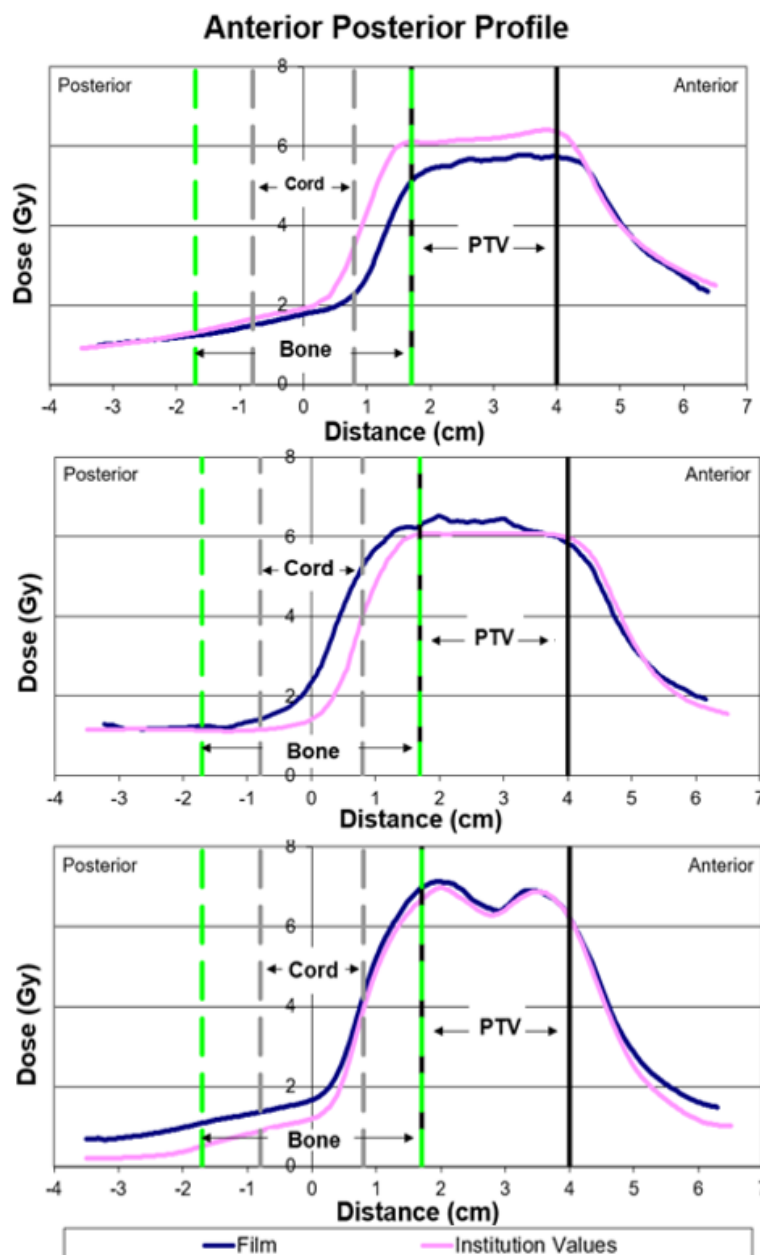
5.      Localization – no motion: dose distribution improperly aligned with target in the left-right direction, in which the target does not move at all.
6.      Global Error: grossly irregular dose distributions. This is illustrated in Fig. 2-2 (middle), where the measured profile is of a different shape, contains an inferior shift and has a lower dose in half of the PTV relative to the planned profile.
7.      Combination Category: contributions from two separate error types, not individually sufficient to each cause a failure, but when combined, caused the irradiation results to fall outside of criteria. Combinations which fell into this category included: SI localization + localization AP, systematic underdose + localization AP, SI localization+ ITV exaggeration (illustrated in Fig. 2-2 (bottom)) and systematic underdose + ITV exaggeration. The combination component “ITV exaggeration” is a side effect (or consequence) of the ITV motion management technique, which affects the dose profile in the shoulder region, causing it to be more susceptible to otherwise negligible dose/localization effects.



**Figure 2.** Lung phantom dose profiles showing superior-inferior (SI) localization error (top), a global error (middle), and a combination error (bottom). This combination error comprises an SI localization + internal target volume (ITV) exaggeration effect. The consequences of the ITV technique can be seen in the shoulder region of the profiles, where the planned dose (pink) is broader than the measured (blue) film profile.

## Spine

1. Systematic dose: uniform overdosing or underdosing of the PTV. This is illustrated in Fig. 2-3 (top) where the measured dose profile has a lower value relative to the institution's planned dose profile.



**Figure 3.** Spine phantom dose profiles showing a systematic underdose (top), a dose fall-off region error (middle), and an organ at risk (OAR) overdose error (bottom). The OAR in this case is the spinal cord.

2. Dose fall-off region: dose error in the steep dose gradient between the PTV and spinal cord. This is illustrated in Fig. 2-3 (middle) where the measured dose profile is of a higher value in the dose fall-off region relative to the institution's planned dose profile.
3. OAR overdose: overdose of the spinal cord structure. This is illustrated in Fig. 2-3 (bottom), where the measured dose is higher than planned in the spinal cord region, essentially causing an overdosing of that OAR.
4. Localization: dose distribution improperly aligned with the target

### **2.3.2 Lung**

Results of the various category assignments are summarized in Table 1. The majority (62%) of the lung phantom failures were due to localization errors; however, almost all of these localization errors (64 out of 79) were in the direction of motion (SI), making up 50% of all lung phantom failures. The average observed error among both superiorly and inferiorly shifted phantoms was  $1.29 \pm 0.54$  cm, with the maximum recorded shift in each direction being 2.5 cm (four cases). Localization error in the anterior-posterior direction of motion was  $0.47 \pm 0.09$  cm on average, which was within the established gamma threshold of 0.5 cm. Systematic dose errors accounted for 28 (22%) of all lung failures, with 19 (16%) of those being due to an underdose and 9 (6%) due to an overdose. The remaining phantom failure modes comprised relatively few cases.

**Table 1.** Categories of failure for IROC lung and SBRT spine phantom irradiations from January 2012 to December 2018, along with head and neck phantom failure data from a previous IROC study<sup>6</sup> (Nov 2014 – Oct 2015).

Error Type	Number of phantoms (%)	95% C.I. (%)
<b>Lung</b>		
Systematic dose	28 (22)	(17-33)
➤ overdose	9 (6)	
➤ underdose	19 (16)	
Local dose	7 (6)	(3-12)
Localization: major motion	64 (50)	(46-64)
➤ superior	26 (20)	
➤ inferior	38 (30)	
Localization: minor motion (ant-post)	8 (6)	(4-13)
Localization: no motion (left-right)	7 (6)	(3-12)
Global	3 (2)	(1-7)
Combination category	10 (8)	(5-15)
<b>Spine</b>		
Systematic dose	25 (60)	(44-73)
➤ overdose	9 (22)	
➤ underdose	16 (38)	
Dose fall-off region	5 (12)	(5-25)
OAR overdose	6 (14)	(7-28)
Localization	6 (14)	(7-28)
<b>Head and Neck <sup>6</sup></b>		
Systematic dose	11 (69)	(44-86)
➤ Overdose	2 (13)	
➤ Underdose	9 (56)	
Localization	2 (12)	(3-36)
Global	3 (19)	(7-43)

The phantom failures are classified by motion management technique in Table 2. In the SI localization category, which was the most prominent error type, the failure rates were markedly elevated for cases irradiated using gated and free-breathing/ITV techniques (Table 2). Failure rates in the SI localization category were 13% and 10% for gated and ITV cases respectively, versus 0.3% ( $p < 0.001$ ) for static phantom cases. In the other two major error categories: dose

and localization in other directions (non-motion), the rates were fairly evenly distributed and were all at or below 5%.

**Table 2.** Main lung phantom error types grouped by respiratory motion management technique.

ERROR TYPE	RESPIRATORY MOTION TECHNIQUE							
	Free breathing (ITV)	95% C.I. (%)	Gating	95% C.I. (%)	Static	95% C.I. (%)	Tracking	95% C.I. (%)
Dose (over, under and local)	$\frac{23}{466}$ (5%)	(3-7)	$\frac{5}{125}$ (4%)	(2-9)	$\frac{7}{391}$ (2%)	(1,4)	$\frac{0}{35}$ (0%)	(0-10)
SI localization	$\frac{47}{466}$ (10%)	(8-13)	$\frac{16}{125}$ (13%)	(8-20)	$\frac{1}{391}$ (0.3%)*	(0,1)	$\frac{0}{35}$ (0%)	(0-10)
Other localization	$\frac{10}{466}$ (2%)	(1-4)	$\frac{3}{125}$ (2%)	(1-7)	$\frac{1}{391}$ (0.3%)	(0,1)	$\frac{0}{35}$ (0%)	(0-10)

*Abbreviations:* CI = confidence interval; ITV = internal target volume; SI = superior-inferior.

Values are represented as a fraction of all irradiated Imaging and Radiation Oncology Core (IROC) lung phantoms irradiated with that type of motion management within the period (2012-2018). Combination and global error categories were excluded due to the specific and unique nature of these failures.

\* $P < .001$ , Fisher's exact test for pairwise comparisons of ITV, gating, and static respiratory motion management techniques for SI localization failures.

### 2.3.3 Spine

The majority (60%) of the spine phantom failures were due to systematic dose errors, with 9 (21%) of them being due to an overdose and 16 (38%) due to an underdose. The phantoms with a systematic underdose error had an average dose difference of  $-7.4\% \pm 2.4$ , and those with a systematic overdose error had an average dose difference of  $4.8\% \pm 1.8$ . The remaining phantoms were nearly evenly distributed among the remaining categories. Of note, the failure rate due to localization errors of the spine phantom (14%) was nearly identical to the failure rate due to localization errors of the lung phantom in the non-motion direction (12%).



### 2.3.4 Additional analysis

Table 3 is a demographic display of the phantoms, grouping them by machine, energy, TPS algorithm, treatment technique and for the lung, respiratory motion management. The number of failing phantoms, and its value as a percentage of the total number of irradiated phantoms within the 2012-2018 timeframe are included. LINACs were categorized by machine classes as defined previously.<sup>16,17</sup> For both phantoms, the majority of irradiations were performed on Varian and Elekta machines; Siemens and Accuray machines were used for the remaining phantoms. The energies used were mostly 6 MV regular, with a few cases of 6 SRS, 6 FFF, 10 MV and 10 FFF beams. Five different TPS algorithms were recorded with the most popular ones being Eclipse AAA and Superposition Convolution. Treatment techniques varied between the two phantoms with Dynamic MLC, Segmental (step & shoot) MLC and VMAT being the most popular techniques among the lung irradiations, while Segmental (step & shoot) MLC and VMAT were most common among the spine irradiations. The lung phantom respiration motion was managed using either the gating, free breathing (ITV) or tracking techniques, or the phantom was irradiated with no motion i.e., static. Tracking is a method specific to Cyberknife users and saw a 100% pass rate. The failure rates for the gating and ITV techniques were much greater than for static cases at 18% each vs 2% respectively, while there were no recorded failures for tracking. Statistical analysis using Fisher's exact test showed that lung phantoms were overall more likely to fail when treated using gating or ITV respiratory motion management techniques over the static technique ( $p < 0.001$ ).

**Table 3.** Demographics of the sample set

Demographic	<i>Lung Phantoms</i>		<i>Spine Phantoms</i>	
<b>Treatment Machine</b>	Number of Failing Phantoms	% of this irradiation type	Number of Failing Phantoms	% of this irradiation type
<b>Varian</b>	<b>94</b>	<b>12%</b>	<b>32</b>	<b>18%</b>
Base class*	43	11%	18	27%
Trilogy	1	6%	4	44%
Truebeam	50	13%	10	10%
<b>Elekta</b>	<b>17</b>	<b>15%</b>	<b>4</b>	<b>25%</b>
Agility	1	13%	0	0%
Infinity	6	16%	1	13%
Synergy	5	12%	0	0%
Versa HD	4	15%	3	38%
Precise	1	50%	0	0%
<b>Siemens</b>	<b>3</b>	<b>7%</b>	<b>1</b>	<b>50%</b>
Artiste	2	15%	1	50%
Oncor	1	17%	0	0%
<b>Accuray</b>	<b>1</b>	<b>4%</b>	<b>5</b>	<b>17%</b>
Cyberknife	0	0%	4	15%
Hi-Art Tomotherapy	1	4%	1	25%
<b>Energy (MV)</b>				
6	91	10%	31	16%
6 SRS	1	6%	4	44%
6 FFF	14	16%	2	50%
10	4	10%	4	20%
10 FFF	6	24%	1	9%
<b>TPS Algorithm</b>				
Eclipse AAA	54	11%	20	17%
Grid-based Boltzmann solvers <sup>†</sup>	19	13%	3	11%
Measured <sup>‡</sup>	0	0%	3	43%
Monte Carlo	5	5%	2	6%
Superposition Convolution	38	13%	14	21%
<b>Treatment Technique</b>				
3D CRT	9	6%	0	0%
Dynamic MLC	23	14%	5	19%
Segmental (Step & Shoot) MLC	16	14%	11	26%
VMAT	67	13%	21	14%
Tomotherapy	1	3%	1	14%
Cyberknife	0	0%	4	15%
<b>Respiration Motion Technique</b>				
Gating <sup>§</sup>	23	18%	N/A	N/A
ITV <sup>§</sup>	84	18%	N/A	N/A
Static	9	2%	N/A	N/A
Tracking	0	0%	N/A	N/A

---

*Abbreviations:* 3D = 3-dimensional; CRT = conformal radiation therapy; FFF = flattening filter free; IROC = Imaging and Radiation Oncology Core; ITV = internal target volume; MLC = multileaf collimator; SRS = stereotactic radiosurgery; TPS = treatment planning system; VMAT = volumetric modulated radiation therapy.

“% this irradiation type” refers to the percentage of all IROC phantoms irradiated under these conditions that failed the irradiation.

\* Clinac 21EX, Clinac 21iX, Clinac 23iX, Clinac iX, Clinac 2300CD

† Acuros dose calculation algorithm

‡ Ray tracing was the only measurement-based algorithm used

§  $p < 0.001$ , Fisher’s exact test for pairwise comparisons of gating, ITV and static respiratory motion management techniques

## 2.3 Discussion

There were 116 moving lung phantom failures placed into 7 categories. The majority of the failing lung phantoms had an SI localization error, which is the direction of major target motion. As shown in Table 2, phantom motion contributed substantially to SI localization errors, as 10% of all free-breathing/ITV and 13% of all gated treatments failed in this way, versus 0.3% of all static treatments. Poor respiratory motion management is likely responsible for these failures, including components like image-guided radiation therapy (IGRT) setup and tumor localization, given that only 1 of 64 phantom failures in this category was stationary during irradiation.

Gating did not improve the phantom SI localization results when compared with ITV, as a similar fraction of phantoms (13% vs 10% respectively) failed in this manner (Table 2). Gating and ITV techniques were previously shown to also perform similarly when compared dosimetrically in lung treatments.<sup>18,19</sup>

Additionally, statistical analysis of the demographic data (Table 2-3) showed that lung phantoms were more likely to fail overall when treated using gating or free-breathing motion management techniques. No statistical significance was found among failure rates due to TPS algorithm, despite previous indications that some algorithms perform better in heterogeneity corrections<sup>20</sup>. This is likely due to low power because of the limited sample size.

It is worthy to note that the failure rates under the minor (AP) and no (LR) motion categories were equal, which makes sense, because the motion in the AP direction is within the bounds of the gamma criteria ( $\pm 0.5$  mm). Treatment error in this direction was therefore more representative of treatment setup errors.

Applications of these clinically relevant findings can be made to similar treatment sites containing motion caused by patient respiration, such as the liver.<sup>21,22</sup> Tumor motion caused by gastrointestinal activity<sup>23</sup> is of even greater concern as this motion is irregular, unlike the very regular motion pattern of the lung phantom. Gastrointestinal motion would require more careful planning and monitoring, for accurate delivery and to avoid irradiating organs at risk.

There were 42 SBRT spine phantom failures placed into 4 categories. These failures, in contrast to those of the lung phantom, were mostly dosimetric in nature, displaying underdosing and overdosing of the PTV by as much as -11% and +10% respectively, and overdosing of OARs and the dose fall-off region. Dose calculation errors have been shown to indicate inaccuracies in institutions' dose calculation software<sup>6</sup>, which is even more crucial for cases such as these highly modulated, high dose plans. In addition, the criteria used in IROC's phantom credentialing processes are loose compared to clinical standards, biological needs, and dosimeter precision.<sup>6</sup> As such, these failures should be considered dramatic underperformances in terms of dose delivery. The results of the spine phantom irradiation are also consistent with those of the IMRT head neck phantom, where 69% of failures were classified as systematic dose errors.<sup>6</sup> Further study is warranted to understand if the same TPS errors are manifesting in both phantoms as head and neck IMRT (comprised of highly modulated but larger fields) tends to be different from spine SBRT (comprised of less modulated but smaller fields).<sup>24</sup>

The IROC phantom serves as a patient surrogate, and allows us to record post treatment doses received by the target, through TLDs and radiographic film. The phantom is therefore the most accurate representation of a patient that we have, and provides us with vital information about the performance and accuracy of our current procedures in actual patient cases among the radiation oncology community. The various categories of failure for each phantom highlight the nature of failures which are most likely to occur for a given treatment type. The majority of the failures occurring in each phantom were errors in the treatment process, and did not appear to be the result of random mistakes or human error. Human error is a likely explanation for localization errors (in the non-motion direction), as this is often explained by a failure to setup to the correct isocenter, and such errors would often (although not always) be caught by clinical image review procedures. For all of the spine, lung, and head and neck phantoms, the rate of this error is very consistent: 12-14% of failures result from this cause. This is a small minority of the causes of failure. The dominant failure modes are ones that are testing the challenge of the irradiation technique being performed, namely SI localization for the moving lung target and systematic dose for the highly modulated spine target. The systematic dose calculation errors, in particular, are hard to imagine as arising from human error. The existence of problems within an institution's treatment process is highlighted by the fact that of the 7 institutions that failed the lung phantom twice, in 4 of them the error type was the same. One institution failed the lung phantom 4 times, and had a similar SI localization error 3 times in a row. Among institutions that recorded multiple failures of the spine phantom, every institution recorded the same error for each phantom failure. The results therefore serve to support that the majority of these errors are not random, but highlight a problem with the radiotherapy process. As such, these problems would likely show up in actual patient treatments. This information can guide quality assurance practices, and alert clinicians and physicists to the components in

the radiation therapy treatment process that are most error prone, to properly guide future rectification efforts.

## **2.4 Conclusion**

The patterns of failure among IROC lung and spine phantoms were investigated for phantoms irradiated from 2012 to 2018. The majority of lung phantom failures were due to a localization error in the direction of major target motion and described the situation of missing the moving target. In contrast, spine phantoms failed mostly because of underdosing of the PTV (target) or overdosing of the organs at risk. These errors are clinically relevant and have high potential to manifest as errors in patient cases. This study can be used as a guide when treating sites with similar parameters, i.e., affected by motion or high modulation. Knowing what is most likely to go wrong for a particular case affords physicists and clinicians the opportunity to exercise precaution in these stages of the radiotherapy process.

## Chapter 3: Quantifying dose calculation errors in IROC Phantoms

This chapter is based on the following publication:

**Edward S.S.**, Glenn M.C., Peterson C.B., Balter P.A., Pollard-Larkin J.M., Howell R.M., Followill D.S., Kry S.F., “Dose calculation errors as a component of failing IROC lung and spine phantom irradiations”, *Med Phys.* 2020;47(9):4502-4508. doi:10.1002/mp.14258

### 3.1 Introduction

Radiation therapy is widely used to treat patients with various types of cancers. Based on the disease site, different challenges exist in meeting the optimization goals of maximum dose to the tumor and minimum dose to the organs at risk (OARs), such as the need to manage motion of the tumor or to work within very tight delivery margins. To verify the ability of institutions to meet these treatment goals, the Imaging and Radiation Oncology Core (IROC) performs end-to-end assessments of various radiation therapy treatment processes.<sup>15</sup> This verification is accomplished using IROC’s end-to-end anthropomorphic phantom program, in which institutions irradiate phantoms either for quality assurance (QA) purposes or clinical trial credentialing.<sup>25</sup>

Various phantoms are used for the unique challenges associated with different disease sites, the most common being the head and neck phantom for intensity-modulated radiation therapy (IMRT), a spine phantom for stereotactic body radiation therapy (SBRT), and a heterogeneous lung phantom for IMRT/SBRT delivery with motion management.<sup>12</sup> For each disease site and corresponding phantom, institutions are instructed to treat these phantoms as they would a patient, which includes taking images, creating a treatment plan, setting up for

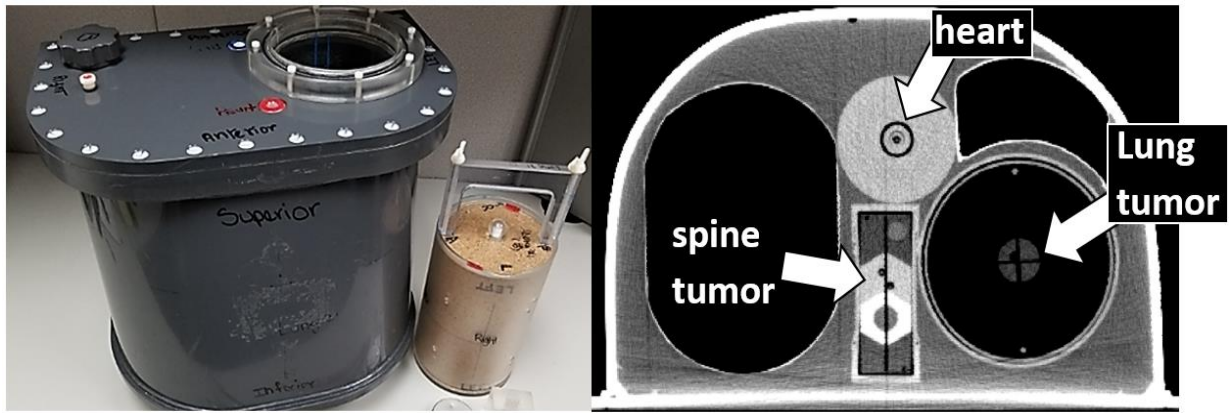
treatment and irradiating the phantom using the treatment plan created. The delivery is evaluated in terms of the simple assessment of whether the delivered dose matches that which was planned.

In recent years, 18% of all phantom irradiations have resulted in institutions failing to meet the acceptance criteria for successful irradiation.<sup>26</sup> These failures are of great concern because the phantoms model actual patient anatomy and clinical workflow, and failure modes appear to be directly clinically relevant, meaning they are likely to manifest in patient cases as well.<sup>6,27</sup> Therefore, there is an urgent need to understand the underlying causes of these phantom failures, so that direct action and remediation can be undertaken.

For IMRT plans, previously investigated using the head and neck phantom, inaccurate dose calculation has been identified as a major contributor to failing phantoms.<sup>7</sup> Suboptimal beam modeling was identified in 67% of institutions that failed the phantom irradiation, and on average, the dose calculation was 3% poorer than expected. However, the relevance and impact of beam modeling errors in other types of treatments have not yet been explored.

For highly modulated treatment plans, such as those used for the SBRT spine,<sup>28</sup> which often have small beam apertures,<sup>24</sup> there is an added level of complexity involved in dose calculation because the target is very close to critical structures. In contrast, centrally located island lung tumors, such as those in the IROC lung phantom, are more commonly treated by using less modulated treatments because the isolated tumor target is not directly abutting critical structures (Fig. 3-1), however, motion of the lung tumor must also be managed.





**Figure 4.** Thorax phantom used for both lung and spine irradiations with the cork lung insert (left). CT cross-sectional image of the phantom shows the lung tumor located inside the cork insert and the spine tumor located on the vertebral foramen, which houses the spinal cord (right).

This study evaluated the presence of dose calculation errors in the IROC lung and spine phantoms irradiated by numerous institutions. The aim was to determine the magnitude of dose calculation error exhibited in the irradiation of these two phantoms and whether that error directly influences the rates of phantom irradiation failure.

## 3.2 Methods

### 3.2.1 Phantom

IROC lung and spine phantom irradiations are performed using the same anthropomorphic thorax phantom shell (Fig. 3-1), which can house either the spine or lung target structures. This phantom, designed to mimic human anatomy, was built with materials chosen to represent the heterogeneity present in and around these two treatment sites, and to be tissue equivalent. The spine target consists of a 22-cm<sup>3</sup> tumor structure that rests on the vertebral body,<sup>12</sup> approximately 0.8 cm from the spinal cord. The lung target structure is 35 cm<sup>3</sup> and is located inside of low-density tissue-equivalent compressed cork, used to mimic lung tissue. The phantom is filled with water before irradiation to represent surrounding soft tissue. Irradiations of the spine phantom target are done in a static configuration. For irradiations of

the lung target, depending on the institution's target motion management system, the phantom rests on a motion platform that simulates patient breathing, using a patient breathing trace that is 2 cm in amplitude in the superior-inferior direction and 0.5 cm in amplitude in the anterior-posterior direction. Various breathing patterns are available for free-breathing/internal target volume (ITV) and gated lung treatments.<sup>27</sup>

To measure the dose to targets and OARs, thermoluminescent dosimeters (TLDs) and radiochromic film are placed inside these structures and read by IROC personnel upon receipt of the phantom.<sup>15</sup> The spine phantom target contains four TLDs, whereas the lung target contains two TLDs. Each TLD is a double-load design, providing powder for two readings. The dose determination from two TLD readings is associated with a measurement uncertainty of 1.6% for both phantoms.<sup>29</sup>

### **3.2.2 Phantom irradiation**

Previously irradiated phantom results were selected for this study. For these irradiations, institutions were instructed to treat the phantom as they would a clinical patient, from imaging through radiation treatment delivery. Upon completion of treatment, the institution returned the phantom to IROC and submitted all associated DICOM RT data. The treatment was considered successful when the following criteria were met: doses measured with TLDs in the target were within  $\pm 7\%$  of the planned dose to a contour covering the TLD, and the average film pass rate was 85% of pixels meeting the specific phantom gamma-index criteria (spine gamma: 5% and 3 mm; lung gamma: 7% and 5 mm).<sup>27</sup>

### **3.2.3 Dose recalculation system**

A dose recalculation system was used to recalculate the plans originally created by the institutions to irradiate each phantom. The dose recalculation system (DRS) that we used was

Mobius 3D (v2.1.2, Varian Medical Systems). Mobius 3D was previously commissioned using data from more than 500 linear accelerators (LINACs)<sup>16,17</sup> collected from various institutions, as part of IROC's monitoring responsibilities. The system contains machine classes that represent the variety of Varian and Elekta manufactured machines, including specific subclasses for energy and multileaf collimator (MLC) configuration (regular vs high definition). Each machine class that was modeled in Mobius 3D was optimized to represent the average of that class in terms of PDD, jaw-defined output factors, MLC-defined output factors, and off-axis factors. Basic dosimetric values, when calculated by these Mobius3D beam models, were different from the reference data values by 0.27% on average.<sup>7</sup> This makes Mobius 3D well suited to perform accurate recalculations, as it is on par with institutions' TPSs which showed discrepancies of 0.36% on average compared to their linacs.<sup>30</sup> The dose algorithm used by Mobius 3D is an independently developed collapsed cone convolution/superposition algorithm.<sup>31</sup>

### **3.2.4 Data collection and analysis**

Phantom plans from over 1000 lung and over 250 spine phantom irradiations performed between July 2013 and July 2019 were selected for analysis from the IROC phantom database. Phantom irradiations before this time were excluded due to the unavailability of DICOM RT data; this exclusion also ensured that the study was more representative of current practice. A total of 258 phantom plans (172 lung and 86 spine) were selected for recalculation. A comparable number of phantom plans was selected for each year. Within each year, the phantom plans were randomly sampled to provide consistent representation over the period observed. The demographics of the selected plans are shown in Table 4.

