The Texas Medical Center Library
DigitalCommons@TMC

Dissertations and Theses (Open Access)

MD Anderson UTHealth Houston Graduate School

12-2011

The Role Of Cell Sterilization In Population Based Studies Of Radiogenic Second Cancers Following Radiation Therapy

Annelise Giebeler

Follow this and additional works at: https://digitalcommons.library.tmc.edu/utgsbs_dissertations

Part of the Clinical Trials Commons, Neoplasms Commons, and the Other Physics Commons

Recommended Citation

Giebeler, Annelise, "The Role Of Cell Sterilization In Population Based Studies Of Radiogenic Second Cancers Following Radiation Therapy" (2011). *Dissertations and Theses (Open Access)*. 214.
https://digitalcommons.library.tmc.edu/utgsbs_dissertations/214

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.

The Role of Cell Sterilization in Population Based Studies of Radiogenic Second Cancers Following Radiation Therapy

By

Annelise Giebeler, M.S.

APPROVED:

_____
Wayne D. Newhauser, Ph.D.
Supervisory Professor

_____
Carol Etzel, Ph.D.

_____
Rebecca Howell, Ph.D.

_____
Anita Mahajan, M.D.

_____
Dragan Mirkovic, Ph.D.

_____
George Coutrakon, Ph.D.

APPROVED:

_____
Dean, The University of Texas
Health Science Center at Houston
Graduate School of Biomedical Sciences

THE ROLE OF CELL STERILIZATION IN POPULATION BASED STUDIES OF RADIOGENIC SECOND CANCERS FOLLOWING RADIATION THERAPY

A

DISSERTATION

Presented to the Faculty of
The University of Texas
Health Science Center at Houston
and
The University of Texas
M. D. Anderson Cancer Center
Graduate School of Biomedical Sciences
in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

by

Annelise Giebeler, M.S.
Houston, Texas

December, 2011

Copyright (c) 2011 Annelise Giebeler, All rights Reserved

Dedication

This work is dedicated to Milla Harmer, Lucille Shetler, Faith Shetler, and Marie Giebeler.

Thank you for paving the way.

Acknowledgements

This project has been the vehicle for an amazing journey which was made possible by the support, guidance and friendship of a great number of people; its success is truly their success. With that, I would like to sincerely thank Dr. Wayne Newhauser for his contribution to this work and my academic progress. His vision and insight have been both invaluable and inspiring. I would also like to thank my co-mentor, Dr. Carol Etzel, for her guidance. Her objective input was a refreshing gift on many levels. In addition, I would like to thank Dr. Rebecca Howell for sharing her photon expertise and her unwavering encouragement. Her support and enthusiasm have been indispensable, particularly as the project faced its completion. Finally, I would like to thank the members of my advisory, exam, and supervisory committees: Dr. Ed Jackson, Dr. Uwe Titt, Dr. George Coutrakon, Dr. Dragan Mirkovic, Dr. Anita Mahajan, and Dr. David David Followill. Their suggestions provided clarity and added new dimensions to this dynamic project.

I would also like to express my gratitude to individuals in the Proton Working Group: Dr. Phil Taddei, Dr. Rui Zhang, Laura Rechner, Dr. Angelica Perez-Andujar, Kenneth Homann, John Eley, and Scharlene Wilson as each member of this amazing group played a vital role in the ultimate success of this work. Additionally, I would like to thank individuals from the Medical Physics program and the Cancer Prevention program who offered kind support through the duration of this project, specifically Georgeanne Moore, Dr. Shine Chang, Dr. Carrie Cameron, and Dee Tello.

The Role of Cell Sterilization in Population Based Studies of Radiogenic Second Cancers Following Radiation Therapy

Publication No. _____

Annelise Giebeler, M.S.

Supervisory Professor: Wayne D. Newhauser, Ph.D.

Advances in radiotherapy have generated increased interest in comparative studies of treatment techniques and their effectiveness. In this respect, pediatric patients are of specific interest because of their sensitivity to radiation induced second cancers. However, due to the rarity of childhood cancers and the long latency of second cancers, large sample sizes are unavailable for the epidemiological study of contemporary radiotherapy treatments. Additionally, when specific treatments are considered, such as proton therapy, sample sizes are further reduced due to the rareness of such treatments. We propose a method to improve statistical power in micro clinical trials. Specifically, we use a more biologically relevant quantity, cancer equivalent dose (D_{CE}), to estimate risk instead of mean absorbed dose (D_{MA}). Our objective was to demonstrate that when D_{CE} is used fewer subjects are needed for clinical trials. Thus, we compared the impact of D_{CE} vs. D_{MA} on sample size in a virtual clinical trial that estimated risk for second cancer (SC) in the thyroid following craniospinal irradiation (CSI) of pediatric patients using protons vs. photons. Dose reconstruction, risk models, and statistical analysis were used to evaluate SC risk from therapeutic and stray radiation from CSI for 18 patients. Absorbed dose was calculated in <u>*two*</u> ways: with (1) traditional D_{MA} and (2) with D_{CE}. D_{CE} and D_{MA} values were used to estimate relative risk of SC incidence (RR_{CE} and RR_{MA}, respectively) after proton vs. photon CSI. Ratios of *RR* for proton vs. photon CSI (RRR_{CE} and RRR_{MA}) were then used in comparative estimations of sample size to determine the minimal number of patients needed to maintain 80% statistical power when using D_{CE} vs. D_{MA}. For all patients, we found that protons substantially reduced the risk of developing a second thyroid cancer when compared to photon therapy. Mean *RRR* values were 0.052±0.014 and 0.087±0.021 for RRR_{MA} and RRR_{CE}, respectively. However, we did not find that use of D_{CE} reduced the number of patents needed for acceptable statistical power (i.e, 80%). In fact, when considerations were

made for *RRR* values that met equipoise requirements and the need for descriptive statistics, the minimum number of patients needed for a micro-clinical trial increased from 17 using D_{MA} to 37 using D_{CE}. Subsequent analyses revealed that for our sample, the most influential factor in determining variations in sample size was the experimental standard deviation of estimates for *RRR* across the patient sample. Additionally, because the relative uncertainty in dose from proton CSI was so much larger (on the order of 2000 times larger) than the other uncertainty terms, it dominated the uncertainty in *RRR*. Thus, we found that use of corrections for cell sterilization, in the form of D_{CE}, may be an important and underappreciated consideration in the design of clinical trials and radio-epidemiological studies. In addition, the accurate application of cell sterilization to thyroid dose was sensitive to variations in absorbed dose, especially for proton CSI, which may stem from errors in patient positioning, range calculation, and other aspects of treatment planning and delivery.

Table of Contents

Signature Page i
Title Page ii
Dedication iii
Acknowledgements iv
Abstract v
Table of Contents vii
1 Introduction and background 1
1.1 General 1
1.2 Statement of the problem 2
1.3 Hypothesis and specific aims 2
2 Methods and materials 4
2.1 Design of the clinical trial 4
2.2 Definitions of organ dose quantities 5
2.2.1 Mean absorbed dose model 5
2.2.2 Cancer equivalent dose model 6
2.3 Patient accrual 6
2.4 Total absorbed dose determined with therapeutic and stray radiation 7
2.5 Therapeutic dose 9
2.5.1 Treatment plans 9
2.5.2 Volume-weighted mean therapeutic dose 13
2.6 Stray Dose 14
2.6.1 Stray dose from proton CSI 14
2.6.2 Stray dose from photon CSI 16
2.7 Equivalent Dose 17
2.8 Risk prediction 18
2.8.1 Risk prediction using mean absorbed dose model 18
2.8.2 Risk prediction using cancer equivalent dose model 20
2.9 Sample size calculations 21
2.10 Uncertainty in predicted risk 22
2.10.1 Uncertainty in absorbed dose, σ_{DX} 23

2.10.2 Uncertainty in the cell sterilization factor, σ_C 26

2.10.3 Uncertainty in the neutron weighting factor, $\sigma_{\overline{w}_R}$ 27

3 Results 29

3.1 Estimation of therapeutic dose 29

3.1.1 Therapeutic dose 29

3.2 Estimation of stray dose 36

3.3 Estimation of total equivalent dose 38

3.4 Risk prediction 45

3.4.1 Relative risk estimates from photon vs. proton CSI 45

3.4.2 Ratio of relative risk estimates for photon vs. proton CSI 49

3.5 Sample size estimations 53

3.5.1 Four factors that affect sample size 58

3.6 Uncertainty analysis 63

3.6.1 Uncertainty in absorbed dose 63

3.6.2 Uncertainty in the cell sterilization factor 65

3.6.3 Uncertainty in the mean radiation weighing factor for neutrons 66

3.6.4 Uncertainty in RRR when D_{CE} vs. D_{MA} are used 67

4 Discussion 69

4.1 Previous literature 69

4.2 Major findings and implications 73

4.3 Study strengths 74

4.4 Study limitations 75

4.5 Future work 76

5 Conclusions 78

Bibliography 80

Vita 88

List of Figures

Figure 1. Illustration of the role of the virtual clinical trial in determining sample size 5
Figure 2. Distribution of age and sex within the CSI cohort. 7
Figure 3. Illustration of the dose distribution for proton CSI. 10
Figure 4. Illustration of the age-specific proton beam range for spinal fields 11
Figure 5. Illustration of the FIF treatment technique for 6 MV photon CSI 12
Figure 6. Illustration of dDVH with dose less than 1 cGy 14
Figure 7. Illustration of scaling volumes (SV)'s. 15
Figure 8. Location of the thyroid relative to isodose lines 17
Figure 9. Thyroid location with respect to treatment field in representative patient 24
Figure 10. Distribution of D_{MA} for proton and proton CSI 30
Figure 11. Sex-specific histograms: size and relative anatomical location of the thyroid 34
Figure 12. Sex-specific thyroid volume and physical depth from skin vs. patient age at exposure 35
Figure 13. Percent difference in D_{MA} between TPS and spread sheet estimates 36
Figure 14. Stray dose from proton CSI 37
Figure 15. Total equivalent dose proton CSI 39
Figure 16. Total equivalent dose photon CSI 40
Figure 17. Absolute difference between H_{MA} and H_{CE} for photon CSI 43
Figure 18. Cell sterilization factor for proton and photon CSI. 44
Figure 19. Relative Risk estimates for photon and proton CSI vs. age at exposure 45
Figure 20. RR estimates for photon and proton CSI vs. age at exposure 46
Figure 21. Comparison of RR_{MA} to RR_{CE} 47
Figure 22. Ratio of relative risk after proton vs. proton CSI 49
Figure 23. Differences in estimates for RRR when D_{MA} vs. D_{CE} were used 52
Figure 24. Percent difference in estimates for RRR when D_{MA} vs. D_{CE} were used 52
Figure 25. Effect of D_{MA} vs. D_{CE} on sample size 57
Figure 26. Effect of power on sample size 59
Figure 27. Effect of standard deviation on sample size 61
Figure 28. Effect of alpha on sample size 62
Figure 29. Thyroid location relative to the treatment field for proton and photon CSI 63
Figure 30. D_{CE} with various correction factors 65

Figure 31. Comparison of dose-response models 72

List of Tables

Table 1. Summary of the methods used to estimate therapeutic and stray dose 8
Table 2. BEIR VII committee's preferred values for modeling thyroid cancer incidence. 19
Table 3. Comparisson of the input parameters used in sample size estimates for the t-test using two freely available statistics packages 21
Table 4. Values for set-up uncertainty per prevailing standards of care for proton and photon CSI. 25
Table 5. Radiation weighting factors for neutrons for uncertainty test patient 28
Table 6. D_{MA} and D_{CE} from our spreadsheet for proton and photon CSI 29
Table 7. D_{MA} from TPS vs. D_{MA} from our spreadsheet for photon CSI. 31
Table 8. D_{MA} from TPS vs. D_{MA} from our spreadsheet for proton CSI. 32
Table 9. Outliers from the comparison of D_{MA} estimates from the TPS vs. spreadsheet for proton CSI. 33
Table 10. Stray dose from proton CSI 37
Table 11. Total equivalent dose proton CSI 38
Table 12. Total equivalent dose proton CSI: summary statistics 39
Table 13. Total equivalent dose photon CSI 41
Table 14. Total equivalent dose photon CSI: summary statistics 42
Table 15. Comparison of therapeutic doses to stray dose 43
Table 16. Ratios of male-to-female values of dose, risk, and risk coefficient 48
Table 17. Ratio of relative risk from photon vs. proton CSI when D_{CE} vs. D_{MA} is applied 50
Table 18. Summary statistics for the ratio of relative risk after photon vs. proton CSI 51
Table 19. Results of t-test to determine if there is a statistically significant difference between RRR_{CE} vs. RRR_{MA} 53
Table 20. Sample size estimation for one sample t-test 54
Table 21. Sample size estimation for two sample t-test 55
Table 22. Input parameters for estimates of sample size relative to differing values for power (1-β), standard deviation of RRR (σ_{RRR}), and α. 58
Table 23. Effect of power on sample size for RRR_{MA} vs. RRR_{CE} 60
Table 24. Absorbed dose and summary statistics for shifts in the thyroid location relative to the original treatment field 64
Table 25. Uncertainty estimates for cell sterilization 66

Table 26. Uncertainty in the mean radiation weighting factor for neutrons.....67
Table 27. Estimates of relative uncertainty in *RRR* when D_{CE} vs. D_{MA} are used 67

Chapter One

1 Introduction and background

1.1 General

In recent decades, there has been notable improvement in the survival of children diagnosed with cancer. Estimates from the SEER program indicate that the overall 5-year survival rate for children aged 0-20 increased from 65% during 1978-1990 to 80% during 1996-2004 (Ries et al. 2006); thus, both the number of survivors and their associated life span are increasing. Reasons for these increases include centralization of care, improvements in chemotherapy and surgery, and key advances in radiotherapy. Radiotherapy is of particular interest here because of its vital role in pediatric treatment design (Taylor 1996).

While proton therapy is noted for its increased ability to produce conformal dose distributions about the tumor, its most notable attribute is its potential to spare healthy tissue. Correspondingly, many have speculated that proton therapy carries relatively lower risk for secondary malignant neoplasms (SMNs), or second cancers that occur in regions of low levels of ionizing radiation. However, there is justified skepticism in that speculation; in particular, there is concern regarding the use of proton radiation therapy because of risk of SMN's from stray neutron irradiation (Hall et al. 2006, Hall et al. 2009; Newhauser et al. 2009a, Newhauser et al. 2009b). Several recent works have investigated stray neutron irradiation in proton therapy (Zheng et al. 2007; Fontenot et al. 2008). These works provided more accurate estimations regarding risks of SMNs after proton therapy. The need for a clear understanding of SMN risk is vital and urgent due to the growing numbers of proton therapy centers (Sisterson 2005).

Regardless of the treatment method, ionizing radiation is known to be carcinogenic (ICRP 1990; Wrixon 2008), and SMN risk has been studied epidemiologically for patients who received radiation therapy (Howlader et al. 2010). In addition, epidemiological studies have examined the risk of SMNs as a result of childhood cancer survival (Linet et al. 1999; Mertens et al. 2001; Neglia et al. 2001; Inskip 2007). However, in the recent literature there have not been reports of epidemiological studies which focus on advanced radiotherapies.

Several groups have reported on stray neutron contributions to dose in both proton and photon modalities (Yan et al. 2002; Kry et al. 2005; Howell et al. 2006; Zheng et al. 2007; Fontenot et al. 2008), and case studies have compared and contrasted the dose-related risk of

radiation induced SMNs after proton and photon treatments (Miralbell et al. 2002; Schneider et al. 2007; Ruben et al. 2008).

Achieving statistically significant results in an epidemiological study that focuses on the use of advanced radiotherapy modalities (such as proton therapy) has been difficult for several reasons. First, childhood cancer is rare; the Surveillance, Epidemiology and End Results program of the NCI reported only 12,400 total first cancers per year in the US (Ries et al. 1999). Second, the latency associated with second cancers is typically long, e.g., typically longer than 5 years for solid tumors (Friedman et al. 2009). Finally, the evolution of radiotherapy equipment is so rapid that it is difficult to accrue a statistically significant result within the life cycle of one generation of treatment technology.

1.2 Statement of the problem

Although there is a clear and urgent need to understand the risk of SMNs following the use of conventional photon vs. proton radiotherapy in pediatric populations, there is still a remarkable paucity of relevant data regarding this subject. Relatively few children develop cancer and even fewer pediatric patients have been treated with proton therapy. The resulting small size of the patient population has thus impeded efforts to accrue a sufficient number of patients for epidemiological correlation of cause and effect. Large sample sizes are typically needed for an epidemiological study because dose across organs is averaged using several points of calculation across the organ of interest (Stovall et al. 2006); as a result, large numbers of patients are needed to offset the large and variable errors in individual dose values. Furthermore, because of the rarity of childhood cancers and the long latencies for SMNs, epidemiologic investigations are challenging. Specifically, years or decades may pass before enough data are available, by which time the relevance of the results is diminished. This is of particular concern given that children are reported to have six times the sensitivity to radiation that adults have (Inskip and Curtis 2007).

1.3 Hypothesis and specific aims

To overcome the obstacle of small sample sizes, we proposed a method to increase the power of a radiation epidemiological study for a given sample size. Our method modified the current practice of using mean absorbed dose with organ-specific SMN risk models by incorporating a cell sterilization factor that accounts for confounding factors such as the non-

linear response of tissue to absorbed dose [20] and non-uniform dose distributions in a specific organ. The hypothesis of this study is that **acceptable statistical power (80%) for an epidemiological study of secondary cancer incidence following radiation therapy can be attained with smaller numbers of patients using "cancer equivalent dose" in place of the traditional "mean absorbed dose."**

To test the hypothesis we conducted a virtual epidemiological study (i.e., a pilot simulation study for possible future radio-epidemiological studies) that used "cancer equivalent dose" and "mean absorbed dose" to estimate the risk of second cancer incidence of the thyroid in survivors treated with craniospinal irradiation (CSI) using proton vs. photon radiotherapies. We used data from this study to determine the number of subjects needed for 80% power in our study. To do this, addressed two specific aims:

Specific Aim 1: Determine the smallest number of subjects needed to achieve 80% power when cancer equivalent dose (D_{CE}) vs. mean absorbed dose (D_{MA}) is used in a virtual epidemiological study of second cancer in the thyroid following CSI using proton vs. photon radiotherapies.

Specific Aim 2: Estimate the uncertainty associated with calculations of risk of second cancer in the thyroid for one representative patient following proton radiotherapy vs. photon radiotherapies when cancer equivalent dose (D_{CE}) and mean absorbed dose (D_{MA}) are respectively applied.

Chapter Two

2 Methods and materials

This chapter describes the methods and materials used in this study. It begins with an overview of the virtual clinical trial and its design (2.1). Next, it presents formalisms for the dose models that were used in this work (2.2), namely the mean absorbed (2.2.1) and cancer equivalent (2.2.2) dose models. Then, it details patient selection methods (2.3) and the definition of total absorbed dose (2.4). Descriptions are given on how values for therapeutic dose (2.5) and stray dose (2.6) were estimated and extracted for proton CSI (2.6.1) and photon CSI (2.6.2), respectively. After that, it includes a section which details how risk was predicted (2.7) using the mean absorbed dose (2.7.1) and cancer equivalent dose (2.7.2). The chapter concludes by describing the methods for estimating statistical power and sample size estimations (2.8) and the uncertainty analysis performed for the study (2.9).

2.1 Design of the clinical trial

We designed a prospective virtual clinical trial to estimate the risk of second cancers in the thyroid following childhood CSI using current radiotherapy techniques. More precisely, (1) we used an established risk model to estimate the risk of a second cancers in the thyroid (NRC 2006), and (2) we used calculated organ and tissue doses. We used this approach to perform an virtual (in-silico) trial in order to overcome the obstacles of small available sample size and the 5-10 year latency time for thyroid cancer after pediatric exposure (Inskip 2001 and references therein). To study the impact of two current radiotherapy techniques, we defined two arms for our trial. In the first arm, we considered CSI with passively scattered protons in the 160 MeV to 200 MeV proton energy interval. In the second arm we considered CSI with 6 MV photons. In both arms CSI was delivered according to the current standards of care (SOC) at our institution (details provided in sections 2.5.1 and 2.5.2). A diagram which illustrates the role of the virtual clinical trial within the overall framework of this project follows (Figure 1).

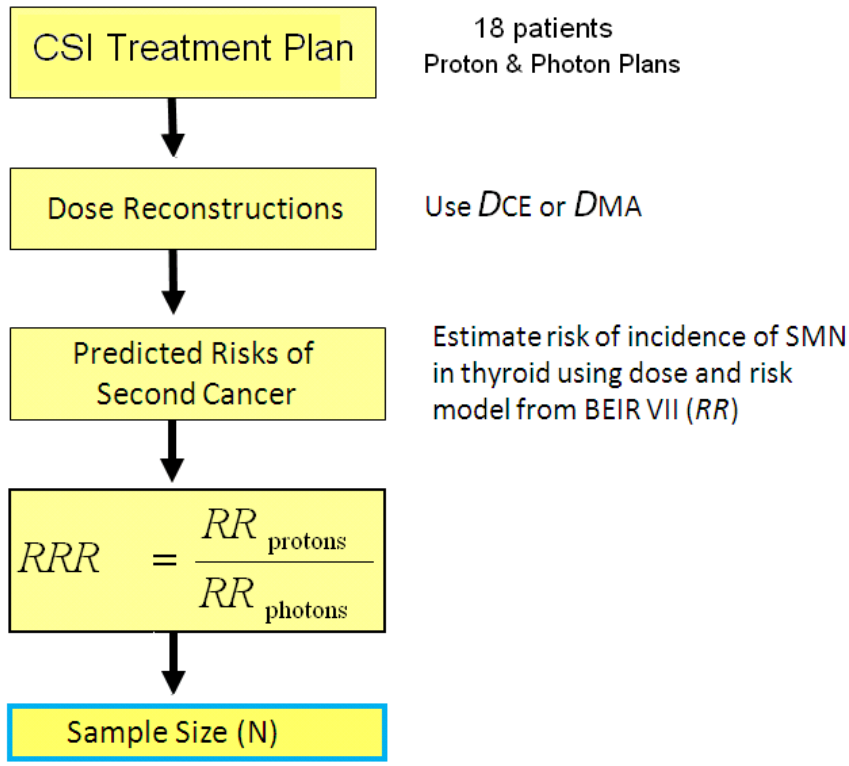

Figure 1. Illustration of the role of the virtual clinical trial in determining sample size
Two dose models, termed D_{MA} and D_{CE}, are applied to the virtual clinical trial to determine their respective effect on the sample size needed and the power achieved in the trial. Details regarding the dose models follow in section 2.2.

2.2 Definitions of organ dose quantities

Dose was calculated in two ways: using the traditional quantity of mean absorbed dose, D_{MA}, (ICRP 1990) and the quantity of cancer equivalent dose, D_{CE}, (Schneider and Kaser-Hotz 2005; Schneider et al. 2005). To ensure that we accounted for biological effects at the location in the thyroid where they were most likely occuring, we applied D_{CE} on a voxel by voxel basis. Then to reduce the number of varibles between use of D_{MA} vs. D_{CE}, we applied D_{MA} on a voxel by voxel basis as well. Details for D_{MA} vs. D_{CE} follow.

2.2.1 Mean absorbed dose model

The mean absorbed dose, D_{MA}, is the mass-weighted average absorbed dose to an organ. It this study it was calculated in the thyroid using

$$D_{MA} = \frac{\sum_{i=1}^{M} D_i m_i}{\sum_{i=1}^{M} m_i} = \frac{\sum_{i=1}^{M} D_i V_i \rho_i}{\sum_{i=1}^{M} V_i \rho_i}, \tag{2.1}$$

where M is the number of dose voxels in the thyroid[1], D_i is the dose at the ith voxel, m_i is the mass at the ith voxel, ρ_i is the mass density of tissue at the ith voxel, and V_i is the volume at the ith voxel (constant for voxels i - j). D_{MA} was averaged over each voxel in the thyroid to account for potential non-uniformities in absorbed dose and mass density.

2.2.2 Cancer equivalent dose model

The cancer equivalent dose, D_{CE} is defined in this study as

$$D_{CE} = \frac{\sum_{i=1}^{M} D_i m_i C_i}{\sum_{i=1}^{M} m_i} = \frac{\sum_{i=1}^{M} D_i V_i \rho_i C_i}{\sum_{i=1}^{M} V_i \rho_i}, \tag{2.2}$$

where M is the number voxels in the thyroid, D_i is the absorbed dose at the ith voxel, m_i is the mass at the ith voxel, ρ_i is the mass density of tissue at the ith voxel, V_i is the volume at the ith voxel (constant for voxels i - j), and C_i is the cell sterilization factor for ith voxel. C_i was calculated according to

$$C_i = e^{-\alpha_{organ} D_i}, \tag{2.3}$$

where α_{organ} is an organ-specific cell sterilization parameter which we defined following the methods of Schneider et al. (2005) and applied to the absorbed dose D_i. Similar to D_{MA}, the cancer equivalent dose was summed over each voxel in the thyroid to provide a corrected dose, or D_{CE}, which allowed for consideration of non-uniformities in absorbed dose, mass density, and cell sterilization.