**Table 4.** Demographics of the phantom sample set

	SPINE PHANTOMS	LUNG PHANTOMS
<b>TREATMENT MACHINE</b>		
Varian Truebeam	49	79
Varian Base Class <sup>a</sup>	19	56
Varian Trilogy	12	19
Elekta Agility	6	18
<b>TREATMENT TYPE</b>		
VMAT	71	114
Dynamic MLC	9	33
Step & Shoot	6	10
3D-CRT	N/A	19
<b>TPS ALGORITHM</b>		
AAA	51	109
Acuros	21	27
Convolution/ Superposition	11	32
Monte Carlo	3	4

<sup>a</sup>Varian base class machines: Clinac 21EX, Clinac 21iX, Clinac 23iX, Clinac iX, Clinac 2300CD<sup>16</sup>

Each plan's DICOM RT data included the dose, plan, structure, and CT image files. Each selected plan was classified by machine type, MLC type, and beam energy and a new dose distribution was recalculated by using the corresponding model in the DRS. This DRS recalculated dose and the institution's treatment planning system (TPS) dose calculation were compared with the measured dose for each TLD. These three values were compared using Equation 1 in order to determine which calculated dose, TPS or DRS, better predicted the actual delivered (TLD) dose to the phantom.<sup>7</sup>

$$D = \left( \left| 1 - \frac{TPS}{TLD} \right| - \left| 1 - \frac{DRS}{TLD} \right| \right) \times 100 \quad \text{Equation 1}$$

The dose difference value, D, was calculated for each TLD in each phantom plan (i.e., two for each lung phantom and four for each spine phantom), as well as an average value for each plan. A positive D value indicated that the DRS calculated value was closer to the

measured TLD value, whereas a negative D value indicated that the institution's TPS calculated value was closer to the measured TLD value.

The magnitude of the D value (i.e., how much better the recalculation performed) was assessed as a fraction of the TPS vs TLD dose deviation (i.e., how well the institution predicted the phantom dose) for each phantom recalculation. This allowed a comparison of the dose difference between planned and measured, versus how much better the DRS was compared to the TPS at accurately calculating that dose. For example, institution A's TPS vs TLD value was 10%. The D value is 7%, which means that on average, the DRS was 7% better at predicting dose to the TLD than the TPS. We can conclude that 7% of the 10% is accounted for by TPS error and so TPS error would be 70% in this case.

This magnitude was determined by using Equation 2.

$$\text{Percentage error} = D / \frac{|TPS - TLD|}{TPS} \times 100 \quad \text{Equation 2}$$

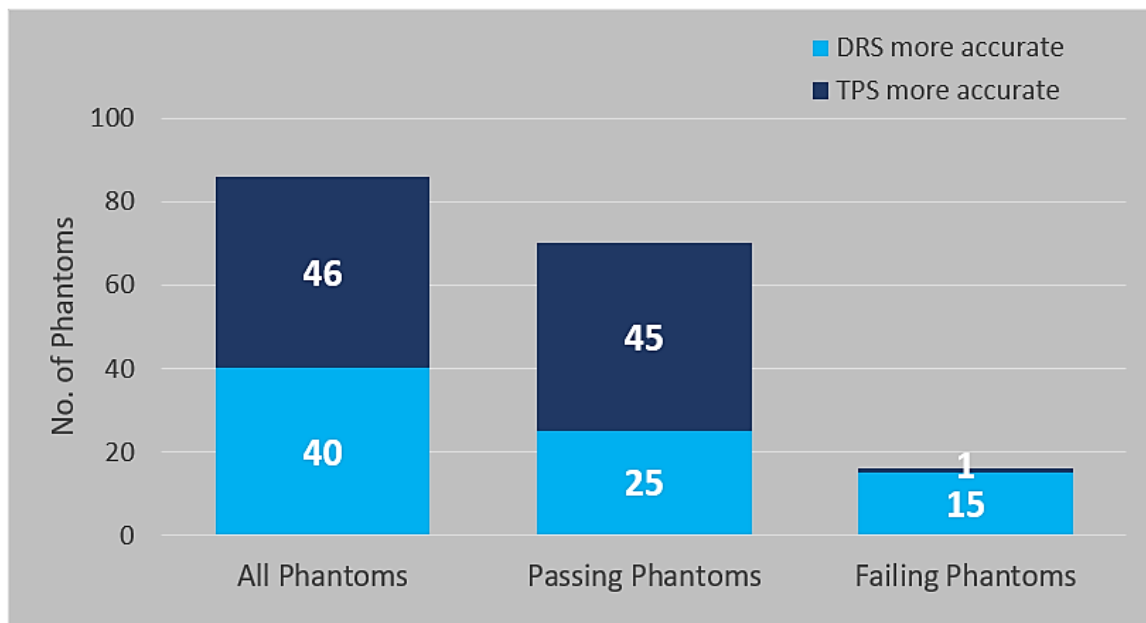
The TLDs with a negative D value, indicating that the TPS performed better, were excluded from this analysis because no dose calculation error was present.

### 3.3 Results

A total of 86 spine and 172 lung phantom treatment plans from irradiations performed between July 2013 and July 2019 were recalculated and assessed. The results were further subdivided into passing and failing plans in order to better understand the effects of dose calculation errors on these subgroups.

### 3.3.1 Spine

Among the 86 spine phantom results, 70 irradiations passed IROC criteria and 16 failed. Although the institution's TPS was expected to far outperform a generic DRS in all cases, a more even split was seen. The TPS was more accurate at calculating dose to the TLD in 46 cases (53%), whereas the DRS was more accurate in 40 cases (47%) (Fig. 3-2). Among passing phantoms, the TPS performed much better than the DRS, producing more accurate dose calculations in 64% of all phantom plans. Among failing phantoms, however, the DRS was more accurate in 94% of cases (Fig. 3-2). The difference in DRS performance between the passing and failing phantom cohorts (35% versus 94%) was highly statistically significant based on Fisher's Exact test ( $p < 0.0001$ ), and implicated beam modeling errors as a major contributor to failing spine phantom irradiations.

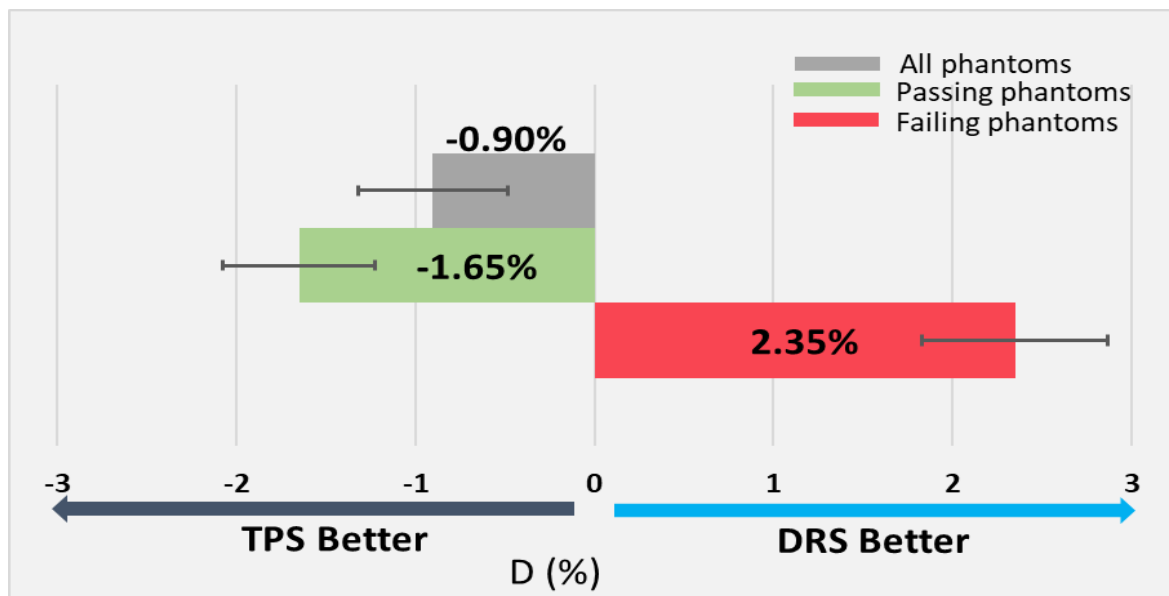


**Figure 5.** Spine phantom results showing the number of phantoms in which the TPS (dark blue) dose calculation was more accurate than the DRS (light blue) dose calculation.

The magnitude of difference between the TPS and DRS was next examined. The overall magnitude of the average dose difference for all spine phantom plans was found to be -0.90%

(Fig. 3-3). This indicates that, on average, institutions' TPSs were more accurate than the DRS was in calculating dose to the TLD by 0.90%. Among passing phantom irradiations, the TPS dose calculations were better than the DRS calculation by an average magnitude of 1.65%. However, among the failing spine phantoms, the DRS was found to be much more accurate at calculating dose to the TLD, by 2.35% on average (Fig. 3-3). A paired t test conducted between the average D values of the passing and failing spine phantom results showed statistical significance between the means ( $p < 0.001$ ).

The phantom TLDs (four in each phantom) were evaluated individually in a waterfall plot showing the dose difference (D) values, distinguishing phantoms that passed or failed the phantom irradiation. The majority of TLDs from phantoms that failed the irradiation criteria had a D value greater than 0, indicating that the DRS was more accurate than the TPS was at calculating the dose to these TLDs (Fig. 3-4).

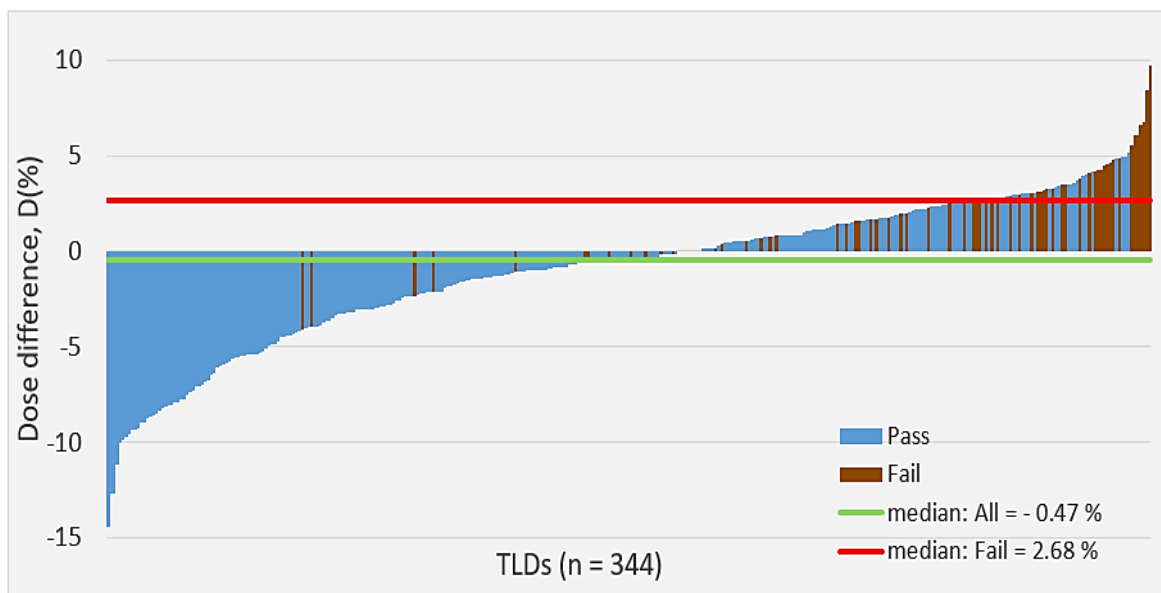


**Figure 6.** Average magnitude of dose difference values (D) present among subcategories of spine phantom plans.

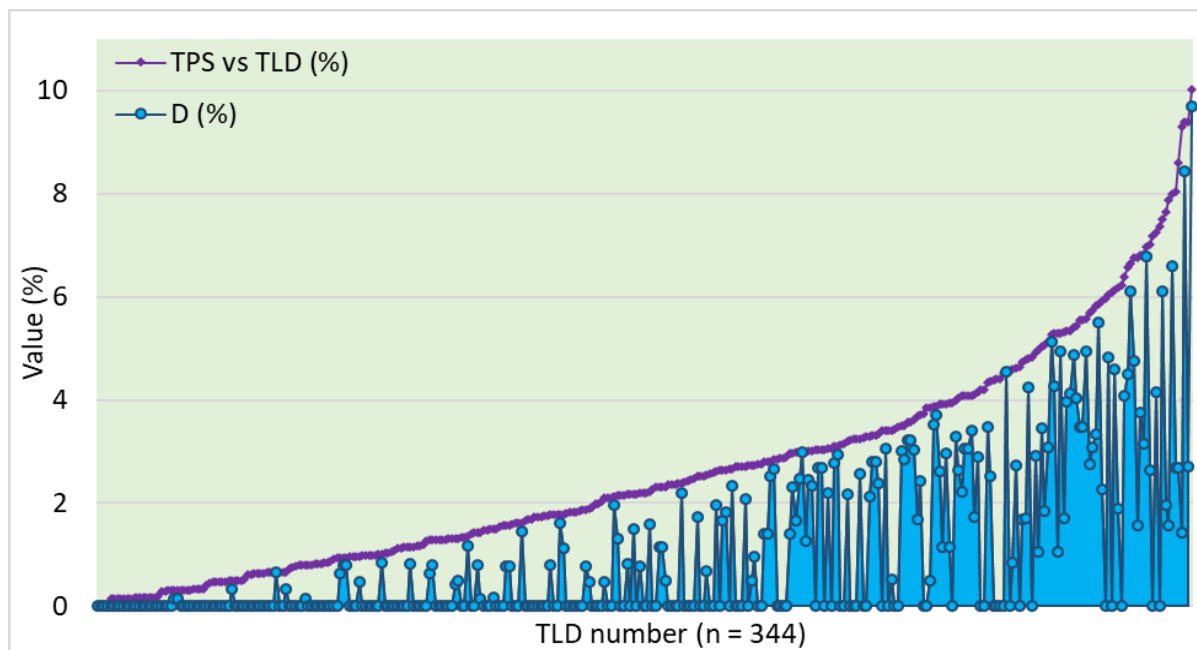
Overall, 52 (81%) of 64 TLDs from failing phantom results had a DRS dose calculation that was more accurate than the corresponding TPS dose calculation.

The fraction of the TPS vs TLD dose difference that could be attributed to inaccurate dose calculation was assessed for all spine phantom TLDs using Equation 2. TLDs with a negative D value were set to zero to indicate no dose calculation error present. The DRS recalculation accounted for 62% of the dose deviation between the TPS and measured TLD values. This can be seen in Fig. 3-5 where the blue shaded area represents 62% of the area under the purple curve. A direct correlation was found between the magnitude of the D value and the TPS vs TLD dose deviation. This is shown in Figure 3-5, where the value of D increased with the TPS vs TLD dose difference for the spine phantom irradiations (i.e., the larger the institution's error in predicting the dose to the phantom, the more improvement was offered by the DRS). A regression analysis was performed on this data and the slope of the relationship was found to be 0.56 ( $p < 0.001$ ), confirming the definite non-zero relationship between these two variables; as the TPS vs TLD dose difference grew, the magnitude of the D value became more positive (meaning DRS value is closer to TLD value).





**Figure 7.** Waterfall plot of spine target TLDs, distinguishing passing and failing individual phantom TLDs. The median for failing phantoms was much higher than that for all phantoms, showing an increase in dose calculation errors among failing spine phantoms.

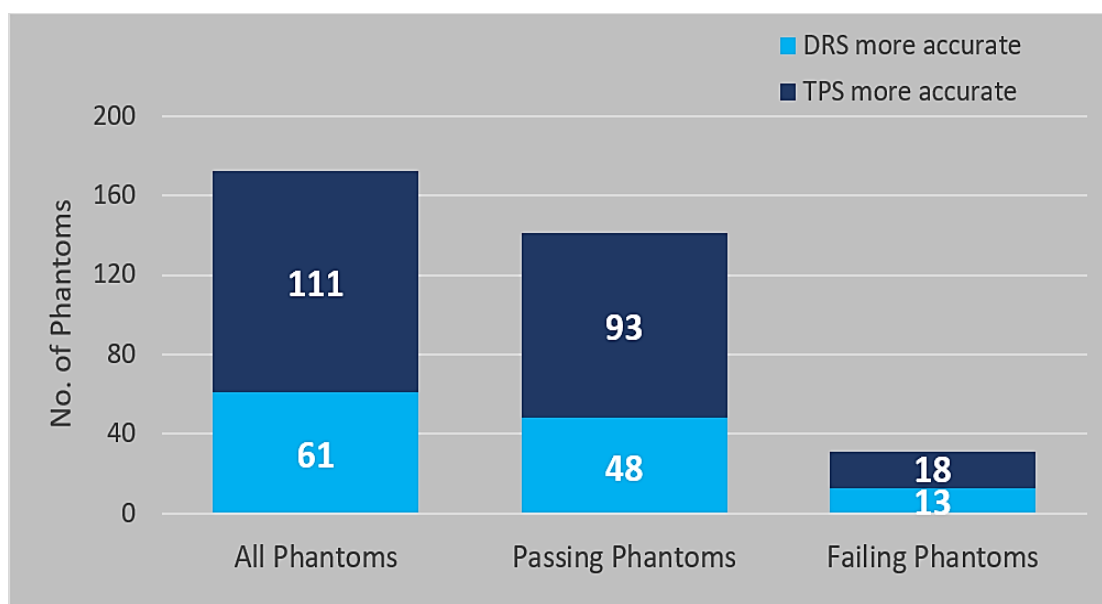


**Figure 8.** A combination plot showing the fraction of the TPS vs TLD dose deviation (area under purple line) that can be attributed to dose calculation error (blue shaded area). A direct correlation can be seen between the TPS vs. TLD dose difference and the magnitude of the dose difference value (D).

Evaluating the results by TPS algorithm, machine type and treatment type identified no statistically significant differences in the patterns of calculation accuracy. Similarly, an evaluation of phantom failing rate and dose calculation error rate as a function of time showed no difference in performance over time, based on a regression analysis ( $p = 0.96$ ).

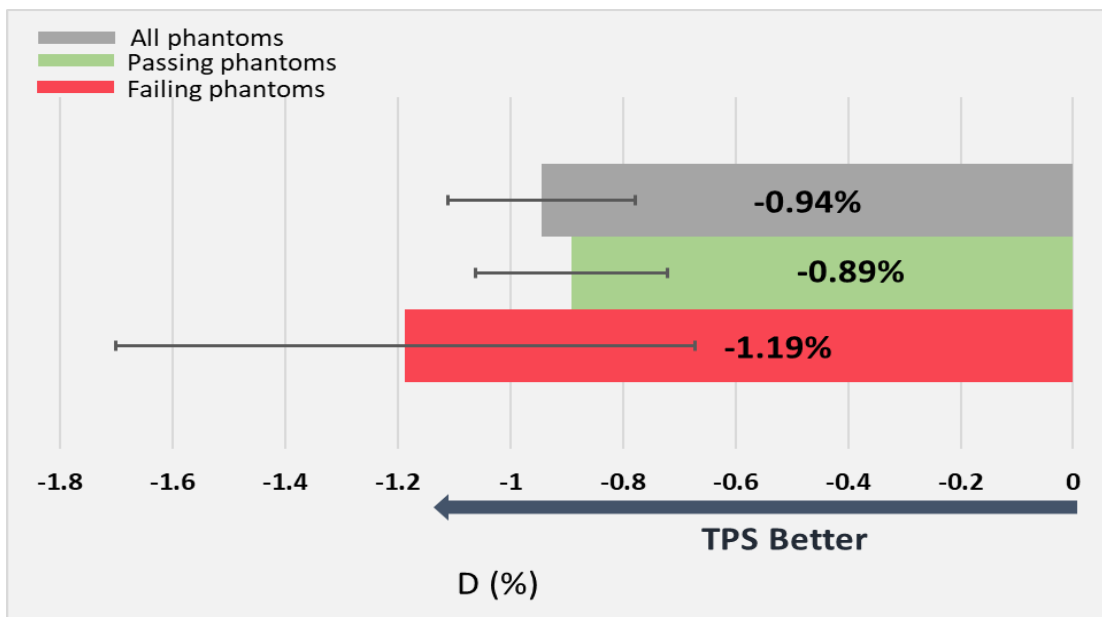
### **3.3.2 Lung**

Of the 172 lung phantom plans evaluated, the TPS was more accurate at calculating dose to the target in 111 plans (65%), whereas the DRS was more accurate in 61 plans (35%) (Fig. 3-6). These percentages were almost identical for the phantom plans with a passing result: the TPS was more accurate in 93 phantoms (66%), whereas the DRS was more accurate in 48 (34%). Notably, among the failing lung phantoms, the same general pattern was still seen. The TPS outperformed the DRS overall, with more accurate calculations in 18 phantom plans (58%) versus the 13 plans (42%) in which the DRS was more accurate, which was quite different from what was observed with the spine phantom. The difference in DRS performance between the passing and failing phantom cohorts (34% versus 42%) was not statistically significant based on a Fisher's Exact test ( $p = 0.41$ ).



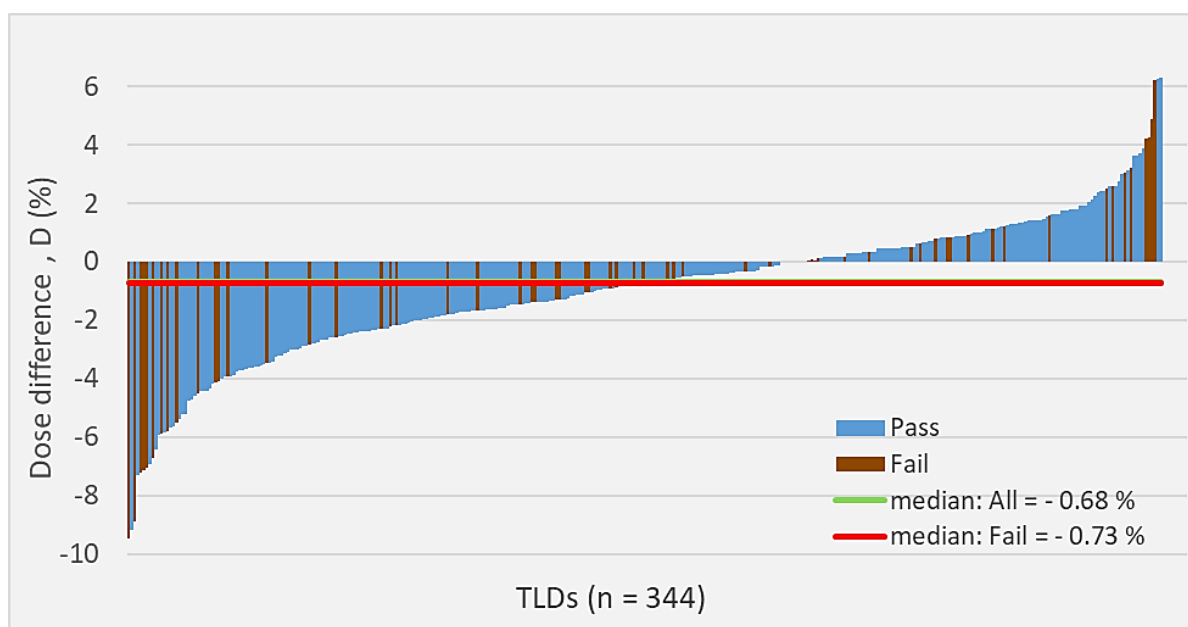
**Figure 9.** Lung phantom results showing the number of phantoms in which the TPS (dark blue) dose calculation was more accurate than the DRS (light blue) dose calculation.

The magnitudes of the D value are shown in Figure 3-7. For each of the lung phantom subcategories of all, passing and failing phantoms, this value was negative, indicating that overall, the TPS outperformed the DRS in calculating dose to lung phantom TLDs. The average magnitudes were - 0.94% for all phantom plans, -0.89% for passing phantoms, and -1.19% for failing phantoms (Fig. 3-7). A paired t test conducted between the average D values of the passing and failing spine phantom results showed no statistical significance between the means ( $p = 0.76$ ).



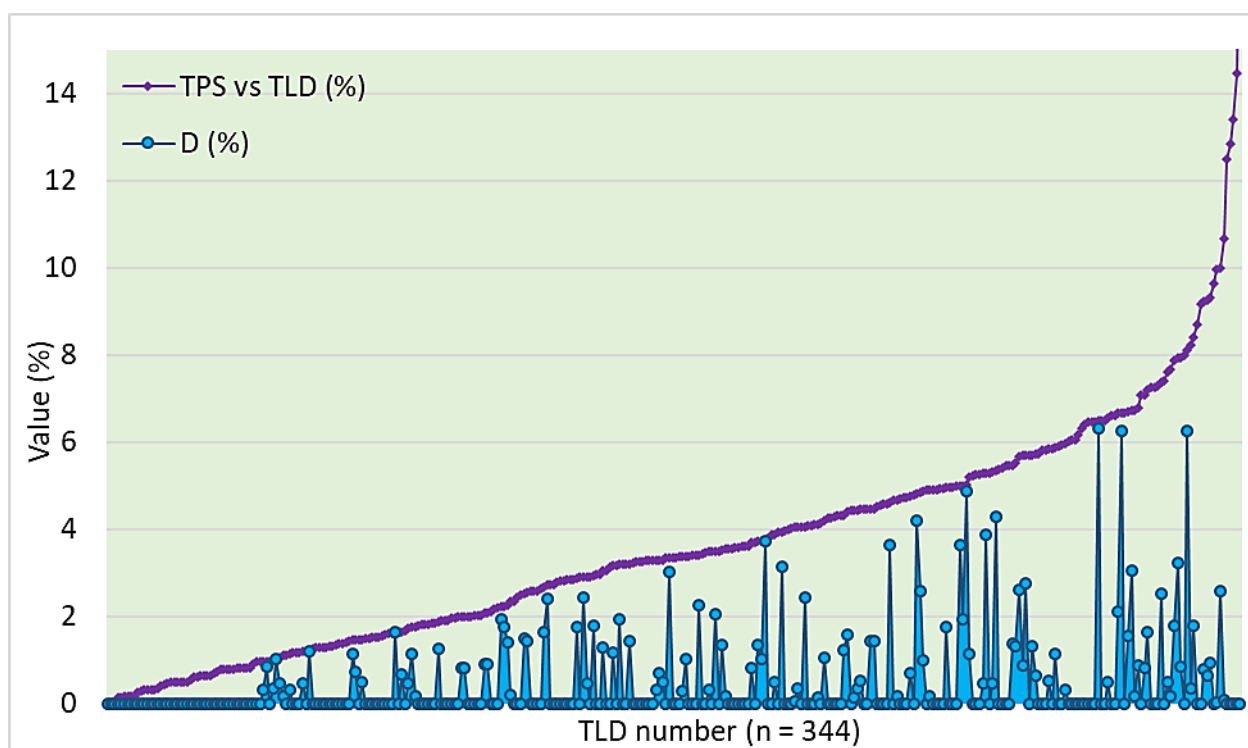
**Figure 10.** Average magnitude of dose difference values (D) present among subcategories of lung phantom plans.

The phantom TLDs (two in each phantom) were evaluated individually in a waterfall plot showing the dose difference (D) values, separated by those phantoms that passed and those that failed. The majority of failing TLDs had a D value less than 0, indicating that the TPSs were more accurate at calculating dose to the TLDs than the DRS was (Fig. 3-8). Upon further evaluation of the failing lung phantom TLDs, we found that the majority of individual TLD results (66%) had a more accurate TPS dose calculation value compared with that of the DRS.



**Figure 11.** Waterfall plot of lung target TLDs, distinguishing passing and failing individual phantom TLDs. The median for failing phantoms was similar to all phantoms, showing no increase in dose calculation errors among failing lung phantoms.

The fraction of the TPS vs. TLD dose difference that could be attributed to inaccurate dose calculation was assessed for lung phantom TLDs using Equation 2. TLDs with a negative D value were set to zero to indicate no dose calculation error present. The DRS recalculation accounted for only 13% of the dose deviation between the TPS and measured TLD values. This can be seen in Fig. 3-9 where the blue shaded area represents 13% of the area under the purple curve. A very weak correlation was found between the two variables, where the value of D did not show much change as the TPS vs. TLD discrepancy increased. A regression analysis of this relationship yielded a slope of 0.08 ( $p \ll 0.001$ ), confirming a weak correlation between these two variables. The magnitude by which the DRS outperformed the TPS in calculating dose to the TLD was therefore unaffected by the magnitude of TPS dose calculation disagreement.



**Figure 12.** A combination plot showing the fraction of the TPS vs TLD dose deviation (area under purple line) that can be attributed to dose calculation error (blue shaded area). A very low correlation can be seen between the TPS vs. TLD dose difference and the magnitude of the dose difference value (D) for individual lung phantom TLDs.

Evaluating the results by TPS algorithm, machine type and treatment type yielded similar patterns of calculation accuracy with no statistically significant differences. We also evaluated failing rate and dose calculation error rate as a function of time; there was also no difference in performance over time based on regression analysis ( $p = 0.27$ ).

### 3.4 Discussion

The effects of TPS dose calculation accuracy on phantom irradiation success, among the IROC lung and spine phantom results, were evaluated in this study. The DRS used is a representation of the average LINAC class. This DRS was commissioned by using data from a large number of machines from various institutions<sup>16,17</sup> and was therefore not specially

optimized for any particular institution's LINAC. This means that the DRS should not outperform any institution's TPS that was presumably commissioned for that specific LINAC, i.e., custom-tailored to their local machine. And overall, this was the case. For both the spine and lung total phantom cohorts, the DRS system performed poorer than the institution's TPS (there was a negative D value, indicating the institution's TPS was more accurate on average). However, this was not the case when failing spine phantoms were examined and the DRS significantly outperformed institutions' TPSs.

The spine phantom results showed that TPS dose calculation accuracy has a direct effect on phantom irradiation success. This was evidenced by the fact that the DRS, the generic average dose calculation system, was significantly more accurate at calculating the dose to the TLD than the institution's own TPS was among the failing phantom plans. The magnitude of this effect was often substantial, indicating possible direct impacts to patient outcome due to these dose discrepancies. The spine phantom dose difference value (D) magnitude showed a direct correlation with the magnitude of an institution's TPS vs measured TLD dose value (Fig. 3-5). This means that the greater the percentage of dose deviation present in a spine phantom irradiation, the better the DRS was at accounting for this difference by performing more accurate TLD dose calculations than the institution's TPS. This reinforces the fact that dose calculation errors are present among spine phantoms and show up more as the phantom irradiation failure becomes more egregious.

The lung phantom results, in contrast, showed very few dose calculation errors throughout, with the institutions' TPS outperforming the DRS in accurately calculating dose in both the failing and passing lung phantom plans. In addition, the dose difference value (D) showed no correlation with the magnitude of the TPS vs measured TLD dose difference (Fig. 3-9). This means that the magnitude of failure seen among lung phantom plan results could be

only slightly attributed to dose calculation inaccuracies.

The presence of dose calculation errors among the spine phantom plans is reasonable. This phantom consists of a target that rests on an OAR, the spinal cord, and thus presents a region of high gradient dose, with more intense optimization requirements. The spine phantom plans are therefore typically very highly modulated and leave more room for error in dose calculation as well as dose delivery. This is in contrast to the lung phantom, which presents a more isolated target structure located far enough away from sensitive OARs and therefore has less stringent optimization requirements.

In previous studies evaluating the types of errors that exist among IROC lung and spine phantoms, the majority of errors associated with the spine phantom were dosimetric in nature. Dosimetric errors made up 60% of all spine phantom failures.<sup>27</sup> In contrast, only 22% of all errors associated with lung phantoms were dosimetric in nature. The majority of lung phantom failures came from localization errors, which is reasonable since more than 75% of lung phantoms were irradiated using a motion table, making localization a major focus for this phantom.<sup>27</sup> The IMRT head and neck phantom was found to have primarily dosimetric errors as the failure mode, accounting for 62% of all head and neck phantom failures.<sup>6</sup> These failures were also overwhelmingly tied to dose calculation errors,<sup>7</sup> which is very similar to those of the SBRT spine phantom, but substantially different from the lung phantom. This comparison reinforces the differences in causes of failure between highly modulated treatments and those with low modulation.

The dose calculation errors that manifested in this study are concerning. These represent inherent errors within the TPS, the same TPS that is used to create and compute patient plans. The dose calculation discrepancies seen in these phantoms are therefore expected to manifest in patient plans, where they carry far greater consequences. Dose calculation errors have been



shown to make up over half of the errors present in the spine and head and neck phantoms. However, more investigation is required in order to understand what aspects of the TPS dose calculation are going awry. Recent work has highlighted errors in basic modeling of dosimetric parameters such as percentage depth doses (PDDs) and off-axis factors,<sup>32</sup> as well as substantial variability in non-dosimetric parameters such as dosimetric leaf gap.<sup>33</sup> Other factors, including non-calculation issues are also involved in phantom failures. Localization, delivery, and machine calibration error may all contribute as well. Ongoing research is being conducted to quantify these other factors in order to create a comprehensive picture of the causes of errors that exist in the IROC phantoms and, correspondingly, in patient treatments.

Being able to distinctly determine the areas that are lacking in the phantom irradiation process will be very beneficial to the radiation oncology community. This information will also be beneficial to IROC, by improving the phantom credentialing process, in terms of troubleshooting when institutions fail to meet criteria, and highlighting which phantoms are able to test specific challenges in the radiation therapy process. This work will also translate to the broader clinical practice by highlighting the areas that require extra care and attention when treating specific disease sites. In particular, it highlights that care in beam modeling is critically important.

### **3.5 Conclusions**

IROC lung and spine phantom plans were recalculated using an independent dose recalculation system, Mobius 3D. Dose calculation deficiencies in institutional treatment plans were found to exist in the spine phantom plans, especially in those that did not meet the irradiation criteria. Although institutions' TPSs performed more accurate dose calculations than

the DRS did among passing spine phantom plans, the DRS performed significantly better among the failing spine phantom plans, calculating dose to the target more accurately in 94% of failing cases. The lung phantom plans displayed a different result, with institutions' TPS values being more accurate than the DRS values among both the passing and failing lung phantom plans. This is an indication that very few dose calculation errors exist among lung phantom plans, and these errors are not directly correlated to the institutions' unsuccessful irradiation of this phantom. These results showed contrasts that emphasize the different requirements for the different plan types: highly modulated spine SBRT vs. less modulated lung treatments.

Particularly for highly modulated treatments, great emphasis should be placed on beam model validation in the TPS in order to ensure accurate dose calculation of patient plans.

## Chapter 4: IROC phantoms accurately detect MLC delivery errors

This chapter is based on the following publication:

**Edward S.S.**, Pollard-Larkin J.M., Balter P.A., Peterson C.B., Howell R.M., Kry S.F., “IROC phantoms accurately detect MLC delivery errors”, To be submitted for publication, 2022

### 4.1 Introduction

The Imaging and Radiation Oncology Core (IROC) phantom program has been used as an end-to-end test of the patient radiation treatment process for decades both in the US and abroad. Several phantoms comprise this program, including the IMRT head and neck (H&N), SBRT spine and moving lung. Upon return of the phantom post irradiation, treatment delivery accuracy is assessed using plan DICOM files and measured dose readings. Recent 2021 IROC phantom results indicate a phantom failure rate of 12% and 22% for the H&N and spine phantoms respectively. This data includes hundreds of institutions mainly across the US, and therefore has the potential to impact the accuracy of RT treatment for thousands of patients.

In previous work, we showed that dose calculation errors played a major role in phantom failures and dose deviations seen among H&N and spine phantom irradiations (although not as prevalent among the lung phantoms<sup>6,34</sup>). In particular, the MLC modeling parameters (DLG or MLC-offset) were critical to accurate dose delivery<sup>35</sup>. While TPS errors are a clear component to poor phantom performance, other failure modes contribute as well. Therefore, in this work we examined how delivery errors, particularly MLC errors, could impact the phantom result.

MLC errors

An MLC leaf error occurs when there is a difference between the intended or planned position and the actual position of the MLC leaf during plan delivery on the machine. To explore the extent to which these errors manifest in the IROC phantoms, we evaluated the following:

First, MLC errors have been documented to be 0.46 mm on average for volumetric modulated arc therapy (VMAT) plans and 0.32 mm for dynamic (DMLC) treatments<sup>36</sup>. Studies have shown that these can cause deviations between the planned and delivered doses in patient cases. Systematic MLC errors, such as offsets in the MLC leaf bank have been shown to cause changes to the planning target volume (PTV) dose of 3.7% and 7.2% for combined offsets of 1 mm and 2 mm respectively<sup>37</sup>. Random leaf error effects were found to be smaller, but nevertheless causing dose errors of up to 1.6% overall<sup>38</sup>. In addition to the magnitude of dose impact from such MLC errors, it was found that the magnitude of MLC error varied by treatment type but remained consistent between institutions<sup>36</sup>. While these delivery errors have been examined in patient cases, it is unclear how such effects would manifest in a standardized credentialing phantom.

Second, IROC phantoms are all irradiated to a single fraction dose of approximately 6 Gy per current IROC protocols<sup>6,26</sup>. This is to optimize the dosimetry, but this fraction size is different from clinical fraction doses which are typically 2 Gy for IMRT H&N and 27 Gy for SBRT spine treatments<sup>39–41</sup>. During the delivery of dMLC and VMAT treatments, the MLC is constantly moving, and this motion is influenced by the number of MUs to be delivered. When the dose per fraction is decreased (for the same treatment plan) the mean MLC speed increases, which could result in a greater chance of MLC positional error. It is unclear how the difference in fraction size corresponds to differences in delivery accuracy for IROC's credentialing phantoms, especially for different generations of accelerator.

This work sought to provide insight into the above two questions. We first evaluated the impact of delivery errors on IROC phantom irradiations. Second, we examined the differences in the IROC phantom fraction size compared to clinical fraction sizes. The results of this experiment will provide a better understanding of the effects of these types of errors on the phantoms, which will help improve the assessment process for institutional irradiations in the phantom program moving forward.

## 4.2 Materials and Methods

### 4.2.1 Phantoms & plans

The 2 phantoms used for this study were the IROC IMRT head & neck (H&N) and SBRT spine anthropomorphic phantoms. Both phantoms are composed of tissue-equivalent materials and are used as an end-to-end test of patient radiation treatment. To improve dosimetric precision, the phantoms were modified so that the inserts held Exradin® A1SL ion chambers in the target and OAR structures instead of the usual TLDs (Figure 13).



**Figure 13.** Phantoms with ion chamber inserts and A1SL chambers inside: Thoracic (spine) (left), IMRT head and neck (right)

The H&N phantom consists of 2 PTVs, both 5 cm long. The primary PTV representing a tumor, is 4 cm in diameter and holds 2 ICs (anterior and posterior). The secondary target representing a lymph node, is 2cm in diameter and holds 1 IC. One chamber is also inserted

into the OAR (cord)<sup>15,26</sup>. The spine phantom represents the human thorax, with a 22 cm<sup>3</sup> target sitting on the vertebral body, approximately 0.8 cm from the spinal cord. The spine insert accommodates 1 IC in the PTV and 1 in the OAR. (Fig. 13).

All plans were created by the same individual, using the Raystation 10B (Raysearch®) treatment planning system. Both step and shoot (sMLC) and VMAT plans were created for each phantom, following the IROC irradiation dose prescription of 6.6 Gy for the H&N and 6 Gy for the Spine. Plans were all created and were optimized to spare organs at risk while meeting the prescription dose to 95% of the PTV.

#### **4.2.2 Introduced delivery errors**

MLC errors were introduced into the phantom plans to explore the impact of delivery errors on the standardized phantom audit. Various combinations of random leaf and whole leaf bank MLC shifts were introduced to each plan using the scripting function in Raystation, based on the Python programming language. A random number generator was used to select random leaves to shift on either leaf bank. The leaf shift distance was then randomly picked from the following: -2, -1, -0.5, 0.5, 1- or 2-mm. Shifts were introduced to either 25%, 50%, 75% or 100% of leaves for each plan iteration, and were implemented only among leaves that span the width of the PTV. This was done to localize these effects to the measurable areas of the phantom (i.e., where the ICs are located). For the sMLC plans, shifts were assigned randomly to leaves in each segment of each beam. Shifts in the VMAT plans were assigned to a random leaf in the first segment and that shift was maintained through all segments of the beam, to simulate a lagging leaf scenario. Whole bank leaf shifts were introduced as either 0.5, 1- or 2-mm total shift outwards (opening of banks), with each bank moving half the distance. All plan doses were recalculated with the new (“erroneous”) MLC positions, resulting in 32 total plans

(8 VMAT and 8 sMLC for each phantom). Average dose values for the PTV and cord ion chambers were recorded from Raystation for both the original and altered plans.

#### **4.2.3 Different prescription sizes**

Additional VMAT plans were created for the Truebeam and Varian 21Ex machines following a typical clinical single fraction prescription of 2 Gy for the H&N and 27 Gy for the Spine<sup>39–41</sup>. To maintain consistency in the plan complexity, the plan doses were then scaled to the IROC prescription and various complexity metrics were verified to ensure that plans remained consistent after scaling. This resulted in 8 total plans (2 for each machine, for each phantom).

#### **4.2.4 Delivery and Analysis**

The two machines used in this work are the Varian Truebeam (Varian Medical Systems, Palo Alto, CA) with high-definition MLCs (HDMLC) and the Varian 21EX linear accelerators. The Truebeam is Varian's newest generation of the linear accelerator and consists of better integrated digital technology than the older C-series machines (which includes the 21EX), such as the Maestro electronic controller system<sup>42</sup>. Both machines are located at the MD Anderson clinic in Houston, Texas, where they are used for patient treatments, hence regularly checked, and serviced as part of routine clinical operations at the clinic. The plans for the first part of this study which examines random vs whole bank MLC errors were irradiated on the Truebeam machine, while the plans with varying prescription sizes were irradiated on both machines, to identify any differences that exist between the two generations of technology determine differences between them.

**Table 5.** MLC characteristics for Varian Truebeam and 21EX machines

<b>Machine</b>	<b>Inner MLC pairs</b>	<b>Outer MLC pairs</b>
<b>Varian Truebeam (HDMLC)</b>	0.25 cm (n = 32)	0.5 cm (n = 28)
<b>Varian 21EX</b>	0.5 cm (n = 40)	1.0 cm (n = 20)

Delivery log files were collected for each plan and processed, which provides details of dose delivered as well as RMS errors detected in the MLCs. The impact of both random and systematic (whole bank) leaf errors on the phantom plans were examined by comparing TPS-original to TPS-perturbed plan doses. The treatments were delivered, and log files processed using the Mobius Fx software (V2.1.2), to confirm these TPS predicted dose deviations. The dosimetric impact of fraction size was examined by comparing measured vs TPS doses and the log files were analyzed to assess MLC RMS error differences between the pairs of plans (original vs scaled).

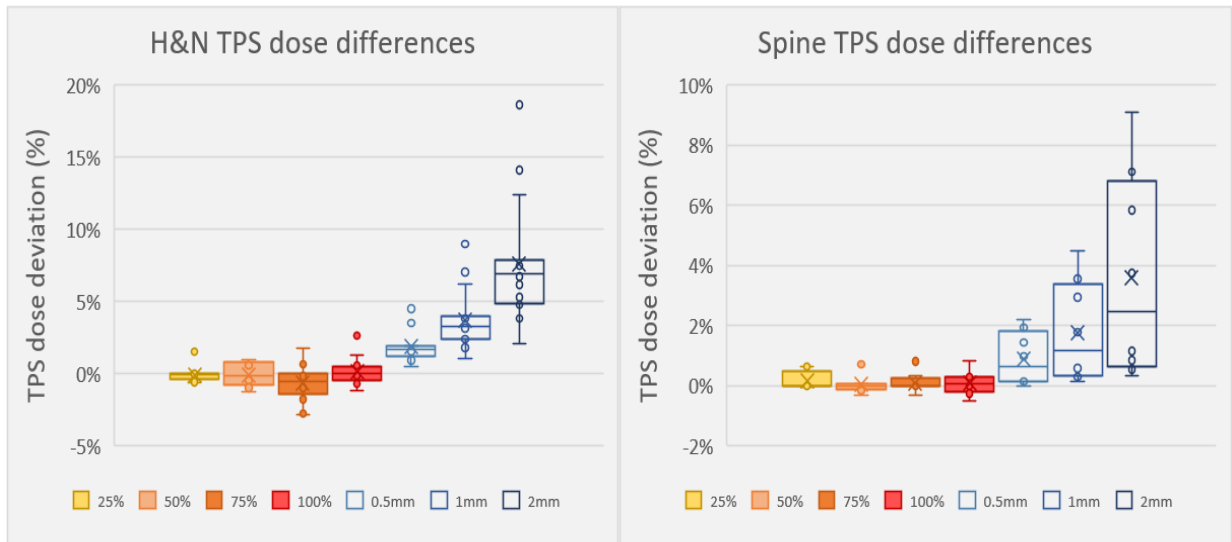
## **4.3 Results**

### **4.3.1 Impact of leaf errors on phantom dose**

TPS dose comparisons between plans showed dose deviations for both error types. Random MLC shifts introduced into the phantom plans resulted in relatively modest dose deviations, particularly in the PTV of up to -2.8% (cord: also -2.8%) for the H&N phantom and 0.7% (cord: 0.8%) for the spine. Systematic shifts of the entire leaf bank resulted in much larger deviations. As the leaf bank offset increased, the phantom showed incrementally increasing doses, with 2 mm shifts leading to PTV dose differences in the H&N phantom of up

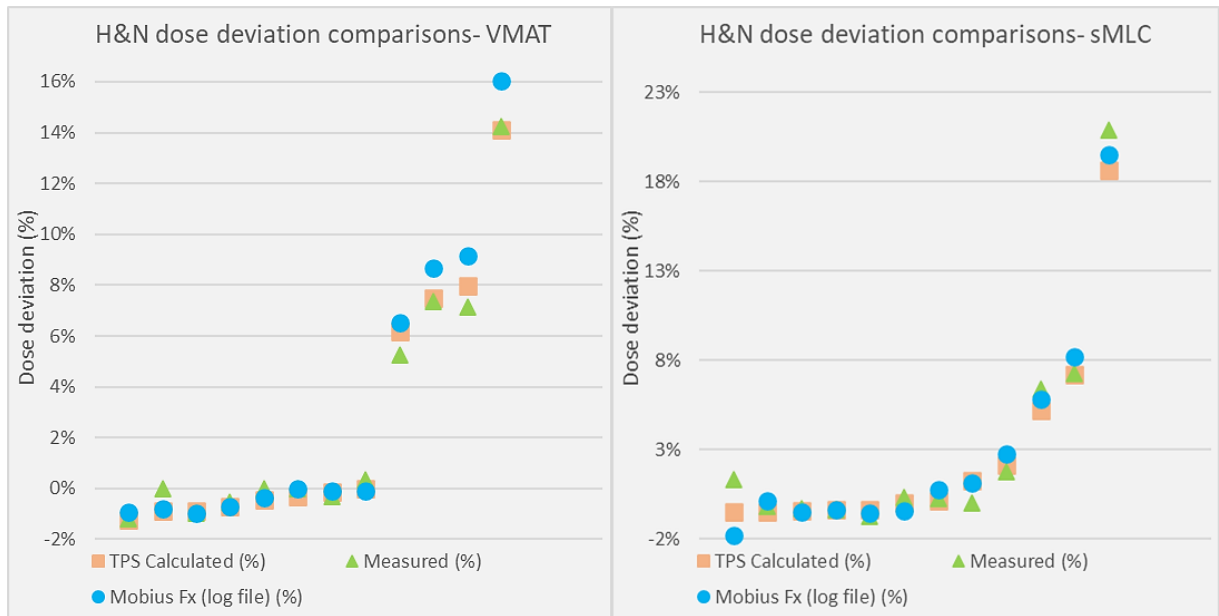


to 8% (cord: 8.7%) and 7.1% (cord: 9.1%) in the spine (Fig 14).

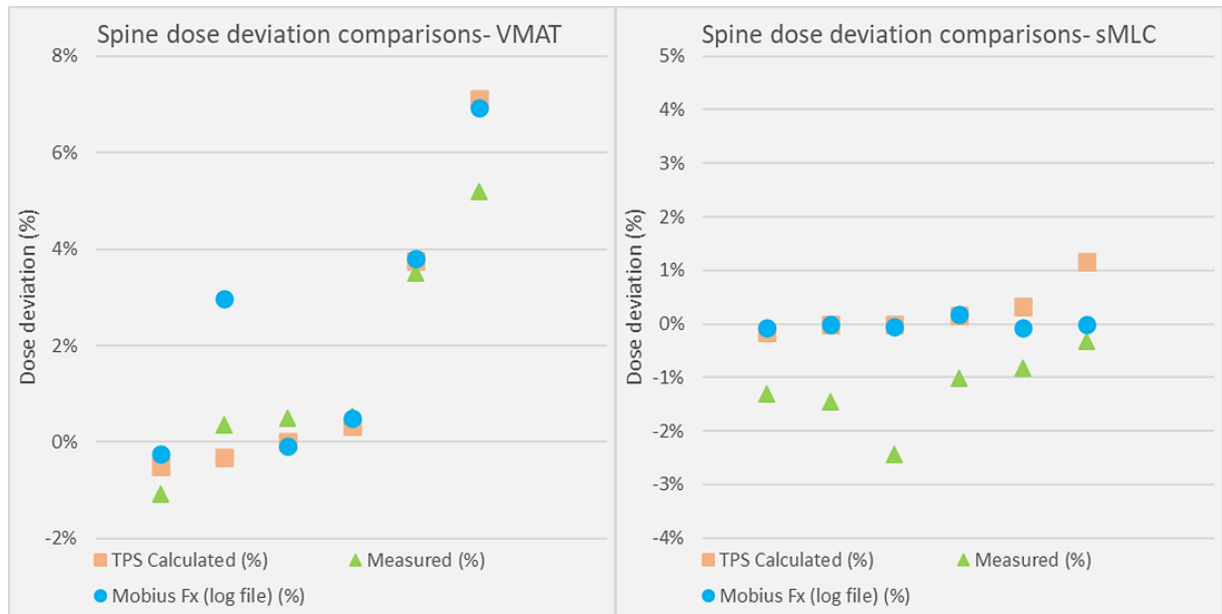


**Figure 14.** Dose deviations of phantom plans due to random and systematic MLC error in the TPS. The solid boxes represent the random shifts, and the hollow ones represent the systematic (whole bank) MLC shifts.

These TPS-predicted dose deviations between the plan pairs (TPS-original to TPS-perturbed) were compared to the measured (IC) and delivery log file (MFx) dose deviations for both phantoms. These comparisons were made for each ion chamber, for each phantom, to get more detailed results across the phantom geometry. For the H&N, dose deviations between TPS, IC and MFx doses were within 2% of each other in 21 of 24 sampled points across the PTVs and OAR (Fig. 15). For the spine, dose deviations were within 2% of each other in 10 of 12 sampled points (PTV & OAR), and all points within the PTV (Fig.16). Overall, the average disagreement between the TPS and measured was 0.72%, while the average disagreement between the delivery log calculation and measured was 0.89%. The agreement between the TPS and measured perturbations indicate that the introduced errors were reasonably well modeled in the planning system. The agreement between these perturbations and the delivery log file predictions emphasizes that these errors were accurately captured with the delivery log system.



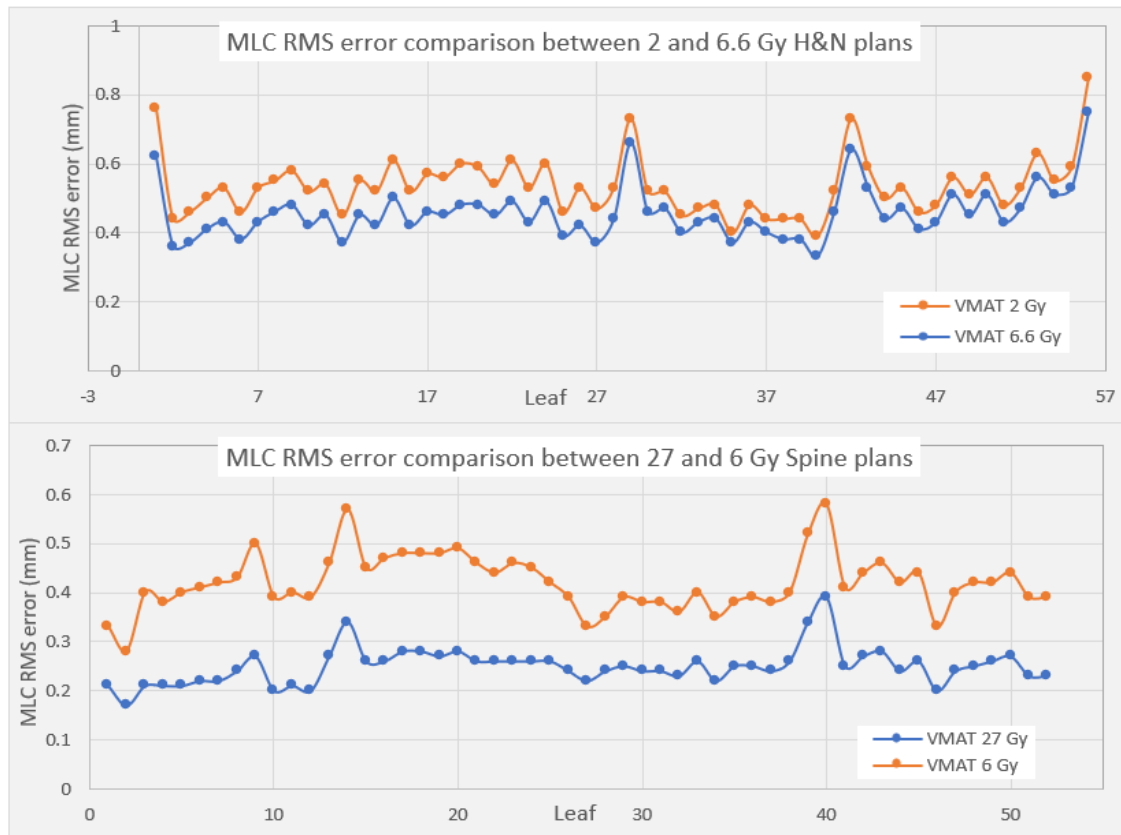
**Figure 15.** Dose comparisons between TPS, measured (ion chamber) and delivered (log file) doses for the H&N phantom, showing differences in recorded dose deviation across all 3 modes. Dose deviations represent difference from the original unperturbed plan, for the 3 plans that were irradiated: 50%, 100% and 2mm MLC shifts. Each datapoint represents an ion chamber point on the phantom; 3 plans each for VMAT and sMLC, with 4 ICs each.