2.3 Patient accrual

Patient data were collected using the consecutive sampling method (Lunsford and Lunsford 1995). The inclusion criteria for the study include: age (2 - 18 y at treatment), treatment at MDACC between 1998 and 2009, treated on protocol (CCGA-9961) or according to the prevailing radiotherapy standard of care (SOC) at MDACC, availability of CT images, and availability of relevant treatment plans. The sample was fairly evenly distributed in age and sex (Figure 2).

[1] 'M' is used for the number of dose voxels in the thyroid because 'N' is used to represent the number of patients in the study. Because the number of dose voxels in the thyroid is constant for each patient, M will be the same for D_{MA} and D_{CE}.

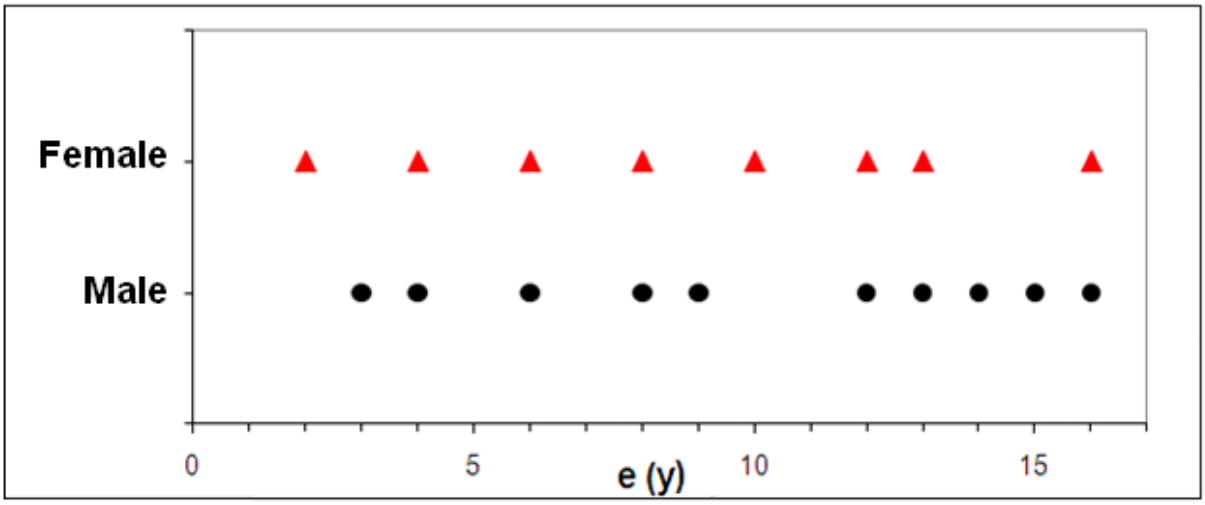

Figure 2. Distribution of age and sex within the CSI cohort.
Male and female patients are plotted for their age at exposure (e). There are 16 males (2-16 y) and 18 females (2-18 y). The maximum difference between the age-paired male and female patients was one year.

For practical purposes, several additional inclusion criteria were used. In particular, (1) patient orientation was required to be supine. (2) Patient immobilization was required to consist of a thermoplastic facemask (AquaplastRT™, WFR/Aquaplast Corp.) and a standard plastic head holder. (3) Spinal fields in the patient's treatment plan for proton CSI were required to be oriented in the posterior-anterior direction through a standardized table top, and finally, (4) to facilitate determination of *total* dose from proton irradiation (section 2.4), the patient's treatment plan was required to pass a project-specific import/export test of the integrity of the electronic data exchange. More specifically, treatment plans were required to successfully export from the clinical radiation treatment planning (RTP) database, import into the research database, support edits while in the research RTP database, successfully export files containing dose and beam set-up information from the research database into the Monte Carlo Proton Radiotherapy Treatment Planning (MCPRTP) system (Newhauser et al. 2007b, Newhauser et al. 2008) and after low-statistic Monte Carlo (MC) simulations (for process testing only), successfully re-import dose files into the research database for display and analysis.

2.4 Total absorbed dose determined with therapeutic and stray radiation

In this study, we defined total absorbed dose to include absorbed dose from therapeutic and stray radiation. While this definition was used in both arms of the study, there were slight

differences in how stray dose was characterized for photon CSI vs. photon CSI. In this section, we briefly summarize those differences and present an overview of the corresponding methods for estimating therapeutic dose and stray dose for proton CSI and photon CSI, respectively.

Therapeutic dose was defined as the absorbed dose in the volume irradiated by the therapeutic beam (inclusive of leakage and scattered radiation). For both arms of the study, therapeutic dose was defined as dose within a 50% isodose region as estimated by a convolution superposition algorithm in a commercially available treatment planning system (TPS) (Eclipse version 8.6, Varian Medical Systems, Palo Alto, CA).

Stray dose was defined as the absorbed dose from radiation which was created as a consequence of delivering the therapeutic beam. Consequently, the methods for estimating stray dose differed for proton CSI and photon CSI, respectively. For proton CSI, stray dose was predominated by neutrons generated in the patient and emanating from the treatment unit (Agosteo et al. 1998; Yan et al. 2002; Taddei et al. 2008, 2009). Because the TPS did not include neutrons in its estimation of stray absorbed dose, we used alternate methods of dose estimation. Specifically, Monte Carlo (MC) simulations were used to determine absorbed dose from the stray neutrons (details provided in section 2.6.1). In contrast, in the photon arm of our study, the source of stray dose was leakage and scattered photon radiation emanating from the treatment unit. Thus, based on the results from Howell et al. (2010a, 2010b) and treatment plans from this work, we determined that the TPS could be used to accurately estimate absorbed dose from stray radiation in our study (details provided in section 2.6.2).

Table 1. Summary of the methods used to estimate therapeutic and stray dose
The methods used to estimate therapeutic and stray dose are listed. 'TPS' indicates that absorbed dose was estimated using the pencil beam algorithm in the TPS, and 'MC' indicates absorbed dose was estimated from Monte Carlo simulations. The parentheses contain section numbers where more detail can be found.

	Therapeutic Absorbed Dose	Stray Absorbed Dose
Proton CSI	TPS (2.5.1)	MC (2.6.1)
Photon CSI	TPS (2.5.2)	TPS (2.6.2)

2.5 Therapeutic dose

2.5.1 Treatment plans

In both the proton and photon arms of the study, therapeutic dose was determined for patient-specific treatment plans which were generated using the clinical TPS. For all patients, proton treatment plans pre-existed in the clinical database. These plans corresponded to the patient's original clinical treatment; however in most cases, due to import/export limitations of the treatment planning system (TPS), modifications to the pre-existing treatment plans were necessary. The modified plans were reviewed by a physician and approved once it was determined that they met the prevailing clinical SOC for the proton arm of the study. In contrast, clinical treatment plans did not exist for the photon arm of the study, as a result, photon treatment plans were developed specifically for this study (using the methods described in section 2.5.2). Similar to the proton treatment plans, the photon treatment plans were approved by the same physician based on the prevailing clinical SOC for photon CSI at our institution.

2.5.1.1 Treatment plans for proton CSI

Treatment plans for therapeutic dose from proton CSI were designed according to the prevailing SOC at our institution, which uses a three to five field approach (Figure 3).

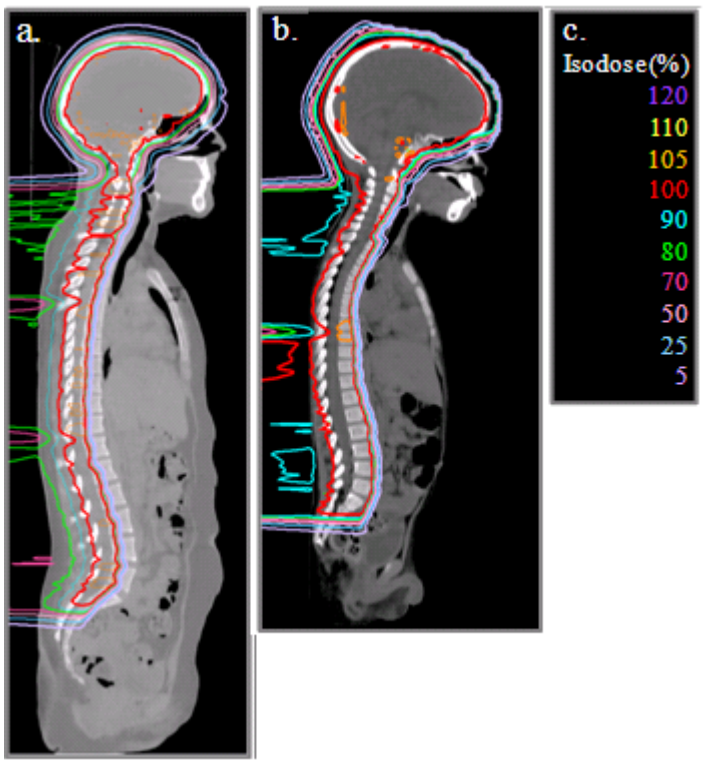

Figure 3. Illustration of the dose distribution for proton CSI.
Typically two fields are used to treat the cranium while one to three fields are used to treat the spinal axis. (a) The larger patient was 16 y-old female, patient 7, which required three spinal fields to cover spinal axis. (b) The smaller patient was 4 y-old male, patient 12, which required only one spinal field to cover spinal axis. (c) Isodose scale for both treatment plans. From Howell et al. (in preparation).

These fields were prescribed to deliver a therapeutic dose of 23.4 Gy (RBE) in 13 1.8 Gy (RBE) fractions. Units of Gy (RBE) were used in accordance with our clinic's SOC which presumes that proton beams have a higher relative biological effectiveness (RBE) than photon beams [22]. Thus, the absorbed dose was estimated, following the ICRU method, as

$$D_{RBE}[\text{Gy (RBE)}] = D_{absorbed}[\text{Gy}] \times \text{RBE} \qquad (2.4)$$

where the RBE value is taken as 1.1 and $D_{absorbed}$ is absorbed dose in Gy.

A detailed methodology for how the proton treatment fields were designed was provided by Howell et al. (in preparation). The methodology is briefly outlined here. (1) Two posterior oblique cranial fields (LPO and RPO) were matched to the most superior spinal field. (2) Spinal fields consisted of two to three posterior fields, depending on the patient's length, which were matched at their superior/inferior junctions, as appropriate, so that the combined spinal field extended from the posterior edge of the cranial fields through the thecal sac. (3)

The posterior-anterior extent of the spinal fields varied with patient age. More specifically, in patients younger than 15 years, the spinal fields were designed to cover the entire vertebral body (Figure 4a). In patients older than 15 years, spinal fields extended anterior to the vertebral canal by only 3 to 6 mm, such that only the posterior-most 3 to 6 mm of the vertebral body were included in the spinal field (Figure 4b). By reducing dose to the vertebral body, dose to bone marrow was reduced. This allowed for higher chemotherapy doses in older patients (Howell et al. in preparation).

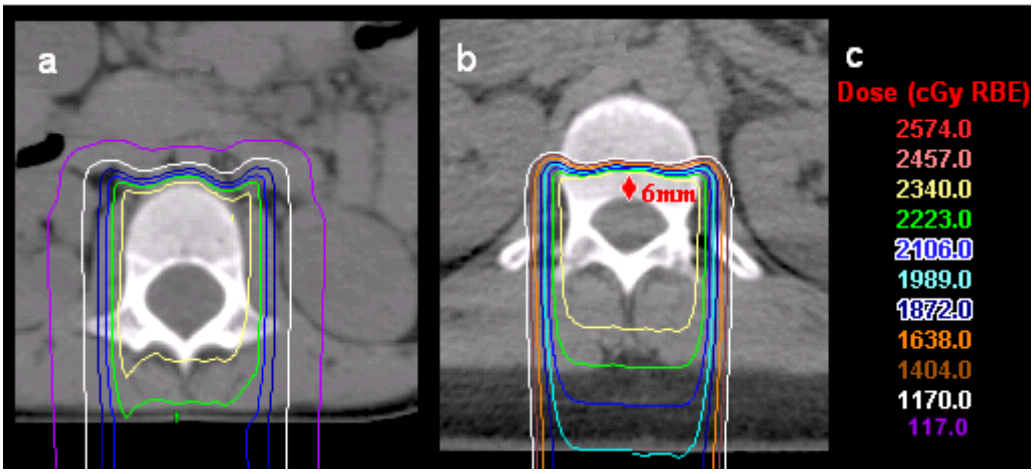

Figure 4. Illustration of the age-specific proton beam range for spinal fields.
Axial view of the isodose lines for a prescribed dose of 2340 cGy (RBE) with proton CSI. (a) Spinal fields for a 2 y-old patient covered the entire vertebral body. (b) Spinal fields for a 15 y-old patient covered the posterior 6 mm of the vertebral body. (c) Isodose scale for both images.

To reduce the MC simulation time, proton treatment plans utilized in this study did not include junction shifts. While this practice differs from the current SOC for proton CSI, i.e., a composite treatment plan which includes three sub-plans with different locations for field junctions, the plans in this study were re-optimized to achieve comparable homogeneity with plans containing junction shifts. These plans were then approved by a board-certified radiation pediatric oncologist. Additional documentation is available elsewhere (Giebeler 2009).

2.5.1.2 Treatment plans for photon CSI

Treatment plans for photon CSI utilized the same CT data, or patient density information, as the proton treatment plans and were developed according to our institution's

prevailing SOC. More precisely, the photon CSI plans utilized 6 MV fields with a field-in-field (FIF) technique to deliver a prescribed dose of 23.4 Gy using 13 fractions of 1.8 Gy apiece. The FIF technique was adapted from its original form (Yom et al 2007) for use with the TPS used in this study by Howell et al. (in preparation). In brief, a typical plan consisted of 3 junction shifted sub-plans. Junctions were shifted by 1 cm between sub-plans to reproduce clinically relevant hot or cold regions at junction locations. Each sub-plan used two laterally opposed brain fields and one to two spine fields. The spine fields were matched to the brain fields and each other at depth, i.e., in soft tissue just posterior to the spinal canal). The location of the match point between fields, or junction location, was varied between sub-plans to avoid hot spots. Additionally, to homogenize dose in high dose areas, the brain and spine fields in each sub-plan were optimized using three to five sub-fields with lower weights. The resulting dose distribution for a 6 MV photon plan for CSI is shown in Figure 5.

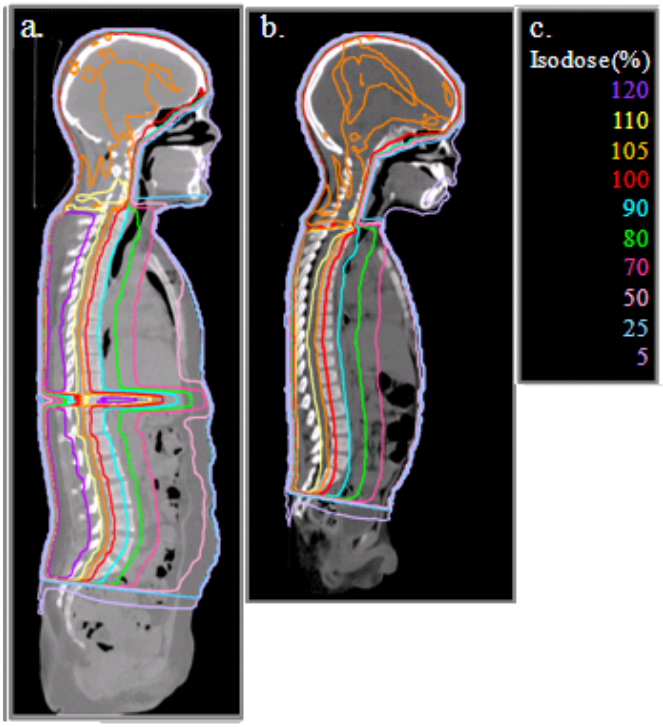

Figure 5. Illustration of the FIF treatment technique for 6 MV photon CSI
(a) The larger patient was 16 y-old female, patient 7, which required two spinal fields to cover the spinal axis. (b) The smaller patient was 4 y-old male, patient 12, which required only one spinal field to cover spinal axis. (c) Isodose scale for both composite treatment plans. From Howell et al. (in preparation).

2.5.2 Volume-weighted mean therapeutic dose

To apply D_{CE} and D_{MA} on a voxelized basis within the thyroid, we exported differential dose volume histograms (dDVH)s from the TPS into a spreadsheet for each patient in the study. The dDVHs reported the volume (cc) of the thyroid which received specific increments of dose (cGy RBE). The increments of dose were termed dose-bins; they spanned the minimum dose in the thyroid to the maximum dose in the thyroid. As a result, each patient's dDVH provided (1) a patient-specific number of dose-bins and (2) the volume (cc) of thyroid which received the corresponding dose to each dose-bin. To determine the mean absorbed therapeutic dose across the thyroid using D_{MA} we utilized equation 2.2 but in the following form[2] where ρ_i is assumed to be constant because the thyroid contains only glandular tissue (as opposed to bone or air pockets),

$$D_{MA} = \frac{\sum_{i-1}^{O} D_i m_i}{\sum_{i=1}^{O} m_i} = \frac{\sum_{i-1}^{O} D_i \rho_i V_i}{\sum_{i=1}^{O} \rho_i V_i} \approx \frac{\sum_{i}^{O} (D_i V_i)}{V_T}, \tag{2.5}$$

where O is the number of dose-bins in the thyroid, D_i is the therapeutic dose (Gy) at the ith dose-bin, m_i is the mass at the ith dose-bin, ρ_i is the mass density of tissue at the ith dose-bin, and V_i is the volume of the ith dose-bin (cc). V_T is total the volume of the thyroid (cc). Similarly for $D_{CE,}$ the cancer equivalent dose across the thyroid was determined using,

$$D_{CE} = \frac{\sum_{i-1}^{O} D_i m_i C_i}{\sum_{i=1}^{O} m_i} = \frac{\sum_{i-1}^{O} D_i \rho_i V_i C_i}{\sum_{i=1}^{O} \rho_i V_i} \approx \frac{\sum_{i}^{O} (D_i V_i C_i)}{V_T}, \tag{2.6}$$

where C_i is the cell sterilization factor (equation 2.3) for the ith dose-bin and N, $D_{i,}$ m_i, ρ_i, V_i, and V_T are as defined for equation 2.5.

On average, the width of the dose-bins in our dDVHs was 1 cGy. However in select cases, bins were re-set to 0.05 cGy before export to avoid errors associated with under-binning of the data. More precisely stated, if the mean absorbed dose in the thyroid was less than 1 cGy smaller bins were needed for accurate resolution of the volume specific dose (Figure 6),

[2] To transition to equation 2.5 from equation 2.2, we applied the density equation, $\rho = m/V$ to the numerator and denominator of equation 2.2 and recognized that the values for m were approximately constant in the thyroid.

and based on accuracy tests for dDVH output, the width of the smaller dose bins was set to 0.05 cGy.

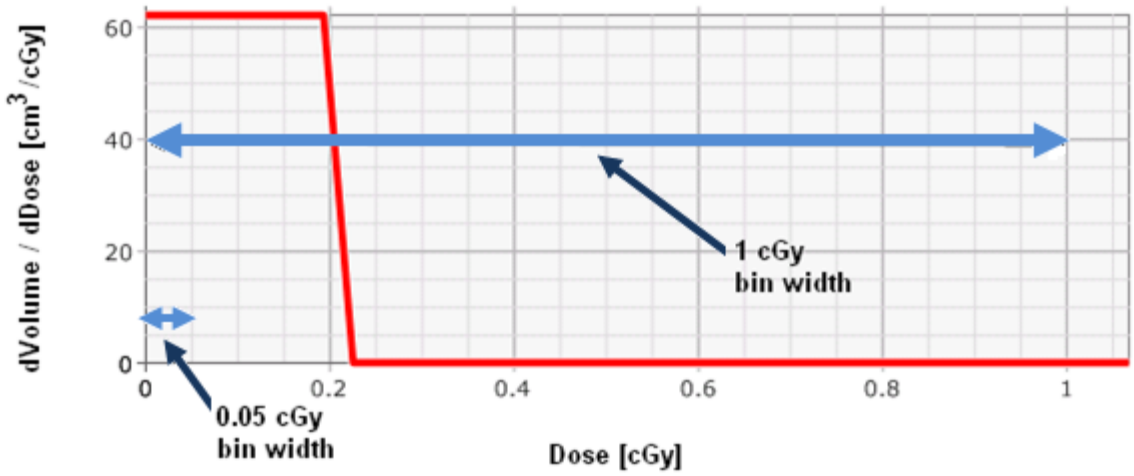

Figure 6. Illustration of dDVH with dose less than 1 cGy
The dDVH for a select case where re-binning of the data from 1 cGy dose bins to 0.05 cGy dose bins was needed. Without re-binning, the volume weighted mean absorbed dose was estimated at 1 cGy as opposed to 0.2 cGy as shown.

2.6 Stray Dose

2.6.1 Stray dose from proton CSI

Stray dose from proton CSI was calculated using methods from previous works on stray dose from neutrons. Detailed descriptions of how stray dose data is calculated and how it contributes to therapeutic dose in pediatric CSI were described previously (Taddei et al. 2010, Taddei et al. 2008, Taddei et al. 2009; Newhauser et al. 2009a). However, for convenience a review of that process follows.

In brief, estimates for absorbed dose from stray neutrons were calculated for each patient using the Monte Carlo Proton Radiotherapy Treatment Planning (MCPRTP) code (Newhauser et al. 2007, 2008) which runs the Monte Carlo code, MCNPX (Monte Carlo N Particle eXtended) (Waters et al. 2002, Pelowitz et al. 2005). Input into MCPRTP included patient specific parameters which were specified by the patients' proton CSI treatment plans (e.g., the CT data set, apertures, range compensators, and beam line parameters), and output from MCPRTP was returned in the form of individual dose files for each cranial and spinal treatment field. After the dose files were generated, they were manually imported into the clinical TPS for evaluation. The evaluation process included visual inspection of the imported dose relative to the patient anatomy and a re-scaling of the imported dose to clinically relevant

levels. Re-scaling was necessitated due to differences in how absorbed dose was normalized between MCPRTP and the clinical TPS. More specifically, MCPRTP uses monitor units (MU) to scale its output into units of Gy. However, because the treatment plans in this study were not clinically delivered, MU's were not assigned to the treatment fields. As a result field-specific MU values were taken from each patient's original clinical treatment and entered into MCPRTP. To address the resulting potential for differences in reported dose, we developed a method which utilized field specific scaling volumes (SV)'s (Figure 7).

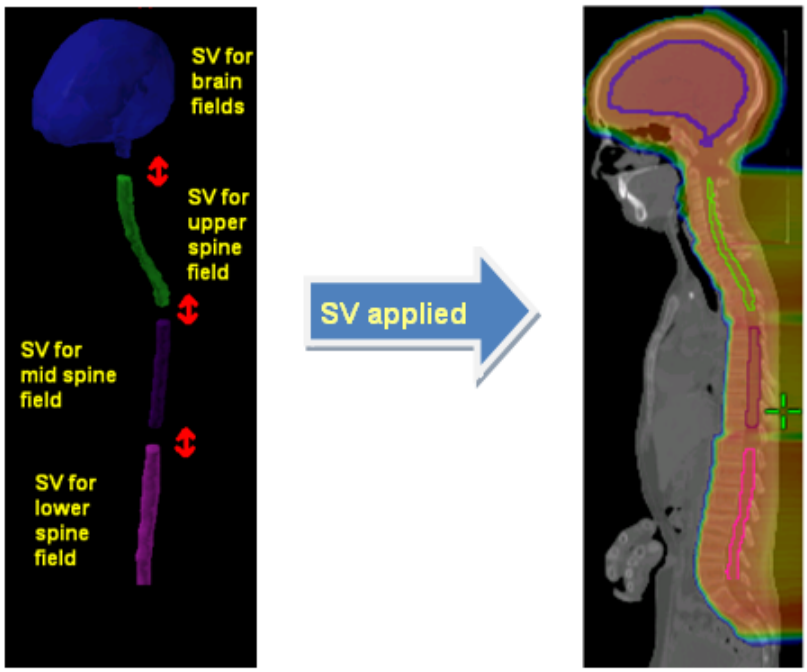

Figure 7. Illustration of scaling volumes (SV)'s.
SV's shown for the brain, upper spine, mid spine, and lower spine fields. The red arrows indicate gaps between the contours which correspond to field edges or regions of high dose heterogeneity.