**Figure 16.** Dose comparisons between TPS, measured (ion chamber) and delivered (log file) doses for the spine phantom, showing differences in recorded dose deviation across all 3 modes. Dose deviations represent difference from the original unperturbed plan, for the 3 plans that were irradiated: 50%, 100% and 2mm MLC shifts. Each datapoint represents an ion chamber point on the phantom; 3 plans each for VMAT and sMLC, with 2 ICs each.

### 4.3.2 Fraction size

For the Truebeam plans, there was very little MLC RMS error found in any plan, regardless of the fraction size. All RMS error was within 0.05 mm for both dose-per-fractions in both the H&N and spine phantoms. The plans delivered on the 2100Ex machine however, showed differences in RMS error between plan pairs. The H&N 2 Gy plan had 17% more RMS error on average than the 6.6 Gy plan and the spine showed an even greater difference, with the 6 Gy plan having 68% more error than the 27 Gy. Leaf RMS error showed the same pattern of distribution throughout in each plan pair, with a distinct increase in the magnitude of error in the plans with a smaller prescription (Fig. 17).



**Figure 17.** Distribution of all MLC RMS errors for the two VMAT plan pairs for the H&N (top) and Spine (bottom) phantoms, irradiated on the Varian 2100EX machine.

The dosimetric impacts of these MLC RMS errors were, however, negligible, in all cases. When assessing the dosimetric accuracy of plan delivery, by comparing the delivered doses (from the log files) to the TPS dose, the 2 H&N plans had a maximum difference in accuracy of 1% and the 2 spine plans had a maximum difference of 0.1%. The details of the dose differences between the low dose-per-fraction and high dose-per-fraction treatments are shown in Table .

**Table 6.** Comparisons of plan delivery accuracy between the two plan prescriptions, showing similar plan delivery accuracy despite varying prescription size.

<b>HEAD &amp; NECK</b>	<b>Delivery dose deviation from TPS</b>		
	<b>6.6 Gy</b>	<b>2 Gy</b>	<b>Difference in accuracy</b>
Primary PTV (1)	0.0%	0.0%	0%
Primary PTV (2)	-0.1%	0.0%	0.1%
Secondary PTV	0.0%	0.0%	0%
Cord	-0.1%	0.9%	1%
<b>SPINE</b>	<b>27 Gy</b>	<b>6 Gy</b>	
PTV	-1.3%	-1.4%	0.1%
Cord	0.9%	0.8%	0.1%

#### 4.4 Discussion

The IROC H&N and Spine phantoms were used with slim line ion chamber inserts to test the IROC H&N and spine phantoms' ability to capture MLC errors during treatment delivery. All the measurements, TPS, and log file recalculations agreed within 2% of each other across all plans and deliveries.

Through measurement and delivery log file analysis, we recorded dose deviations for both random and whole bank MLC leaf shifts in phantom plans. Random errors showed very little dose deviations, as was the case with clinical plans in the literature. Our study found average dose variations in the PTV of -2.8% to 2.7%, which corresponds reasonably with the results from prior studies, where dose variations of -1.5% to 1.6% were found in the pCTV for H&N plans<sup>38</sup>. In contrast, substantial dose deviations were recorded for whole bank shifts. For the H&N plans, systematic MLC offsets introduced errors of 8% in the PTV and up to 18% in the cord. In clinical scenarios, similar errors were found by two separate studies which showed dose differences of up to 7.6%<sup>38</sup> and 8.8%<sup>37</sup> for H&N plans with 2mm total MLC whole bank shifts.

Spine SMLC whole bank shifts recorded much smaller dose deviations in this study than the H&N plan. This was due to the nature of the aperture shapes in the plans, where most leaves were either almost fully open or leaves from one bank closed all the way across the PTV. Therefore, when whole bank shifts were introduced to the plan, they did not have much of a dosimetric effect on the PTV or Cord doses. The pattern of agreement with clinical cases validates the phantom's ability to not only broadly identify when the delivered dose is different from planned, but also associate the degree of dose deviation with a certain type of MLC error.

The delivery log file analysis in Mobius Fx recorded dose deviations similar to the measured and TPS doses and can be reliably utilized as a secondary form of analysis for the IROC phantoms. Limitations do exist however, where log files may not detect errors caused by poor machine calibration, or sub-optimal motor performance as was previously found in a separate study<sup>43</sup>. These scenarios may result in the machine assuming that the leaves are correctly positioned, which would cause them to be logged as such in the delivery log files, thereby missing the error. If these errors were to occur, they would manifest in the measured

phantom dose deviations but cause disagreement with the delivery log files. This highlights the need for proper calibration of machine components such as MLCs, and independent end-to-end testing to verify accuracy.

Analysis of VMAT plans with varying prescription doses demonstrated increased MLC rms errors for the plans with smaller prescriptions. This was due to the MLC leaves having to move more quickly because of lower MUs and shorter delivery time for the gantry to circumnavigate the entire arc. Speeds were found to be up to 4 times faster for the spine 6 Gy versus the 27 Gy plan, and that resulted in 68% more rms error. Results suggest that when using older Varian machines, irradiating the SBRT phantoms such as the spine, at the 6 Gy IROC prescription dose is a more rigorous and robust test of the clinical prescription. This however would be the opposite for the IMRT H&N plans that are usually administered clinically at much lower fraction sizes of 2 Gy. While these rms errors can be captured using the IROC phantoms, the dose effects were negligible for both plans and therefore did not create enough of a dose discrepancy to cause clinical impacts. Given that these plans were irradiated on a well-maintained machine, these results represent the best-case scenario for errors. Dose deviations may be greater on machines with a less rigorous QA and maintenance schedule, leading to greater errors that may have clinical impacts.

This analysis of delivery errors can be used as an additional tool during IROC's phantom assessment to help pin-point specific causes of phantom failure related to plan delivery. A dose recalculation based on the log files can be used as a further step of investigation, when institutions have dose discrepancies in their phantom results.

## **4.5 Conclusion**

We tested the IROC H&N and spine phantoms' ability to detect MLC errors, both manually introduced and as a result of varying MLC speeds caused by smaller VMAT prescription doses. Dose deviations were assessed in the TPS, ion chamber measurements and delivery log files. Dose deviations observed in the phantom were consistent with reported deviations observed in clinical cases, reinforcing that this phantom is a suitable patient surrogate and appropriate for clinical trial credentialing and QA. The fraction size for phantom delivery versus clinical delivery had an impact on the MLC positioning accuracy; greater RMS error was recorded in phantom plans with a smaller prescription dose (particularly for older machines). Dosimetrically however, on well-maintained machines at least, these differences in MLC positions did not cause clinically significant dosimetric changes.

## Chapter 5: Quantifying delivery errors in IROC phantoms

This chapter is based on the following publication:

**Edward S.S.**, Peterson C.B., Pollard-Larkin J.M., Balter P.A., Howell R.M., Kry S.F.,  
“Treatment delivery errors as a component of IROC phantom dose deviations”, To be  
submitted for publication, 2022

### 5.1 Introduction

Radiation therapy is a complex process. Numerous steps are involved as a patient goes from imaging via CT or MRI, to receiving radiation treatment, and as with any multi-stage process, there is the ability for errors to occur at each stage. At IROC, the magnitude of these errors is explored using the end-to-end anthropomorphic phantom program<sup>15</sup>. This program, which uses tissue-equivalent phantoms to represent various patient anatomy, can assess an institution’s ability to accurately deliver radiation treatment to a patient. Thousands of phantom irradiations have been made over the past two decades by hundreds of institutions all over the world. Analysis of this data not only allows us to assess the accuracy of treatments, but to also examine individual components of the treatment process, to find the root causes of dose deviations seen among these phantoms.

Previous work has examined the extent of treatment planning system dose calculation errors among the IROC phantoms and the role that this type of error plays in phantom performance. Dose calculation errors were found to be the cause of 60% of dose deviations among SBRT spine phantoms, 30% of moving lung phantoms<sup>34</sup> and 69% of IMRT head and neck (H&N) phantoms<sup>6</sup>. While this partially explains some of the discrepancies that exist, there is still a portion of the dose deviation that remains unaccounted for.



The AAPM TG100 report found that delivery errors associated with treatment machine failures pose the 3<sup>rd</sup> highest risk of failure to a radiation therapy treatment program. They also suggested that the recommended monthly MLC leaf positioning assessment recommendations may not be adequate enough to catch errors, especially in treatments with a few fractions, such as SBRT<sup>44</sup>. In this work, we explore how much of this remaining dose deviation is due to delivery errors.

A separate study using the IROC H&N and spine phantom geometries fitted with ion chamber inserts, investigated the ability of the phantoms to detect delivery error. The results of that study confirmed that the IROC H&N and spine phantoms are capable of capturing delivery errors, primarily caused by MLC RMS error in the form of measured dose deviations to the planning target volume (PTV) and organs at risk (OARs). Random error caused dose deviations up to 2.8% in the average PTV dose, whereas whole bank leaf errors of 2 mm caused deviations of up to 7.9%. An analysis of the corresponding log files for the H&N phantom showed that in most cases, discrepancies in delivered dose was captured by the delivery log files to within 1% of the TPS and measured dose discrepancies (chapter 4, Fig. 16). These results were also found clinically where random errors caused changes to the PTV of 1.6%<sup>38</sup>, and whole bank shifts of 2mm total caused PTV dose deviations of up to almost 8%.<sup>37,38</sup>

Although delivery errors have been shown to affect patient dose, the prevalence of delivery errors among the wider radiation therapy community, and the corresponding dosimetric impact, is largely unknown. In this work, we evaluated the delivery log files from various institutions who irradiated the IROC lung and IMRT H&N phantoms during a specified time frame, to determine the magnitude of delivery error that exists among these phantoms, as well as the relationship between delivery errors and phantom performance.

## 5.2 Materials and Methods

IROC phantoms are used for end-to-end QA and clinical trial credentialing for centers across the US and abroad. The IMRT H&N phantom is used to assess IMRT treatments and contains a primary and secondary PTV as well as an organ at risk (OAR). The PTVs and OAR hold six double loaded thermoluminescent dosimeters (TLDs) in total.<sup>6</sup> The lung phantom contains an island lung tumor structure with two double-loaded TLDs contained within the PTV. This phantom sits on a moving platform that mimics patient breathing, and is usually used to assess motion management and heterogeneity RT treatments.<sup>27</sup>

In the normal IROC workflow, the institution irradiates the phantom and returns it along with their plan DICOM files. IROC personnel read the TLDs contained within the phantom to determine dose received by the target and critical structures. Radiochromic films, also located within the target, are analyzed to determine the gamma of the dose distribution within the structures. These results are compared with the treatment plan to determine any differences between the planned and delivered doses. The TLD passing criteria for both phantoms is that the measured dose be within  $\pm 7\%$  of the planned dose. The gamma threshold for pixels passing is 85% for both phantoms with 7%/4 mm for H&N<sup>6</sup> and 7%/5mm for the lung<sup>27</sup>.

### 5.2.1 Delivery Log files

In January of 2019, IROC began requesting submission of delivery log files from institutions who irradiated the IROC phantoms. Due to accessibility of these log files, we limited collection to only Varian Truebeam and Base (Varian 2100, EX, iX) class model machines. To ensure the correct log file was submitted, we verified that the treatment plan had

been delivered at the same time and on the same machine as the log files. Log files were collected through to January 2022.

### **5.2.2 Mobius FX**

Mobius FX is the component of Mobius 3D (Varian Medical Systems™) that determines dose delivered when a phantom or patient is treated on a treatment machine. Mobius Fx uses a GPU-accelerated collapsed cone dose algorithm and a generic model for each linac type to calculate 3D delivered dose from the treatment delivery log files.<sup>45</sup> The software outputs delivered dose values for each target structure within the treatment plan, as well as root mean square (RMS) error values for the gantry, collimator, MLCs as well as X and Y jaw position values throughout the treatment.

### **5.2.3 Plan analysis: delivery component**

Each phantom result was assessed individually to compare the TPS dose to the delivered dose to determine the discrepancy between the two values. DICOM files were used to compute TPS dose and log files were used to find the dose delivered by the machine.

Plans were analyzed using the following methods:

- i. The treatment plans were classified and recalculated according to their corresponding machine models pre-established in our dose recalculation system(DRS): Mobius3D.<sup>7,16</sup> Since we are processing plans from hundreds of various machines, these classes allow us to group plans according to the machine type that the phantom was irradiated on. Mobius 3D recomputes the plan dose using the submitted DICOM files, and the machine parameters based on the class that machine falls under.

- ii. The treatment log files were then processed via the Mobius FX component of the DRS which pairs the log files with the previously recomputed treatment plan. This produces a TPS dose as well as a delivered dose.
- iii. The delivered dose was recorded and compared to the TPS recalculated plan dose. This provided us with a direct comparison between the dose calculation and delivery analysis for each plan and ruled out any possible TPS calculation errors that may have been present.

#### 5.2.4 Predicted dose value

The predicted dose was used to represent what the dose to the phantom TLDs would have been if dose differences due to delivery error were considered in the calculation. The predicted dose value was calculated for each phantom target TLD. This was performed for each individual TLD to avoid averaging effects across the phantoms, which each contain multiple TLDs in each target structure.

$$\text{Predicted dose} = \text{TPS dose} + (\text{MFx Delivered dose} - \text{M3D calculated dose}) \quad \text{Equation 3}$$

Predicted dose was calculated using Equation 3. The TPS dose is the institution's original dose calculation and is obtained from their DICOM files that are submitted along with the phantom. The M3D dose is the plan dose recalculated on Mobius3D using the same DICOM files. The MFx delivered dose is the dose that the MFx software determines was delivered by the machine, based on the treatment log files.

We contrasted both this predicted dose, and the institution's TPS dose, with the TLD measured dose using equation 4. These two ratios allowed us to assess whether our predicted dose value was more accurate at determining measured phantom TLD dose. A positive D result

indicates that the predicted dose, which accounts for delivery error, was closer to the TLD dose, whereas a negative result indicates that the institution's TPS dose was closer to the TLD dose.

$$D = \left( \left| 1 - \frac{\text{TPS}}{\text{TLD}} \right| - \left| 1 - \frac{\text{Predicted}}{\text{TLD}} \right| \right) \times 100 \quad \textbf{Equation 4}$$

### 5.2.5 Sample analysis

An example of the complete analysis using a single phantom result is shown below.

Predicted dose calculation:

Our predicted dose for TLD 1, using equation 3 would be:

$$\text{Predicted dose} = \text{TPS dose} + (\text{MFx Delivered dose} - \text{M3D calculated dose})$$

$$\text{Predicted dose} = 6.83 + (6.88 - 6.87) = 6.84 \text{ Gy}$$

The delivery dose error (D):

$$D = \left( \left| 1 - \frac{\text{TPS}}{\text{TLD}} \right| - \left| 1 - \frac{\text{Predicted}}{\text{TLD}} \right| \right) \times 100$$

$$D = \left( \left| 1 - \frac{6.83}{7.25} \right| - \left| 1 - \frac{6.84}{7.25} \right| \right) \times 100 = 0.14\%$$

According to the predicted dose, when we incorporate the delivery error into the TPS calculation, the dose to TLD 1 was 0.01 Gy higher than the TPS calculated, bringing us 0.14% closer to the measured TLD value of 7.25 Gy. The summary of the remaining TLD calculations for this example phantom are shown in Table 7. The average delivery error is 0.29% for this phantom, indicating that on average, the predicted dose based on delivery error consideration, was 0.29% closer to the TLD than the institution's TPS.

**Table 7.** Results for an example H&N phantom showing how delivery error is calculated using each individual TLD before finding the average for the phantom.

	<b>TLD 1</b>	<b>TLD 2</b>	<b>TLD 3</b>	<b>TLD 4</b>	<b>TLD 5</b>	<b>TLD 6</b>	<b>Avg</b>
<b>TPS, Gy</b>	6.83	6.77	6.70	6.95	5.74	5.68	6.45
<b>M3D, Gy</b>	6.87	6.77	6.71	6.87	5.77	5.66	6.44
<b>Delivered (MFx), Gy</b>	6.88	6.78	6.72	6.88	5.81	5.69	6.46
<b>TLD (measured), Gy</b>	7.25	7.00	6.95	7.24	5.96	5.89	6.72
<b>Predicted dose, Gy</b>	6.84	6.78	6.71	6.96	5.78	5.71	6.46
<b>Delivery error (D)</b>	0.14	0.14	0.14	0.14	0.67	0.51	<b>0.29</b>

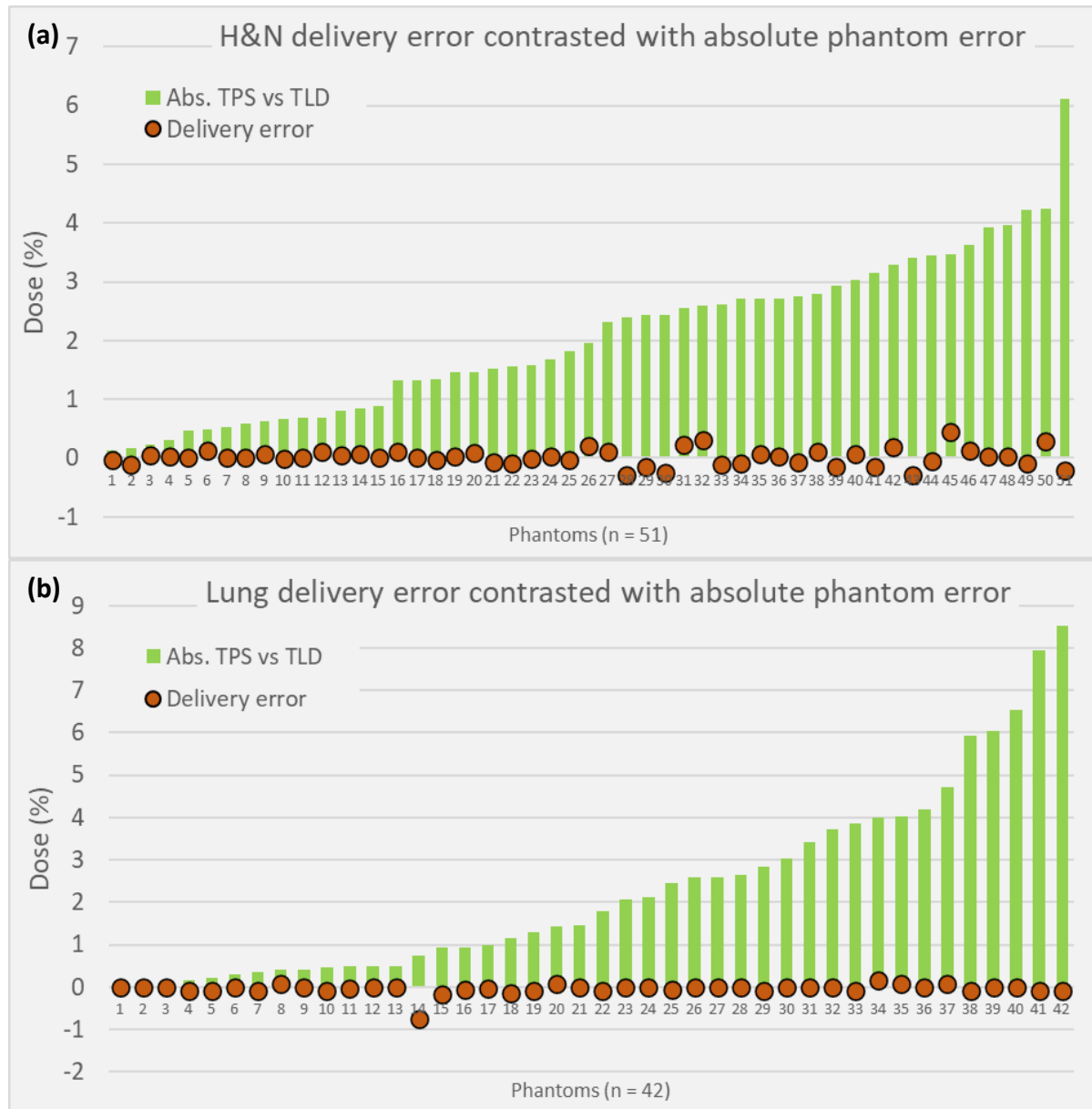
### 5.3 Results

We processed log files for 51 H&N phantoms (1 repeat institution), 20 of which were irradiated on a Varian base class machine: 21EX, 21iX, 23EX, iX and Trilogy, with the other 31 being Truebeam (including Edge and Vitalbeam) irradiations. There were 42 lung phantoms (2 repeat institutions) made up of 12 Varian base irradiations and 30 Truebeam.

#### 5.3.1 Phantom results

There were 51 H&N and 42 lung phantoms evaluated. The H&N phantom had an average absolute dose deviation (TPS vs TLD) of 2.1%, with a maximum deviation of 6.1%. The delivery error was minimal in comparison, with an average of 0.02% and ranging from -0.3% to 0.5%. Figure 17(a) shows the phantom delivery error versus the overall phantom dose deviation. A very weak correlation is shown between these two variables, indicating that delivery error was not a major contributing factor to overall phantom performance ( $r = -0.04$ ,  $p = 0.40$ ) (Pearson correlation).

The lung phantom results were similar. The lung had an average dose deviation of 2.3%, with a maximum deviation of 8.5%. The delivery error for this phantom was -0.04% on average and ranged from -0.8% to 0.2%. In Figure 17(b), an even weaker correlation between phantom dose deviation and delivery error is seen ( $r = 0.13$ ,  $p = 0.8$ ).

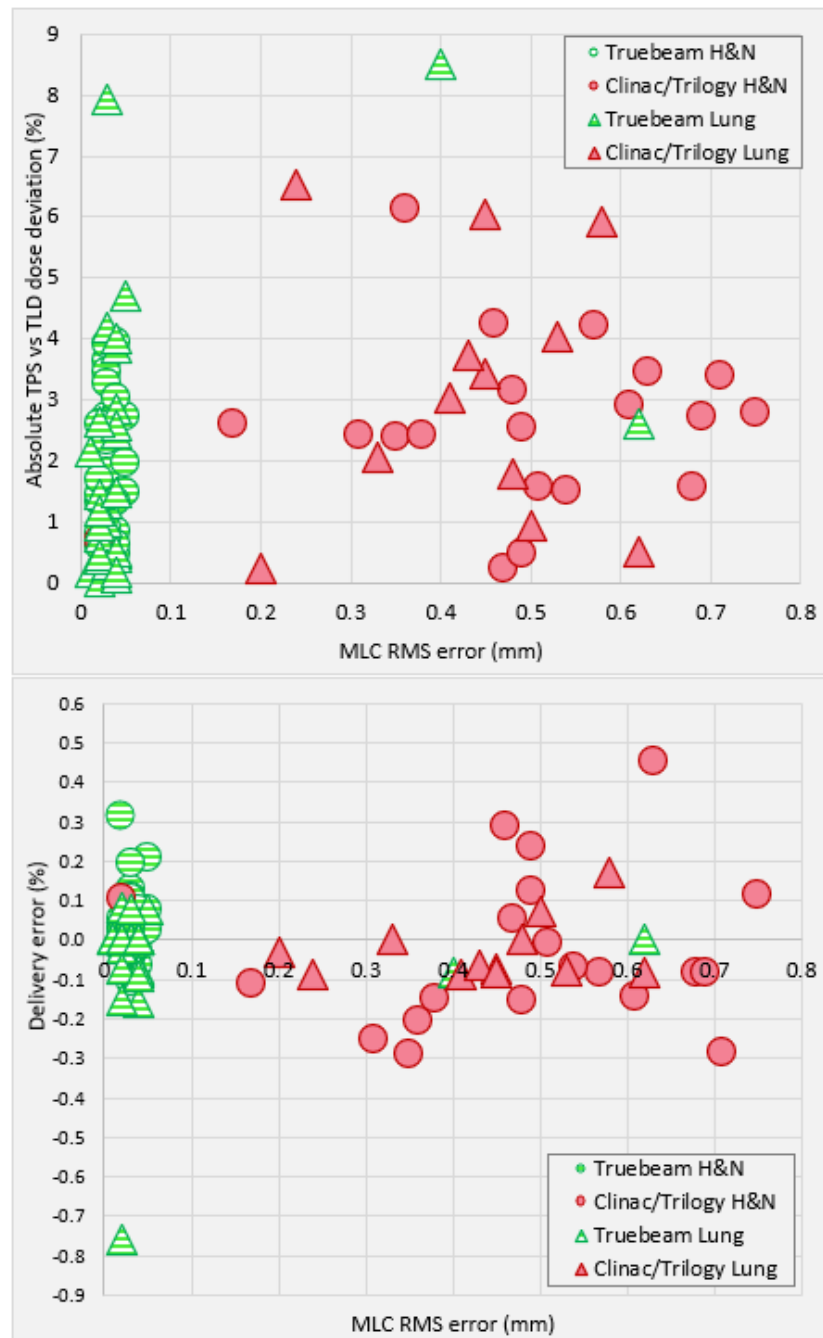


**Figure 18.** Distribution of delivery error compared with TPS vs TLD dose deviation (phantom performance), showing that one does not increase with the other. Delivery error remained relatively unchanged as TPS vs TLD dose deviation increased (phantom performance worsened), indicating minimal contribution from this error type.

### 5.3.2 MLC error

MLC RMS error results for the two Varian machine classes: Truebeam (including Edge) and the older Base model (iX, EX & Trilogy)<sup>2</sup>, showed that the Truebeam machine MLCs performed better among both phantoms. The average MLC RMS error for both phantoms was 0.02 mm for the Truebeam machines and 0.30 mm for the base models. The distribution of these errors in comparison to absolute phantom dose deviation is shown in figure 18. Despite the RMS errors being so different, the average dose deviation for the two sub-groups of phantoms irradiated on each machine type were within 1.2% of each other (Truebeam = 2%, Base model = 3.18%).





**Figure 19.** Distribution of MLC rms error for the two Varian machine classes (Truebeam and Clinac) in relation to phantom performance, represented by absolute TPS vs TLD dose deviation (top), and delivery error (bottom). MLC rms error was found to be machine dependent rather than based on phantom type, magnitude of dose deviation or delivery error.

## 5.4 Discussion

This work investigated the presence of log file recorded delivery error among IROC head & neck and lung phantom irradiations. Delivery error was found to contribute very little to the dose deviations in either phantom, with average delivery error being 0.02% of H&N dose deviations and -0.04% of lung (meaning zero contribution). Additionally, no correlation was found between MLC rms error and either phantom performance or delivery error.

Delivery error was found to have a very small effect on phantom plan accuracy. Even among the 1 H&N and 5 lung phantoms that failed to meet the IROC irradiation criteria, there was no indication of delivery error playing a part in their poor performance. Delivery errors may be caused by various components of the treatment machine, such as jaws, collimators and multi-leaf collimators (MLCs) to name a few. Of all these components, the MLCs are the most prone to error due to their frequent and continuous motion, which is necessary for treatment modulation in dynamic treatments, which made up 96% of plans in this study.

MLC rms error was found to only be machine dependent, with more technologically advanced Varian Truebeam machines (and similar models) outperforming previous Varian machine models. This was not surprising, because technology in treatment machines has continually improved, and the more technologically advanced Varian Truebeam models speak to these improvements, as a 10-fold decrease in delivery error was seen among Truebeam machines when compared with the base model.

The use of Mobius Fx to assess MLC rms errors was previously tested and verified in a separate study. Manually introduced random and whole bank MLC errors were detected in the form of dose deviations, through both measurements and log file analysis of the phantoms (chapter 4). This previous study confirms that the results found here, are a true representation

of the delivery error that exists among these phantoms. The small and negligible differences in delivered dose compared with plan dose, indicate the presence of random MLC errors but not systematic (whole MLC bank shifts) which would have resulted in larger, more clinically significant dose errors<sup>37,38</sup>.

Limitations do exist within this study with regards to the ability of log files to detect true errors in the plan delivery. Log files have been shown to not detect all errors, especially when these errors stem from incorrect machine calibration and faulty mechanisms within the machine. The Varian machines record the MLC leaf positions by the number of turns of the motor, so any issues with the leaf to motor connection could result in inaccurate leaf positioning records and hence incorrectly recorded doses<sup>43</sup>.

This work has demonstrated the negligible effects of delivery errors on phantom irradiations. However, the cohort of phantoms assessed represent clinical trial participants, who are especially invested in performing well in these assessments and delivering the best RT. Hence, they can be considered as generally above average performing and may not provide an accurate picture of all error that exists within the wider community.

Greater focus and emphasis still needs to be applied to errors in other aspects of the treatment process, such as treatment planning, which was previously found to be prevalent among H&N and spine phantom treatments<sup>34</sup>, and motion management which remains a major issue for lung treatments.<sup>27</sup> Through continued efforts, the aim is for all error contributions to be minimal at best, to improve the accuracy of patient treatments within the radiation therapy community.

## **5.5 Conclusion**

The presence and effects of log file recorded delivery error were investigated among IROC H&N and lung phantoms. This delivery error was found to be negligible among these phantom results and had no effect on the accuracy of the irradiations. MLC RMS errors were found to be greater among plans irradiated on older Varian model machines while the newer Varian Truebeam machines showed more accurate MLC motion across all plans. Further work is needed to understand the extent of the contributions to dose deviations from other sources of error among phantom and patient plans.

## **Chapter 6: Quantifying the magnitude of dose deviation caused by various major sources of error present in IROC phantom irradiations**

This chapter is based on the following publication:

**Edward S.S.**, Howell R.M., Balter P.A., Peterson C.B., Pollard-Larkin J.M., Kry S.F.,

“Quantifying the magnitude of dose deviation caused by various major sources of error present in IROC phantom irradiations”, To be submitted for publication, 2022

### **6.1 Introduction**

Hundreds of institutions, both in the US and abroad, irradiate the Imaging and Radiation Oncology Core (IROC) phantoms each year as an end-to-end test of their radiation therapy (RT) processes. IROC has various phantoms for this purpose, which represent different anatomical sites such as head and neck (H&N), lung and spine. These phantoms are made up of tissue-equivalent materials and contain various structures which represent tumors and organs at risk<sup>12,15</sup>. Radiation doses received by the phantom is assessed through the use of thermoluminescent dosimeters (TLDs) placed inside each structure. This phantom program has been in existence for over 2 decades, and although pass rates have improved overtime, current 2021-2022 pass rates lie between 78% to 88% for the H&N, lung, and spine phantoms.

Phantoms are used to assess the accuracy of RT treatment, which is a multi-step process, involving a team of experts working together to achieve the best treatment possible for their patients. Accuracy in treatment planning, patient setup and even quality assurance of the treatment machines play very important roles in the overall success of treatment. Since the IROC phantom test is an end-to-end remote assessment tool, it is difficult to pinpoint exactly what goes wrong within the entire process for any given phantom irradiation, to cause a failing

or subpar result. A previous study analyzed the various reasons for failure of these phantoms, and showed numerous causes such as systematic and local dose errors and localization errors in various directions<sup>27</sup>. Dose errors were primarily due to beam modeling errors<sup>46</sup> whereas localization errors were caused by incorrect phantom setup, or poor motion management techniques in the case of a moving phantom<sup>27,47</sup>. This indicates that issues exist at various stages of the treatment process and are worth individually investigating to better understand the reasons behind the poorer results that are seen among phantom populations.

In this study, we investigated 3 major, measurable causes of error and quantified their contribution to the overall dose deviation of the phantoms.

#### **6.1.1 Dose calculation error**

Treatment planning is one of the key components in the RT process. This is done using a treatment planning system (TPS), which is commissioned by an institution with data from their linear accelerators and other treatment machines. Through TPS calibration, physical and dosimetric parameters of the machines are modeled in the TPS, in order to accurately predict the dose that will be delivered once the plan is transferred to the machine and delivered to the patient. These calibration parameters do not change and therefore have an impact on every plan that is created using the TPS. Previous studies have shown that IROC phantom results contain dose calculation errors<sup>6,7</sup>, which are more pronounced for poorer phantom irradiations. These errors also appear to be more prevalent among the phantom irradiations with more highly modulated plans, such as the head & neck and spine.

### 6.1.2 Delivery error

A treatment plan is created for each patient and essentially tells the machine exactly what, when and how to deliver radiation to the patient. It determines gantry, collimator, couch, and multi-leaf collimator (MLC) positions, as well as monitor unit (MU) settings. For the duration of treatment, the LINAC records the actual positioning of these components and the monitor units delivered, during delivery of the plan with millisecond precision. Using the log files generated, we can compare the planned doses from the TPS to the actual delivered doses from the machine to assess the differences between the two. Studies have shown that errors in machine positioning, in particular, MLC positioning, can cause deviations between the planned and delivered dose. Systematic MLC errors, such as offsets in the MLC leaf bank have shown changes to the planning target volume (PTV) dose of 3.7% and 7.2% for offsets of 0.5mm and 1mm respectively<sup>37</sup>. Random error effects were found to be smaller, only causing a change in dose of up to 1.6% overall<sup>38</sup>. Log files are likely to flag random errors during delivery but are not likely to show systematic errors as these would result from a miss-calibration of the machine that is not detectable with internal monitoring. Although these effects are known, their prevalence among the RT community is unknown. Considering these potential effects on patient treatments, it is imperative to use these log files to determine the magnitude of the contributions of machine delivery accuracy to overall phantom irradiation error.

### 6.1.3 Machine output error

The American Association of Physicists in Medicine (AAPM) published reports that set guidelines for linear accelerator (LINAC) output, indicating that it should be maintained to within 1-2% of the expected/nominal value.<sup>48,49</sup> Deviations from the nominal machine output contribute systematically to differences between the planned and delivered doses to the phantoms during an irradiation. Output errors also pose potential problems for all patients who are treated using that machine and therefore have far-reaching implications within the clinic. Previous results indicate that out of approximately 20,000 beam output checks performed at various institutions, 2.4% (480) of results were outside of the IROC's 5% tolerance level<sup>50</sup>. Investigating the effects of these errors on phantom irradiation success will tell us the direct impact of these errors to treatment plans. Quantifying the contribution of output errors to dose deviation will allow us to evaluate machine performance without the added layers of complexity that accompany plan design in many phantom and patient plans. This is a more straightforward measurement of the monitor unit to dose ratio of the LINAC.

These phantom results are indicative of the quality of radiation therapy that is being administered to patients and so the errors found have the potential to impact the quality of RT treatment for their patients. A better understanding of the nature and magnitude of these individual errors will help IROC to provide better feedback to institutions with subpar results, with the goal of improving the accuracy of RT among the radiation oncology community.



## **6.2 Methods and Materials**

The IROC phantoms in this study are the IMRT head and neck, moving lung and SBRT spine phantoms. They are all composed of tissue-equivalent materials and contain both targets and organs at risk (OARs) structures. Institutions irradiating these phantoms are instructed to treat them like patients, and IROC personnel are able to determine dose received by the various structures through analysis of the thermoluminescent dosimeters (TLDs) placed inside them<sup>6,15,27</sup>.

In this study, we analyzed IROC phantom data for three different error types. Each error type was quantified individually, and the results combined to deduce the overall error that contributed to TPS vs TLD dose deviations among the phantom population. TLDs were also evaluated individually because a successful irradiation is dependent on every TLD within the phantom meeting the irradiation criteria. Therefore, having a single TLD out of tolerance would result in the phantom failing the irradiation.

### **6.2.1 Dose calculation error**

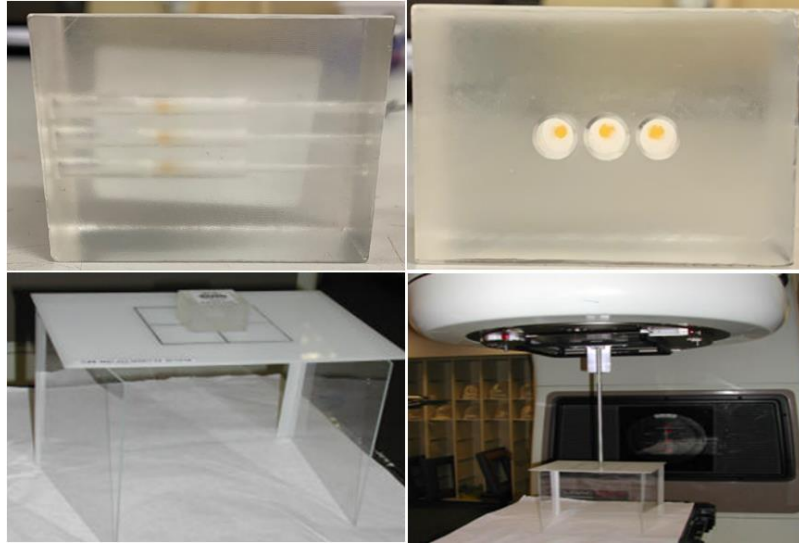
DICOM files for the phantom plan were returned to IROC along with the phantom. We used these files to recompute the phantom plan using our independent dose recalculation system (DRS), Mobius3D (Varian Medical Systems®). The DRS was commissioned using data from over 500 linear accelerators from various institutions, and therefore represents an average performing machine by community standards<sup>16,17</sup>. In cases where the DRS more accurately computes dose to an institution's phantom TLD than their TPS, this indicates a dose calculation error or inaccuracy with the institution's TPS, which should ideally be more closely matched to their machine than the DRS. Further details of the methods used in this portion of the study are outlined in our previous work<sup>34</sup>.

### **6.2.2 Delivery error**

Dose deviations that occur during delivery of the phantom plan were assessed using delivery log files collected with each irradiated phantom. Log file collection was not previously part of the IROC phantom procedure, and so these were specifically requested from institutions irradiating the phantoms using Varian machines from 2019 to 2022. These files were then processed using Mobius FX (Varian Medical Systems®)(V2.1.2), and deviations in dose and MLC rms error were recorded. Further details of the methods used in this portion of the study are outlined in our previous work (chapter 5.2).

### **6.2.3 Output error**

Every IROC phantom is accompanied by a water-equivalent acrylic block containing three TLDs (Fig. 6-1), which is used to assess the machine output prior to phantom irradiation. The institution is instructed to deliver 300 cGy to their machine's "output specification point", which is the point at which the calibration dose rate for that machine is specified. Additionally, institutions are required to submit their own machine output measurement on that day, which is classified as machine drift. This is the amount by which their machine is off from the ideal calibration of 1 (1MU = 1cGy), but still within their clinically acceptable limits. This measurement is oftentimes performed with a farmer ion chamber or the daily output check device.



**Figure 20.** IROC acrylic output block containing three double loaded TLD's (top). Output block setup at the LINAC with pointer measuring distance to the machine's output specification point (bottom)

Upon return of the acrylic output block along with the phantom, IROC compares the output block TLD dose to the planned dose of 300cGy that was supposed to be delivered by the institution. Based on this value, the TLD vs institution ratio is generated, which is the ratio of the institution reported dose to the actual TLD dose. For this study, we deduced the true output of the machine by comparing the institution reported output value to the TLD vs reported dose ratio as shown in equation 5.

$$\text{True machine output} = \text{institution reported output} \times \text{TLD vs institution ratio}$$

### Equation 5

For example, if an institution reported an output of 1.02, indicating that their machine output was 2% high, and their output TLD vs institution ratio was 0.97, the true machine output would be  $1.02 \times 0.97 = 0.99$ , indicating that the machine output was truly 1% low on that day. This true output value was used to assess output error for each phantom irradiation analyzed.

## 6.2.4 Error analysis

### 1. Predicted dose

A predicted dose was derived for each of the 3 error types for each phantom assessed in this study. The predicted dose represents an adjustment of the institution's TPS dose number, by the difference resulting from a dose discrepancy due to dose calculation, output or delivery magnitude inaccuracies found for that phantom result. Using the output example above, the predicted dose would be computed by lowering the TPS dose by 1%, to account for machine output being 1% low on that day.

### 2. Error

The dose difference value (D) represents error magnitude for each error type and is computed by comparing the TPS and predicted doses, to identify which was a better estimate of the phantom TLD dose. Equation 6 compares the ratio of  $TPS/TLD$  and  $Predicted/TLD$ , where a positive result means that the independent dose recalculation performed better than the institution's TPS, and a negative result indicates the opposite. The magnitude of this result also determines how much better the prediction was, where positive x% indicates the presence of x% contribution from the error type analyzed.

$$D_{\text{output}} = \left( \left| 1 - \frac{TPS}{TLD} \right| - \left| 1 - \frac{Predicted}{TLD} \right| \right) \times 100$$

**Equation 6.** Example of output error calculation

### 3. 3.2% dose deviation

The TLDs used in the phantoms are double loaded and therefore carry a measurement uncertainty or standard deviation of 1.6%<sup>29</sup>. For a 95% confidence level,

we assessed phantom deviations at the  $2\sigma$  level of 3.2%. This level represents a 95% chance that the dose deviations between the institution's TPS and the measured phantom TLDs are beyond the level of measurement uncertainty and represent true error in the measured TLD dose deviation. Phantom TLDs were evaluated individually, and results were split into two groups for assessment: (1) absolute TPS vs TLD dose deviation below 3.2%, or better performing TLDs and (2) absolute TPS vs TLD dose deviation greater than 3.2%, or poor-performing TLDs. Absolute values were used because the magnitude of the deviation is what mattered, versus what direction it was in. Equation 6 also accounts for differences in direction allowing error to be determined whether the phantom was over or underdosed.

## 6.3 Results

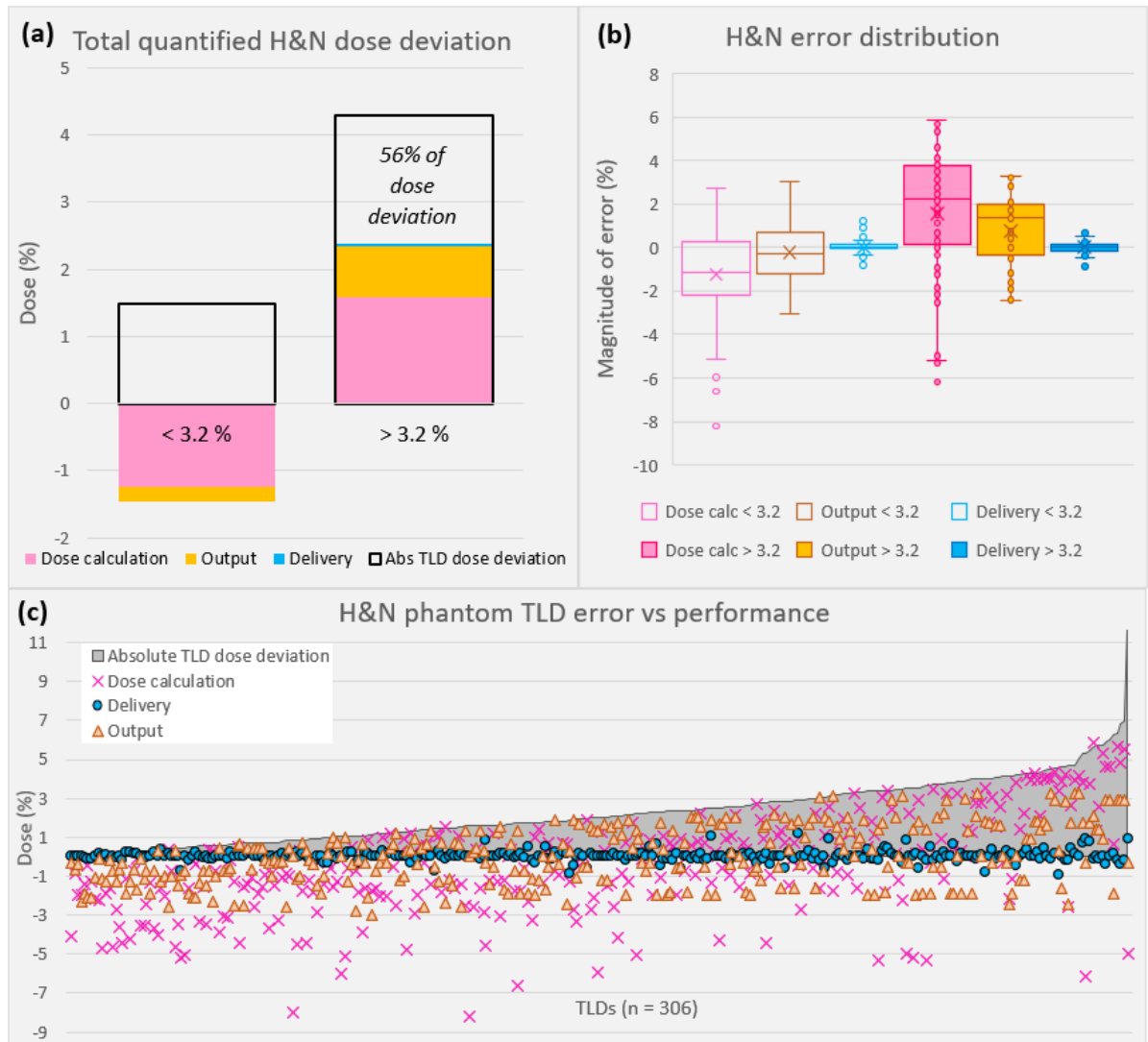
### 6.3.1 All error types

We evaluated all 3 error types for 51 H&N and 42 lung phantoms and evaluated dose calculation and output error for 63 spine phantoms. The number of TLDs and the distribution between those with dose deviation less than and greater than 3.2% is shown in Table 8.

**Table 8.** Distribution of dose deviation results for phantom TLDs

	Number of TLDs		
Dose deviation	<b>H&amp;N</b>	<b>Spine</b>	<b>Lung</b>
<b>&lt; 3.2%</b>	222	161	59
<b>&gt; 3.2%</b>	84	91	23

**6.3.1 (i) H&N:** We quantified 56% of dose deviations among H&N phantom TLDs with greater than 3.2% dose deviation. Dose calculation error was the greatest contributor to the H&N phantom error, representing 37% of the dose deviation among poorer performing phantom TLDs. Output error made up 18% of deviations and delivery, only 1% (Fig.20 (a) & (b)). Dose calculation error also had the strongest correlation with phantom performance of 0.59 ( $p < 0.01$ ) (Pearson correlation). Figure 20 (c) shows the distribution of the individual H&N TLD dose deviations, compared to the corresponding magnitude of each error found within the TLDs.



**Figure 21.** (a) Illustration and (b) distribution of the total dose deviation quantified for the two subgroups of H&N phantom TLD results; 56% of dose deviation was quantified in TLDs with >3.2% dose deviation. This is shown in the bar graph where the black empty column represents dose deviation, and the colored bars account for this deviation with various error types (c) Comparison of absolute TPS vs TLD dose deviation (gray area) to each error type. As TLD dose deviation increases (from left to right), phantom performance gets worse. Dose calculation error was the most prevalent and had the strongest correlation (0.59,  $p < .01$ ) with phantom performance, as can be seen from the large number of positive values for this error type towards the right of the graph. Output error was less positively correlated (0.43,  $p < .01$ ) with TLD performance and delivery error remained constant (-0.02,  $p = 0.8$ ) throughout.

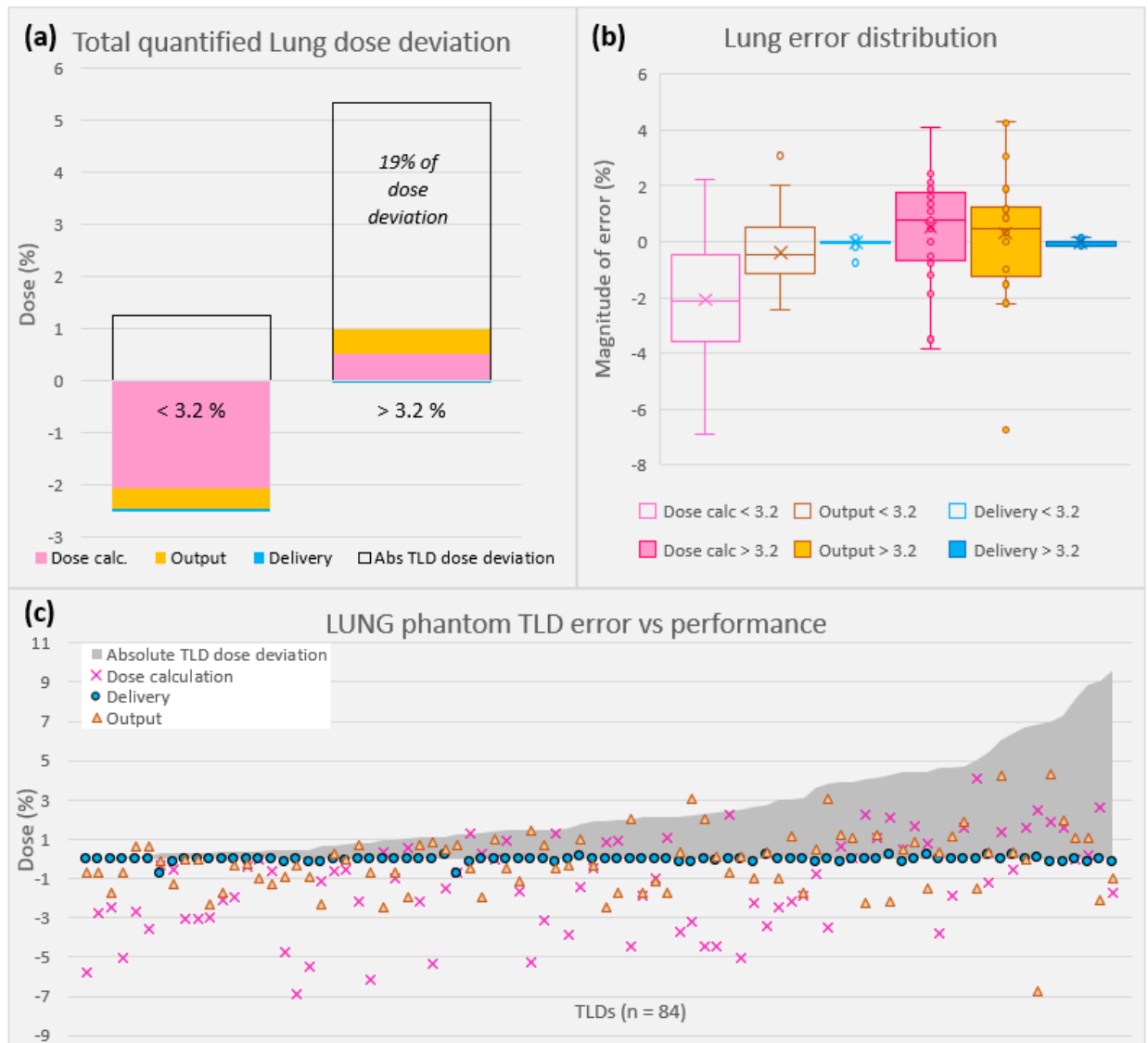
**6.3.1 (ii) Spine:** We quantified 68% of dose deviations among spine TLDs with greater than 3.2% dose deviation. Dose calculation error was the greatest contributor to the spine phantom error, representing 37% of the dose deviation among poorer performing TLDs. Output error also made up a substantial portion of the error, accounting for 31% of deviations. There was no delivery error contribution analyzed for this phantom due to insufficient irradiations during the log file collection period (Fig.21 (a) & (b)). Dose calculation error had the strongest correlation with phantom performance of 0.65 ( $p < 0.01$ ) (Pearson correlation). Figure 21(c) shows the distribution of the individual spine TLD dose deviations, compared to the corresponding magnitude of each error found within the TLDs.





**Figure 22.** (a) Illustration and (b) distribution of the total dose deviation quantified for the two subgroups of spine phantom TLD results; 68% of dose deviation was quantified in TLDs with >3.2% dose deviation. This is shown in the bar graph where the black empty column represents dose deviation, and the colored bars account for this deviation with various error types. (c) Comparison of absolute TPS vs TLD dose deviation (gray area) to each error type. As TLD dose deviation increases (from left to right), phantom performance gets worse. Dose calculation error was the most prevalent and had the strongest correlation (0.65,  $p < .01$ ) with phantom performance, as can be seen from the large number of positive values for this error type towards the right of the graph. Output error was less positively correlated with TLD performance (0.47,  $p < .01$ ).