The SV's were designed to correspond to regions of uniform dose. As a result, they did not include field junctions or extend to the full range of their respective treatment fields. Once the SV's were defined, estimated values of mean dose in the SV from MCPRTP and the clinical TPS were compared on a field by field basis using,

$$S_j = \frac{\overline{D_{\mathrm{TPS,SV}}}}{\overline{D_{\mathrm{MC,SV}}}}, \tag{2.7}$$

where S_j is the field specific scaling factor for dose generated by the MCPRTP vs. clinical TPS in the jth field, i.e., j = 1,2,3,…,p. $\overline{D_{\mathrm{TPS,SV}}}$ is the mean dose across SV as reported by the TPS and $\overline{D_{\mathrm{MC,SV}}}$ is the mean dose across the SV as reported by MCPRTP. Once S_j was determined for each field, we estimated the mean equivalent dose in our organ of interest, the thyroid. This was done using

$$H = \sum_{j=1}^{\mathrm{M}} H_j \,, \qquad (2.8)$$

where equivalent dose in the thyroid, H, was determined on a field-by-field basis for fields 1 through p using,

$$H_j = \frac{\overline{w_{\mathrm{R}}}}{10} \cdot S_j \cdot \frac{\overline{D_{\mathrm{TPS,T}}}}{1.1} \,, \qquad (2.9)$$

where $\overline{D_{\mathrm{TPS,T}}}$ is the mean absorbed dose calculated by MCPRTP in the thyroid in units of cGy, 10 converts from cGy to mGy because H is in units of mSv, 1.1 converts the absorbed dose units from cGy (RBE) to cGy, S_j is defined by Equation 2.7, and $\overline{w_{\mathrm{R}}}$ is the mean radiation-weighting factor for neutrons (details in section 2.7).

2.6.2 Stray dose from photon CSI

Recent work by Howell et al. (2010a, 2010b) found that the 6MV photon dose calculations for this TPS were accurate at doses as low as 5% of the prescribed dose. In this study, for photon CSI, the thyroid was included within the 5% isodose region generated by the TPS for all patients in the study (N = 18). Thus, in the case of photon CSI, the thyroid was considered to be an in-field organ, and estimations of *total* dose by the TPS were considered to be accurate (Howell et al. 2010a, Howell et al. 2010b).

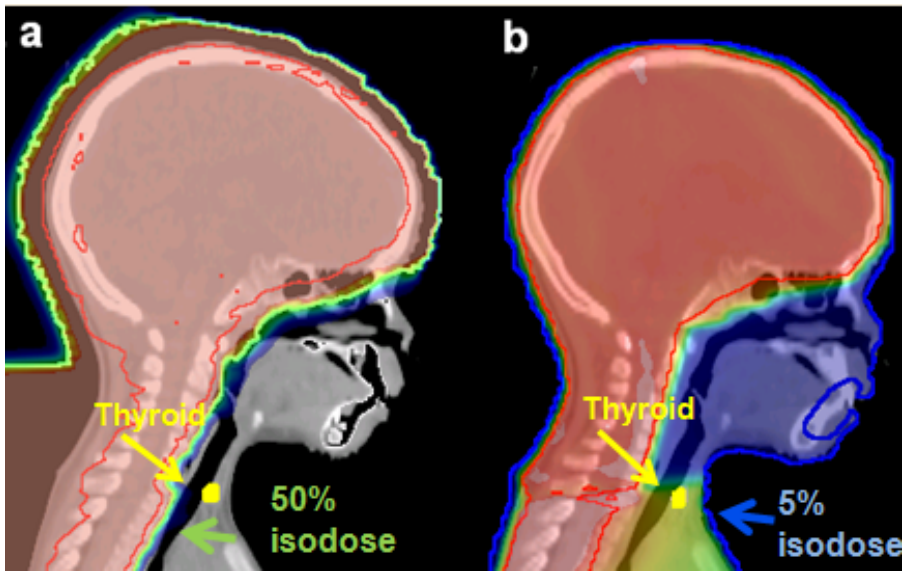

Figure 8. Location of the thyroid relative to isodose lines
Sagittal image of the isodose lines for a 4 year-old male (patient #12). The thyroid location relative to isodose lines (a) proton CSI and (b) photon CSI are shown.

2.7 Equivalent Dose

Because our estimations for total absorbed dose (section 2.4) included contributions from photon, proton, and neutron irradiation, we utilized the concept of equivalent dose as defined by the International Commission on Radiological Protection, ICRP. The ICRP (ICRP 1991) defined equivalent dose for a tissue, T, as

$$H_T = \overline{w_R} D_T \tag{2.10}$$

where D_T is the absorbed dose in a tissue, T, and $\overline{w_R}$ is the mean radiation weighting factor.

For photon irradiation, the value for $\overline{w_R}$ was taken as 1 (ICRP 2003). Thus for photon CSI, the equivalent dose in Sv is,

$$H_T = 1 * D_T \tag{2.11}$$

where D_T is the absorbed dose in Gy in the thyroid[3].

For proton irradiation, we consulted ICRP Publication 103 (2007) which suggests that $\overline{w_R}$ can be estimated using the mean quality factor, $\overline{Q}$, which is averaged over the effective interval of LET values for the field. Thus, equation for the equivalent dose in Sv for the thyroid from therapeutic protons becomes,

$$H = \overline{w_R} D \approx \overline{Q} D, \tag{2.12a}$$

[3] For simplicity, we drop the subscript 'T' for all subsequent equation because we considered only one tissue, thyroid, in this project. Thus, all equations following equation 2.11 implicitly describe thyroid tissue only.

where $\overline{Q}$ for therapeutic protons is 1.1 (section 2.5.1.1) and D is defined as in equation 2.7. Thus for proton CSI,

$$H_{\text{therapeutic}} = 1.1 * D, \tag{2.12b}$$

and the equation for the equivalent dose for the thyroid from stray neutrons becomes

$$H_{\text{stray}} = \overline{w}_{\text{R}} \times D, \tag{2.14}$$

where D is defined as in equation 2.7 and $\overline{w_{\text{R}}}$ is the energy summed weighting factor for neutrons. We used field specific values for $\overline{w_R}$ which were taken from a study by Newhauer et al. (2009).

In brief, Newhauser et al. (2009) performed organ-specific simulations of neutron spectral fluence in an anthropomorphic computational phantom which received proton CSI. Then they calculated values for $\overline{w_R}$ following recommendations in ICRP Publication 92 (2003) for cranial, superior spinal and inferior spinal treatment fields. We averaged their values for $\overline{w_R}$ across all organs, so in this work, $\overline{w_R}$ for the cranial fields, superior spinal field, and inferior spinal field were, 7.75, 8.09, and 8.17, respectively. For the middle spinal field we used the average of the superior and inferior spinal fields, 8.13.

Once $H_{\text{therapeautic}}$ and H_{stray} were defined, equations 2.14 and 2.12b, the total equivalent dose for proton irradiation was defined as,

$$H_{\text{total}} = H_{\text{therapeutic}} + H_{\text{stray}} \tag{2.15}$$

where H_{total} is in units of Sv.

2.8 Risk prediction

The estimated relative risk of SMN incidence in the thyroid, *ERR*, was estimated in <u>*two*</u> ways: using the traditional D_{MA} (Equation 2.1) and the proposed D_{CE} (Equation 2.2). The model for risk of radiogenic SMN incidence in the thyroid was taken from the report of the BEIR VII Committee (NRC 2006).

2.8.1 Risk prediction using mean absorbed dose model

When D_{MA} was used to estimate total absorbed dose, the estimated relative risk of SMN incidence, or ERR_{MA}, was calculated according to

$$ERR_{\mathrm{MA}} = \sum_{i}^{N} H_{\mathrm{MA}\,i} * \frac{V_i}{V_{\mathrm{T}}} * \left(\frac{ERR}{H}\right)_{\mathrm{T}}, \qquad (2.16)$$

where $H_{\mathrm{MA}i}$ is the total equivalent dose (Equation 15) in Sieverts in the dose-bin i when D_{MA} is applied. V_i is the volume of the dose-bin (cc); V_{T} is total the volume of the thyroid (cc); and $\left(\frac{ERR}{H}\right)_{\mathrm{T}}$ is the factor which describes the risk for SMNs in the thyroid per unit of equivalent dose (Sv) for a specific sex and age at the time of exposure. As such, $\left(\frac{ERR}{H}\right)_{\mathrm{T}}$ was defined for males and females, respectively, using,

$$\left(\frac{ERR}{H}\right)_{\mathrm{T}}^{\mathrm{M}} = \beta_{\mathrm{M}} e^{\gamma(e^*)} a^{\eta}, \qquad (2.17a)$$

$$\left(\frac{ERR}{H}\right)_{\mathrm{T}}^{\mathrm{F}} = \beta_{\mathrm{F}} e^{\gamma(e^*)} a^{\eta}, \qquad (2.17b)$$

and

$$e^*: \{(e - 30 | e < 30), (0 | e > 30)\} \qquad (2.18)$$

where e^* is defined as the age at time of exposure in years; β_{M} is the age-specific, and organ-specific instantaneous *ERR*/Sv value for males; β_{F} is the age-specific, and organ-specific instantaneous *ERR*/Sv value for females; is the where e is the age at time of exposure in years; a is attained age (years); γ accounts for the per-decade increase in age at exposure over the range of zero to 30 years, and η represents the exponent of attained age (NRC 2006). Thyroid specific values of these variables follow in Table 2 (below).

Table 2. BEIR VII committee's preferred values for modeling thyroid cancer incidence. Thyroid specific values of variables in the BEIR VII model are based on analyses by Ron et al. (1995) and a report of the NIH (2003). Confidence intervals, CI, are shown for the age-specific and organ-specific values, β_{M} and β_{F}. They are based on standard errors of non-sex-specific estimates across the studies pooled in the work by Ron et al. (NRC 2006).

Variable	Thyroid specific Value (NRC, 2006)	Confidence Interval
β_{M}	0.53	(0.14, 2.0)
β_{F}	1.05	(0.28,3.9)
γ	-0.83	-
η	0	-

Once ERR_{MA} was calculated, an additional term, the relative risk, RR_{MA}, of SMN incidence in the thyroid was calculated using

$$RR_{\mathrm{MA}} = ERR_{\mathrm{MA}} + 1. \tag{2.19}$$

Then, to determine the ratio of relative risk for incidence of SMN in the thyroid following CSI with proton vs. photon radiotherapies when D_{MA} was applied, we used,

$$RRR_{\mathrm{MA}} = \frac{RR_{\mathrm{MA,\,photons}}}{RR_{\mathrm{MA,\,protons}}} \tag{2.20}$$

where RRR_{MA} is the ratio of relative risk for lifetime incidence of thyroid cancer, $RR_{\mathrm{MA,\,photons}}$ and $RR_{\mathrm{MA,\,protons}}$ correspond to the application of equation 2.19 in the case of photon and proton CSI, respectively.

2.8.2 Risk prediction using cancer equivalent dose model

When D_{CE} was used to estimate total dose, the corresponding value for estimated relative risk of SMN incidence, or ERR_{CE}, was calculated using

$$ERR_{\mathrm{CE}} = \sum_{i}^{N} H_{\mathrm{CE}i} * \frac{V_i}{V_{\mathrm{T}}} * \left(\frac{ERR}{H}\right)_{\mathrm{T}}, \tag{2.21}$$

where $H_{\mathrm{CE}i}$ is the total equivalent dose in sieverts in the ith equivalent dose-bin when D_{CE} is applied. The definitions for V_i, V_{T}, and $\left(\frac{ERR}{H}\right)_{\mathrm{T}}$ are the same as they were in the previous section (equation 2.16).

Also similar to the previous section, we calculated the relative risk, RR_{CE}, of SMN incidence in the thyroid using,

$$RR_{\mathrm{CE}} = ERR_{\mathrm{CE}} + 1 \tag{2.22}$$

and the ratio of relative risk for lifetime incidence of thyroid cancer using,

$$RRR_{\mathrm{CE}} = \frac{RR_{\mathrm{CE,\,photons}}}{RR_{\mathrm{CE,\,protons}}} \tag{2.23}$$

where $RR_{\mathrm{CE,\,photons}}$ corresponds to the application of equation 2.22 in the case of photon and proton CSI.

2.9 Sample size calculations

We performed three separate sample size calculations using data from our estimations of predicted risk, RRR_{MA} and RRR_{CE}, for our cohort of 18 patients. Our first two calculations were used to estimate the number of patients needed for our virtual clinical trial when (1) RRR_{MA} was used and (2) when RRR_{CE} was used. Our third calculation estimated the number of patients needed for a clinical trial which directly compares use of D_{MA} vs. D_{CE} to predict risk. To do this we used several freely available software programs (Dattalo 2009). In particular, the two we most commonly used were a downloadable power and sample size calculator (PS version 3.0.43, Dupont and Plummer, 1990) and a web-based calculator (Brant 2011). We used these calculators to perform (1) a two sided, one sample t-test which compared RRR_{MA} against the trial's null hypothesis, (2) a two sided, one sample t-test which compared RRR_{CE} against the trial's null hypothesis, and (3) a two sided, two sample t-test which compared, RRR_{MA} vs. RRR_{CE}. (Details for each sample size calculation are provided in subsequent paragraphs.) After we performed the sample size calculations, we inter-compared the results and verified them against commercial software (S-Plus version 7.0, Insightful Corp., Seattle.). The required input parameters for PS and the web-based calculator are summarized below in Table 3.

Table 3. Comparisson of the input parameters used in sample size estimates for the t-test using two freely available statistics packages

Input parameters slightly differed between statistics packages. Both packages required entries for the experimental standard deviation of *RRR*, α (the probability of rejecting the null hypothesis when it is true), and power or (1-β) (the probability of not rejecting the null hypothesis when it is false). However, PS required the difference between population means, d, and the web-based calculator used $\overline{RRR}$. In the case of a one sample t-test using PS, d = 1-$\overline{RRR}$. In the case of a one sample test using the web based calculator, the test population was compared against $\overline{RRR}$= 1. In the case of the two sample t-test using PS, d = $\overline{RRR}_{CE}$.- $\overline{RRR}_{MA}$. In the case of the web based calculator, $\overline{RRR}_{CE}$ and $\overline{RRR}_{MA}$ were entered.

Input Parameters	Statistics Package	
	PS	Web-based Calculator
$\overline{RRR}$	No	Yes
Experimental Standard Deviation of *RRR*	Yes	Yes
d	Yes	No

$M = \frac{\text{\# Control Subjects}}{\text{\# Experiemental Subjects}}$	Yes	No
α	Yes	Yes
$(1-\beta)$	Yes	Yes

In our final estimations of sample size, we used PS. As a result, for the estimations which used the one sample t-test, we defined 'd' as the difference between our mean values of RRR_{MA} and RRR_{CE}, respectively, against the null hypothesis for our virtual clinical trial. Our null hypothesis (H_0) was that the ratio of relative risk for SMN in the thyroid between proton vs. photon radiotherapies is unity, i.e., $RRR = 1$. The alternate hypothesis (H_a) was that there is increased predicted risk for SMNs in the thyroid when photon radiotherapy is used for CSI; thus the ratio of relative risk for SMN between proton vs. photon radiotherapies is less than one, i.e., H_a: $RRR<1$.

For estimations of sample size which used the independent two sample t-test, we compared mean values of RRR_{MA} and RRR_{CE} against one another. While this did not address the question posed by the virtual clinical trial, it addressed the important intermediate and related question of whether or not we had enough subjects in our sample group to compare RRR_{MA} to RRR_{CE}. Values of our input parameters are identical to those summarized in Table 3 with the distinction that 'd' compares RRR_{MA} and RRR_{CE} to one another instead of unity, so a 'd' value of zero was the null for the two sample t-test.

2.10 Uncertainty in predicted risk

To address specific aim two, we estimated the uncertainty associated with the mean *RRR* in the thyroid from proton CSIvs. photon CSI for one patient when *RRR* was determined using D_{CE} and D_{MA}, respectively. More specifically, we first identified the predominant contributors to uncertainty in our estimates of *RRR*. Then, we determined the sensitivity of *RRR* to these factors for one representative patient, a 10 year-old female whose age, thyroid size, and thyroid location approximate the corresponding mean values of our study population.

Our estimates of relative uncertainty in *RRR* utilized standard error propagation techniques, meaning they utilized the following equation:

$$\left(\frac{\sigma_{RRR_X}}{RRR_X}\right)^2 = \left(\frac{\sigma_{ERR_{X,\text{protons}}}}{ERR_{X,\text{protons}}}\right)^2 + \left(\frac{\sigma_{ERR_{X,\text{photons}}}}{ERR_{X,\text{photons}}}\right)^2 - 2\frac{\sigma_{ERR_{X,\text{protons}}}\sigma_{ERR_{X,\text{photons}}}}{ERR_{X,\text{protons}}ERR_{X,\text{photons}}}, \qquad (2.24)$$

where RRR_X is the ratio of relative risk defined by Equation 2.20 when X = MA and by Equation 2.23 when X = CE. $ERR_{X,\text{protons}}$ is the excess relative risk following proton CSI as defined in Equation 2.16 when X = MA and in Equation 2.21 when X = CE, and similarly, $ERR_{X,\text{photons}}$ is the excess relative risk following photon CSI for the case when X = MA or CE . σ_{RRR_X} is the uncertainty RRR_X; $\sigma_{ERR_{X,\text{protons}}}$ is uncertainty in ERR_X from proton CSI; and $\sigma_{ERR_{X,\text{photons}}}$ is uncertainty in ERR_X from photon CSI. Finally, the last term in Equation 2.24 represents covariance. The expanded equation for uncertainty in ERR_X is determined using Equations 2.16 and 2.21. Thus,

$$\left(\frac{\sigma_{ERR_{X,Q}}}{ERR_{X,Q}}\right)^2 = \left(\frac{\sigma_C}{C}\right)^2 + \left(\frac{\sigma_{\overline{w_R}}}{\overline{w_R}}\right)^2 + \left(\frac{\sigma_{D_X}}{D_X}\right)^2, \qquad (2.25)$$

where $\frac{\sigma_{ERR_{X,Q}}}{ERR_{X,Q}}$ represents uncertainty in the age and gender specific risk model for SMNs in the thyroid from BEIR VII when the mean absorbed or cancer equivalent dose concepts are applied (X = MA or CE), respectively, for CSI modalities (Q = proton or photon), σ_C represents uncertainty in the cell sterilization factor, $\sigma_{\overline{w_R}}$ represents uncertainty in the radiation weighting factor for neutrons, and σ_{D_X} represents uncertainty in absorbed dose when the mean absorbed or cancer equivalent dose concepts are applied (X = MA or CE, respectively). The methods for determining specific values of σ_C, $\sigma_{\overline{w_R}}$, and σ_{D_X} are outlined in the following sections.

2.10.1 Uncertainty in absorbed dose, σ_{D_X}

In this section we present our methodology for estimating uncertainty in absorbed dose in the thyroid due to variations in patient-set up for proton vs. photon CSI. Recall that our estimates for risk of SMNs in the thyroid are determined based on voxelized regions of dose, consequently, small inter-fractional shifts in the location of the thyroid relative to the radiation field can potentially impact estimates of D_{MA} and D_{CE}. The location of the thyroid for our representative patient is shown in Figure 9.

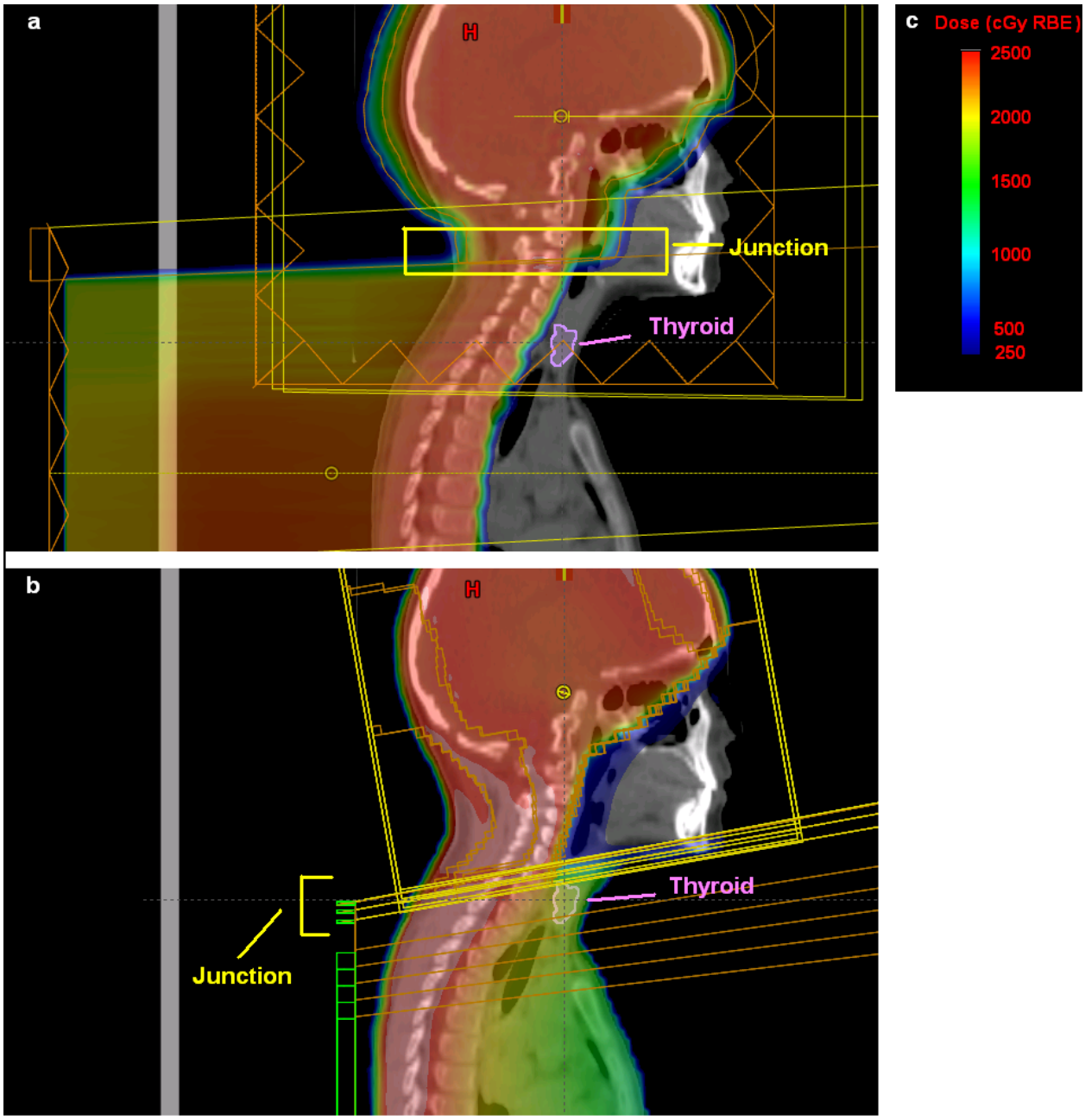

Figure 9. Thyroid location with respect to treatment field in representative patient
Sagittal images were taken of the left lobe of the thyroid for a 10 year-old female patient (patient #14) with a prescribed dose of 2340 cGy (RBE) from proton (a) and photon (b) CSI. The thyroid, outlined in pink, is located well away from the junction of the cranial and upper spine field junction (yellow box). (a) Field edges and dose (cGy RBE, RBE factor of 1.1)) are shown for proton CSI. (b) Field edges and dose (cGy, RBE factor of 1.0) are shown for photon CSI. (c) Dose scale for (a) and (b).

To estimate uncertainty in absorbed dose, we estimated the effect of inter-fraction shifts, between the therapeutic treatment field and the thyroid, on therapeutic dose in the thyroid. Briefly, we calculated the differences in therapeutic dose in the thyroid after representative shifts (in the anterior, posterior, superior, inferior, and lateral directions). Then we used the largest interval in dose to represent uncertainty in absorbed dose. The magnitude of the inter-fraction shifts was selected using clinically determined values for uncertainty in patient set-up (Table 4). These values represent the prevailing SOC for proton and photon CSI, respectively, at our institution, at the time of this study.

Table 4. Values for set-up uncertainty per prevailing standards of care for proton and photon CSI.
Selected set-up uncertainties used in proton and photon CSI. In the bottom row, are the values we used for our uncertainty analysis. They correspond to the uncertainty associated with the spinal fields when the practice of daily imaging is followed (per the SOC for pediatric CSI).

Values for set-up uncertainty	Proton CSI (cm)	Photon CSI (cm)
Face mask and head holder	0.3	0.3
Alpha cradle or Med-Tec device	0.5	0.5
Index to treatment couch	0.7	0.7
Daily imaging + index to treatment couch	0.5	0.5
Values selected for uncertainty analysis (spinal field)	0.5	0.5

Table 4 includes uncertainties for the brain and spine fields and varies according to immobilization type. In our sample, immobilization for brain fields included use of a thermoplastic face mask and head holder. For spine fields immobilization included indexing to the treatment couch (on a Styrofoam pad). Additionally, per the current SOC for pediatric patients, daily imaging was included in the patient set-up. We selected an uncertainty value of 0.5 cm. This value is conservative because it is based on the larger uncertainty associated with the spinal field rather than the smaller uncertainty of the cranial fields.

We used the value of 0.5 cm for set-up uncertainty in proton and photon CSI to simulate possible variations in the position of the thyroid relative to the planned treatment

fields. This was done for one representative proton treatment plan and one representative photon treatment plan for the same patient. Simulation of thyroid motion relative to its respective treatment fields was achieved by shifting its contour within the treatment planning system by the maximum set-up uncertainty value in the superior, inferior, left, right, anterior, and posterior directions. The mean absorbed dose was recorded for each contour location, and the maximum difference in absorbed dose was used to estimate the associated uncertainty in absorbed dose associated with variations in thyroid location due to patient set-up for proton and photon CSI.

2.10.2 Uncertainty in the cell sterilization factor, σ_C

In this section we present our methodology for determining the sensitivity of our risk predictions to uncertainty in the cell sterilization factor, *C* (Equation 2.3), for proton vs. photon CSI. Uncertainty in *C* is due to the uncertainty of a thyroid specific factor alpha, α, which was defined following the methods of Schneider et al. (2005). Estimates for α were reported using data from Hodgkin's patients (Schneider et al. 2005). However, there has been some debate over the applicability of data from Hodgkin's patients to all patient populations (Kry and Salehpour 2006). In particular, there is concern that Hodgkin's patients may be more susceptible to thyroid cancer than the general population due to genetic or environmental risk factors (Ronckers et al. 2006). Yet, there are epidemiological results which suggest that cell-sterilization occurs in the thyroid similarly for non-Hodgkin's patients as well as Hodgkin's patients (Sigurdson et al. 2005). As a result, there is a numerically unknown, yet inherent, uncertainty associated with α. To address this in our work, we estimated the sensitivity of our estimates of risk to an interval of α values which were defined using the following (Schneider et al. 2005):

$$\alpha(\mathrm{Sv}^{-1}) = \{0.033 \mid \mathrm{CI}: 0.020 - 0.047\}. \qquad (2.26)$$

where the minimum value is 0.020 Gy^{-1} or α_{MIN} , the nominal value of α is 0.033 Gy^{-1} or α_*, and the maximum value was 0.047Gy^{-1} or α_{MAX}. Then, we applied the interval of α values to the following to determine the corresponding cell sterilization factor, C for proton and photon CSI using (adapted from Equation 2.3),

$$C_x = e^{-\alpha_x H_{\mathrm{total}}}, \qquad (2.27)$$

where the subscript '*x*' indicates MIN, nominal, or MAX and H_{total} is defined from Equations

2.12b and 2.15. Next, we used the interval between C_{MIN} and C_{MAX} to conservatively estimate the sensitivity of our risk estimates to uncertainty in α using,

$$\sigma_{\mathrm{C}} \approx C_{\mathrm{MAX}} - \mathrm{C}_{\mathrm{MIN}} \,. \tag{2.29}$$

where σ_{C} represents uncertainty in the cell sterilization factor. After that, we divided by the nominal correction factor, C.

2.10.3 Uncertainty in the neutron weighting factor, $\sigma_{\overline{w_{\mathrm{R}}}}$

This section describes the method used to estimate the uncertainty in the mean radiation weighting factor for neutrons, $\sigma_{\overline{w_{\mathrm{R}}}}$. In this project we used values of $\overline{w_{\mathrm{R}}}$ that were published by Newhauser et al. (2009), but to maintain the validity of our study results we also determined the sensitivity or our results to uncertainty in $\overline{w_{\mathrm{R}}}$. To do this, we used the results of previous works by Yan et al. (2002) and Newhauser et al. (2009) to determine the most representative values for scaling $\overline{w_{\mathrm{R}}}$. More precisely, we scaled $\overline{w_{\mathrm{R}}}$ by 0.5 and 2 to represent a bracket of uncertainty in $\overline{w_{\mathrm{R}}}$. Once we scaled our values of $\overline{w_{\mathrm{R}}}$, we applied these values to our estimates for stray neutron dose. This created an interval of values of H_{total} (Equation 2.15) which resulted in a corresponding interval of values of RRR_X where X = MA or CE. We then used the resulting intervals for RRR_X to define an uncertainty factor for $\overline{w_{\mathrm{R}}}$ in proton CSI as follows:

$$\sigma_{\overline{w_{\mathrm{R}}},X} = |RRR_X(\overline{w_R} = 0.5) - RRR_X(\overline{w_R} = 2)|, \tag{2.30}$$

where $\sigma_{\overline{w_{\mathrm{R}}},X}$ represents uncertainty in the neutron weighting factor when RRR is determined using D_{MA} or D_{CE}, i.e., X = MA or CE. Finally, because we wanted a general estimate for uncertainty in the field-averaged $\overline{w_{\mathrm{R}}}$, as opposed to a field-specific estimate, we divided by the mean value for $\overline{w_{\mathrm{R}}}$ across all treatment fields (last entry in Table 5 below) to satisfy the term for the neutron weighting factor, $\frac{\sigma_{\overline{w_{\mathrm{R}}}}}{\overline{w_{\mathrm{R}}}}$, from Equation 2.25. This was done for one patient receiving proton CSI; the same patient identified in section 2.9.1 above. As a result, there was no mid spine field. A similar test was *not* performed for photon CSI because neutrons are not produced for 6 MV photon beams.

Table 5. Radiation weighting factors for neutrons for uncertainty test patient
Neutron weighting factors which were used in the case of our representative patient are listed. The mid spine field was not required for full coverage of the spinal axis. Also listed is the mean of the field-specific values of $\overline{w_R}$.

Field	$\overline{w_R}$ value
Cranial (LPO)	7.75
Cranial (RPO)	7.75
Upper Spine	8.09
Mid Spine	-
Lower Spine	8.13
Mean	7.93

Chapter Three

3 Results

This chapter presents the results of this study. Ultimately, it details our findings for the effect of cell sterilization on sample size in a study of the predicted risk of a SC in the thyroid following pediatric CSI with proton vs. photon therapy. To do this, we first estimated the effects of D_{MA} vs. D_{CE} on therapeutic dose (3.1), stray dose (3.2), the combination of therapeutic and stray dose in terms of total equivalent dose (3.3), and the predicted risk of SC incidence in the thyroid (3.4). Then, using our values of predicted risk, we determined the corresponding sample size needed for statistical significance (3.5). In addition, this chapter provides the results of our uncertainty analysis of predicted risk (3.6).

3.1 Estimation of therapeutic dose

3.1.1 Therapeutic dose

We used dDVH information and a spreadsheet to estimate values of D_{MA} (equation 2.5) and D_{CE} (equation 2.6) using the therapeutic dose (Gy) in each patient's thyroid (Table 6). Stray dose (Sv) (section 3.2) was added in section 3.3 to form equivalent dose (Sv).

Table 6. D_{MA} and D_{CE} from our spreadsheet for proton and photon CSI
Estimates of D_{MA} and D_{CE} from our spreadsheet are shown. The estimates were made using therapeutic dose in the thyroid.

				PROTON		PHOTON	
Patient Index	Sex	Age (e)	Volume of thyroid (cc)	D_{MA} (cGy RBE)	D_{CE} (cGy RBE)	D_{MA} (cGy)	D_{CE} (cGy)
1	F	13	4.3	45.7	35.6	1598	913
2	M	16	14.3	0.2	0.2	1772	983
3	M	13	12.6	0.1	0.1	1896	1003
4	M	15	10	0.1	0.1	1929	1020
5	M	14	7.8	1.1	1.1	1616	911
6	M	12	11.7	1.1	1.1	1887	1008
7	F	16	7.7	1.4	1.3	1856	1001
8	F	12	1.7	1.0	0.9	1674	944
9	M	9	5.4	6.1	5.9	1436	851
10	M	8	2.3	6.4	6.1	2074	1059
11	M	6	4.1	11.7	10.5	1273	776
12	M	4	3.9	3.9	3.8	1359	820
13	M	3	1.5	3.0	2.8	1027	699
14	F	10	3.8	1.4	1.4	1850	1006
15	F	8	4.5	0.2	0.2	1172	732
16	F	6	3.3	2.2	2.1	1477	880
17	F	4	3.2	0.6	0.6	1100	695
18	F	2	1.9	4.1	3.9	812	583

The distribution of values for D_{MA} and D_{CE} in Table 6 is shown in Figure 10.

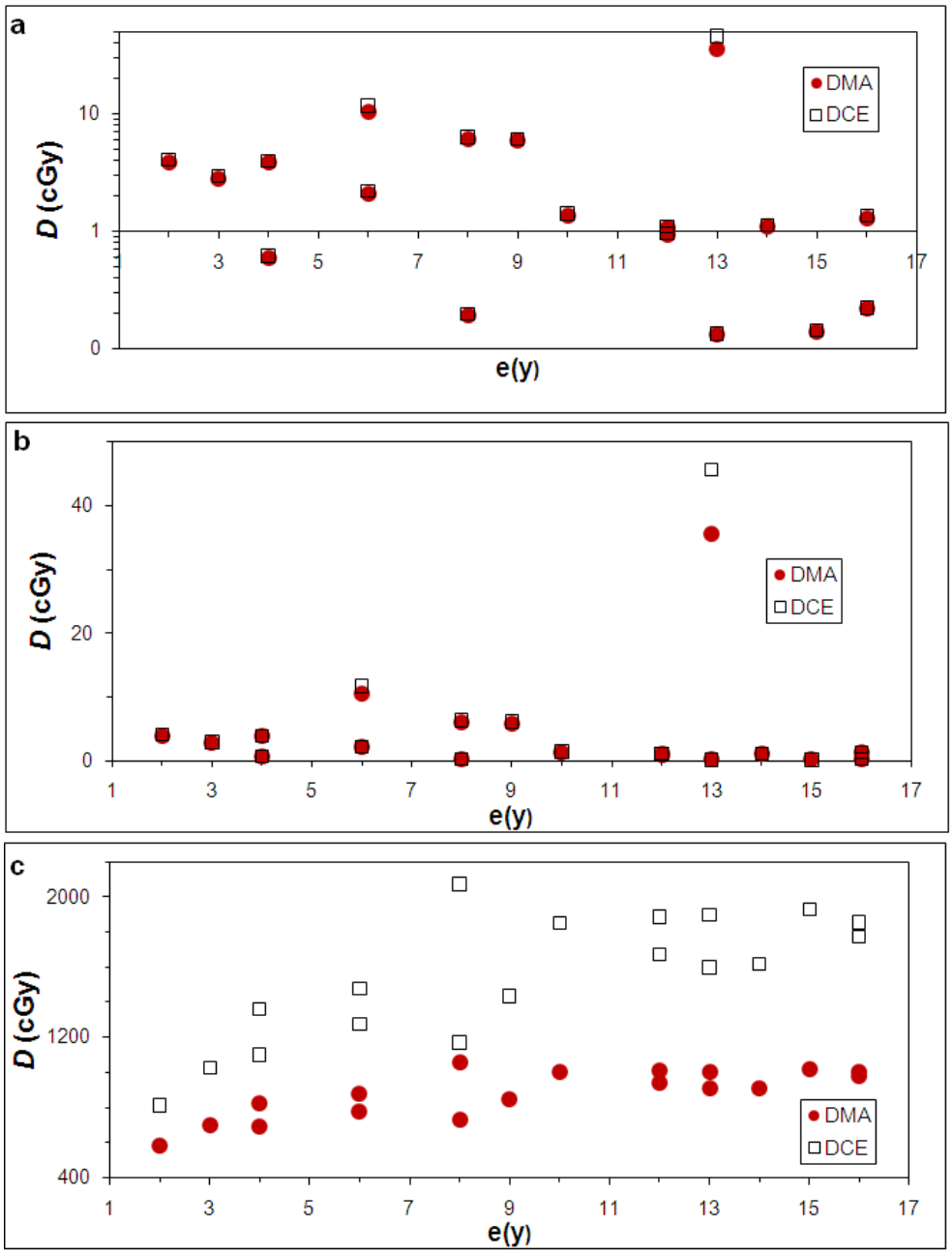

Figure 10. Distribution of D_{MA} for proton and proton CSI
(a) D_{MA} and D_{CE} in the thyroid for therapeutic dose from proton CSI (b) D_{MA} and D_{CE} in the thyroid for therapeutic dose from proton CSI on a semi logarithmic plot (c) D_{MA} and D_{CE} in the thyroid for photon CSI.

To cross check our spreadsheet (Table 6), we compared the distribution of D_{MA} values in the spreadsheet against the corresponding D_{MA} values calculated by the TPS. The results of the comparison are presented in Tables 7 and 8 below.

Table 7. D_{MA} from TPS vs. D_{MA} from our spreadsheet for photon CSI.
Descriptive statistics for the distribution of D_{MA} in the thyroid from photon CSI for our sample of 18 patients are provided. Data which correspond to TPS estimations of D_{MA} are found in column A. Data which correspond to spreadsheet estimations of D_{MA} are found in column B. Column C lists the absolute value of the difference between data in columns A and B. Column D lists the percent difference between values from columns A and B. Negative values in columns C and D indicate that spreadsheet estimates of D_{MA} are larger than those from the TPS. Conversely, positive values in columns C and D indicate that TPS estimates of D_{MA} are larger than those from the spreadsheet.

	A	B	C	D
	From value obtained using TPS	From value obtained using Spreadsheet	Difference A-B	Percent Difference (A-B)* 100/A
	(cGy RBE)	(cGy RBE)	(cGy RBE)	(%)
$D_{MA, MAX}$	2035.7	2073.6	-37.9	-1.9
$D_{MA, MIN}$	829.4	812.3	17.1	2.1
$D_{MA, MAX}$ - $D_{MA, MIN}$	1206.3	1261.3	-55.0	-4.6
Mean D_{MA}	1545.6	1544.8	0.8	0.0
Standard Deviation of the Experimental Mean	84.6	85.0	-0.4	-0.4
Experimental Standard Deviation	359.0	359.8	-0.8	-0.2

From Table 7 for photon CSI, we saw that for our sample of 18 patients, the percent difference between mean values of D_{MA} from the TPS vs. spreadsheet was less than 1%. Similarly, there was less than 1% percent difference in the experimental standard deviation and the standard deviation of the experimental mean of D_{MA} when the TPS vs. spreadsheet was used. Additionally, when the maximum estimate of D_{MA}, $D_{MA, MAX}$, from the TPS was compared to $D_{MA, MAX}$ from the spreadsheet, the spreadsheet estimate was 2% higher. Conversely, when the minimum estimate of D_{MA}, $D_{MA, MIN}$, from the TPS was compared to $D_{MA, MIN}$ from the spreadsheet, the spreadsheet estimate was 2% lower. The interval between $D_{MA, MAX}$ and $D_{MA, MIN}$ was almost 5% higher for the spreadsheet vs. TPS estimates. Thus, while there were differences in rounding and truncation between the spreadsheet and TPS, estimates for D_{MA} they were considered acceptable.

Table 8. D_{MA} from TPS vs. D_{MA} from our spreadsheet for proton CSI.
Descriptive statistics for the distribution of D_{MA} in the thyroid from proton CSI for our sample of 18 patients are provided. Data which correspond to TPS estimations of D_{MA} are found in column A. Data which correspond to spreadsheet estimations of D_{MA} are found in column B. Column C lists the absolute value of the difference between data in columns A and B. Column D lists the percent difference between values from columns A and B. Negative values in columns C and D indicate that spreadsheet estimations of D_{MA} are larger than those from the TPS.

	A	B	C	D
	From value obtained using TPS	From value obtained using Spreadsheet	Difference A-B	Percent Difference (A-B)* 100/A
	(cGy RBE)	(cGy RBE)	(cGy RBE)	(%)
$D_{MA, MAX}$	45.3	45.7	-0.4	-0.9
$D_{MA, MIN}$	0.1	0.1	0.0	-34.1
$D_{MA, MAX}$ - $D_{MA, MIN}$	45.2	45.6	-0.4	-0.8
Mean D_{MA}	4.9	5.0	-0.1	-1.6
Standard Deviation of the Experimental Mean	2.5	2.5	0.0	-0.9
Experimental Standard Deviation	10.5	10.6	-0.1	-0.9

In Table 8, we found that the population averaged D_{MA} from proton CSI was approximately 400 times smaller than those for photon CSI (Table 7). Also, in Table 8 we found that for proton CSI, the percent difference between mean values of D_{MA} from the TPS vs. spreadsheet was less than 2% with the spreadsheet values being larger. Similar to photon CSI, there was less than 1% percent difference in the experimental standard deviation and the standard deviation of the experimental mean of D_{MA} when the TPS vs. spreadsheet was used. Additionally, when the maximum estimate of D_{MA}, $D_{MA, MAX}$, from the TPS was compared to $D_{MA, MAX}$ from the spreadsheet, the spreadsheet estimate was almost 1% higher. However, when $D_{MA, MIN}$ from the TPS was compared to $D_{MA, MIN}$ from the spreadsheet, the spreadsheet estimate was 34% lower. To understand the reason for this difference, we reviewed the original TPS and spreadsheet distributions of D_{MA} for proton CSI (distributions values not shown here). From the distributions we found that while there the mean difference in estimates for D_{MA} from the TPS vs. spreadsheet is small, -0.08 cGy (RBE), there are specific instances where differences as low as -0.04 cGy (RBE) can result in a 44% percent difference.

To determine the potential impact of these small differences on our estimates of thyroid dose for proton patients, we classified patients with a percent difference greater than 10% as outliers. Three of 18 patients were qualified as outliers. We compared their D_{MA} values as well as any physical attributes such as the size and anatomical location of the thyroid, gender, or age that could potentially explain the differnces in D_{MA}. Data for these patients is summarized in Table 9 and Figures 11 and 12 below.

Table 9. Outliers from the comparison of D_{MA} estimates from the TPS vs. spreadsheet for proton CSI

Values of therapeutic mean absorbed dose from proton CSI are compared for three patients. Column A lists values determined by the treatment planning system (TPS), and column B lists values that were determined by taking the volume weighted average of dose values in our spreadsheet. Column C lists the difference between columns A and B, and column D reports the percent difference between columns A and B. Additional physical characteristics are listed in columns E-I. For comparison mean values for the entire sample (N = 18) are listed in the last row.

	A	B	C	D	E	F	G	H	I
Patient Index	From value obtained using TPS (cGy RBE)	From value obtained using spreadsheet (cGy RBE)	Difference (A-B) (cGy RBE)	Percent Difference (A-B)* 100/A (%)	Distance to center of thyroid (cm)	Distance to anterior edge of thyroid (cm)	Volume of contour (cc)	Sex	age (y)
3	0.1	0.13	0.03	-34.10	8.93	10.15	12.6	male	13
4	0.1	0.14	0.04	-44.20	12.96	14.60	10.0	male	15
2	0.2	0.22	0.02	-11.86	14.82	16.64	14.3	male	16
Mean values for sample (N = 18)	4.9	5.02	0.08	-6.23	8.8	9.9	5.8	-	9.5

From Table 9, we found two main trends. (1) The three outliers have comparatively small D_{MA} values when compared to the mean D_{MA} values of 5.0 cGy (RBE) for the entire sample (Table 8). (2) The three outliers have comparatively large thyroid volumes (≥ 10 cc) when compared to the mean thyroid volume of 5.8 cc for the entire sample. We stratified our sample by the size and relative anatomical location of the thyroid and patient sex (Figures 11 and 12 below).

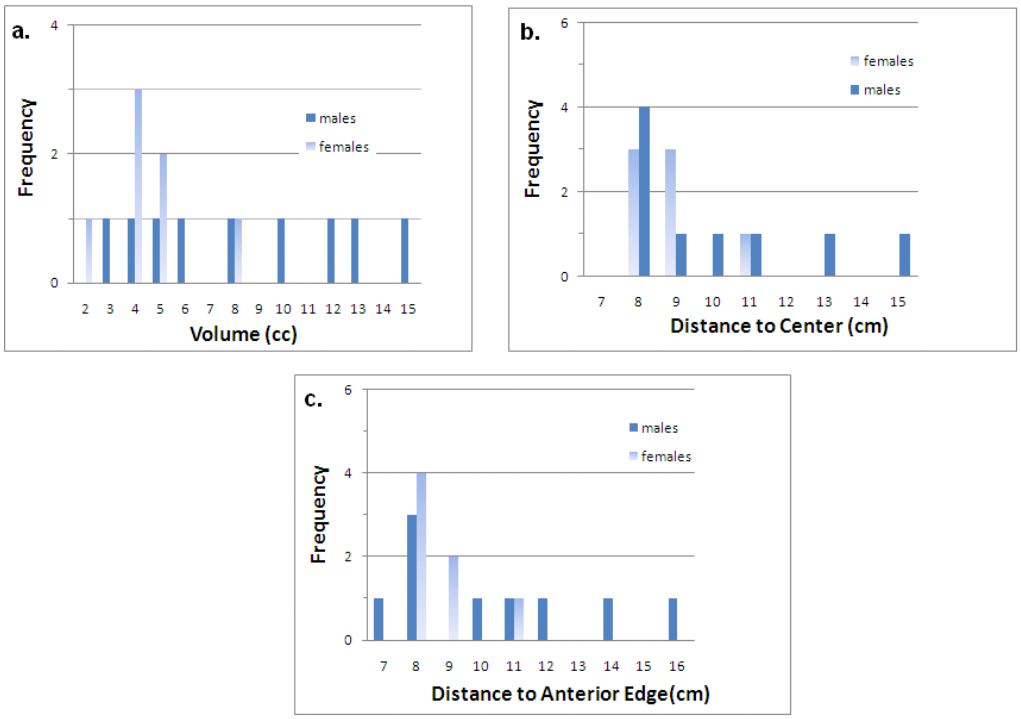

Figure 11. Sex-specific histograms: size and relative anatomical location of the thyroid
(a) Frequency distribution of thyroid volume. (b) Frequency distribution of distance along the treatment axis from the surface of the patient's skin to the center of the patient's thyroid. (c) Frequency distribution of physical distance along the treatment axis from the surface of the patient's skin to the anterior-most point of the patient's thyroid.

Figure 11 reveals that there are potential sex-related trends in our sample. For example, in Figure 11a, the inter-patient variation in thyroid volume was smaller for females than males. In the male patients, thyroid volume was fairly evenly distributed from 3 to 15 cc. Figures 11b and 11c which plot the inter-patient variation in the location of the thyroid show that the variation was also smaller for females than males. To further investigate these trends, we plotted our values of thyroid volume and depth within the patient with respect to the patient age at exposure (Figure 12).

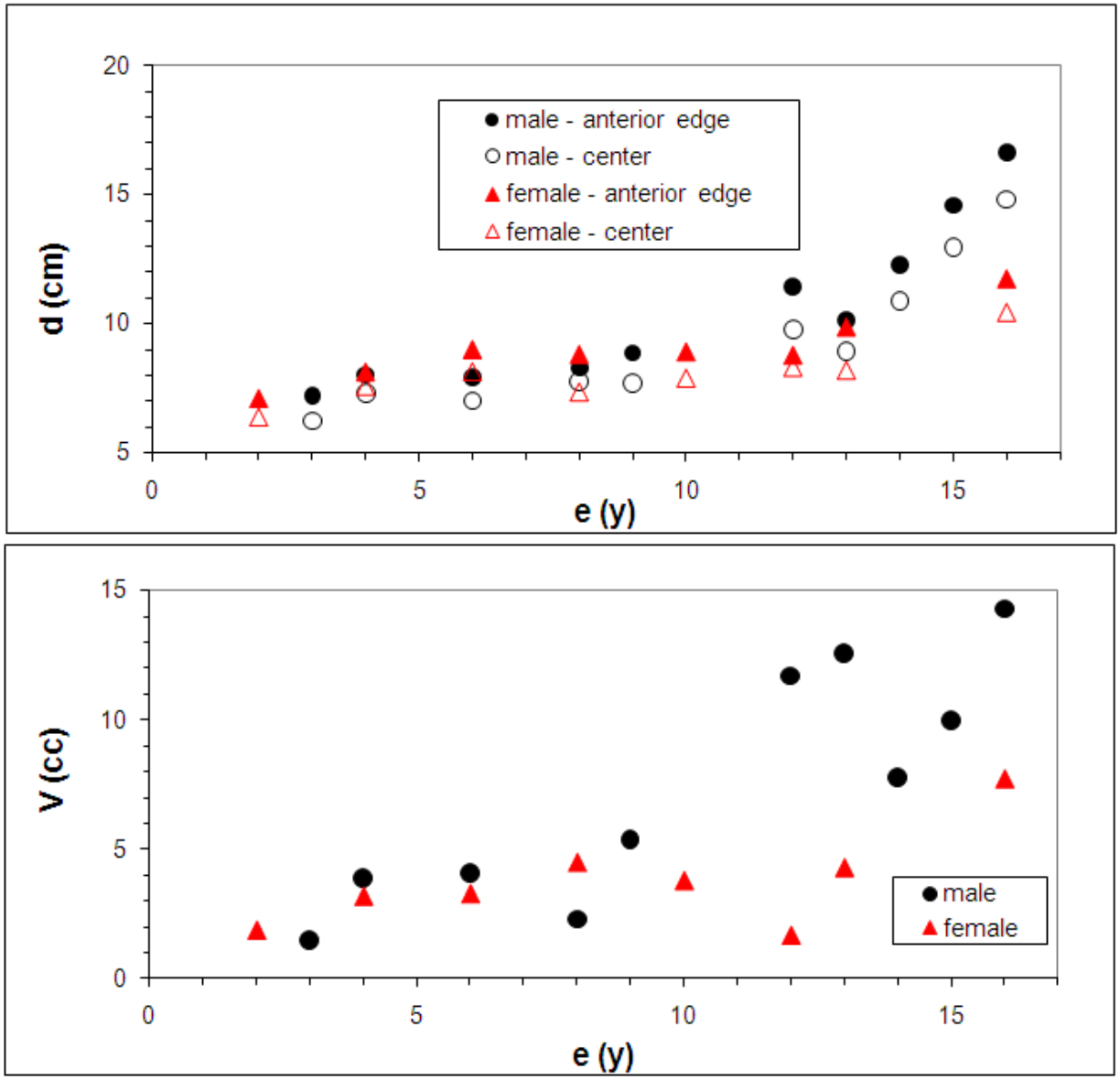

Figure 12. Sex-specific thyroid volume and physical depth from skin vs. patient age at exposure
(a) Distance along the treatment axis from the surface of the patient's skin to the center and anterior-most surface of the thyroid vs. patient age at exposure (e). Patient information is stratified according to sex. (b) Thyroid volume (V) vs. patient age at exposure (e).