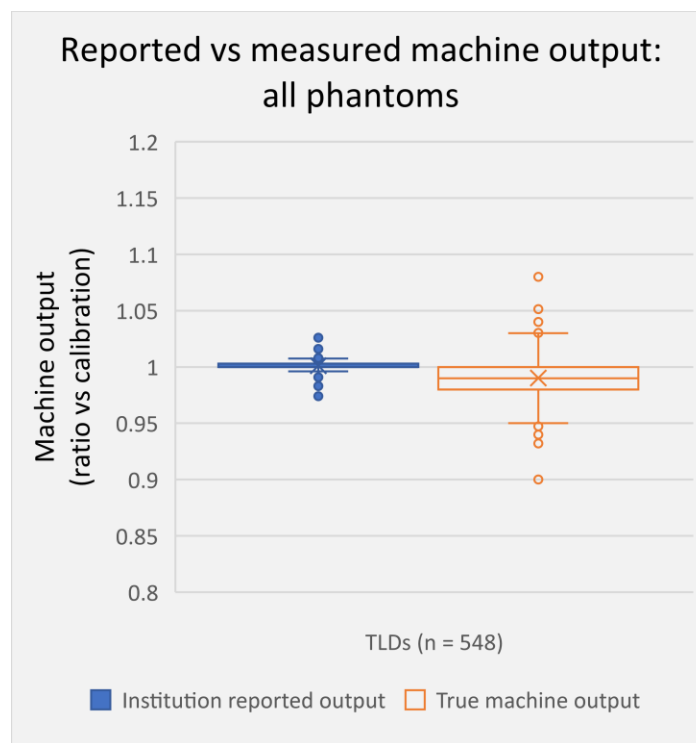
**6.3.1 (iii) Lung:** We quantified 19 % of dose deviations among lung TLDs with greater than 3.2% dose deviation. Both dose calculation and output error contributed almost equally to the error in the lung phantom at 10% and 9% respectively (Fig.22 (a) & (b)). Delivery error contributed -0.3%, which in this case equates to a zero contribution. Dose calculation error had a correlation with phantom performance of 0.48 ( $p < .01$ ) (Pearson correlation), while output was 0.27 ( $p < .05$ ) and delivery 0.02 ( $p = 0.9$ ). Figure 22 (c) shows the distribution of the individual lung TLD dose deviations, compared to the corresponding magnitude of each error found within the TLDs.



**Figure 23.** (a) Illustration and (b) distribution of the total dose deviation quantified for the two subgroups of lung phantom TLD results; 19% of dose deviation was quantified in TLDs with  $> 3.2\%$  dose deviation. This is shown in the bar graph where the black empty column represents dose deviation, and the colored bars account for this deviation with various error types. (c) Comparison of absolute TPS vs TLD dose deviation (gray area) to each error type. As TLD dose deviation increases (from left to right), phantom performance gets worse. Dose calculation error had the strongest correlation ( $0.48, p < .01$ ) with phantom performance, and output had a correlation of  $0.27(p < .05)$  with delivery error remaining relatively unchanged throughout.

### 6.3.2 Machine output

A total of 548 phantom output measurements were evaluated across the H&N (n = 315), spine (n = 70) and lung (n = 163) phantoms. Of 548 output TLD results, 400 (73%) had an institutional reported value which was higher than TLD measured:  $1.001 \pm 0.5\%$  reported vs  $0.983 \pm 1.3\%$  measured. There were 106 (19%) results with a lower institutional reported value than TLD measured:  $1.001 \pm 0.46\%$  reported vs  $1.013 \pm 1.34\%$ . Figure 23 shows the comparison between reported and measured machine output for all phantoms evaluated.



**Figure 24.** Output results for 548 institutions showing that institutions on average reported a value for measured machine output that was higher than the true output value.

## 6.4 Discussion

An evaluation of dose calculation, delivery, and machine output errors among IROC phantom results was performed to quantify the amount of each kind of error that contributed to dose deviations in these phantom irradiation results. Among the TLD results with dose deviation beyond measurement uncertainty of 3.2%, the total error quantified was 56% for H&N, 68% for spine and 19% for lung phantom results. Dose calculation errors were most prevalent among the SBRT spine and IMRT H&N phantoms. Although only 14 spine phantoms (22%) in this study had a failing result, the results still closely mirror those of a previous study which examined the reasons for failure of the spine and lung phantoms. It was found that 60% of spine phantoms failed due to systematic dose errors<sup>27</sup>, which stem from the treatment planning system (TPS) parameter selection and dose calculation algorithm<sup>46</sup>. A precursor to this work which examined dose calculation error specifically, found that on average, the DRS calculation showed a 2.35% improvement in dose accuracy when compared to the institution's TPS<sup>34</sup>. The DRS represents a stock model or average performing dose calculation algorithm, so its ability to outperform an institution's TPS which is customized to their treatment machine, indicates the presence of error within that TPS. The H&N phantom, having similar plan modulation complexity to the spine, has been shown to have similar results. A study of the H&N phantom results showed that 58% of failures were caused by systematic dose errors<sup>6</sup>, coinciding with our findings of dose calculation errors contributing the most to this phantom's dose deviations. Lung phantom plans generally have less modulation than the other two phantoms in this study, and generally test heterogeneity and motion management. Motion management was previously found to be a major concern for treatment accuracy as localization errors were found to be the cause of 50% of lung phantom irradiations. Systematic dose errors

only caused 21% of failures and therefore does not have as great of an impact on this phantom<sup>27</sup>.

Delivery errors stem from discrepancies between the TPS calculated dose and what the machine delivered to the phantom. This was assessed by analyzing the corresponding log files for each treatment and MLC RMS in particular, was the factor that was further investigated to determine its effects on phantom accuracy. For some plans the MLC leaves constantly move during treatment delivery to modulate dose to the phantom, so the extent of these errors was worth investigating further. A precursor to this study showed that MLC rms error was more present among Varian 2100 machines versus the more technologically advanced Truebeam model. There was a slightly positive correlation between phantom performance and MLC rms error, however this correlation was less than 0.5 (Spearman rank correlation) for both phantoms and not statistically significant for the H&N (chapter 5.3.2). While the log files can be used for additional evaluation of the IROC phantoms, there are limitations that exist with regard to the errors that can be identified by the log files. The log files have been shown to miss errors that other evaluation methods were able to identify<sup>43</sup>, and errors identified by the log files did not have clinically significant impacts on dose to the phantoms (chapter 4).

Output deviations are not taken into account when analyzing the IROC phantom results as part of the evaluation procedure. Machine output deviation may thus be affecting the results of the phantom TLD measurements. Given that machine output was lower, on average, than the expected value, systematic underdosing of phantoms may stem from this. Of the 64 systematic dose phantom failures examined across all phantoms in Edward 2020<sup>27</sup>, 44 (69%) dose error failures were due to underdosing of the target structure. In this current study, 73% of measured output results were lower than reported by the institutions. These two findings suggest a possible connection between output error and phantom dose outcomes.

## 6.5 Conclusion

Three different error types: dose calculation, delivery and output were investigated and their contribution to the IROC phantom results was quantified. Dose calculation error was the greatest contributor, as it was the reason for 37%, 37% and 10% of dose deviation among H&N, spine and lung phantoms respectively. Output error made up 18%, 30% and 9% of error in H&N, spine and lung respectively, and majority (73%) of institutions underreported their machine dose, contributed to these inaccuracies. These results highlight the need for a more comprehensive assessment of phantom results and give us further insight into the specific reasons for failing and poor-performing phantom irradiations.

## **Chapter 7: Discussion and Conclusions**

### **7.1 Project summary**

This main objective of this work was to quantify the magnitude of various types of errors that contribute to IROC phantom irradiation results. Through the IROC phantom program results, we have found that on average, for all phantoms, 15% of institutions failed to accurately deliver the intended radiation prescription to the phantom. Given that the IROC irradiation criteria is much less strict than the typical clinical standard, this 15% represented only the most egregious cases, and does not encompass all poor-performing results. To better understand and more precisely determine the reasons for these poor performances, we identified and quantified various causes of error that were measurable as part of IROC's remote phantom program. Preliminarily, we conducted a study on the causes of failure of different IROC phantoms to get a general overview of the types of errors that exist. Under aim 1, we quantified dose calculation errors among the phantoms. Aim 2 was used to evaluate delivery errors by first determining the IROC phantom's accuracy with capturing these errors and then quantifying the magnitude of existing errors among the phantom population. Under this aim, we also evaluated machine output/calibration errors and quantified its contribution to the overall phantom dose deviations. Finally, in aim 3, we combine the various errors found and categorized them into better and poorer-performing phantom results to summarize the total errors that contribute to dose deviations among IROC phantom results.

#### **7.1.1 Preliminary study**

Prior to quantifying error, we performed a retrospective analysis on the IROC moving lung and SBRT spine phantoms. The spine phantom generally has more complex and highly modulated treatment plans due to the nature of the small target structure (tumor) which sits on



the organ at risk (OAR), the spinal cord. The lung phantom has a simpler “island” target structure, which represents a tumor situated within lung tissue. This phantom additionally sits on a motion platform to mimic patient breathing. We analyzed failing phantom irradiations between January 2012 to December 2018 by individually assessing the TLD point dose agreements, dose profiles and gamma analyses of each phantom. Failure mode categories were created based on the various causes of failure recorded and each phantom was placed into the appropriate category. Majority (60%) of the spine failures were due to systematic dose errors which means the irradiation was administered in the correct position, but the magnitude of dose was either too high or too low. Underdosing of the target was more prevalent than overdosing, occurring among 69% of systematic dose error cases, at an average of  $-7.4 \pm 2.4\%$ . Majority (50%) of the lung failures were caused by localization errors in the direction of major breathing motion, where the target was partially missed by up to 2.5 cm in some cases (chapter 2).

Systematic errors plaguing the spine generally stem from inaccuracies in the TPS and the localization errors that caused most of the lung failures were due to poor motion management strategies, as both ITV and gating yielded similar results (Table 2-2). Other failure mode categories included local dose failures and global errors which represented dose discrepancies in either one or several areas of the PTV respectively. These errors could be caused by a number of factors related to phantom setup and delivery on the treatment machine. Using these results, we decided to further investigate the magnitude of dose calculation errors and delivery errors among IROC phantom results. Although localization errors played a major role among lung failures, it would be difficult to quantify this error without being present during the time of treatment setup and delivery or conducting a more thorough assessment of each institution’s motion management procedures.

### 7.1.2 Specific Aim 1

#### **Quantify the magnitude of dose deviation caused by dose calculation error in poor performing IROC phantom irradiations**

In this aim, we quantified the magnitude of dose deviations among IROC spine and lung phantoms that were caused by dose calculation errors. This was done through independent recalculation of the DICOM files for each phantom irradiation plan, using an independent dose recalculation system (DRS). The DRS represents an average performing machine and therefore indicated the presence of error when the DRS outperformed an institution's TPS in terms of accurately calculating dose to the phantom TLDs (chapter 3). Similarly to the failures seen in the preliminary study, dose calculation errors were more prominent in the spine phantoms than the lung. The DRS outperformed the TPS in 47% of spine phantom cases and 93% of failing cases (Fig. 5), with an average improvement of 2.35% (Fig. 6). Dose calculation errors accounted for 62% of the dose deviation between the TPS and TLD (Fig. 8).

Among the lung results, the DRS only outperformed the TPS in 35% of phantoms and 42% of failing phantoms (Fig. 9). Dose calculation errors accounted for only 13% of the dose deviation between the TPS and TLD (Fig 12). Comparing these results to those of a previous head and neck (H&N) phantom study, the spine and H&N phantoms were found to have similar amounts of dose calculation error. This indicates that highly modulated phantom (and patient) treatments are more susceptible to errors from TPS dose calculations than plans with low modulation such as the lung.

### 7.1.3 Specific Aim 2

#### **Quantify the magnitude of dose deviation caused by delivery error in poor performing IROC phantom irradiations**

This aim was designed to assess delivery errors that influence the IROC phantom results. The first part of this aim was done by conducting a study to assess the ability of the IROC phantoms to accurately detect MLC delivery errors (chapter 4). Various sMLC and VMAT plans were created for the two phantoms using the Raystation planning system. Random and whole bank MLC leaf shifts were then introduced via Raystation's scripting function, to model error that may occur in the MLC leaves during treatment delivery. The plans were then delivered on a Truebeam machine at MD Anderson, to the H&N and spine phantoms, custom fitted with slimline ion chamber inserts to allow ease of multiple measurements. The log files of the plan delivery were then collected and analyzed using Mobius FX, an independent log file analysis tool. The random MLC shifts caused PTV dose deviations of up to -2.8% for the H&N and 0.7% for the spine plans. Whole bank MLC shifts resulted in PTV dose deviations of up to 8% for the H&N and 7.1% for the spine, for a shift of 2 mm (section 4.3.1). These results were very similar to dose deviations found in the literature for similar clinical plans and demonstrate the phantom's ability to catch and identify these error types accurately.

The effect of fraction size on MLC leaf motion error was also investigated in this study (section 4.3.2). VMAT plans were made for the same phantoms, one at the IROC prescription of 6.6 Gy for the H&N and 6 Gy for the spine, and another at the clinical fraction size of 2 Gy for the H&N and 27 Gy for the SBRT spine. These plans were delivered on both the Truebeam machine and the Varian 21EX, and the treatment log files were analyzed for MLC rms error. The less technologically advanced 21EX recorded a greater amount of MLC rms error than the Truebeam plans. On this machine, the 2 Gy H&N plan had 17% more rms error than the 6.6 Gy

plan and the 6 Gy spine plan had 68% more rms error than the 27 Gy plan. Despite these differences in error, the dose deviations were negligible between the two fraction sizes. The H&N had a maximum dose accuracy difference of 1.3% and the spine had a maximum of 0.2%, between the two pairs of plans.

Given the results of this study, we then quantified the magnitude of delivery error that contributes to dose deviations among IROC phantom results. This involved requesting log files from phantom irradiations, which was previously not part of the IROC phantom analysis process. Log files were collected for H&N and lung irradiations between January 2019 and January 2022 and assessed using Mobius FX. Delivered doses were evaluated for dose deviation from the true TLD dose and MLC rms error was assessed to determine impact on phantom performance. Delivered dose error ranged between -0.3% and 0.5% for the H&N phantom and -0.8% to 0.2% for the lung results. Additionally, MLC rms error was small, and found to be machine dependent, where the Varian 21EX had an average of 0.3 mm which was tenfold greater than the Truebeam at 0.02 mm. Delivery error had an overall negligible contribution to dose deviations among the H&N and lung phantoms. This dataset is limiting in the fact that we only investigated plans irradiated on Varian machines and log files have been shown to not capture all errors that may exist with the MLCs (section 5.4).

The final component of delivery error evaluated was machine output and calibration error. To quantify this error, detailed in chapter 6, we used the institution's reported output of the treatment machine and the measured TLD output dose from their irradiation of the IROC output block, to compute the magnitude of output error, i.e., how far off the machine output was from a value of 1 (1cGy/1MU). We evaluated 548 output results across the H&N, spine and lung phantoms and found that majority (73%) of institutions reported a machine dose, that was higher than the true output dose calculated based on the TLDs, by an average of 1.8%. The

73% of over reported dose coincides closely with the amount of failing phantoms from the first study (chapter 2), where underdosing represented most failures that occurred from systematic dose errors. Underdosing represented 69%

of these failures across all phantoms while overdoing was responsible for the other 31% (Table 2-1). This could be due to institutions overestimating their machine output and possibly delivering a dose during treatment that is lower than intended.

### **7.1.4 Specific Aim 3**

#### **Combine all error contributions to determine the total magnitude of dose deviation accounted for among poor-performing IROC phantom irradiations**

In this aim we combine the results of the dose calculation and delivery errors quantified in aims 1 and 2 respectively. The results were categorized into two groups representing better and poorer performing phantom irradiation results. The better performing group of phantoms were ones with TPS vs TLD dose deviations less than 3.2% and the poorer performing phantoms had deviations greater than 3.2%. We used 3.2% as the threshold because this represents TLD measurement uncertainty at the  $2\sigma$  level, beyond which we are 95% certain that the dose deviation is as a result of a true error. Overall, we were able to quantify 56% of dose deviations in the H&N, 68% in the spine and 19% in the lung phantom irradiation results (section 6.3). Dose calculation and output errors comprised majority of these errors: 37% and 18% for H&N, 37% and 31% for spine, 10% and 9% for lung respectively, and showed positive correlations with the magnitude of phantom dose deviation (section 6-3). Delivered dose error (assessed through the log files) was minimal in its contribution and had no impact on phantom dose deviation.

## 7.2 Evaluation of the Hypothesis

The hypothesis of this project was that the sources of error evaluated would comprise on average, at least 50% of the magnitude of total dose deviation that is currently seen among all poorer performing IROC phantom irradiations. For the subset of phantoms having all errors evaluated, the magnitude of quantified error in the poor performing phantoms is as follows:

### 7.2.1 H&N phantoms

Error	Magnitude (%)
Dose calculation	37
Delivered dose	1
Machine output	18
Total	56

We quantified 56% of the total dose deviation for the poor performing phantom TLDs in the H&N irradiation results. Our hypothesis was therefore met for this phantom since we were able to quantify majority of the error among these phantom results.

### 7.2.2 Spine phantoms

Error	Magnitude (%)
Dose calculation	37
Delivered dose	N/A
Machine output	31
Total	68

We quantified 68% of the total dose deviation for the poor performing phantom TLDs in the Spine irradiation. Our hypothesis was also met for this phantom since we were able to quantify majority of the error among these phantom results. Despite not collecting enough log files to perform analysis on this phantom, the rest of the error is unlikely due to delivered dose error given that very little of this error type was found in either of the other phantoms.

### 7.2.3 Lung phantoms

Error	Magnitude (%)
Dose calculation	<b>10</b>
Delivered dose	<b>-0.3(<math>\approx 0</math>)</b>
Machine output	<b>9</b>
Total	<b>19</b>

We quantified 19% of the total dose deviation for the poor performing phantom TLDs in the lung irradiation. This is 31% shy of the 50% magnitude that we hypothesized. This low magnitude of error is not surprising for the lung phantom. The preliminary study (chapter 2) indicated that most of the error causing lung phantom failures was due to poor motion management. Motion management error was not quantified as part of this study due to the complexity and limitations involved with assessing this error remotely. This large amount of unquantified error creates a great opportunity for further investigation into this problem.

### **7.3 Future Research and Applications**

This study was a success in that it provided us with the reasons for over half of the dose deviations among the spine and H&N phantom results. A small portion of the lung dose deviations were quantified, and there is a great opportunity for improving on these results. As mentioned before, the lung phantom primarily fails due to motion management errors. This error is not easily quantifiable as the phantom program is a remote audit process, so it would be difficult to assess each institution's motion management program in terms of setup and execution at the machine, without physically being there. Additionally, a correlation between localization error (magnitude and direction) and dose deviation would have to be developed in order to computationally determine how much dose deviation is caused by each shift, whether it be superior-inferior or anterior-posterior (the two directions of motion for this phantom). This work could also be expanded to include other IROC phantoms such as the SRS H&N which is similar in geometry but different in treatment type to the IMRT H&N.

The results from this work will help IROC to better inform program participants of their phantom results (whether failing or passing), by utilizing the error modes investigated to provide a more detailed breakdown of their results. Currently, institutions are provided with a binary “pass/fail” result, along with their TPS vs TLD dose deviations, dose profiles and gamma analyses. When a failing result occurs, the institution oftentimes re-irradiates the phantom (in some cases up to 3 times) and reports any changes made to IROC, especially if the subsequent irradiation results in a pass. This approach may solve some of the issues that cause a failing result but may require multiple repeat irradiations to figure it out. Instead, providing descriptions of what areas need to be worked on, such as “machine output was too low” or “beam modeling might be sub-optimal” would better guide physicists in fixing or improving their results. This method has already been tested in a handful of cases at IROC in 2022. One



institution irradiated the IMRT H&N phantom, and although obtaining a passing result (TLD dose within  $\pm 7\%$ ) , their TPS vs TLD dose deviation was 5.5%. In addition to the current IROC procedures, we performed an additional evaluation of this phantom using the methods in this study. Dose calculation error contributed 3.9% and machine output contributed 1.8% for a combined dose deviation error of 5.7%, only 0.2% off from the total dose deviation seen for this phantom. Unfortunately, delivery error was not evaluated in this case due to the lack of delivery log files. This institution can then use this information to verify machine output and reevaluate their TPS beam modeling parameters to improve their future phantom and clinical outcomes. This example shows the extent of the detail that IROC would be able to provide as part of the evaluation and feedback to future program participants. Therefore, a critical next step would be implementing this work into the current IROC workflow.

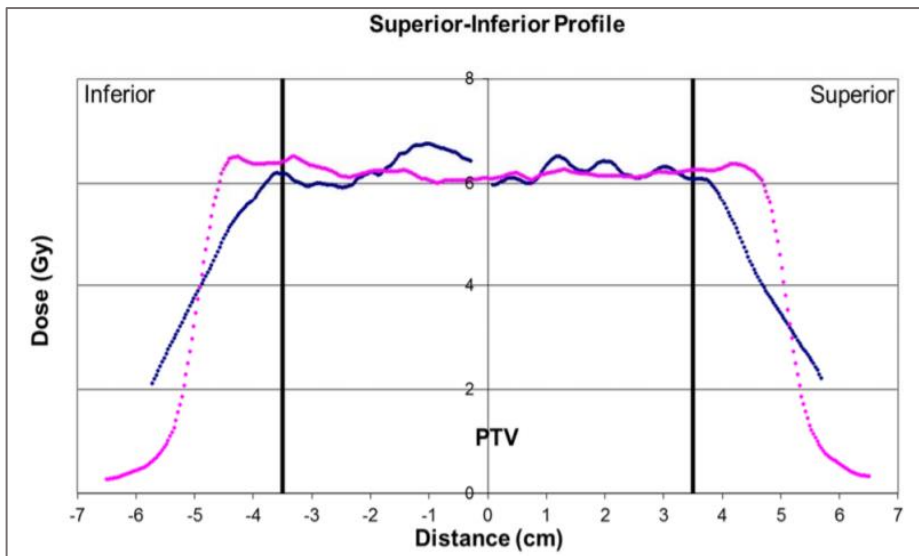
## Appendix A: Supplement to Chapter 2

This appendix serves as the supplement to

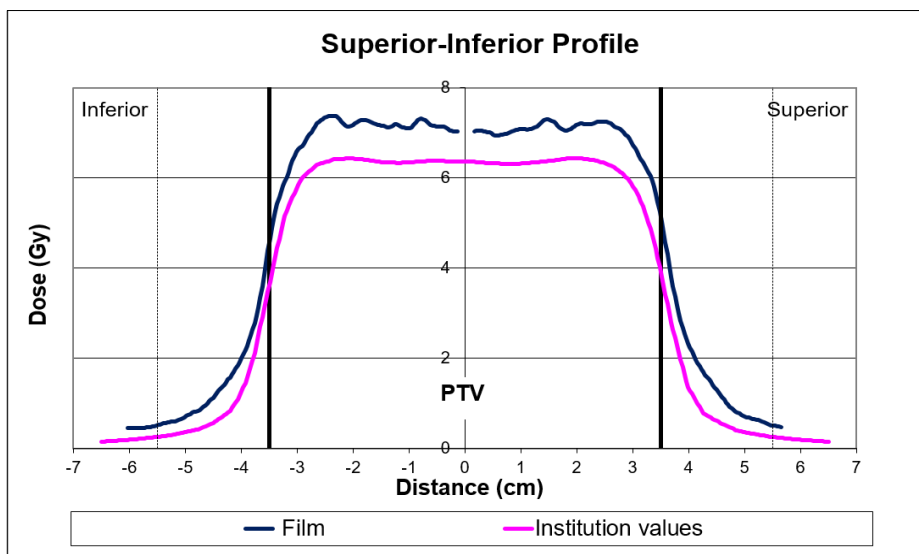
Chapter 2: Differences in the patterns of failure between IROC lung and spine phantom

irradiations

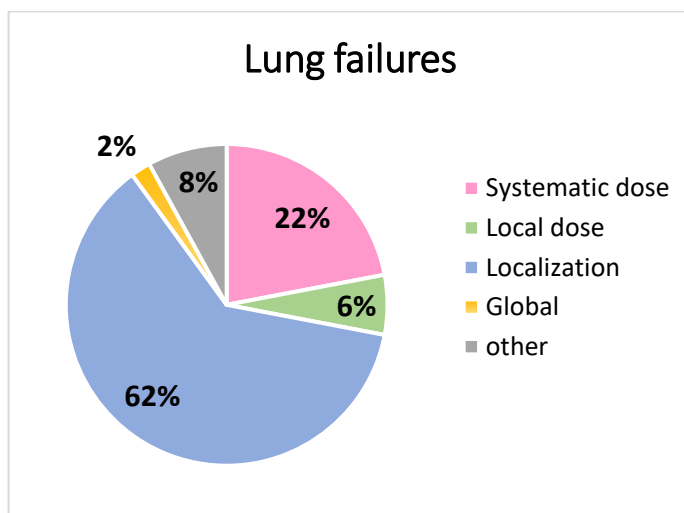
Additional failure category illustrations not included in manuscript:



**Figure 25.** Lung phantom local dose failure

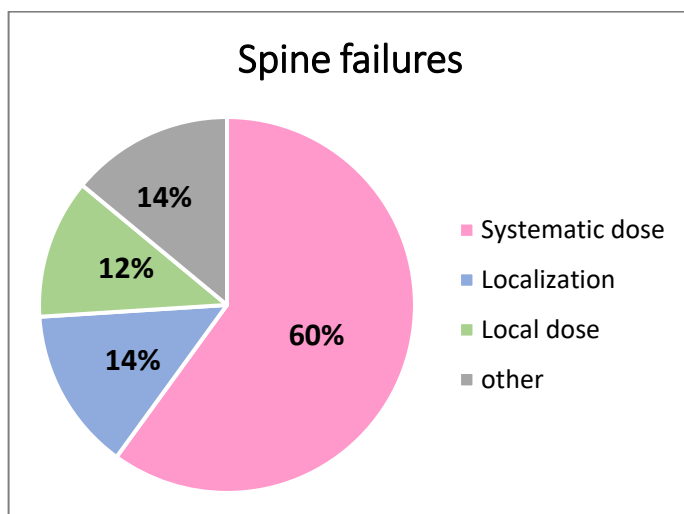


**Figure 26.** Lung phantom systematic overdose



**Figure 27.** Major categories of failure for lung phantoms

The “combination” category is listed as “other” for simplification.



**Figure 28.** Major categories of failure for spine phantoms

OAR overdose and localization errors are grouped as “other” for simplification.