Figure 12 indicates that there may be correlations with respect to sex and thyroid location. Thyroid depth and volume increased with respect to age at exposure for both sexes, and this increase was more pronounced for males than females.

Using the data in Figures 11 and 12, we performed one final analysis. We created a scatter plot to test for a statistically significant correlation between thyroid volume (cc) and the difference between D_{MA} values from the TPS and our spreadsheet (Figure 13).

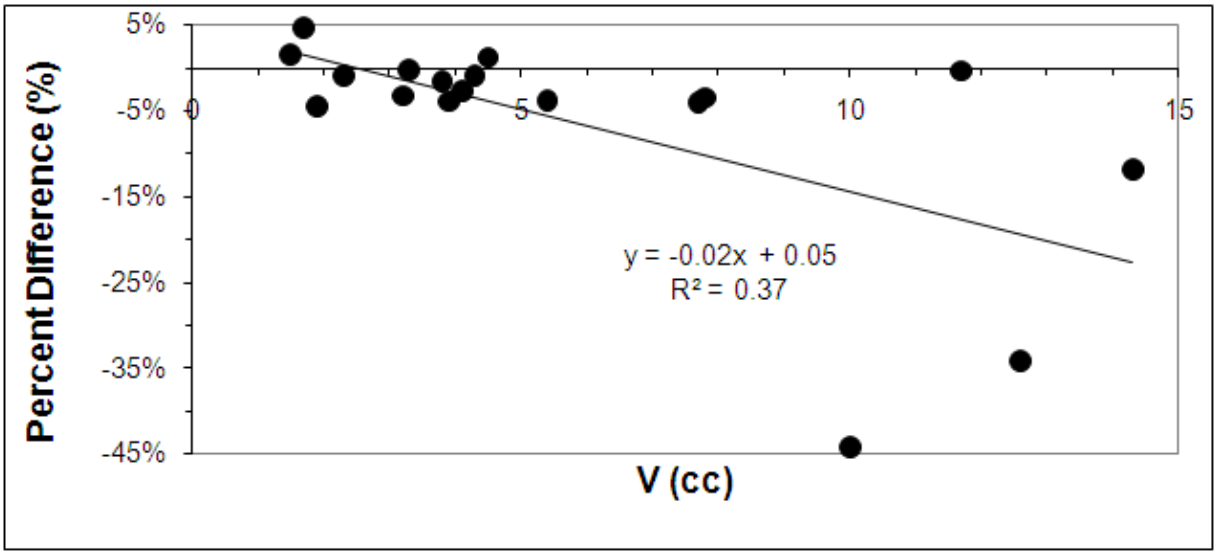

Figure 13. Percent difference in D_{MA} between TPS and spread sheet estimates
Thyroid volume (V) vs. percent difference between the TPS and spreadsheet based estimates of mean absorbed dose. A trend line and coefficient of determination (R^2) are also shown, and from R^2, the correlation coefficient (R) is 0.61.

From Figure 13, we see that a strong positive correlation between differences in D_{MA} and thyroid volume was not observed. While visual inspection may suggest that a thyroid volume that is larger than 10 cc may correspond to a larger percent difference between the TPS and spreadsheet estimates for mean absorbed dose, there is too much variation within our data to make a conclusive statement. Essentially, a larger sample size is needed to conclusively explain the difference between the TPS vs. spreadsheet estimates for D_{MA} in terms of partial volume effects, age at exposure, and sex. However, because the number of outliers, 3, was relatively small when compared to our overall sample size of 18 patients, we elected to continue use of the spreadsheet to estimate values of D_{MA} and, in turn, D_{CE}.

3.2 Estimation of stray dose

The results of our MC simulations for stray dose to the thyroid are summarized in Figure 14 and Table 10 below.

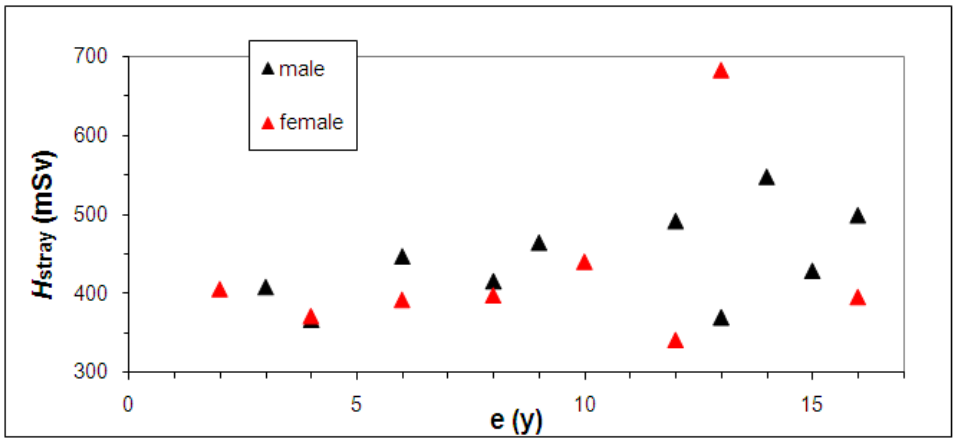

Figure 14. Stray dose from proton CSI
Calculated stray dose to the thyroid vs. age at exposure (e).

From Figure 14, we see that for 17 of the 18 patients, the majority of H_{stray} are between 400 mSv and 500 mSv. One patient, patient index 1, had H_{stray} closer to 700 mSv. Patient #1 was a 13 year-old female patient whose therapeutic dose from proton CSI was 17.6 times that of the mean therapeutic dose for our sample (Tables 6 and 7, Figure 10). To determine the impact of this patient's therapeutic dose on the estimation of stray dose for our sample, we reported summary statistics for several sub-groups within our sample: a group which includes patient 1, a group which excludes patient 1, all males, and all females (Table 10).

Table 10. Stray dose from proton CSI
Summary statistics for patients shown in Figure 14. Column A lists summary statistics for all patients shown (N = 18). Column B lists summary statistics for all patients except patient #1 (N = 17). Column C lists summary statistics for all males (N = 10). Column D list summary statistics for all females, including patient #1 (N = 8).

	A	B	C	D
	H_{stray} (Male & Female) (mSv)	H_{stray} - Patient #1 Removed (Male & Female) (mSv)	H_{stray} (Male) (mSv)	H_{stray} (Female) (mSv)
$D_{MA,MAX}$	683	547	547	683
$D_{MA,MIN}$	341	341	367	341
$D_{MA,MAX}$ - $D_{MA,MIN}$	342	206	180	342
Mean D_{MA}	437	423	444	428
Standard Deviation of the Experimental Mean	19	13	18	38
Experimental Standard Deviation	81	54	58	107

From Table 10 above, the mean D_{MA} for the sample was 437 mSv, but when D_{MA} for patient 1 was removed from the sample, the mean D_{MA} decreased by 4% to 423 mSv. This suggests the importance and impact of patient-specific variations on descriptive statistics of a small sample group. Additionally, the maximum values for mean dose dropped by 20% (i.e., from 683 mSv to 547 mSv) when the high dose patient was removed. However, we also saw that, the mean value of stray dose was slightly higher for males, 1.6%, than females, even when the female population contained an identified high dose patient.

3.3 Estimation of total equivalent dose

Table 11 lists the total equivalent dose, H_{total}, from proton CSI; it is the combined contribution of stray and therapeutic radiation in the thyroid. H_{total} values were calculated using D_{MA} and D_{CE}.

Table 11. Total equivalent dose proton CSI
Values are shown for H_{total}, when D_{CE} vs. D_{MA} is applied, resulting in predicted values of H_{CE} and H_{MA}, respectively. H_{CE} and H_{MA} are listed for 18 patients in columns A and B. Columns C and D compare the results in columns A and B via absolute difference and percent absolute difference.

	A	B	C	D
Patient Index	H_{MA} (mSv)	H_{CE} (mSv)	Difference (A-B)	Percent Difference (A-B)* 100/A (%)
1	1143	1051	91	8
2	501	493	8	2
3	372	367	5	1
4	431	425	6	1
5	560	549	10	2
6	503	495	8	2
7	410	405	6	1
8	347	343	4	1
9	527	517	10	2
10	472	464	8	2
11	567	549	18	3
12	405	399	6	1
13	429	422	7	2
14	460	453	7	2
15	400	394	5	1
16	412	407	6	1
17	372	368	5	1
18	443	436	7	2

Figure 15 (below) plots the values of total equivalent dose, H_{total}, listed in Table 11.

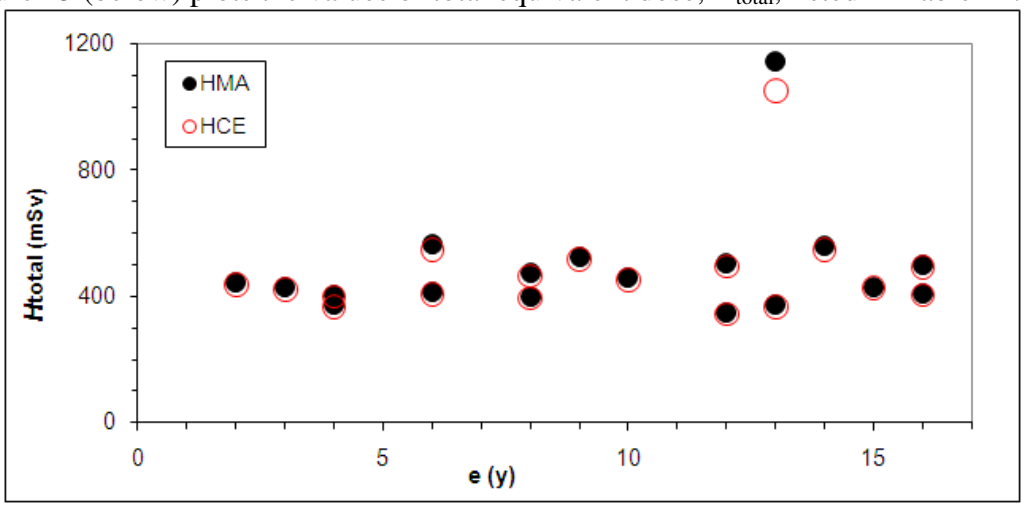

Figure 15. Total equivalent dose proton CSI
Total equivalent dose, H_{total}, verses age at exposure (e). H_{total} values were calculated using D_{MA} and D_{CE}. They are plotted as H_{MA} and H_{CE}.

Table 12. Total equivalent dose proton CSI: summary statistics
Summary statistics are listed for our sample of 18 patients. Columns A and B present results for the total equivalent dose when D_{MA} and D_{CE} were used, respectively. Columns C and D compare the results in columns A and B via absolute difference and percent absolute difference.

	A	B	C	D
	$H_{total, MA}$ (mSv)	$H_{total, CE}$ (mSv)	Difference (A-B) (mSv)	Percent Difference (A-B)* 100/A (%)
Maximum	1143	1051	91	8
Minimum	347	343	4	1
Interval	796	708	87	11
Mean	486	474	12	2
Standard Deviation of the Experimental Mean	41	37	5	11
Experimental Standard Deviation	176	156	19	11
Percent Standard Deviation of the Experimental Mean	8.5%	7.8%	1%	-
Percent Experimental Standard Deviation	36.1%	32.9%	3%	-

In Figure 15, we saw no evidence of a relationship between H_{total} and e. Also, in Table 12, we see that for proton CSI, the mean difference between values of H_{MA} vs. H_{CE} was only 2%. However, there is an 11% larger experimental standard deviation and an 8% higher maximum value for H_{MA} vs. H_{CE}. This supports the patient specific results (Table 11) which suggested that there were specific instances where use of D_{CE} did make a difference. For example, for patients 1 and 11 there was a 91 mSv and 18 mSv decrease in dose, and for patients 9 and 5 there was a 10 mSv decrease (Table 11). Additionally, while patients 1, 11, 9, and 5 had the four largest H_{MA} and the four largest differences between H_{MA} and H_{CE}, the difference between H_{MA} and H_{CE} did not directly scale with dose. Patient 9 and patient 5 each had a 10 mSv decrease from H_{MA} to H_{CE}, but their H_{MA} was 560 mSv and 527 mSv, respectfully. In the case of patient with the largest H_{MA}, patient 1, H_{MA} was 1143 mSv, and H_{CE} was 1051 mSv, thus the difference between H_{MA} and H_{CE} for the high dose patient was 91 mSv. This is 8 times larger than the mean difference for the sample, i.e., 12 mSv. Thus, use of D_{CE} normalized the effect of small differences in dose on H_{total}.

In Figure 16, we switch the focus from proton CSI to photon CSI. Similar to Figure 15 for proton CSI, values of total equivalent dose are presented in graphical and tabular form. In both cases, we compare values of H_{total} when D_{MA} vs. D_{CE} are applied to our sample population, i.e., we compare H_{MA} vs. H_{CE}.

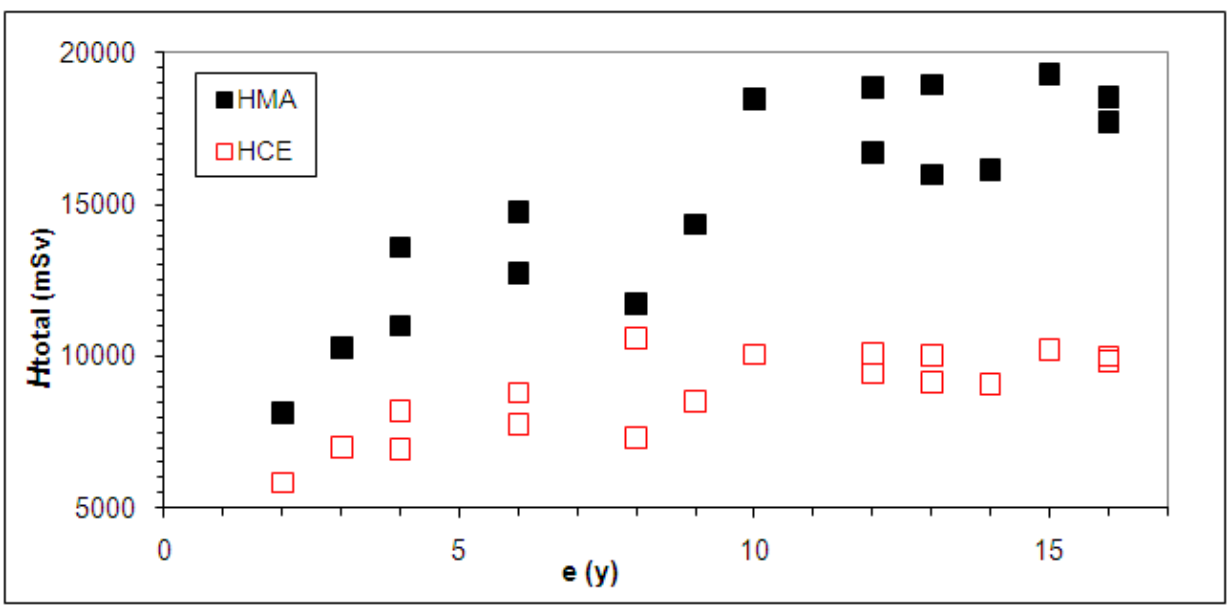

Figure 16. Total equivalent dose photon CSI
Total equivalent dose, H_{total}, is plotted against age at exposure (e).

Table 13. Total equivalent dose photon CSI

Values are shown for H_{total}, when D_{CE} vs. D_{MA} is applied, resulting in predicted values of H_{CE} and H_{MA}, respectively. Estimations for H_{CE} and H_{MA} are provided for 18 patients in columns A and B. Columns C and D compare the results in columns A and B via absolute difference and percent absolute difference.

	A	B	C	D
Patient Index	*H*MA (mSv)	*H*CE (mSv)	Difference (A-B)	Percent Difference (A-B)* 100/A (%)
1	15,981	9,129	6,852	43
2	17,723	9,826	7,898	45
3	18,965	10,028	8,936	47
4	19,291	10,200	9,090	47
5	16,157	9,105	7,052	44
6	18,868	10,083	8,785	47
7	18,559	10,009	8,550	46
8	16,736	9,437	7,299	44
9	14,359	8,509	5,850	41
10	20,736	10,591	10,145	49
11	12,730	7,757	4,973	39
12	13,588	8,204	5,384	40
13	10,270	6,992	3,278	32
14	18,498	10,062	8,436	46
15	11,716	7,324	4,392	37
16	14,769	8,805	5,965	40
17	11,003	6,952	4,051	37
18	8,123	5,827	2,296	28

Table 14. Total equivalent dose photon CSI: summary statistics
Summary statistics are shown for our sample of 18 patients. Similar to Figure 15, columns A and B present results for the total equivalent dose when the mean absorbed and cancer equivalent dose concepts are applied, respectively. Columns C and D compare the results in columns A and B via absolute difference and percent absolute difference.

	A	B	C	D
	$H_{total,\ MA}$ (mSv)	$H_{total,\ CE}$ (mSv)	Difference (A-B) (mSv)	Percent Difference (A-B)* 100/A (%)
Maximum	20736	10591	10145	49
Minimum	8123	5827	2296	28
Interval	12613	4764	7849	62
Mean	15448	8824	6624	43
Standard Deviation of the Experimental Mean	848	325	523	62
Experimental Standard Deviation	3598	1377	2220	62
Percent Standard Deviation of the Experimental Mean	5.5%	3.7%	2%	-
Percent Experimental Standard Deviation	23.3%	15.6%	8%	-

From Figure 16, we see that values of H_{total} decrease dramatically when D_{CE} is applied, and that the interval in H_{total} when D_{CE} vs. D_{MA} are used increases with age at exposure. Table 14 reveals that the difference between mean values of H_{total} is 6,6264 mSV with photon CSI vs. 12 mSV with proton CSI. Additionally, with photon CSI, the percent difference in the interval of H_{total} when D_{CE} vs. D_{MA} are used is 62% vs. 11% for proton CSI (Table 12). On the other hand, the experimental standard deviation of H_{MA} vs. H_{CE} is smaller for photon CSI vs. proton CSI, i.e., 23.3% and 15.6%, respectively, for proton CSI vs. 36.1% and 32.9%, respectively, for photon CSI. Thus, there was less variation in the mean values of H_{MA} and H_{CE} for photon CSI than for proton CSI (Tables 12 and 14).

To determine if there were additional trends in the difference between H_{MA} and H_{CE} for photon CSI, we plotted the absolute difference in H_{MA} vs. H_{CE} with respect to age at exposure (Figure 17).

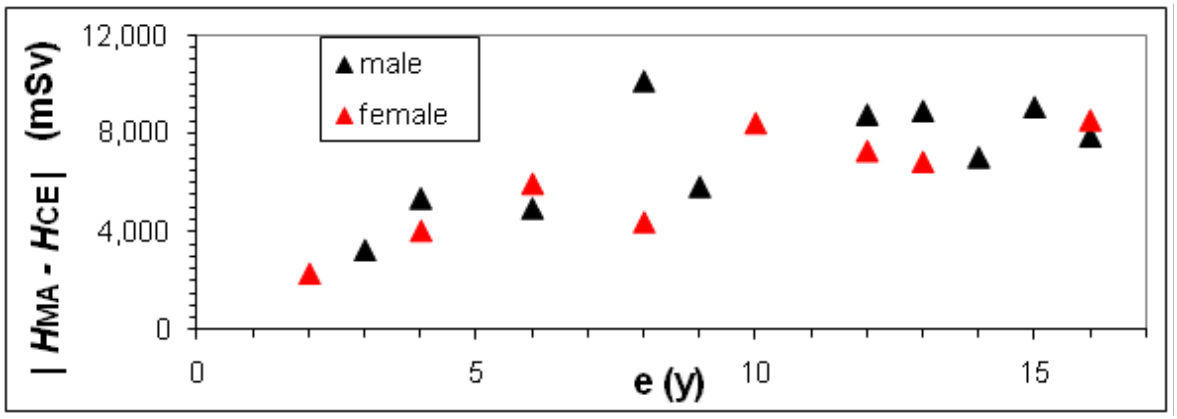

Figure 17. Absolute difference between H_{MA} and H_{CE} for photon CSI
The absolute difference between H_{MA} and H_{CE} is plotted against age, *e*, for male and female patients.

Figure 17, reveals that $| H_{MA} - H_{CE}|$ appears to level off to 8000 mSv at about 10 years of age of exposure. However, we would need more data-points for patients that are more than 10 years-old to make a definitive statement regarding this dose-age trend. Finally, there did not seem to be a significant sex-specific difference in $| H_{MA} - H_{CE}|$ values.

To further characterize the contribution of stray dose within our study, we compared the values of stray dose to therapeutic dose (Table 15 below) in the cases of proton and photon CSI.

Table 15. Comparison of therapeutic doses to stray dose
Summary statistics for our sample of 18 patients are listed. The column for stray dose was not populated in the case of photon CSI because the thyroid was in the exit region of the field.

	Proton CSI			Photon CSI		
	$H_{therapeutic}$ (mSv)	H_{stray} (mSv)	$H_{therapeutic}$ (%)	$H_{therapeutic}$ (mSv)	H_{stray} (mSv)	$H_{therapeutic}$ (%)
Maximum	457	683	67	20,736		100
Minimum	1	341	0	8,123		100
Interval	456	342	-	12,613	thyroid in exit region	100
Mean	50	437	11	15,448		100
Standard Deviation of the Experimental Mean	25	19	-	848		-
Experimental Standard Deviation	106	81	-	3,598		-

From Table 15, we found that for proton CSI, the mean $H_{therapeutic}$ was 50 mSv while the mean H_{stray} was 437 mSv. As a result, the mean $H_{therapeutic}$ was 11% of the mean H_{stray}. Correspondingly, the minimum $H_{therapeutic}$ was less than 1% of the minimum H_{stray} , and the maximum value of $H_{therapeutic}$ was 67% that of the maximum value for stray dose. Thus for proton CSI, the importance of including stray radiation is underscored. To further understand the difference between H_{MA} and H_{CE} for photon vs. proton CSI, we focused on their fundamental main difference, the cell sterilization factor, *C,* (Equation 2.3). Because *C* was estimated on a voxel by voxel basis, we determined the mean cell sterilization factor for each patient's thyroid, $\bar{C}$, and plotted those values against age at exposure, *e*, for proton and photon CSI (Figure 18).

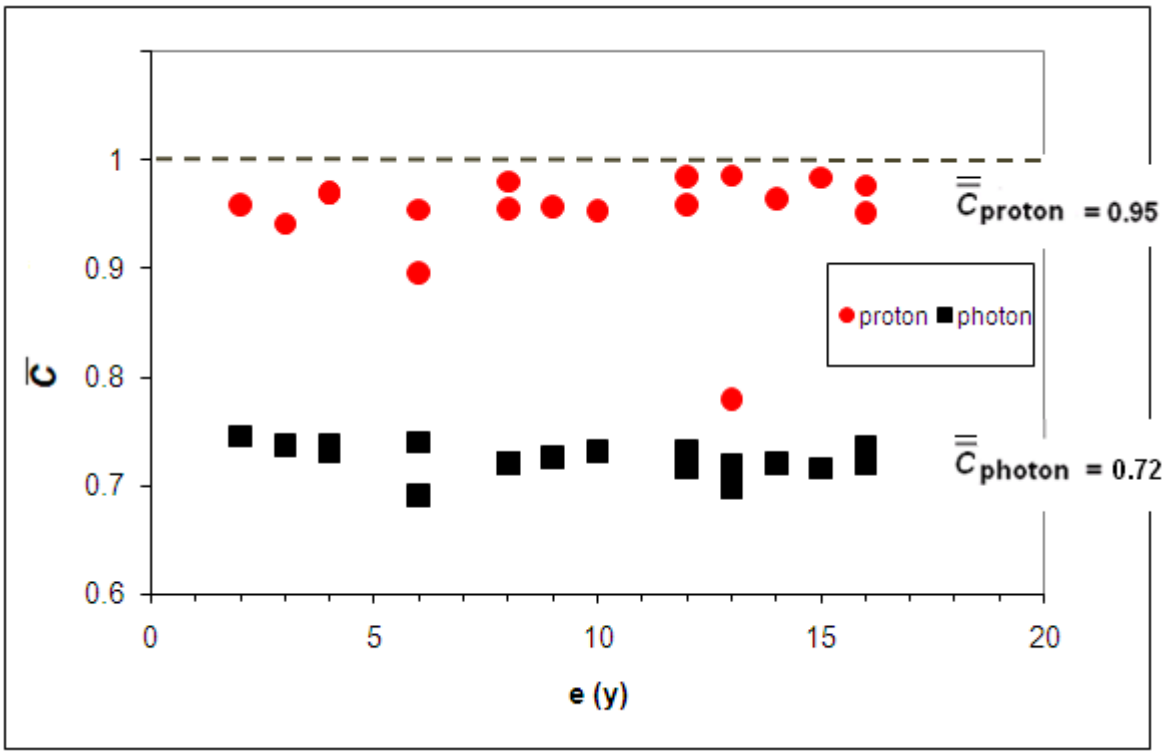

Figure 18. Cell sterilization factor for proton and photon CSI.
Mean cell sterilization factor, $\bar{C}$, was plotted against age at exposure for proton and photon CSI. The population averaged $\bar{C}$ for proton CSI was 0.95. The overall mean $\bar{C}$ for photon CSI was 0.72. The dashed line indicates a correction factor of 1.0.

In Figure 18, we see the same underlying patterns as we did in Figures 15a and 16a. For the case of proton CSI, Figure 18 shows that $\bar{C}$ was approximately 1 for all patients except

patient 1, the patient who received the highest dose to the thyroid. For the case of photon CSI, $\bar{C}$ for the high dose patient was 0.78 which is closer to the mean $\bar{C}$ for photons, 0.72.

3.4 Risk prediction

In this section, we present results of estimates of the risk of second cancers in the thyroid. To do this we applied the model for risk of SMNs in the thyroid from the report of the BEIR VII Committee (NRC 2006). The model was applied to H_{MA} and H_{CE}.

3.4.1 Relative risk estimates from photon vs. proton CSI

Figure 19 plots the distribution of values of the predicted relative risk, RR_{MA} and RR_{CE}, of second cancers in the thyroid after photon vs. photon CSI (Equations 2.19 and 2.22).

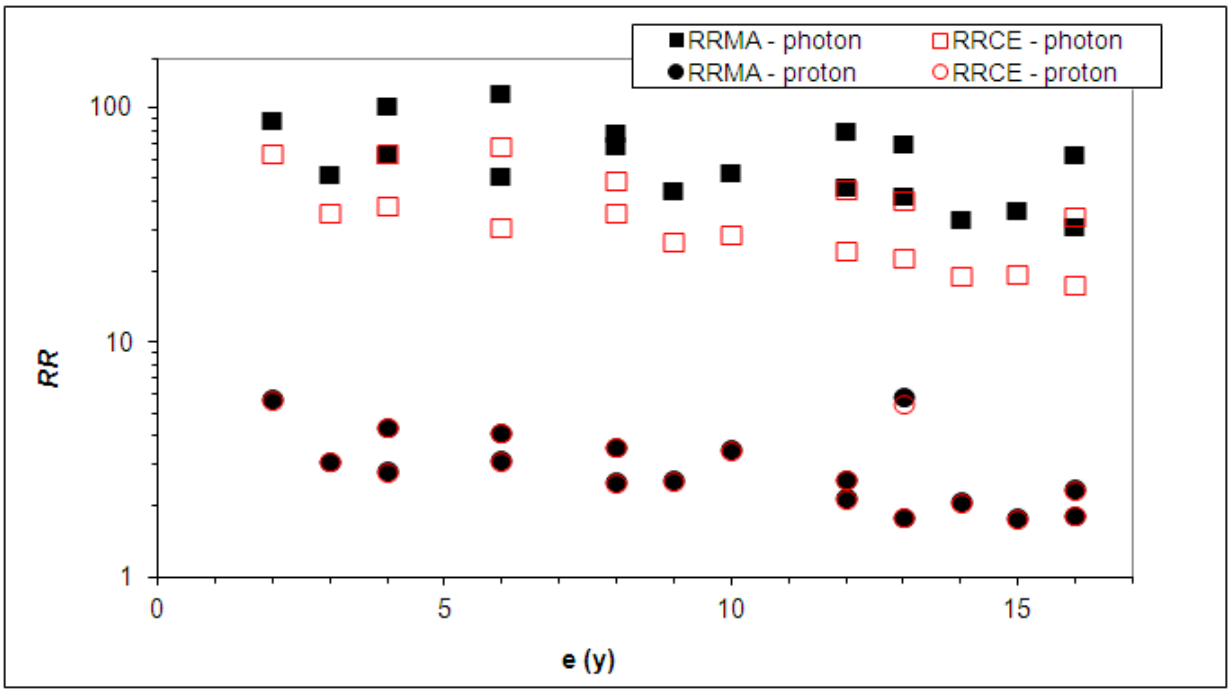

Figure 19. Relative Risk estimates for photon and proton CSI vs. age at exposure
The relative risk, *RR*, of second cancers in the thyroid from photon CSI and proton CSI is plotted on a logarithmic scale against age at exposure, e. RR_{MA} is the estimate of *RR* when D_{MA} was used to estimate total equivalent dose, and RR_{CE} is the estimate of *RR* when D_{CE} was used to estimate total equivalent dose.

Figure 19 indicates that after photon and proton CSI, *RR* values decrease with age at exposure. Also, similar to the graphs of H_{total} and *C*, we see that, in general, our results for photon CSI are more sensitive to use of D_{MA} vs. D_{CE} than our results for proton CSI. To elucidate the differences in *RR* for photon vs. proton CSI, we plotted the results from Figure 18 on a semi-log scale (Figure 20).

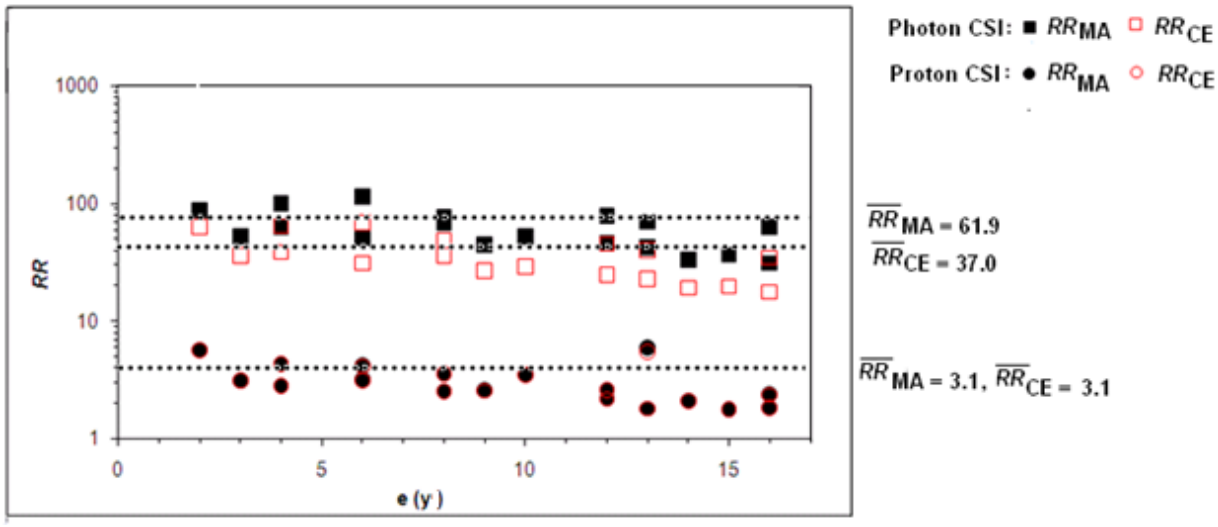

Figure 20. $\overline{RR}$ estimates for photon and proton CSI vs. age at exposure
The $\overline{RR}$ from photon CSI and proton CSI are overlaid on the plot from figure 19.

Figure 20 shows that $\overline{RR}$ for photon CSI is larger than $\overline{RR}$ for proton CSI by factors of 12 and 20 when D_{MA} and D_{CE} are applied, respectively. Additionally, Figure 19 demonstrates that use of D_{CE} reduces $\overline{RR}$ by almost 60% for photon CSI and has negligible effect for proton CSI (high dose patient, patient 1, included). To further analyze our results for RR_{MA} and RR_{CE}, we plotted the difference between RR_{MA} and RR_{CE} against patient age in years (Figure 21a) and the ratio of RR_{CE} to RR_{MA} against patient age at exposure (Figure 21b).

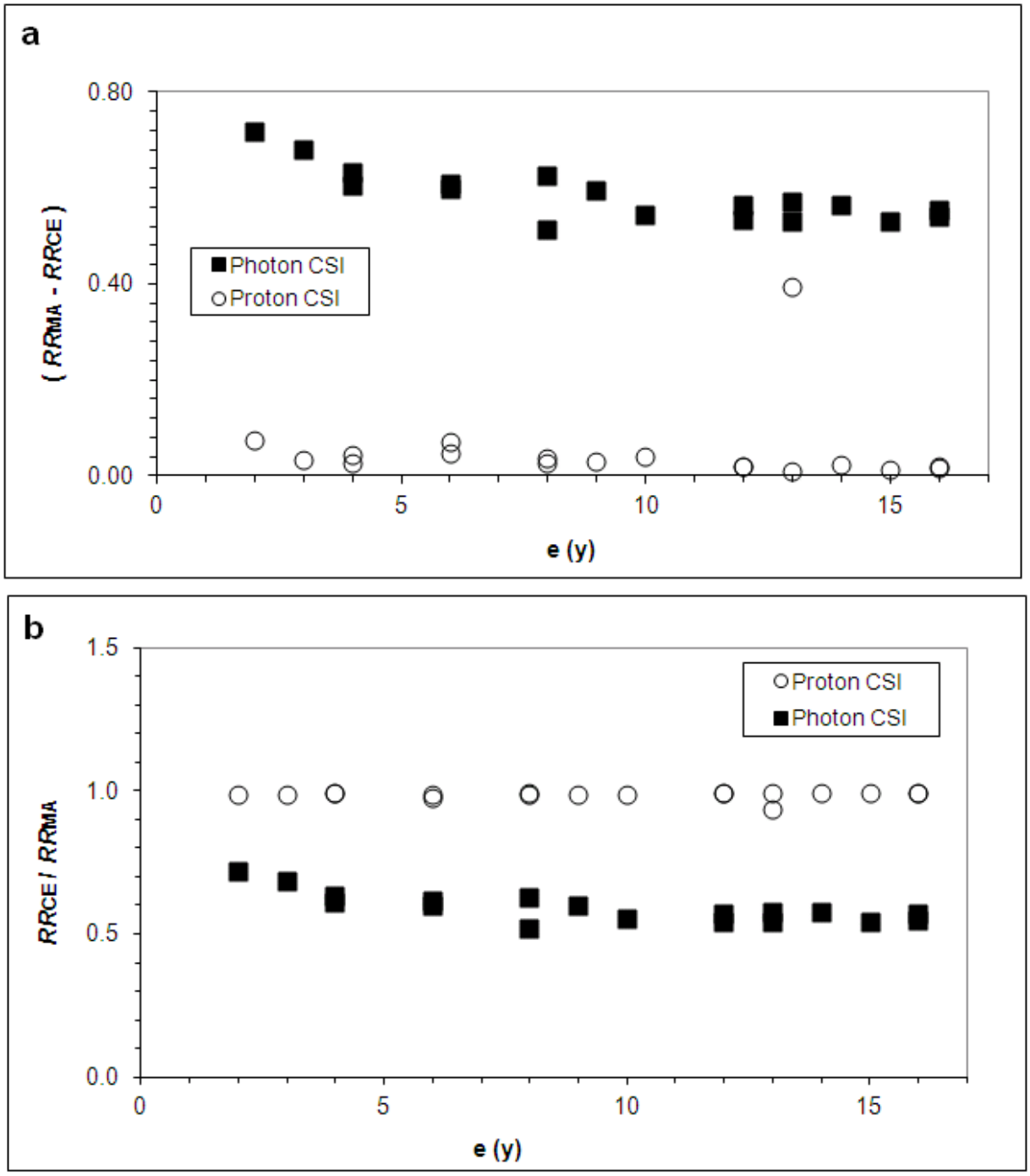

Figure 21. Comparison of RR_{MA} to RR_{CE}
In (a) the difference between estimates for *RR* for each patient (N = 18) when D_{MA} and D_{CE} are applied is plotted against patient age at exposure in years, *e*, for proton and photon CSI. In (b) the ratio of estimates for *RR* for each patient (N = 18) when D_{MA} and D_{CE} are applied is plotted against patient age in years, *e*, for proton and photon CSI, i.e., the ratio of RR_{CE} to RR_{MA} is plotted against '*e*' for proton and photon CSI.

Figure 21 reveals that the relationship between RR_{MA} and RR_{CE} is fairly constant with respect to age at exposure. There is a slight increase in the difference between RR_{MA} and RR_{CE} (Figure 21a) for patients that are less than 5 years old. However, there are not enough data points in that region to provide conclusive results.

3.4.1.1 Sex specific trends for relative risk estimates

When the risk model (Equation 2.16 or 2.21) is applied to our estimates of H_{MA} or H_{CE}, there is an inherent difference in how it is applied to male vs. female patients. In particular the values of the age-specific and organ-specific values, β_M and β_F, differ for males and females, respectively (Table 2). The ratio of β_M to β_F is 0.5. As a result, we expected that regardless of the radiation source, proton CSI or photon CSI, we would see a ratio of *RR* in males to *RR* in females of approximately 0.5. Our results from this comparison are listed in Table 16 below.

Table 16. Ratios of male-to-female values of dose, risk, and risk coefficient
The ratio of mean values in male vs. female patients for: $\overline{D}_{therapeutic}$, $\overline{H}_{stray}$, $\overline{H}_{total}$, $\overline{RR}$ using D_{MA}, $\overline{RR}$ using D_{CE}, and the age-specific and organ-specific risk models from BEIR VII are listed for proton CSI and photon CSI. Data for the female patient receiving the highest dose from proton CSI was excluded from this analysis.

Ratio Male / Female	Proton CSI	Photon CSI
$\overline{D}_{therapeutic}$	2.2	1.1
$\overline{H}_{stray}$	1.1	n/a
$\overline{H}_{total}$	1.1	1.1
$\overline{RR}_{MA}$	0.8	0.8
$\overline{RR}_{CE}$	0.8	0.7
β_M/β_F	0.5	0.5

The data in table 16 reveals that for our sample, female patients received almost twice the therapeutic dose from proton CSI to the thyroid as males and 10% more stray dose. Because the contribution from stray dose dominated total equivalent dose in the case of proton therapy (Table 15), the combined effect of therapeutic and stray dose resulted in only 10% more total equivalent dose from proton CSI for females than males.

A similar result was found for photon CSI. Correspondingly, the ratio of male to female values of *RR* was 0.7 to 0.8, or approximately 60% higher than expected from β_M / β_F.

While these results seem to support the analysis in section 3.1.1, due to our limited sample sizes (8 females, 10 males), they did not reveal a significant relationship between sex and dose or predicted risk of second cancers in the thyroid.

3.4.2 Ratio of relative risk estimates for photon vs. proton CSI

In this section we compare the *RR* from photon and proton CSI by taking their ratio. We termed this the ratio of relative risk or *RRR*. Because this was done for *RR* values which were estimated using D_{MA} and D_{CE} (Equations 2.20 and 2.23), our corresponding terminology for *RRR* was RRR_{MA} and RRR_{CE}, respectively.

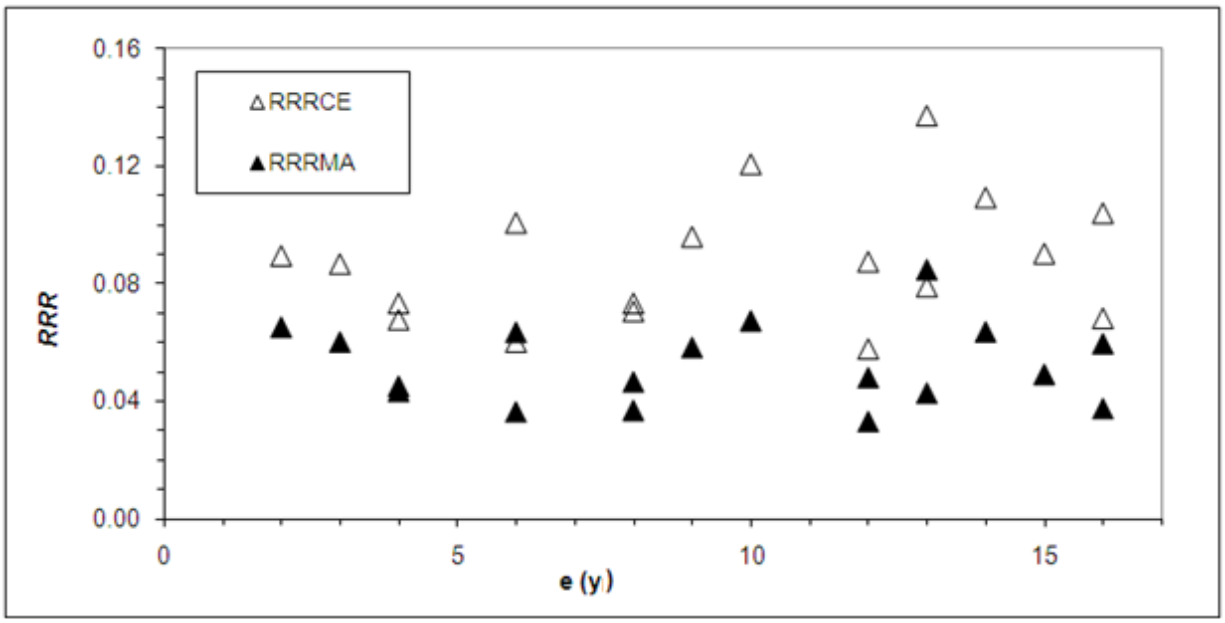

Figure 22. Ratio of relative risk after proton vs. proton CSI
Predicted ratio of relative risk (*RRR*) between proton vs. proton CSI are plotted against age at exposure, *e*. The first estimation (open triangles) utilizes D_{CE} (RRR_{CE}), and the second (closed triangle) utilizes D_{MA} (RRR_{MA}).

Figure 22 indicates that *RRR* is typically higher when D_{CE} (RRR_{CE}) is applied vs. when D_{MA} (RRR_{MA}) is applied. This finding restates our previous results that RR_{CE} was consistently lower than RR_{MA} in the case of photon CSI and approximately equal in the case of photon CSI (Figure 20). More specifically, from Figure 20, we saw that for photon CSI the mean value for RR_{CE} was 60% lower than the mean value for RR_{MA}, and for proton CSI there was a negligible difference. The impact of these differences on the predicted *RRR* is shown in Table 17 for our patient sample.

Table 17. Ratio of relative risk from photon vs. proton CSI when D_{CE} vs. D_{MA} is applied
Values are shown for *RRR* when D_{CE} vs. D_{MA} is applied, resulting in predicted values of RRR_{CE} and RRR_{MA}, respectively. Estimates of RRR_{CE} and RRR_{MA} are provided for 18 patients in columns A and B. Columns C and D compare the results in columns A and B via absolute difference and percent absolute difference.

	A	B	C	D
Patient Index	RRR_{MA}	RRR_{CE}	Difference (A-B)	Percent Difference (A-B)* 100/A (%)
1	0.085	0.137	-0.052	-62
2	0.060	0.104	-0.044	-74
3	0.043	0.079	-0.036	-84
4	0.049	0.090	-0.041	-83
5	0.064	0.109	-0.046	-72
6	0.048	0.087	-0.039	-82
7	0.038	0.068	-0.031	-81
8	0.033	0.058	-0.025	-74
9	0.058	0.096	-0.038	-64
10	0.037	0.071	-0.034	-91
11	0.064	0.101	-0.037	-58
12	0.045	0.073	-0.028	-62
13	0.060	0.087	-0.026	-44
14	0.067	0.121	-0.053	-79
15	0.047	0.073	-0.027	-57
16	0.036	0.060	-0.024	-65
17	0.043	0.068	-0.024	-56
18	0.065	0.089	-0.024	-37

Table 18. Summary statistics for the ratio of relative risk after photon vs. proton CSI
Summary statistics for Table 17 are provided. Columns A and B present results for the ratio of relative risk (*RRR*) from photon vs. proton CSI when the mean absorbed and cancer equivalent dose concepts are applied, respectively. Columns C and D compare the results in columns A and B via absolute difference and percent absolute difference.

	A	B	C	D
	RRR_{MA}	RRR_{CE}	Difference (A-B)	Percent Difference (A-B)* 100/A (%)
Maximum	0.085	0.137	-0.052	-62
Minimum	0.033	0.058	-0.025	-74
Interval	0.052	0.079	-0.028	-54
Mean	0.052	0.087	-0.035	-67
Standard Deviation of the Experimental Mean	0.003	0.005	-0.002	-55
Experimental Standard Deviation	0.014	0.021	-0.008	-55
Percent Standard Deviation of the Experimental Mean	6.2%	5.8%	0.4%	-
Percent Experimental Standard Deviation	26%	24%	2%	-

From Tables 17 and 18, the mean value of *RRR* increased by 67% when it was estimated using D_{CE} vs. D_{MA}. The percent differences between maximum values and minimum values of *RRR* are 62% and 74%, respectively for RRR_{CE} vs. RRR_{MA}. This reflects the sensitivity of the *RRR* to patient specific variation. Visual representation of the data for columns A and B is found in Figures 23 and 24 below.

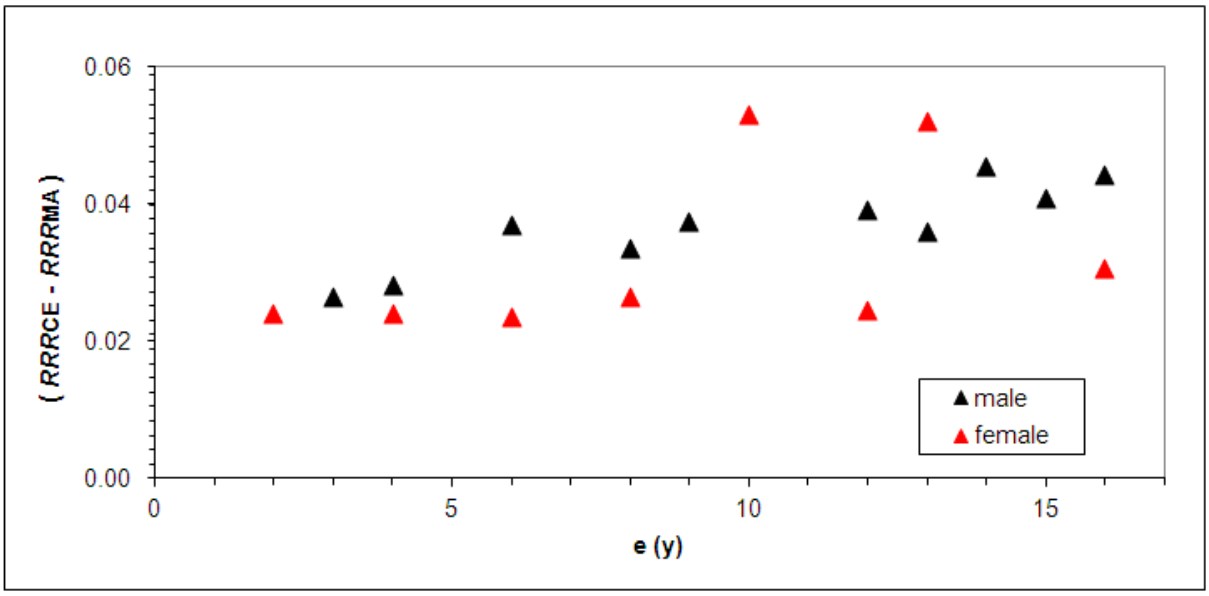

Figure 23. Differences in estimates for *RRR* when D_{MA} vs. D_{CE} were used
The difference in estimates for *RRR* when the D_{CE} vs. D_{MA} concept is applied, i.e., RRR_{CE} and RRR_{MA}, respectively, is stratified by sex and plotted against patient age, *e*.