The spreadsheet containing all data analysis is located on the IROC Houston network drive at

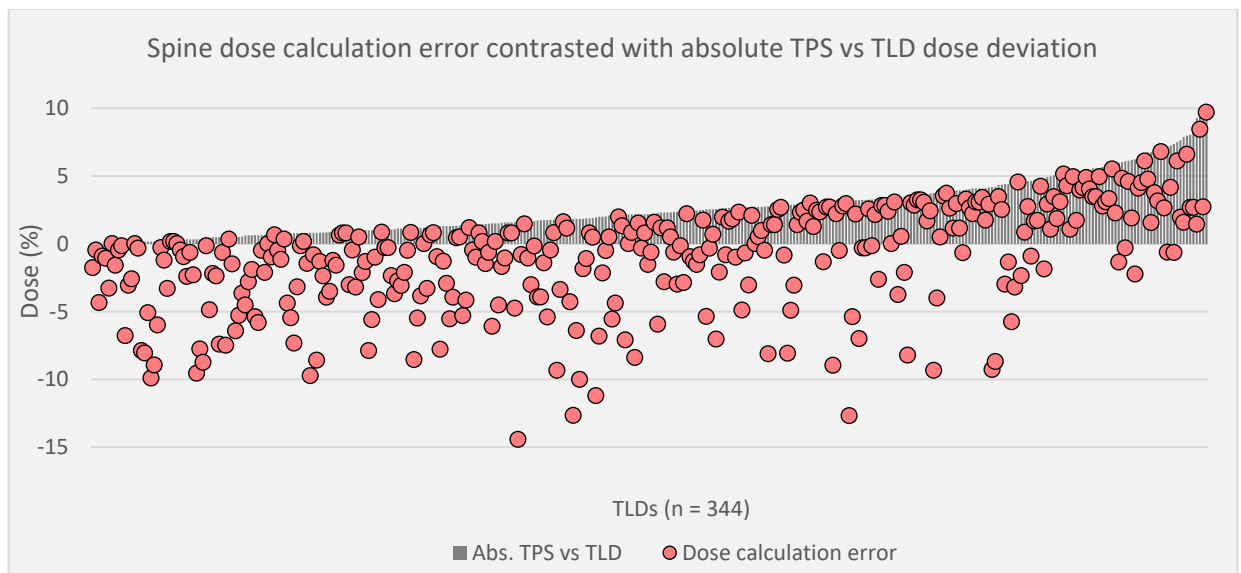
**J:\Everyone\Sharbacha\Dissertation\Appendix A**

## Appendix B: Supplement to Chapter 3

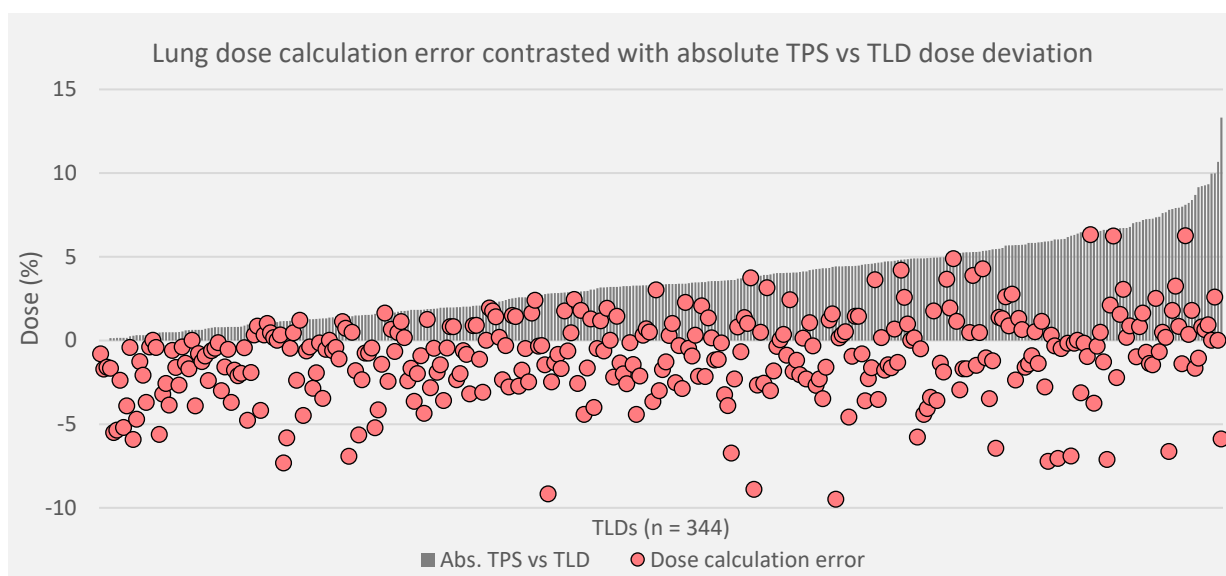
This appendix serves as the supplement to

### Chapter 3: Quantifying dose calculation errors in IROC Phantoms

The following figures are an extension of the results from the spine and lung phantom dose recalculations, showing correlations between dose calculation errors and TPS vs TLD dose deviations.

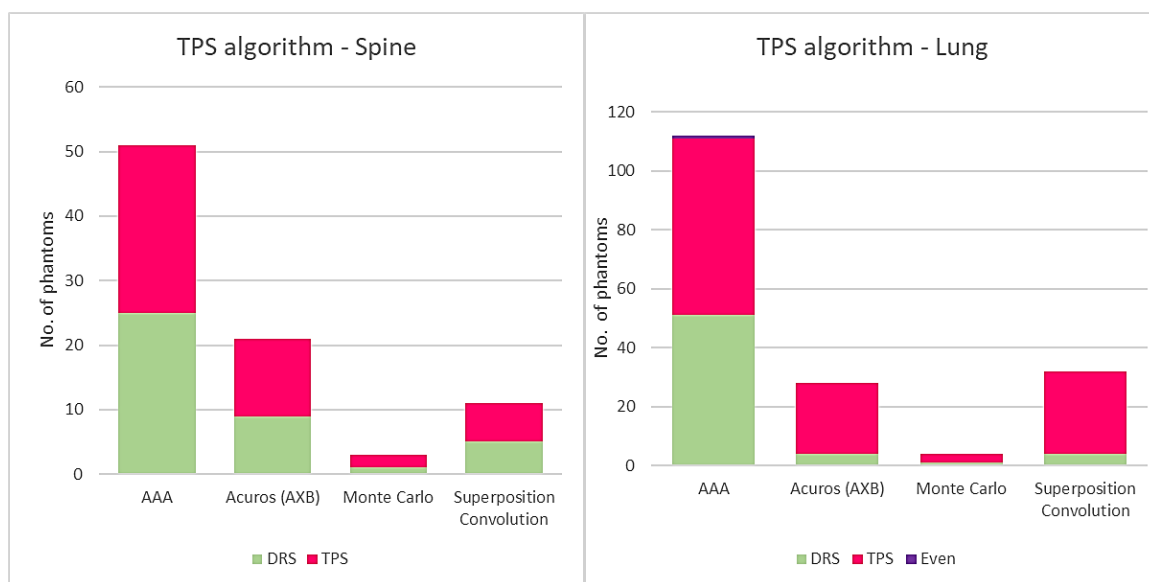


**Figure 29.** Magnitude of dose calculation error vs Absolute TPS vs TLD dose deviation for the spine phantom results.

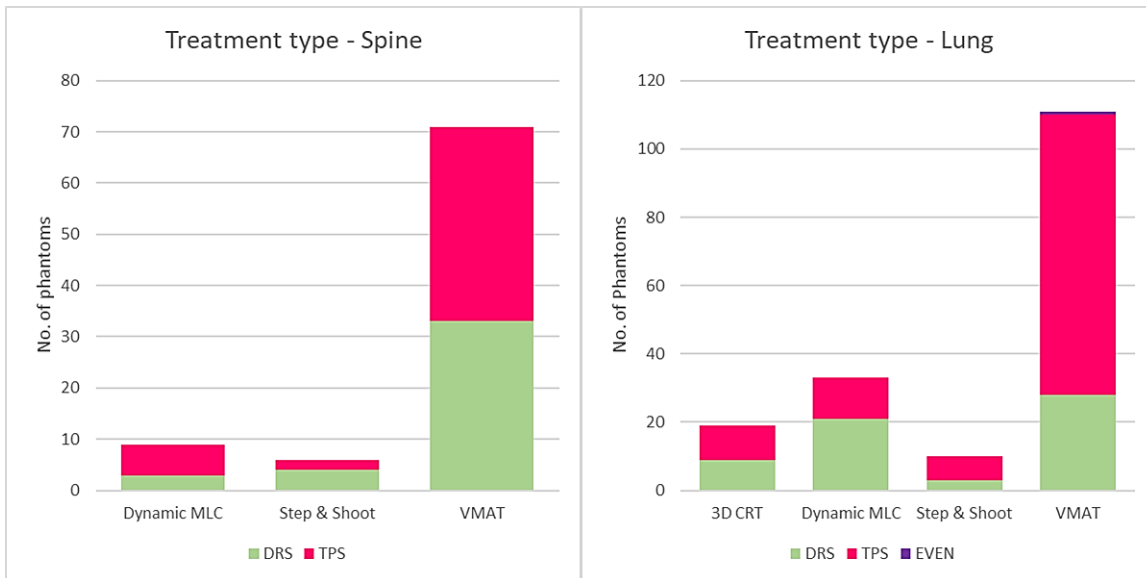


**Figure 30.** Magnitude of dose calculation error vs Absolute TPS vs TLD dose deviation for the lung phantom results.

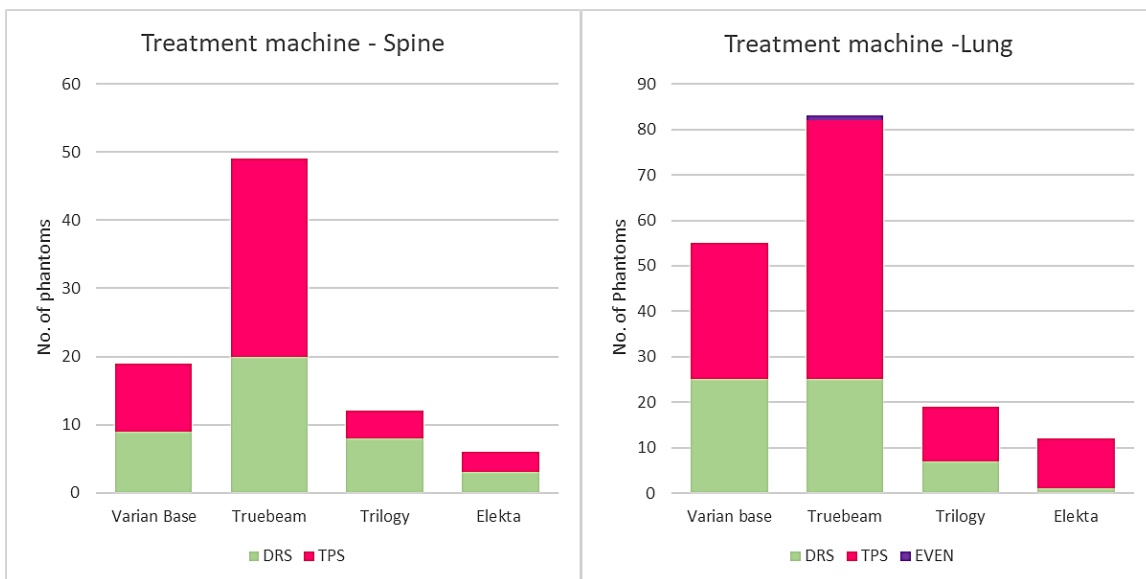
The figures below show DRS vs TPS performance based on TPS algorithm, treatment machine and treatment type for both phantoms, and additionally, respiratory management method for the lung phantom.



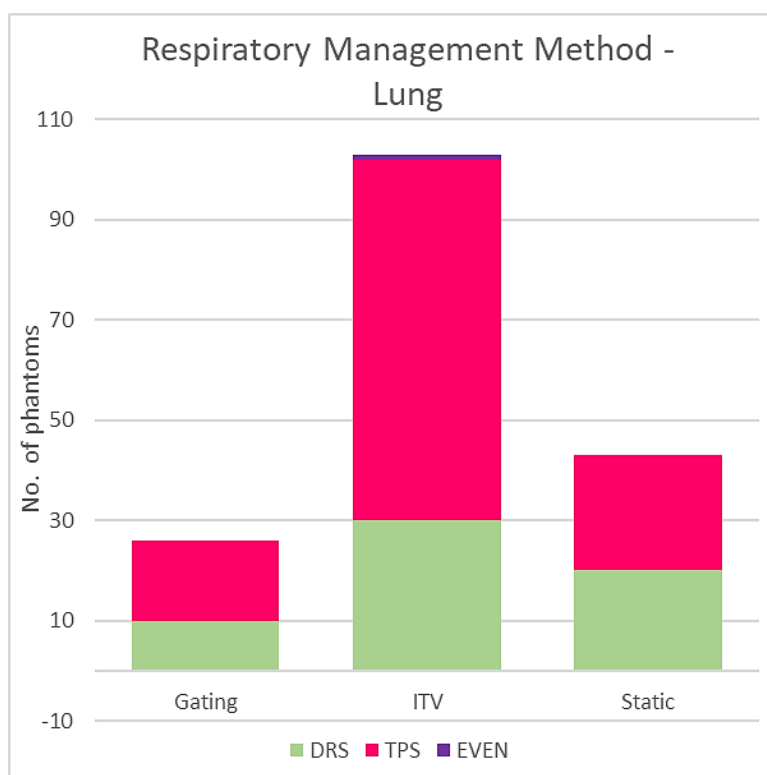
**Figure 31.** Phantom results categorized by TPS algorithm



**Figure 32.** Phantom results categorized by treatment type



**Figure 33.** Phantom results categorized by machine type



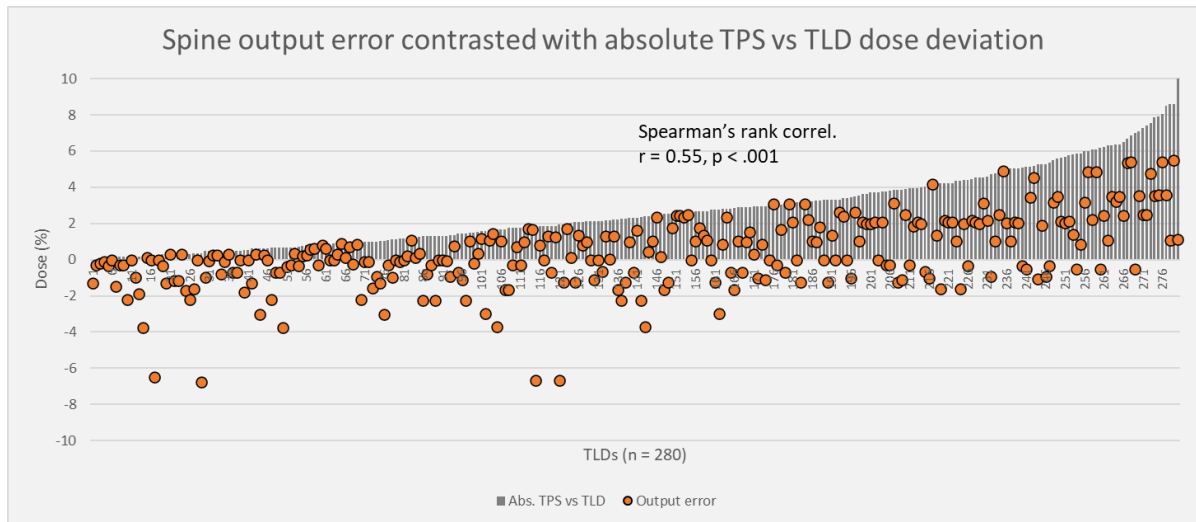
**Figure 34.** Lung phantoms categorized by respiratory management method

The spreadsheet containing all data analysis is located on the IROC Houston network drive at

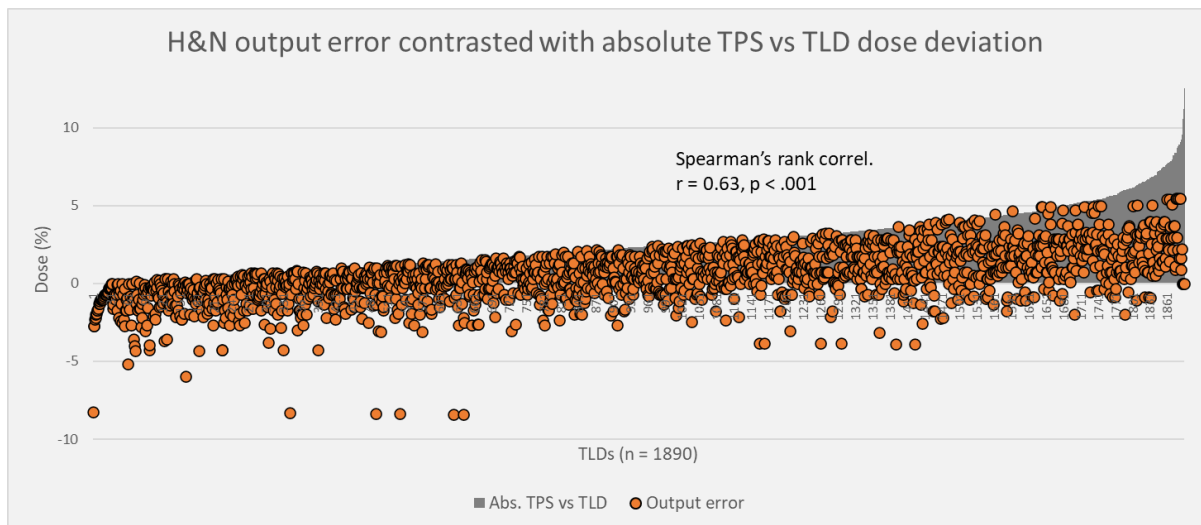
**J:\Everyone\Sharbacha\Dissertation\Appendix B**

## Appendix C: Supplement to Chapter 6

This appendix contains figures from the defense presentation of Output error.

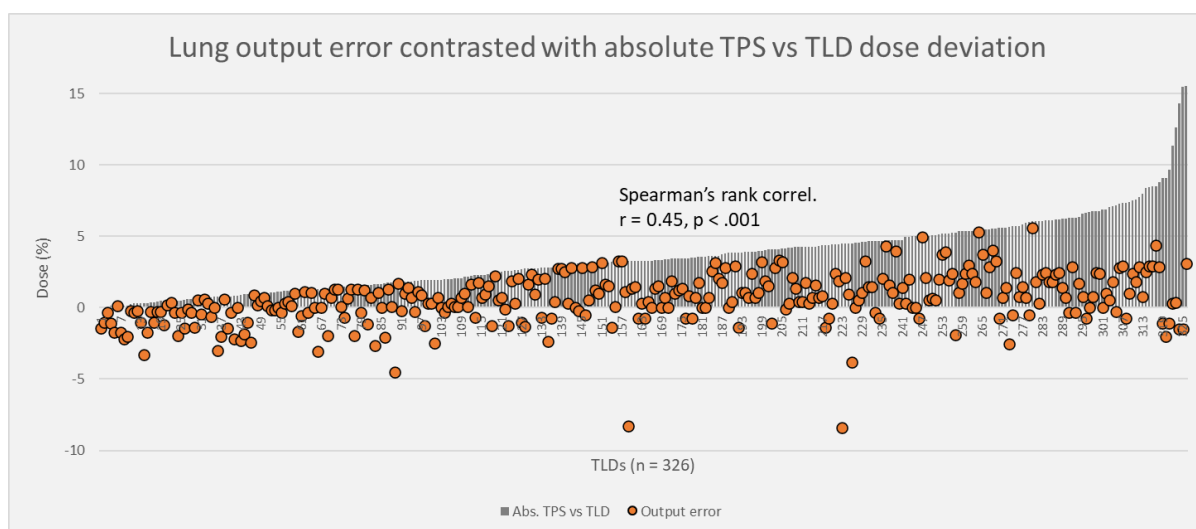


**Figure 35.** Magnitude of output error vs Absolute TPS vs TLD dose deviation for the spine phantom results.



**Figure 36.** Magnitude of output error vs Absolute TPS vs TLD dose deviation for the H&N phantom results.





**Figure 37.** Magnitude of output error vs Absolute TPS vs TLD dose deviation for the lung phantom results.

## Appendix D: Individual Cases & Additional Materials

### Sample cases: IMRT H&N

#### Case 1

**Table 9.** Raw TLD data for Case 1

	<b>TLD 1</b>	<b>TLD 2</b>	<b>TLD 3</b>	<b>TLD 4</b>	<b>TLD 5</b>	<b>TLD 6</b>	<b>Avg</b>
<b>TPS dose (Gy)</b>	6.7	6.8	6.7	6.8	5.6	5.6	6.3
<b>DRS dose (Gy)</b>	6.4	6.4	6.4	6.4	5.3	5.3	6.0
<b>TLD dose (Gy)</b>	6.3	6.3	6.4	6.3	5.2	5.2	5.9

**Table 10.** Analyzed TLD data for Case 1

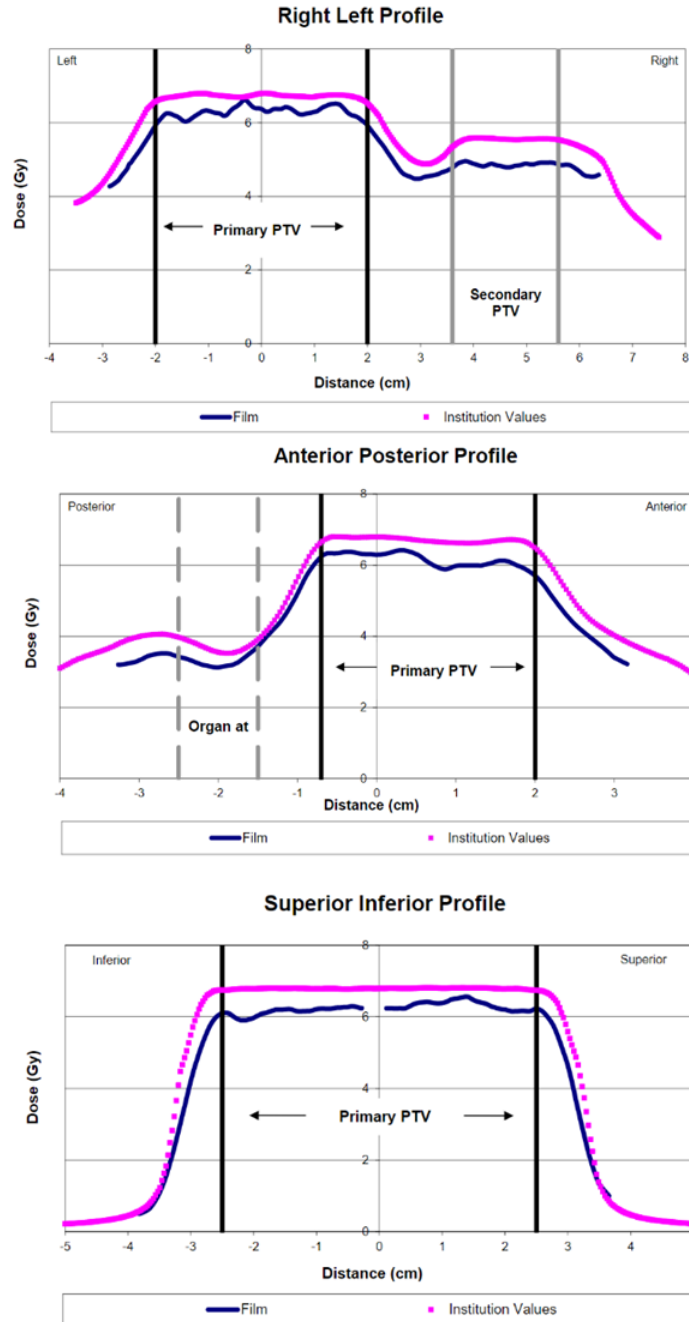
	<b>TPS vs TLD dose deviation (%)</b>	<b>Absolute TLD dose deviation (%)</b>	<b>Dose calculation error (%)</b>	<b>Delivery error (%)</b>	<b>Output error (%)</b>
TLD 1	-6.0	6.0	4.6	0.0	2.9
TLD 2	-6.4	6.4	5.7	-0.2	2.9
TLD 3	-4.6	4.6	4.2	-0.2	2.8
TLD 4	-6.9	6.9	5.5	-0.2	2.9
TLD 5	-6.8	6.8	4.8	-0.4	2.9
TLD 6	-5.9	5.9	4.6	-0.4	2.9
<b>Average</b>	<b>-6.1</b>	<b>6.1</b>	<b>4.9</b>	<b>-0.2</b>	<b>2.9</b>

The TPS vs TLD dose deviations in this phantom were primarily computational or dosimetric in nature. We were able to quantify this error because of this, with the major error contributions coming from dose calculation and machine output error.

- This phantom passed the IROC irradiation
- Machine output was measured as 2.7% low on that day
  - Output error assessments improved TPS to TLD agreement by 2.9%
- TPS calculated dose was 6.3 Gy on average, vs DRS calculated dose of 6.0 Gy.

- Dose calculation error improved TPS to TLD agreement by 4.9% on average.
- We found error to account for 7.8% total out of 6.1% TPS vs TLD dose deviation. This value puts us 1.7% above actual measured TLD value, but still much closer than the TPS was at 6.1% under measured value.

Dose profiles from this case are shown below:



**Figure 38.** Dose profiles of IMRT H&N Case 1 showing systematic underdosing of the PTVs, indicating dosimetric inaccuracies.

## Case 2

**Table 11.** Raw TLD data for Case 2

	<b>TLD 1</b>	<b>TLD 2</b>	<b>TLD 3</b>	<b>TLD 4</b>	<b>TLD 5</b>	<b>TLD 6</b>	<b>Avg</b>
<b>TPS dose (Gy)</b>	6.8	6.8	6.7	7.0	5.7	5.7	6.4
<b>DRS dose (Gy)</b>	6.9	6.8	6.7	6.9	5.8	5.7	6.4
<b>TLD dose (Gy)</b>	7.3	7.0	7.0	7.2	6.0	5.9	6.7

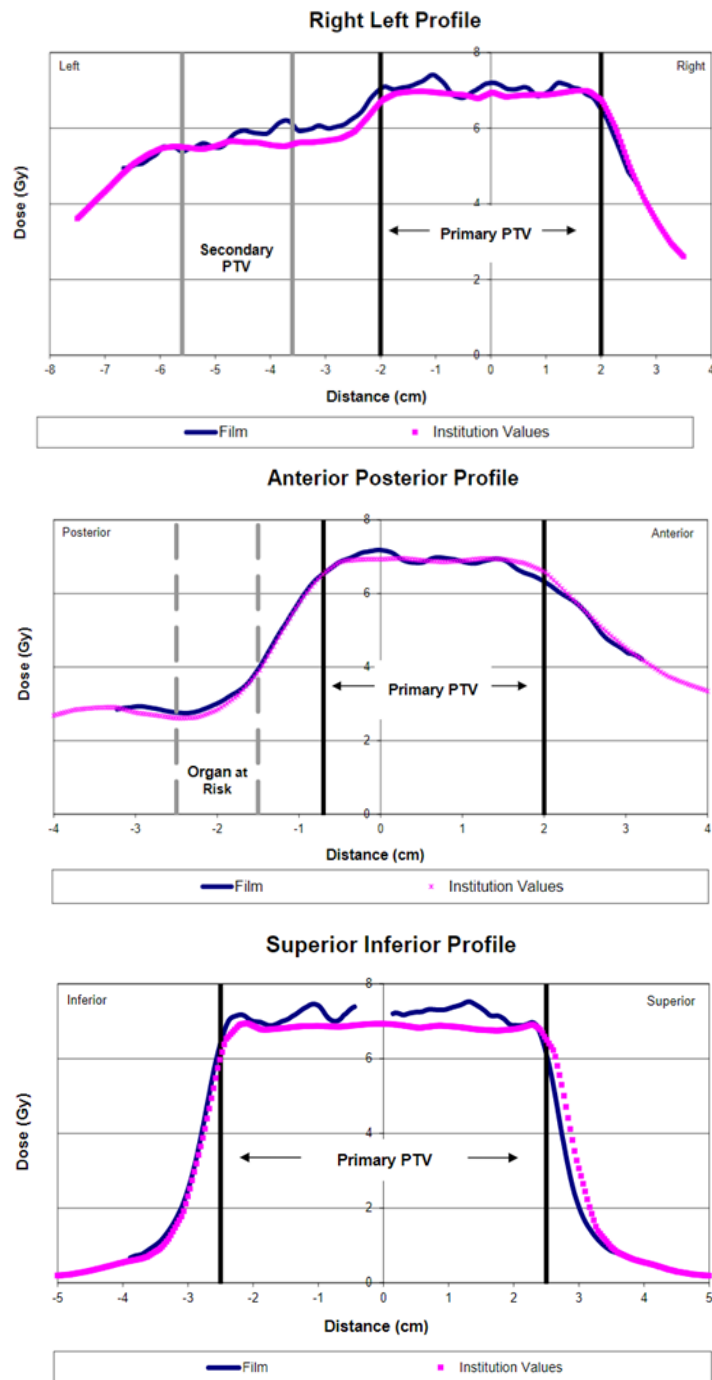
**Table 12.** Analyzed TLD data for Case 2

	<b>TPS vs TLD dose deviation (%)</b>	<b>Absolute TLD dose deviation (%)</b>	<b>Dose calculation error (%)</b>	<b>Delivery error (%)</b>	<b>Output error (%)</b>
TLD 1	6.3	6.3	0.7	0.1	-1.9
TLD 2	3.6	3.6	0.1	0.1	-1.9
TLD 3	3.9	3.9	0.3	0.1	-1.9
TLD 4	4.2	4.2	-1.1	0.1	-1.9
TLD 5	3.8	3.8	0.5	0.7	-1.9
TLD 6	3.7	3.7	-0.3	0.5	-1.9
<b>Average</b>	<b>4.2</b>	<b>4.2</b>	<b>0.0</b>	<b>0.3</b>	<b>-1.9</b>

- This phantom passed the IROC irradiation
- The TPS vs TLD dose deviations in this phantom case were not dosimetric in nature.