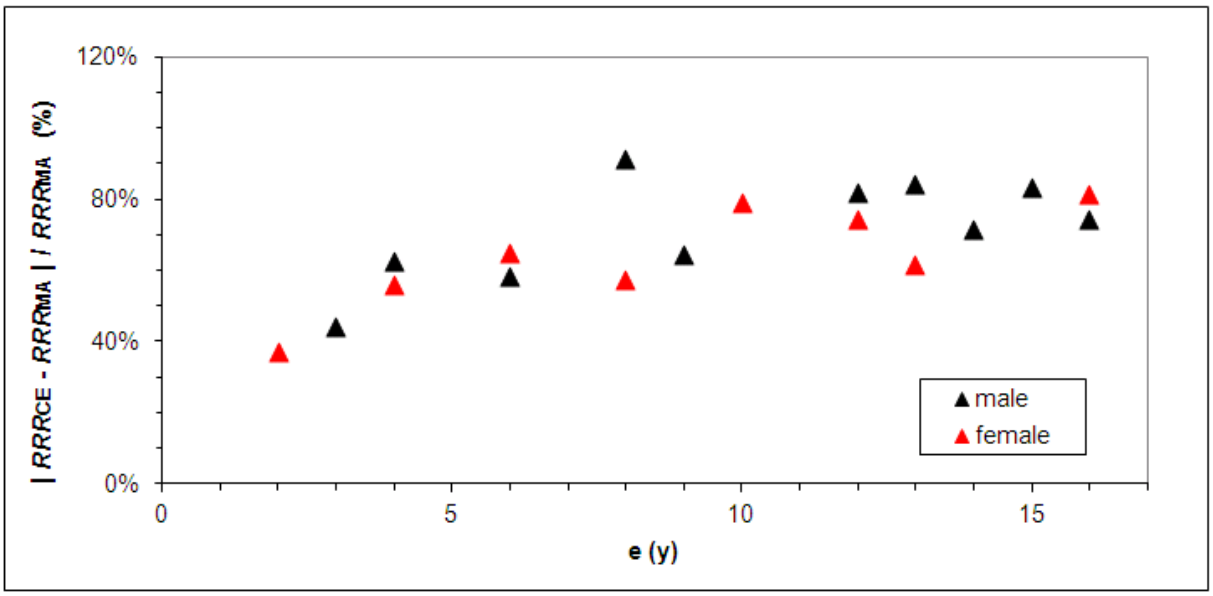

Figure 24. Percent difference in estimates for *RRR* when D_{MA} vs. D_{CE} were used
The percent difference in estimates for *RRR* when D_{CE} vs. D_{MA} is applied, i.e., RRR_{CE} and RRR_{MA}, respectively, is stratified for sex and plotted against patient age, *e*.

In a final analysis of our estimates for RRR_{CE} and RRR_{MA} we performed three t-tests using S-Plus. These tests were performed to determine if our estimates for RRR_{CE} vs. RRR_{MA} were statistically different from the null hypothesis ($RRR = 1$) and one other. In our first test,

we compared the mean value of RRR_{MA} against $RRR = 1$. In the second test, we compared the mean value of RRR_{CE} against $RRR = 1$, and in the third test, we used a paired t-test to compare the mean value for RRR_{CE} against the mean value for RRR_{MA} (Table 19).

Table 19. Results of t-test to determine if there is a statistically significant difference between RRR_{CE} vs. RRR_{MA}

Column A shows results of a t-test for one sample in which the mean value for RRR_{MA} is compared to $RRR = 1$. Column B shows results of a t-test for one sample in which the mean value for RRR_{CE} is compared to $RRR = 1$. Column C shows results of a paired t-test for two sample means: RRR_{CE} and RRR_{MA}. All tests were performed for our patient sample (N = 18). The results for the two-tail test are highlighted in yellow because they are most relevant to the virtual clinical trial in this work.

	A	B	C
N	18	18	18
P(T<=t) one-tail	3.28E-33	1.08E-29	1.05E-11
t Critical one-tail	1.74E+00	1.74E+00	1.74E+00
P(T<=t) two-tail	6.55E-33	2.16E-29	2.09E-11
t Critical two-tail	2.11E+00	2.11E+00	2.11E+00

Table 19 provides results for a one-tail and two-tailed t-test. For either test and in all columns, the reported p-values are much less than 0.05, thus there is very strong statistical evidence that both RRR_{MA} and RRR_{CE} are significantly different from unity. Similarly, there is very strong evidence of statistically significant difference between RRR_{CE} and RRR_{MA}.

3.4.2.1 Sex specific trends for the ratio of relative risk for photon vs. proton CSI

Our estimates for RRR_{CE} and RRR_{MA} show no identifiable trend with respect to sex. Additionally, comparisons of RRR_{CE} and RRR_{MA} (Figures 23 and 24) also show now difference with respect to sex. This is an expected result because according to Equations 2.20 and 2.22, the sex specific terms in the risk model, β_M and β_F, cancel when the ratio of RR values is taken.

3.5 Sample size estimations

We performed three separate analyses of sample size. Our first two analyses were the most relevant to our virtual clinical trial because they estimated the number of patients needed to determine a statistically significant difference in predicted risk from proton vs. photon CSI

for two cases: (1) the case when D_{MA} was used to determine predicted risk and (2) the case when D_{CE} was used to determine predicted risk. Our third analysis addressed the related question of whether or not we had enough subjects to compare estimates for predicted risk using D_{MA} vs. D_{CE}.

In our first analysis we considered the case in which a patient population was characterized using the mean value and experimental standard deviation of RRR_{CE} as determined in our work. In our second analysis we considered the case in which a patient population was characterized using the mean value and experimental standard deviation of RRR_{MA} as determined in our work. In these first two analyses we used the one sample t-test in which we compared our mean values of RRR_{MA} and RRR_{CE}, respectively, against the null hypothesis for our virtual clinical trial, i.e., H_0: $RRR = 1$.

Table 20. Sample size estimation for one sample t-test
Results are shown for three sample size estimators, two freely available statistical packages and S-Plus. Input into the estimators included mean *RRR* for our patient population (N = 18), the experimental standard deviation (σ_{RRR}) for our patient population (N = 18), and the desired values of α and power (1-β) which were 5% and 80%, respectively. Output from the estimators was given as the minimum sample size needed to achieve 5% for alpha and 80% power. Differences in estimates for the required sample size between software packages are most likely the result of differences in the software's rounding, i.e., a sample size of 0 vs. 1, and truncation of reported results.

Input Parameters	Statistics Package					
	PS Software		Web-based Calculator		S-Plus	
	RRR_{MA}	RRR_{CE}	RRR_{MA}	RRR_{CE}	RRR_{MA}	RRR_{CE}
$\overline{RRR}$	0.052	0.087	0.052	0.087	0.052	0.087
σ_{RRR}	0.014	0.021	0.014	0.021	0.014	0.021
α	0.05	0.05	0.05	0.05	0.05	0.05
(1-β)	0.8	0.8	0.8	0.8	0.8	0.8
Required Sample Size	0	0	1	1	1	1

The results for our sample size estimates in Table 20 indicate that the sample size required for 80% power is at most 1, regardless of whether D_{CE} or D_{MA} was applied. One reason for this finding is that our estimates for RRR_{MA} and RRR_{CE} are 20 to 12.5 times lower than that of our null hypothesis ($RRR = 1$). Additionally the variation or experimental standard deviation of our sample was very small. As a result, the estimators calculated that at most one patient was needed to show that the risk of second cancers in the thyroid is reduced with proton

CSI vs. photon CSI. Thus, according to our estimation for sample size, use of D_{CE} to reduce sample size becomes an un-necessary step.

In our second analysis of sample size, we estimated the number of patients needed if one were to perform a study which directly compared predicted risk using D_{MA} vs. D_{CE}. To do this we used a two sample t-test in which we compared our mean values of RRR_{MA} to our mean values of RRR_{CE}. More precisely, we were no longer comparing our estimates for *RRR* against the null hypothesis: we were instead comparing RRR_{MA} to RRR_{CE}. Thus, in this estimation, the estimates for sample sizes represent the number of patients needed to see a significant difference between RRR_{MA} and RRR_{CE} (Table 21).

Table 21. Sample size estimation for two sample t-test

Results are shown for two sample size estimators (freely available statistical packages). Sample size was estimated for two data sets. In the first data set we included all patients in the study (N = 18), In the second data set, we excluded the patient receiving a notably high dose from proton CSI, patient #1, (N = 17). Input for each data set included values of mean *RRR,* the experimental standard deviation (σ_{RRR}), and our desired values of (α)and power (1-β) which were 5% and 80%, respectively. Output from the estimators was given as the minimum sample size needed to achieve 5% for alpha and 80% power.

Input Parameters	Statistics Package			
	PS Software		Web-based Calculator	
	RRR_{MA}	RRR_{CE}	RRR_{MA}	RRR_{CE}
High dose proton patient *included* (N = 18):				
$\overline{RRR}$	0.052	0.087	0.052	0.087
σ_{RRR}	0.014	0.021	0.014	0.021
α	0.05	0.05	0.05	0.05
(1-β)	0.8	0.8	0.8	0.8
Required Sample Size	4	7	6	3
High dose proton patient *excluded* (N = 17):				
$\overline{RRR}$	0.05	0.084	0.05	0.084
σ_{RRR}	0.011	0.018	0.011	0.018
α	0.05	0.05	0.05	0.05
(1-β)	0.8	0.8	0.8	0.8
Required Sample Size	3	6	5	2

In Table 21, estimates of required sample size increased relative to the estimates for the one sample t-test (Table 20). One reason for the increase is that when we switched to a two sample t-test, statistical variation within each sample was increased. This occurred because each sample no longer served as its own control. However, even with the increase in required sample size, Table 21 revealed that our sample size (N = 18) is large enough to demonstrate that RRR_{MA} and RRR_{CE} were statistically different from each other when the t-test is used. Thus, in future studies, it is reasonable to expect that two samples, in which the predicted RRR is determined using different methods, can be directly compared.

To generalize the trends from Tables 18 (RRR) and 20 (corresponding sample size), we performed a more broadly-applicable analysis. That is, we used the input parameters in Table 18 (restated in Table 22 below) and the PS software (Dupont and Plummer 1990) to generate curves for sample size for RRR_{MA} and RRR_{CE} for specific statistical power values, i.e. 50%, 60%, 80% and 90%. These curves were determined using a two-sided t-test with two independent samples, thus, our values of $\overline{RRR}_{\mathrm{MA}}$ and $\overline{RRR}_{\mathrm{CE}}$ were compared against the null hypothesis, H_o: $RRR = 1$. The results of this analysis as it relates to our project goal of achieving 80% power is shown in Figure 25 below.

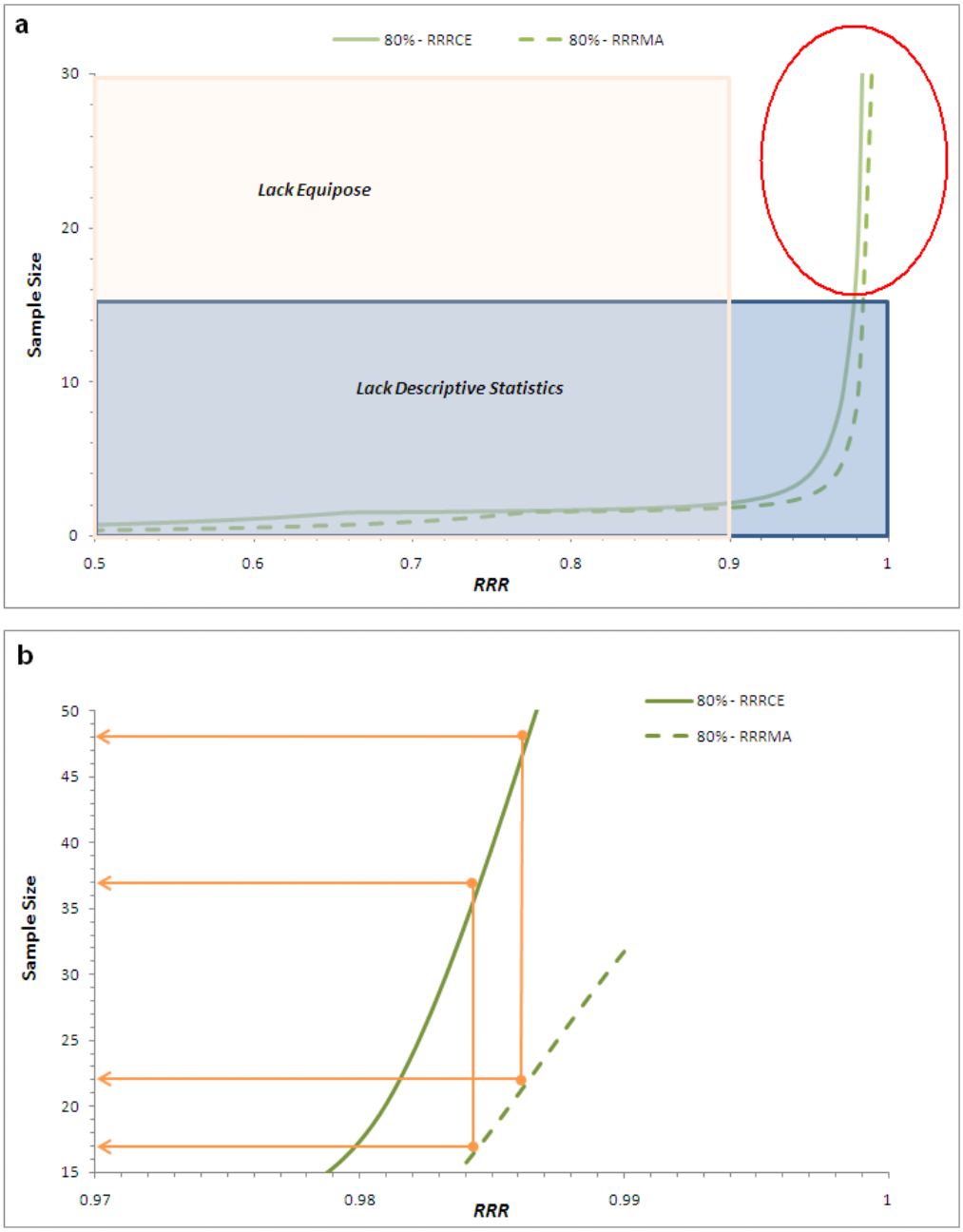

Figure 25. Effect of D_{MA} vs. D_{CE} on sample size

(a) Plot of sample sizes when 80% power is achieved for a population characterized by our values of RRR_{MA} and RRR_{CE} ($\overline{RRR}_{MA}$ = 0.052, standard deviation for RRR_{MA} = 0.014, $\overline{RRR}_{CE}$ = 0.087, standard deviation for RRR_{CE} = 0.021). The red circle indicates the region on the graph that is not limited by concerns of equipoise due to low *RRR* values or lack of descriptive statistics due to low sample sizes. (b) The region within the red circle in figure (a) is enlarged and demonstrates the relationship between *RRR* values and sample size when *RRR* is calculated using D_{MA} vs. D_{CE}. The orange arrows indicate the sample sizes for RRR_{MA} and RRR_{CE} at *RRR* = 0.986 (48 vs. 22 patients) and *RRR* = 0.984 (17 vs. 37 patients).

Figure 25a demonstrates the relationship between sample size and RRR_{MA} or RRR_{CE} when 80% statistical power was achieved. The curves for RRR_{MA} or RRR_{CE} were characterized using values from our patient sample (N = 18), i.e., $\overline{RRR}_{MA}$, $\overline{RRR}_{CE}$, and values of experimental standard deviations of RRR_{MA} and RRR_{CE}, respectfully. In addition, Figure 25 demonstrates the effect of factors which limit the intervals of acceptable sample size and *RRR,* i.e., the number of subjects required to achieve descriptive statistics and the range of *RRR* in which concerns regarding equipoise are avoided. When the aforementioned limitations are addressed, the intervals of acceptable sample size and the *RRR* shrank (Figure 25b). Within that window, we found that for 80% power and a *RRR* value of 0.984, 37 patients vs. 17 were needed if *RRR* is calculated using D_{CE} vs. D_{MA}. For a *RRR* value of 0.986, 48 patients vs. 22 were needed if *RRR* is calculated using D_{CE} vs. D_{MA}.

3.5.1 Four factors that affect sample size

To further understand the interplay between estimates for *RRR* and sample size, we used PS to estimate sample size for three general relationships. In particular, we followed a procedure similar to the one used to create Figure 25 to illustrate the influence of (1) power, (2) variance (or standard deviation), and (3) α on sample size using data from the virtual clinical trial. Thus, our estimations for sample size probed the impact of differing values of power, variance (or standard deviation), and α for two populations, i.e., one characterized by our values of $RRR_{MA,}$ and one characterized by our values of, RRR_{CE} (Figures 26-28). The corresponding input parameters for these three tests are listed in Table 22 below.

Table 22. Input parameters for estimates of sample size relative to differing values for power (1-β), standard deviation of *RRR* (σ_{RRR}), and α.

Input Parameters	RRR_{MA}	RRR_{CE}
$\overline{RRR}$	0.052	0.087
σ_{RRR}	0.014	0.021
α	0.05	0.05
(1-β)	0.8	0.8

Figure 26 represents our first generalized estimation. That is, it compares the effect of differing values for power on sample size.

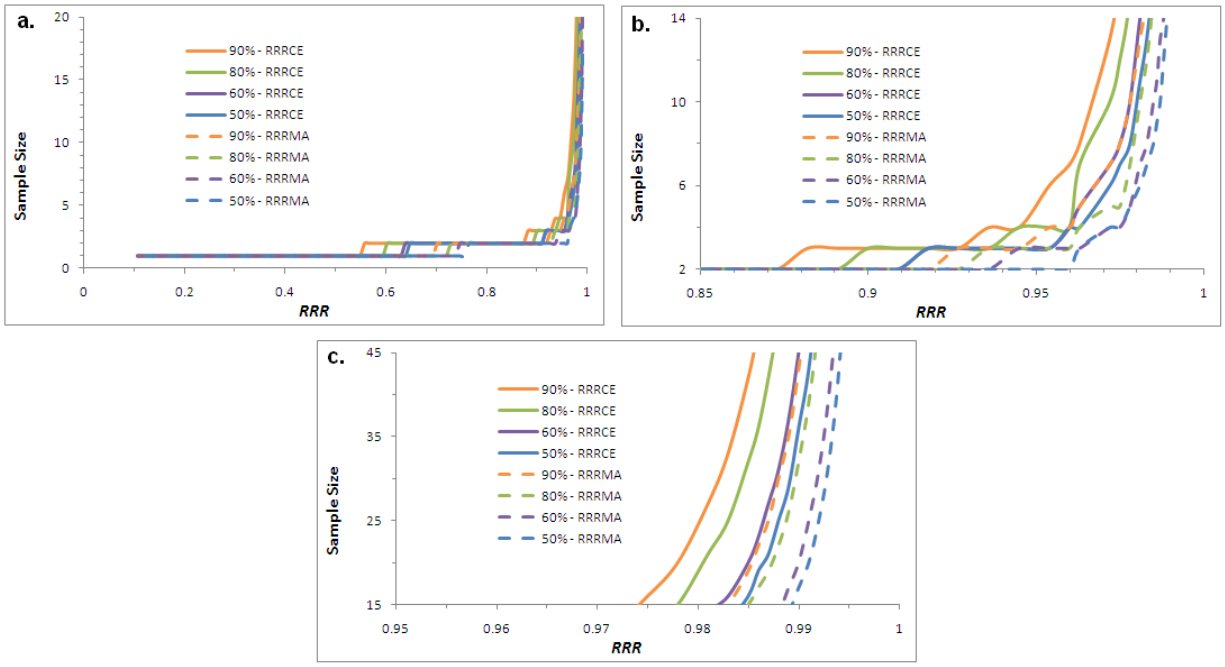

Figure 26. Effect of power on sample size
Sample size is plotted against *RRR* for two populations (input parameters in Table 22). In one population (solid lines), *RRR* was calculated using D_{CE}, i.e., $RRR_{CE,}$ and in the other (dashed lines) *RRR* was calculated using D_{MA}, i.e., RRR_{MA}. Power curves are shown in solid lines for RRR_{CE} vs. sample size and in dashed lines for RRR_{MA} vs. sample size. Alpha was set to 5% for all curves.

Figure 26 demonstrates that in order to achieve 50% to 90% power, a larger sample size is required when D_{CE} is used to estimate *RRR*. It also shows that the difference in sample sizes is most applicable when small samples are used, i.e., sample sizes of 5 or less. This finding is expanded upon in Table 23 (below) which presents an analysis based on data from Figure 26.

Table 23. Effect of power on sample size for RRR_{MA} vs. RRR_{CE}

Values for the absolute difference and the percent absolute difference between RRR_{MA} and RRR_{CE} are shown for the 80% and 50% power curves at specific sample sizes.

	$\lvert RRR_{CE} - RRR_{MA} \rvert$		$\frac{\lvert RRR_{CE} - RRR_{MA} \rvert * 100}{RRR_{MA}}$ (%)	
Sample Size	80% Power	50% Power	80% Power	50% Power
1	0.14	0.13	19	17
2	0.04	0.03	4	3
3	0.02	0.02	2	2
5	0.01	0.01	0.9	1
10	0.01	0.005	0.9	0.5
20	0.004	0.006	0.4	0.6

Table 23 highlights the result that use of D_{MA} *vs.* D_{CE} when estimating *RRR* has a more notable difference at the level of individual comparisons of risk for an individual patient, i.e., as opposed to a sample of patients. In other words, estimates of *RRR* differed by as much as 19% when RRR_{CE} was used in place of RRR_{MA} in a case study (sample size of 1) or 4% when a sample size of 2 is used. When the sample size increases beyond 5, the percent difference in estimates for RRR_{MA} vs. RRR_{CE} drops to less than 1%.

To further investigate use of RRR_{MA} vs. $RRR_{CE,}$ we focused on the potential impact of differences in standard deviation on sample size. Recall that in our virtual clinical trial, the experimental standard deviation was 0.014 for RRR_{MA} and 0.021 for RRR_{CE} . Figure 27 plots the impact of the standard deviation on sample size.

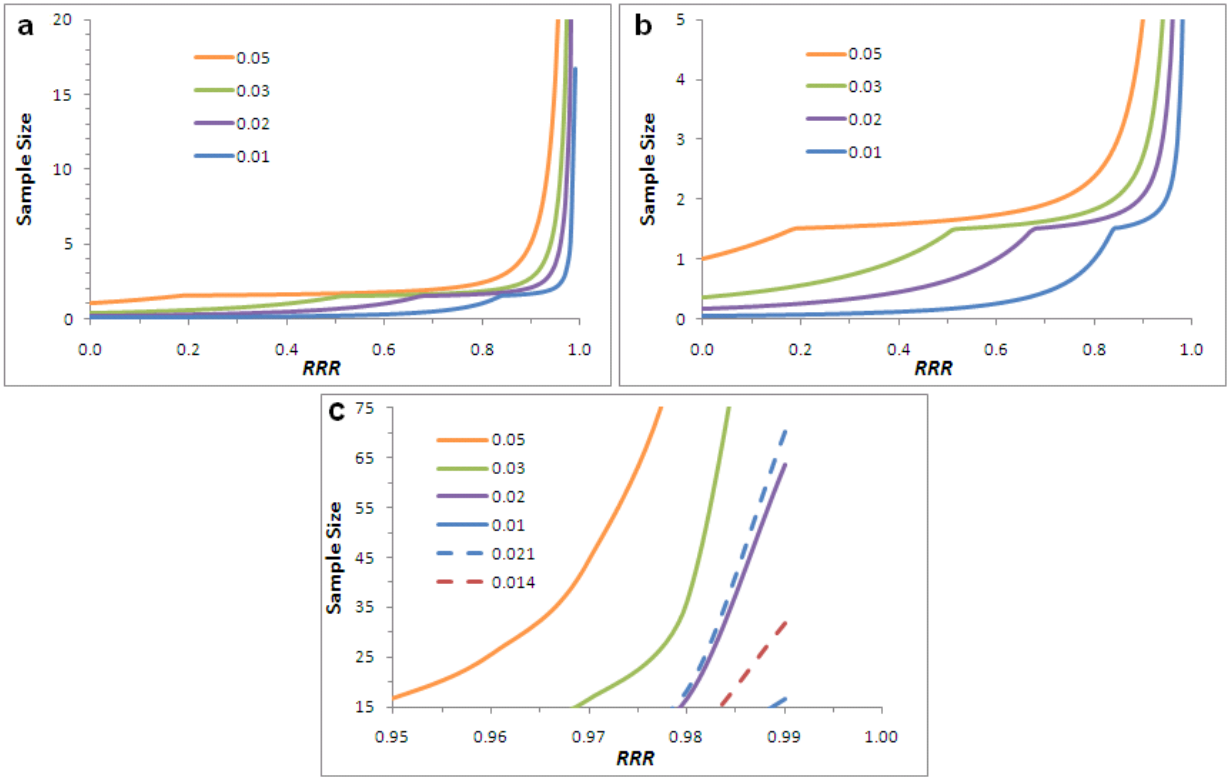

Figure 27. Effect of standard deviation on sample size
Curves for estimations of the *RRR* are plotted against sample size for 80% power and an alpha value of 0.05. In (a) and (b) curves are shown for the cases in which the standard deviations of estimates for *RRR* are 0.05, 0.03, 0.02, and 0.01. In (a) the general shape of the plot is shown. Enlarged images of the relationship between curves for (b) sample sizes less than 5 and (c) sample sizes larger than 15 are shown. Also in (c) standard deviations from the virtual clinical trial were added (dashed lines). Data for standard deviations of 0.01. 0.014, and 0.02 were not available for sample sizes less than 16, 32, and 64, respectively.