The dose profiles (below) indicate local dose discrepancies or errors along the SI and LR profiles. Greater delivery error effects were seen in the secondary PTV(TLDs 5 and 6) compared to the primary PTV:

- Primary PTV delivery error = 0.1%
- Secondary PTV delivery error = 0.6%



**Figure 39.** Dose profiles of IMRT H&N Case 2 showing local dose errors in various parts of the SI and LR profiles.

## Sample cases: Lung

### Case 3

**Table 13.** Raw TLD data for Case 3

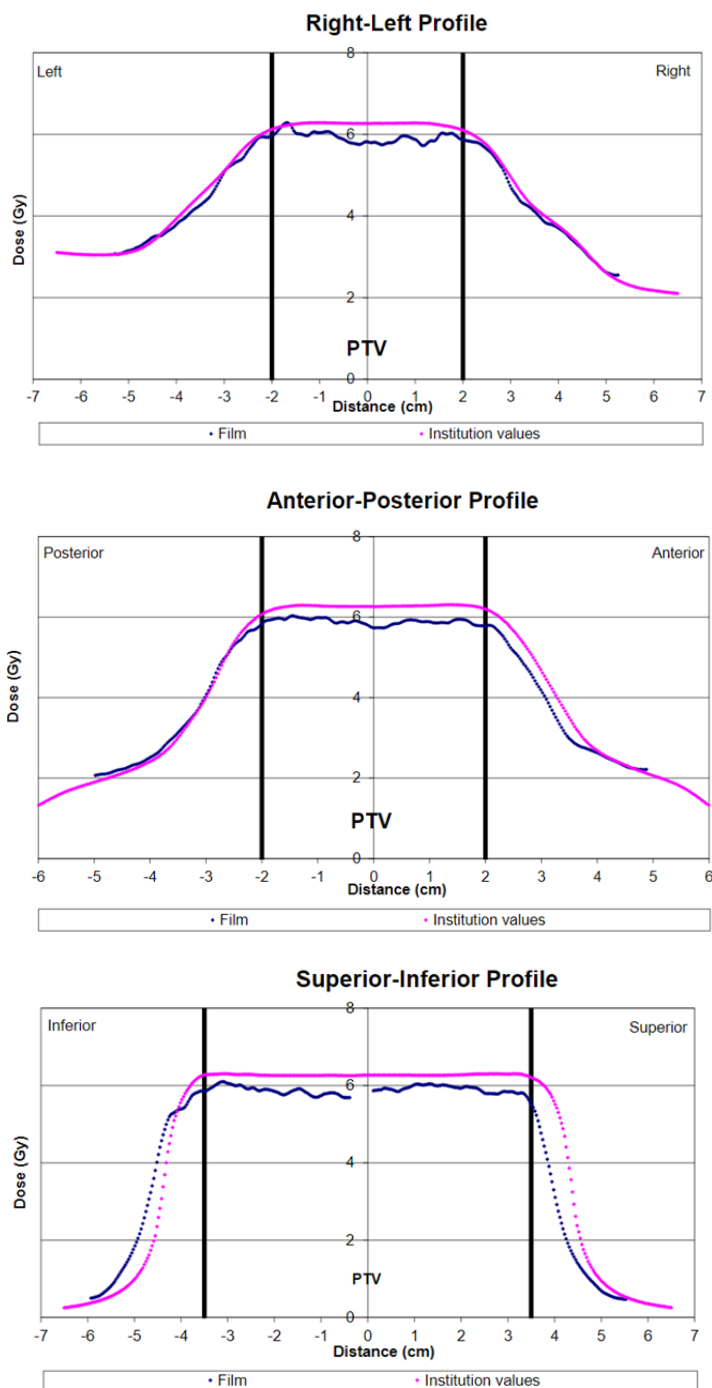
	<b>TLD 1</b>	<b>TLD 2</b>	<b>Avg</b>
<b>TPS dose (Gy)</b>	6.3	6.3	6.3
<b>DRS dose (Gy)</b>	6.2	6.2	6.2
<b>TLD dose (Gy)</b>	5.8	5.9	5.9

**Table 14.** Analyzed TLD data for Case 3

	<b>TPS vs TLD dose deviation (%)</b>	<b>Absolute TLD dose deviation (%)</b>	<b>Dose calculation error (%)</b>	<b>Delivery error (%)</b>	<b>Output error (%)</b>
TLD 1	-7.0	7.0	1.9	-0.2	4.3
TLD 2	-6.1	6.1	1.4	0.0	4.2
<b>Average</b>	<b>-6.5</b>	<b>6.5</b>	<b>1.6</b>	<b>-0.1</b>	<b>4.3</b>

The TPS vs TLD dose deviations in case 3 were dosimetric in nature.

- This phantom passed the IROC irradiation
- Machine output was measured as 4% low on that day
  - Output error assessments improved TPS to TLD agreement by 4.3% on average
- The TPS dose was calculated as 6.28 Gy on average and the DRS was 6.18 Gy. The TLD average dose was measured as 5.8Gy.
  - Dose calculation error assessment improved TPS to TLD dose agreement by 1.6% on average
- The total dose deviation quantified in this case was  $4.3\% + 1.6\% = 5.9\%$  out of a total 6.3% dose deviation on average. This equates to 90% of dose quantified for this phantom



**Figure 40.** Dose profiles of lung Case 3 showing systematic underdosing of the PTVs, indicating dosimetric inaccuracies.



## Case 4

**Table 15.** Raw TLD data for Case 4

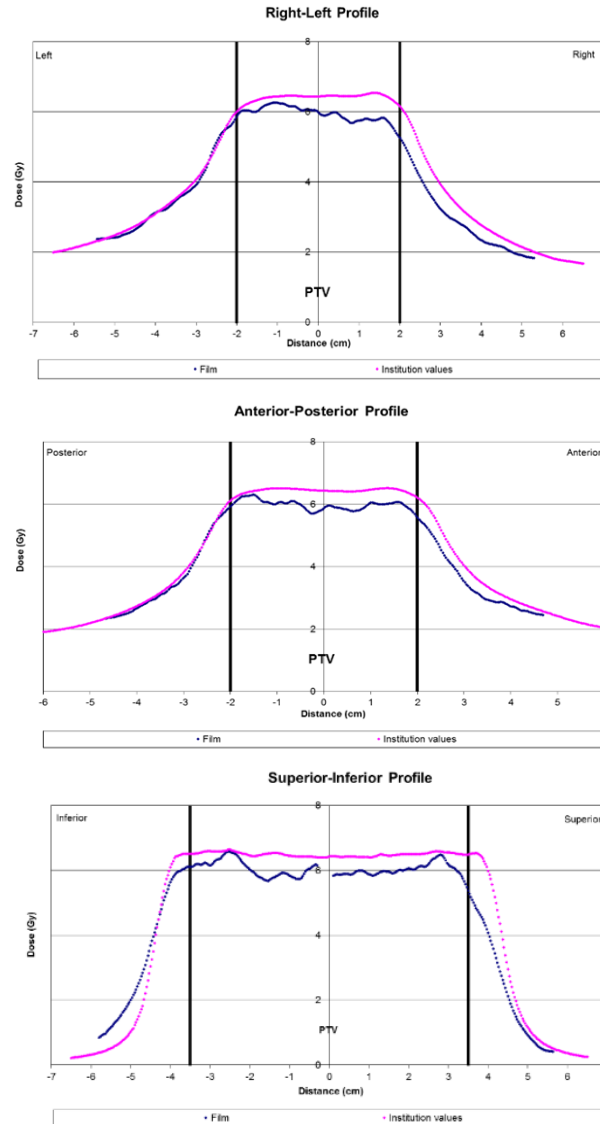
	<b>TLD 1</b>	<b>TLD 2</b>	<b>Avg</b>
<b>TPS dose (Gy)</b>	6.4	6.4	6.4
<b>DRS dose (Gy)</b>	6.2	6.3	6.3
<b>TLD dose (Gy)</b>	5.8	6.0	5.9

**Table 16.** Analyzed TLD data for Case 4

	<b>TPS vs TLD dose deviation (%)</b>	<b>Absolute TLD dose deviation (%)</b>	<b>Dose calculation error (%)</b>	<b>Delivery error (%)</b>	<b>Output error (%)</b>
TLD 1	-9.0	9.0	2.6	0.0	-2.1
TLD 2	-6.9	6.9	2.4	0.0	-6.8
<b>Average</b>	<b>-7.9</b>	<b>7.9</b>	<b>2.5</b>	<b>0.0</b>	<b>-4.4</b>

The TPS vs TLD dose deviations in case 4 were NOT dosimetric in nature.

- This phantom failed the IROC irradiation, TLD 1 was outside of the  $\pm 7\%$  dose tolerance.
- There were small improvements seen through dose calculation assessments: 2.5% on average.
- Output analysis did not cause an improvement. In fact, output assessments brought us further away from the true TLD dose: -4.4%.
- The dose profiles indicate an ITV or motion management related error



**Figure 41.** Dose profiles of lung Case 4 indicating ITV errors in the SI profile, where the planned dose (pink) appears broader than the film (blue) dose in the shoulder regions of the profile.

Additional phantom analysis data for all error types are located on the IROC Houston network drive at **J:\Everyone\Sharbacha\Dissertation\Appendix D**.

This contains:

- MLC perturbation Python code for use in Raystation:
  - both random and whole bank leaf shifts
- Spreadsheets of all remaining error analysis

## References

1. Institute NC. Statistics at a Glance: The Burden of Cancer in the United States. Published 2018. <https://www.cancer.gov/about-cancer/understanding/statistics>
2. Delaney G, Jacob S, Featherstone C, Barton M. The role of radiotherapy in cancer treatment. *Cancer*. 2005;104(6):1129-1137. doi:10.1002/CNCR.21324
3. Hodapp N. The ICRU Report No. 83: Prescribing, recording and reporting photon-beam intensity-modulated radiation therapy (IMRT). *Strahlentherapie und Onkol*. 2012;188(1):97-99. doi:10.1007/s00066-011-0015-x
4. Olch A. G. S. Ibbott, J. R. Anderson, J. Deye, T. J. Fitzgerald, D. S. Followill, M. T. RK, Gillin J. R. Palta, J. A. Purdy, M. M. Urie. MSH. Quality Assurance for Clinical Trials: A Primer for Physicists, Prepared by AAPM Subcommittee on QA for Clinical Trials. *AAPM Rep No 86 Am Inst Phys*. Published online 2004.
5. Ibbott GS. QA in Radiation Therapy: The RPC Perspective. *J Phys Conf Ser*. 2010;250:12001. doi:10.1088/1742-6596/250/1/012001
6. Carson ME, Molineu A, Taylor PA, Followill DS, Stingo FC, Kry SF. Examining credentialing criteria and poor performance indicators for IROC Houston's anthropomorphic head and neck phantom. *Med Phys*. 2016;43(12):6491-6496. doi:doi:10.1118/1.4967344
7. Kerns JR, Stingo F, Followill DS, Howell RM, Melancon A, Kry SF. Treatment Planning System Calculation Errors Are Present in Most Imaging and Radiation Oncology Core-Houston Phantom Failures. *Int J Radiat Oncol*. 2017;98(5):1197-1203. doi:<https://doi.org/10.1016/j.ijrobp.2017.03.049>
8. Baskar R, Lee KA, Yeo R, Yeoh KW. Cancer and radiation therapy: current advances and future directions. *Int J Med Sci*. 2012;9(3):193-199. doi:10.7150/ijms.3635
9. Ohri N, Shen XL, Dicker AP, Doyle L, Harrison A, Showalter T. Radiotherapy Protocol

- Deviations and Clinical Outcomes: A Meta-analysis of Cooperative Group Clinical Trials. *Jnci-Journal Natl Cancer Inst.* 2013;105(6):387-393. doi:10.1093/jnci/djt001
10. Peters L, O’Sullivan B, Giralt J, Fitzgerald T, Trotti A, Bernier J, Bourhis J, Yuen K, Fisher R, Rischin D. Critical Impact of Radiotherapy Protocol Compliance and Quality in the Treatment of Advanced Head and Neck Cancer: Results From TROG 02.02. *J Clin Oncol.* 2010;28(18):2996-3001. doi:10.1200/JCO.2009.27.4498
  11. Ibbott GS, Followill DS, Molineu HA, Lowenstein JR, Alvarez PE, Roll JE. Challenges in Credentialing Institutions and Participants in Advanced Technology Multi-institutional Clinical Trials. Accessed August 29, 2022. <http://www.qarc.org>
  12. Followill DS, Evans DR, Cherry C, Molineu A, Fisher G, Hanson WF, Ibbott GS. Design, development, and implementation of the Radiological Physics Center’s pelvis and thorax anthropomorphic quality assurance phantoms. *Med Phys.* 2007;34(6Part1):2070-2076. doi:doi:10.1118/1.2737158
  13. Kry S, Alvarez P, Molineu A, Amador C, Galvin J, Followill D. Algorithms Used in Heterogeneous Dose Calculations Show Systematic Error as Measured With the Radiological Physics Center’s Anthropomorphic Thorax Phantom Used for RTOG Credentialing. *Int J Radiat Oncol • Biol • Phys.* 2012;84(3):S128-S129. doi:10.1016/j.ijrobp.2012.07.134
  14. Shoales J, Followill D, Ibbott G, Balter P, Tolani N. WE-D-T-617-07: Development of An Independent Audit Device for Remote Verification of 4D Radiotherapy. *Med Phys.* 2005;32(6Part19):2138. doi:10.1118/1.1998567
  15. Molineu A, Followill DS, Balter PA, Hanson WF, Gillin MT, Huq MS, Eisbruch A, Ibbott GS. Design and implementation of an anthropomorphic quality assurance phantom for intensity-modulated radiation therapy for the Radiation Therapy Oncology

- Group. *Int J Radiat Oncol*. 2005;63(2):577-583.  
doi:<https://doi.org/10.1016/j.ijrobp.2005.05.021>
16. Kerns JR, Followill DS, Lowenstein J, Molineu A, Alvarez P, Taylor PA, Stingo FC, Kry SF. Technical Report: Reference photon dosimetry data for Varian accelerators based on IROC-Houston site visit data. *Med Phys*. 2016;43(5):2374-2386.  
doi:10.1118/1.4945697
  17. Kerns JR, Followill DS, Lowenstein J, Molineu A, Alvarez P, Taylor PA, Kry SF. Reference dosimetry data and modeling challenges for Elekta accelerators based on IROC-Houston site visit data. *Med Phys*. 2018;45(5):2337-2344. doi:10.1002/mp.12865
  18. Kim J, Wu Q, Zhao B, Wen N, Ajlouni M, Movsas B, Chetty IJ. To gate or not to gate - dosimetric evaluation comparing Gated vs. ITV-based methodologies in stereotactic ablative body radiotherapy (SABR) treatment of lung cancer.(Report). *Radiat Oncol*. 2016;11(1). doi:10.1186/s13014-016-0699-2
  19. Kim J, Zhao B, Ajlouni M, Movsas B, Chetty IJ. SU-C-210-01: Are Clinically Relevant Dosimetric Endpoints Significantly Better with Gating of Lung SBRT Vs. ITV-Based Treatment?: Results of a Large Cohort Investigation Analyzing Predictive Dosimetric Indicators as a Function of Tumor Volume and Motion A. *Med Phys*. 2015;42(6):3204.  
doi:10.1118/1.4923846
  20. Rana S, Rogers K, Pokharel S, Cheng C. Evaluation of Acuros XB algorithm based on RTOG 0813 dosimetric criteria for SBRT lung treatment with RapidArc. *J Appl Clin Med Phys*. 2014;15(1):118-129. doi:10.1120/jacmp.v15i1.4474
  21. Wu QJ, Thongphiew D, Wang Z, Chankong V, Yin FF. The impact of respiratory motion and treatment technique on stereotactic body radiation therapy for liver cancer. *Med Phys*. 2008;35(4):1440-1451. doi:10.1118/1.2839095

22. Osborn VW, Lee A, Yamada Y. Stereotactic Body Radiation Therapy for Spinal Malignancies. *Technol Cancer Res Treat*. 2018;17:1533033818802304. doi:10.1177/1533033818802304
23. Abbas H, Chang B, Chen ZJ. Motion management in gastrointestinal cancers. *J Gastrointest Oncol*. 2014;5(3):223-235. doi:10.3978/j.issn.2078-6891.2014.028
24. Du W, Cho SH, Zhang X, Hoffman KE, Kudchadker RJ. Quantification of beam complexity in intensity-modulated radiation therapy treatment plans. *Med Phys*. 2014;41(2):21716. doi:10.1118/1.4861821
25. Followill DS, Urie M, Galvin JM, Ulin K, Xiao Y, Fitzgerald TJ. Credentialing for participation in clinical trials. *Front Oncol*. 2012;2:198. doi:10.3389/fonc.2012.00198
26. 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). doi:10.1118/1.4773309
27. Edward SS, Alvarez PE, Taylor PA, Molineu HA, Peterson CB, Followill DS, Kry SF. Differences in the patterns of failure between IROC lung and spine phantom irradiations. *Pract Radiat Oncol*. Published online May 13, 2020. doi:10.1016/j.prro.2020.04.004
28. Petrocchia HM, Malajovich I, Barsky AR, Ghiam AF, Jones J, Wang C, Zou W, Teo BKK, Dong L, Metz JM, Li T. Spine SBRT With Halcyon™: Plan Quality, Modulation Complexity, Delivery Accuracy, and Speed. *Front Oncol*. 2019;9(319). doi:10.3389/fonc.2019.00319
29. Kirby TH, Hanson WF, Johnston DA. Uncertainty analysis of absorbed dose calculations from thermoluminescence dosimeters. *Med Phys*. 1992;19(6):1427-1433. doi:10.1118/1.596797
30. Kerns JR, Followill DS, Lowenstein J, Molineu A, Alvarez P, Taylor PA, Kry SF.

- Agreement Between Institutional Measurements and Treatment Planning System Calculations for Basic Dosimetric Parameters as Measured by the Imaging and Radiation Oncology Core-Houston. *Int J Radiat Oncol Biol Phys*. 2016;95(5):1527-1534. doi:10.1016/j.ijrobp.2016.03.035
31. Childress N, Stevens E, Eklund D, Zhang M. Mobius3D White Paper: Dose Calculation Algorithm. *Mobius Med Syst LP*. Published online 2013.
  32. Kry SF, Dromgoole L, Alvarez P, Leif J, Molineu A, Taylor P, Followill DS. Radiation Therapy Deficiencies Identified During On-Site Dosimetry Visits by the Imaging and Radiation Oncology Core Houston Quality Assurance Center. *Int J Radiat Oncol Biol Phys*. 2017;99(5):1094-1100. doi:10.1016/j.ijrobp.2017.08.013
  33. Glenn MC, Peterson CB, Followill DS, Howell RM, Pollard-Larkin JM, Kry SF. Reference dataset of users' photon beam modeling parameters for the Eclipse, Pinnacle, and RayStation treatment planning systems. *Med Phys*. Published online October 30, 2019:mp.13892. doi:10.1002/mp.13892
  34. Edward SS, C. Glenn M, Peterson CB, Balter PA, Pollard-Larkin JM, Howell RM, S. Followill D, Kry SF. Dose calculation errors as a component of failing IROC lung and spine phantom irradiations. *Med Phys*. 2020;47(9):4502-4508. doi:10.1002/mp.14258
  35. Glenn M, Followill D, Howell R, Pollard-Larkin J, Peterson C, Kry S. *Sensitivity Analysis of Common Beam Modeling Parameters in the Eclipse Treatment Planning System on IROC Head and Neck Phantom Results*.
  36. Kerns JR, Childress N, Kry SF. A multi-institution evaluation of MLC log files and performance in IMRT delivery. *Radiat Oncol*. 2014;9(1). doi:10.1186/1748-717X-9-176
  37. Nithiyanantham K, Mani GK, Subramani V, Mueller L, Palaniappan KK, Kataria T. Analysis of direct clinical consequences of MLC positional errors in volumetric-

- modulated arc therapy using 3D dosimetry system. *J Appl Clin Med Phys*. 2015;16(5):296-305. doi:10.1120/jacmp.v16i5.5515
38. Mu G, Ludlum E, Xia P. Impact of MLC leaf position errors on simple and complex IMRT plans for head and neck cancer. *Phys Med Biol*. 2008;53(1):77-88. doi:10.1088/0031-9155/53/1/005
  39. Delishaj D, Ursino S, Pasqualetti F, Matteucci F, Cristaudo A, Soatti C Pietro, Barcellini A, Paiar F. Set-up errors in head and neck cancer treated with IMRT technique assessed by cone-beam computed tomography: a feasible protocol. *Radiat Oncol J*. 2018;36(1):54-62. doi:10.3857/roj.2017.00493
  40. Gupta T, Sinha S, Ghosh-Laskar S, Budrukkar A, Mummudi N, Swain M, Phurailatpam R, Prabhash K, Agarwal JP. Intensity-modulated radiation therapy versus three-dimensional conformal radiotherapy in head and neck squamous cell carcinoma: Long-term and mature outcomes of a prospective randomized trial. *Radiat Oncol*. 2020;15(1):1-9. doi:10.1186/S13014-020-01666-5/FIGURES/3
  41. Tseng CL, Eppinga W, Charest-Morin R, Soliman H, Myrehaug S, Maralani PJ, Campbell M, Lee YK, Fisher C, Fehlings MG, Chang EL, Lo SS, Sahgal A. Spine Stereotactic Body Radiotherapy: Indications, Outcomes, and Points of Caution. *Glob Spine J*. 2017;7(2):179. doi:10.1177/2192568217694016
  42. TrueBeamArchitecture\_ProductBrief\_RAD10152\_Oct2010.pdf. Accessed October 24, 2022. [https://varian.widen.net/view/pdf/vdqrgwfh0w/TrueBeamArchitecture\\_ProductBrief\\_RAD10152\\_Oct2010.pdf?u=wefire](https://varian.widen.net/view/pdf/vdqrgwfh0w/TrueBeamArchitecture_ProductBrief_RAD10152_Oct2010.pdf?u=wefire)
  43. Agnew A, Agnew CE, Grattan MWD, 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. doi:10.1088/0031-9155/59/9/N49
44. Saiful Huq M, Fraass BA, Dunscombe PB, Gibbons JP, Ibbott GS, Mundt AJ, Mutic S, Palta JR, Rath F, Thomadsen BR, Williamson JF, Yorke ED. The report of Task Group 100 of the AAPM: Application of risk analysis methods to radiation therapy quality management. Published online 2016. doi:10.1118/1.4947547
  45. Fontenota JD. Evaluation of a novel secondary check tool for intensity-modulated radiotherapy treatment planning. *J Appl Clin Med Phys*. 2014;15(5):207-215. doi:10.1120/jacmp.v15i5.4990
  46. Glenn MC, Peterson CB, Howell RM, Followill DS, Pollard-Larkin JM, Kry SF. Sensitivity of IROC phantom performance to radiotherapy treatment planning system beam modeling parameters based on community-driven data. *Med Phys*. 2020;47(10):5250-5259. doi:10.1002/mp.14396
  47. Taylor PA, Alvarez PE, Mehrens H, Followill DS. Failure Modes in IROC Photon Liver Phantom Irradiations. *Pract Radiat Oncol*. 2021;11(3):e322-e328. doi:10.1016/J.PRRO.2020.11.008
  48. Kutcher GJ, Coia L, Gillin MT, Hanson WF, Leibel S, Morton RJ, Palta JR, Purdy JA, Reinstein LE, Svensson GK, Weller M, Wingfield L. AAPM Report of TG 40 Comprehensive QA for Radiation Oncology. *Med Phys*. Published online 1994.
  49. Klein EE, Hanley J, Bayouth J, Yin FF, Simon W, Dresser S, Serago C, Aguirre F, Ma L, Arjomandy B, Liu C, Sandin C, Holmes T. Task group 142 report: Quality assurance of medical accelerators. *Med Phys*. 2009;36(9):4197-4212. doi:10.1118/1.3190392
  50. Alvarez P, Kry SF, Stingo F, Followill D. TLD and OSLD dosimetry systems for remote audits of radiotherapy external beam calibration. doi:10.1016/j.radmeas.2017.01.005

## **Vita**

Sharbacha Sherby Norma Edward was born in Vieux-Fort, Saint Lucia to Norman and Sonia Edward. She earned the highest grade on the island at the secondary school (high school) entrance exam and went on to graduate from the Vieux-Fort Comprehensive Secondary School (VFCSS). She then completed her A levels in Mathematics, Chemistry and Physics and proceeded to teach Mathematics at her alma mater, VFCSS, for three years. In 2010, Sharbacha moved to Chicago, Illinois to pursue her bachelor's degree in Physics at the Illinois Institute of Technology. She then worked for 2 years at the Siteman Cancer Center in St. Louis as a clinical physics assistant. In July 2017, she moved to Houston Texas to pursue her PhD in medical physics at the MD Anderson Cancer Center UTHHealth Houston Graduate School of Biomedical Sciences.

Permanent address:

Savannes Bay, Vieux-Fort

Saint Lucia, W.I.