In Figures 27a and 27b, we see that with smaller standard deviations, fewer samples are needed to achieve 80% power and a 5% alpha value. In Figure 27c we see an application of this concept within the frame work of our virtual clinical trial because curves which represent the experimental standard deviation from our clinical trial are shown. Additionally in Figure 27c, the curves for standard deviation are shown within the window of sample size values that is not limited by a lack of equipoise or descriptive statistics (Figure 27c). From Figure 27c, we see that there can be a very subtle affect of standard deviation on sample size and that it is most pronounced at values of *RRR* which approach unity. More specifically in a comparison of the curves for standard deviation values: 0.020 vs. 0.021, there is a maximum difference of 6 in sample size at *RRR* = 0.99.

Figure 28 plots the effect of differing alpha values (α) on estimations for sample size, i.e., each α corresponded to differing probabilities of Type I errors which impacted the required sample size.[4]

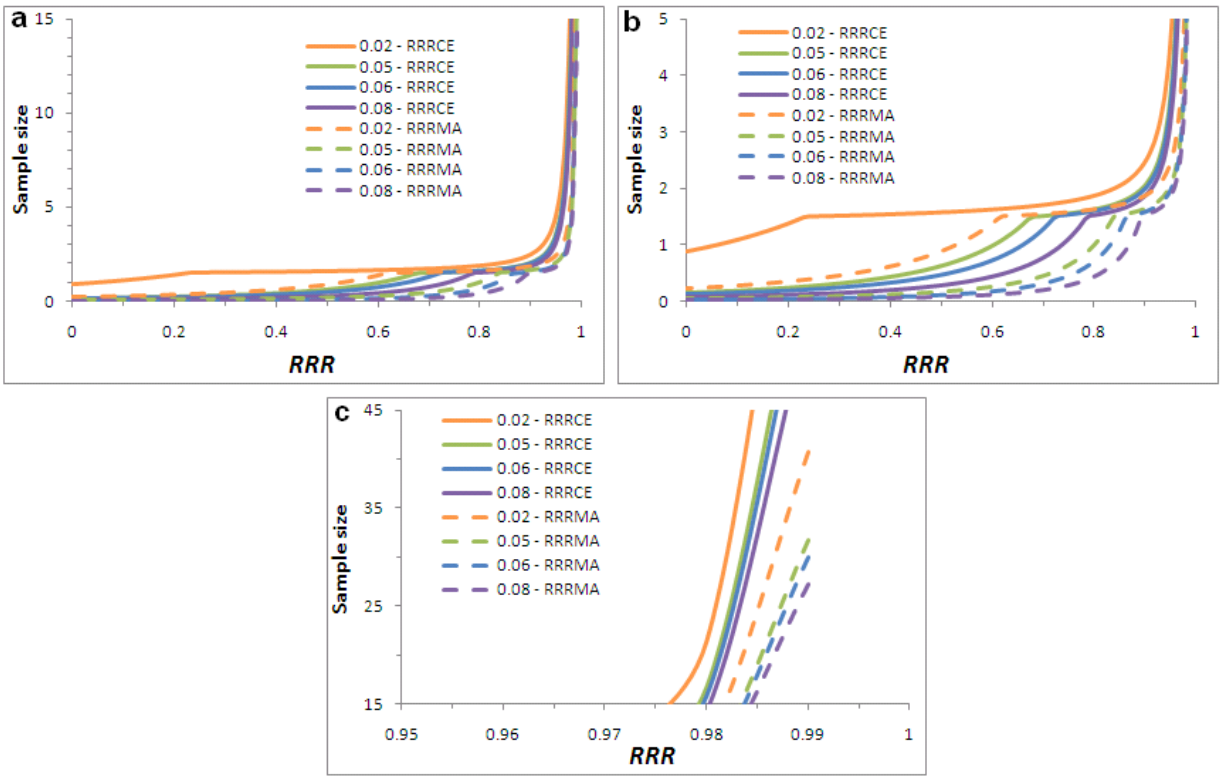

Figure 28. Effect of alpha on sample size
Curves for varying alpha values are shown in solid lines for RRR_{CE} vs. sample size and in dashed lines for RRR_{MA} vs. sample size. Input parameters were given in Table 22. In (a) the general shape of the plot is shown. Enlarged images are shown for (b) sample sizes less than 5 and (c) sample sizes larger than 5.

In Figure 28, we see that larger alpha values correspond with smaller sample sizes. We also see in Figures 28a and 28b that at sample sizes less than 5 there is a potential for overlap in curves which are characterized by different input parameters, i.e., in the region of $RRR = 0.6$ to $RRR = 0.8$, the curve for an alpha value of 0.02 for the population characterized by input from RRR_{MA} crosses several alpha curves (alpha = 0.05, 0.06, and 0.08) belonging the population characterized by input from RRR_{CE}. However when the plot is limited to a range of RRR and sample size that satisfy the need for equipoise and descriptive statistics, as in Figure 28c, there

[4] Type I errors represented errors associated with the assumption that for our sample there was no difference in predicted risk between proton and photon CSI when, in reality, a larger population would show a difference in predicted risk between proton and photon CSI.

is no chance for overlap between curves. Thus, as expected, Figure 28 demonstrates that the potential for introducing a Type I error is reduced at higher sample sizes.

3.6 Uncertainty analysis

To estimate uncertainty in *RRR* when D_{CE} and D_{MA} are respectively used, we performed uncertainty analysis for the predominant factors which contributed to uncertainty in *RRR*. Our analysis of uncertainty in absorbed dose, the cell sterilization factor, and the neutron weighting factor, are presented in sections 3.5.1, 3.5.2, and 3.5.3, respectively. In section 3.5.4, we combine the data from sections 3.5.1 - 3.5.3 to estimate the corresponding uncertainty in *RRR* when D_{CE} and D_{MA} were respectively used. Finally in section 3.5.5, we estimated the impact of uncertainty in *RRR* on sample size.

3.6.1 Uncertainty in absorbed dose

In this section we present the results of our estimation of uncertainty in absorbed dose in the thyroid due to variations in patient-set up for proton vs. photon CSI. More specifically, we estimated the mean dose in the thyroid when it was shifted relative to its original position, i.e., its location at the time of treatment planning. Shifts on the order of 0.5 cm were made in the superior, inferior, anterior, posterior, right, and left directions. An illustration of the anterior and posterior shifts is shown below in Figure 29.

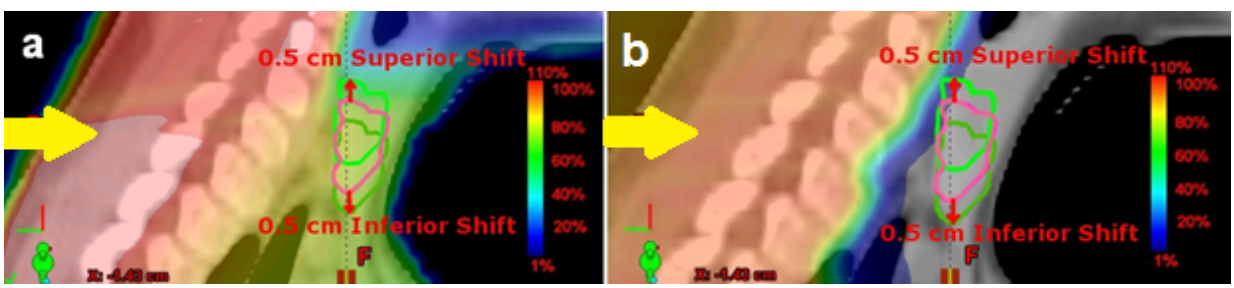

Figure 29. Thyroid location relative to the treatment field for proton and photon CSI
The original position of the thyroid (pink) is shown for a 10 year-old female. Also shown are its positions when shifted by 0.5 cm superiorly (light green) or inferiorly (dark green). The beam direction is indicated by a yellow arrow, and dose distributions are shown in color wash. In (a) the dose distribution is shown for photon CSI, and in (b) the dose distribution is shown for proton CSI. In both images dose is shown in units of cGy (RBE) where RBE is set to 1.1 for proton CSI and 1 for photon CSI.

From Figure 29, we see that shifts in thyroid location in the anterior – posterior direction may potentially impact the mean absorbed dose in the thyroid; however, we also see that shifts in specific directions may have more impact than others. Quantitative results for the mean absorbed dose in each shifted location are presented in Table 24 below.

Table 24. Absorbed dose and summary statistics for shifts in the thyroid location relative to the original treatment field

	Mean absorbed dose			
	Proton CSI		Photon CSI	
	D_{CE} (cGy RBE) RBE = 1.1	D_{MA} (cGy RBE) RBE = 1.1	D_{CE} (cGy RBE) RBE = 1	D_{MA} (cGy RBE) RBE = 1
Shift Direction				
Superior	0.120	0.120	9.720	17.078
Inferior	0.003	0.003	10.098	18.747
Anterior	0.001	0.001	9.833	17.544
Posterior	0.393	0.398	10.162	19.060
Left	0.027	0.027	9.984	18.209
Right	0.011	0.011	10.032	18.428
No Shift	*0.014*	*0.014*	*10.019*	*18.367*
Summary Statistics (shifted volumes)				
N	6	6	6	6
Maximum	0.393	0.398	10.162	19.060
Minimum	0.001	0.001	9.720	17.078
Interval	0.392	0.397	0.441	1.982
Mean	0.092	0.093	9.972	18.178
Standard Deviation of the Mean	0.063	0.064	0.068	0.304
Experimental Standard Deviation	0.154	0.156	0.166	0.745

From Table 24, we see that shifts in the posterior direction corresponded with the largest change in mean absorbed dose for both proton and photon CSI. More specifically, the estimated interval in absorbed dose values due to 0.5 cm shifts in the thyroid location was 39.3 cGy (RBE) and 39.8 cGy (RBE) for proton CSI when D_{CE} and D_{MA} are respectively applied, and for photon CSI the estimated intervals are 44.1 cGy and 198.2 cGy when D_{CE} and D_{MA} were respectively applied. These intervals represent to our estimated uncertainty in absorbed dose due to variations in patient set-up, or our value for σ_{D_x} in Equation 2.25. As a result, when we took a ratio of the our values of σ_{D_x} and the mean absorbed dose in the case of no shifts, 1.4 cGy (RBE) for proton CSI and 1001.9 cGy or 1836.7 cGy for photon CSI when D_{CE}

or D_{MA} were respectively applied, our values of $\frac{\sigma_{D_X}}{D_X}$ in Equation 2.25 were 28.0 or 28.4 for proton CSI when X = CE or MA and 0.044 or 0.018 for photon CSI when X = CE or MA. These findings support our earlier observations (see Figure15) which demonstrated that proton CSI is highly sensitive to fluctuations in thyroid dose.

3.6.2 Uncertainty in the cell sterilization factor

We estimated uncertainty in the cell sterilization factor, σ_C, by determining its interval when values for the thyroid-specific cell sterilization factor, α (Equation 2.27) are varied. To visualize the effect of variations in α on C, we plotted a range of absorbed dose values, D (Gy), against the corresponding cancer equivalent dose, D_{CE} (Gy). This was done for the minimum value for α (Equation 2.26), α_{MIN}; the value for α that was used in this study, α_*; and the maximum value for α (Equation 2.26), α_{MAX}; i.e., we plotted D (Gy) vs. D (Gy)* $C(D,\alpha)$ (Figure 30).

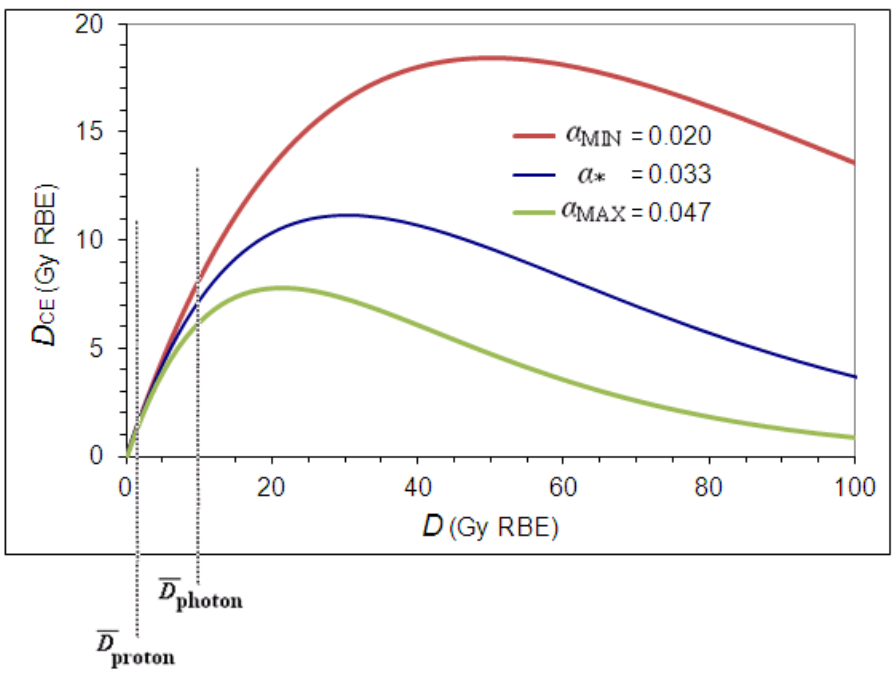

Figure 30. D_{CE} with various correction factors
Absorbed dose, D (Gy), was plotted against a several curves for cancer equivalent dose, D_{CE} (Gy). The red curve represents the case where the minimum value for α was applied, i.e., $C(D, \alpha = 0.020)$. The blue curve represents the case where the value for α that was used in this study was applied, i.e., $C(D, \alpha = 0.033)$. The green curve represents the case where the maximum value for α was applied, i.e., $C(D, \alpha = 0.047)$. Inflection points for the curves occur at 50 Gy, 30 Gy, and 21 Gy, respectively. The dashed lines show the locations of the study's mean values of absorbed dose in the thyroid from proton CSI (Table 7) and photon CSI (Table 8).

In Figure 30, we found, that for absorbed dose values which are less than 7.5 Gy, there was little sensitivity to variation in α, but for absorbed dose greater than 10 Gy there is greater sensitivity. This provides a visual demonstration of what previous sections revealed: the larger values of absorbed dose from photon CSI are more sensitive to the cell sterilization factor than the smaller values from proton CSI. We also found that the curves for D_{CE} had inflection points at 50 Gy, 30 Gy, and 21 Gy for α_{MIN}, α_*, and α_{MAX}, respectively. The location of our inflection point for α_* agrees with the results of Sigurdson et al (2005) in which the inflection point for observing cell sterilization in the thyroid was observed at 30 Gy. This agreement supported our use of 0.033 for α in the study; however to determine the potential uncertainty in our estimates, we used the minimum and maximum values of α to determine the maximum interval for C (Equation 2.29). Our results for proton and photon CSI are found in Table 25 below.

Table 25. Uncertainty estimates for cell sterilization

In the first three columns values are reported for mean absorbed dose from the TPS, $\overline{D}$, stray dose from neutrons, H_{stray}, and the total effective dose, H_{total}, for a 10 year-old female (patient # 14) receiving 23.4 Gy (RBE) from proton vs. photon CSI where the RBE values are 1.1 and 1, respectively. In the next three columns, values are listed for the thyroid-specific cell sterilization factor, α, i.e. the minimum value for α, the value for α that was used in this study (α_*), and the maximum value for α; the corresponding correction factors, C; and the corresponding values of cancer equivalent dose, H_{CE}. Finally in the last two columns, we present the interval for C, σ_C, and $\frac{\sigma_C}{C_*}$, where C_* is the value of C when α_* is used.

	$\overline{D}$ (Gy RBE)	H_{stray} (Sv)	H_{total} (Sv)	α (1/Sv)		C	H_{CE} (Sv)	σ_C	$\frac{\sigma_C}{C_*}$
Proton CSI	0.014	0.429	0.443	MIN	0.020	0.991	0.439	0.012	0.012
				*	0.033	0.985	0.436		
				MAX	0.047	0.979	0.434		
Photon CSI	18.367	n/a	18.367	MIN	0.020	0.693	12.721	0.271	0.496
				*	0.033	0.545	10.019		
				MAX	0.047	0.422	7.747		

3.6.3 Uncertainty in the mean radiation weighing factor for neutrons

We estimated uncertainty in $\overline{w_R}$ by scaling our values of $\overline{w_R}$ by 0.5 and 2 and applying those values to our estimates for RRR_X where X = MA or CE. We then used these the scaled values of RRR_X to determine a representative bracket of uncertainty for $\overline{w_R}$. Results for this process are shown in Table 26 below.

Table 26. Uncertainty in the mean radiation weighting factor for neutrons

$\overline{w_R}$ scaling factor	RRR_{CE}	RRR_{MA}	$\sigma_{\overline{w_R}}(RRR_{CE}) = \lvert \overline{w}_{R(2)} - \overline{w}_{R(0.5)} \rvert$	$\sigma_{\overline{w_R}}(RRR_{MA}) = \lvert \overline{w}_{R(2)} - \overline{w}_{R(0.5)} \rvert$	$\frac{\sigma_{\overline{w_R}}}{\overline{w_R}}(RRR_{CE})$	$\frac{\sigma_{\overline{w_R}}}{\overline{w_R}}(RRR_{MA})$
1	0.085	0.062				
0.5	0.056	0.041	0.086	0.065	0.011	0.008
2	0.142	0.105				

Table 26 reveals that scaling $\overline{w_R}$ did have an impact on our estimates for RRR_{MA} and RRR_{CE}, and as a result, our estimates for uncertainty in $\overline{w_R}$ differed accordingly. More specifically uncertainty in $\overline{w_R}$ was 24% less when D_{MA} was applied to our estimates for *RRR* than or when the D_{CE} was applied. Nonetheless, when these respective values of uncertainty were divided by the mean value for $\overline{w_R}$ across all treatment fields (7.93 from Table 5), the distinction between use of D_{MA} vs. D_{CE} was reduced. In particular, the relative uncertainty in $\overline{w_R}$ was approximately 0.01, regardless of its method of estimation.

3.6.4 Uncertainty in *RRR* when D_{CE} vs. D_{MA} are used

Once the results from the individual uncertainty analyses, i.e., the uncertainty analyses for absorbed dose, cell sterilization, and the neutron weighting factor, were combined (Equation 2.25), we determined their combined uncertainty through our estimates of *RRR* (Equation 2.24). Results of this analysis are listed in Table 27 below.

Table 27. Estimates of relative uncertainty in *RRR* when D_{CE} vs. D_{MA} are used

Values are listed for the terms in Equations 2.24 and 2.25 which are used to estimate relative uncertainty in RRR_{MA} and RRR_{CE}. Entries for $\frac{\sigma_C}{C^*}$ are blank when D_{MA} was used because a cell sterilization factor was not applied in that case. Also, the entry for $\frac{\sigma_{\overline{w_R}}}{\overline{w_R}}$ is blank for photon CSI because the mean radiation weighting factor for neutrons did not apply.

	$\frac{\sigma_{D_{CE}}}{D_{CE}}$	$\frac{\sigma_{C}}{C_{*}}$	$\frac{\sigma_{\overline{w}_R}}{\overline{w}_R}$	$\frac{\sigma_{ERR_{CE}}}{ERR_{CE}}$	$\frac{\sigma_{RRR_{CE}}}{RRR_{CE}}$
Proton CSI	27.999	0.012	0.011	27.999	27.986
Photon CSI	0.044	0.496	-	0.498	

	$\frac{\sigma_{D_{MA}}}{D_{MA}}$	$\frac{\sigma_{C}}{C_{*}}$	$\frac{\sigma_{\overline{w}_R}}{\overline{w}_R}$	$\frac{\sigma_{ERR_{MA}}}{ERR_{MA}}$	$\frac{\sigma_{RRR_{MA}}}{RRR_{MA}}$
Proton CSI	28.357	-	0.008	28.357	28.243
Photon CSI	0.108	-	-	0.108	

Table 27 reveals that relative uncertainty in *RRR* is approximately the same regardless of when D_{CE} was applied or when D_{MA} was applied, i.e., the percent difference between $\frac{\sigma_{RRR_{CE}}}{RRR_{CE}}$ and $\frac{\sigma_{RRR_{MA}}}{RRR_{MA}}$ is less than 1%. This stems from the finding that uncertainty in *RRR* is largely dominated by relative uncertainty in absorbed dose, namely the uncertainty in absorbed dose from proton CSI. Moreover, the disproportionately large values of $\frac{\sigma_{D_X}}{D_x}$ from proton CSI vs. photon CSI indicate that the predicted risk of SC in the thyroid from proton CSI is notably more patient-specific than that of photon CSI. Additionally, with large values of $\frac{\sigma_{D_X}}{D_x}$, uncertainty in predicted risk for adjacent organs is increased.

Chapter Four

4 Discussion

We used a virtual clinical trial to determine the impact of cell sterilization effects on estimations of the relative risk for a second cancer in the thyroid following proton vs. photon CSI. Our goal was to determine if use of D_{CE} would reduce the sample size needed to achieve 80% power. However, data from our virtual trial revealed that achieving 80% statistical power (with an alpha value of 0.05) required a maximum sample size of 1 patient. Because this finding occurred regardless of whether D_{CE} or D_{MA} was used, it highlighted several important trends in our data. In particular, it confirmed that the predicted risk of second cancer in the thyroid is greater after CSI with photons vs. protons. Accordingly, because the predicted risk increased with the equivalent dose in the thyroid, use of D_{CE} reduced total equivalent dose for photon CSI by approximately 44%, and while the total equivalent dose in proton CSI was generally insensitive to D_{CE},we did find that in one patient the use of D_{CE} reduced total equivalent dose for proton CSI by as much as 22%. This revealed the potential importance of patient-specific analysis for proton CSI and hints that the greatest factor contributing to uncertainty in our estimates for *RRR* was uncertainty in the absorbed dose from proton CSI. Additionally, this work we analyzed risk in a population of patients (N = 18) and studied the influence of power, experimental standard deviation, and alpha values on required sample size. Also, this work highlighted the potential impact of cell sterilization factors on future clinical trial design. More specifically, it revealed that sample size can increase with the use of a cell sterilization factor and that failure to account for this can lead to trial designs that will have larger than expected probabilities of Type II and Type I errors. This is important because it can have clinical repercussions with respect to the additional resources and time that would be required to acquire additional patients. This is particularly relevant in the case of rare cancers in small populations.

4.1 Previous literature

When compared to previous works, this study is unique in its application of dosimetic metrics. Specifically, this study is the first to apply a cell sterilization factor to a population of patients in which estimates for second cancer risk are based on patient-specific estimates of stray and therapeutic dose. As a result, we are unable to make direct comparisons of our

estimates for second cancer risk with previous works, but we were able to compare specific aspects of our results with those previous works, specifically those of Newhauser et al. (2009) and Brodin et al. (2011).

In their work, Newhauser et al. considered the risk of second cancer of the thyroid from therapeutic and stray radiation in the case of pediatric CSI as we did. In particular, Newhauser et al. estimated lifetime attributable risk following the recommendations of ICRP Publication 60 (1990) and they reported that the predicted lifetime risk of developing a second cancer of the thyroid at 0.35%. In this work, we estimated *RR* following the recommendations in the report of the BEIR VII Committee (NRC 2006). Our estimates of stray dose, 437 mSv, are in good agreement with those of Newhauser et al, 443 mSv. Newhauser et al. (2009) used the therapeutic absorbed dose from Miralbell et al. (2002), and we used therapeutic doses which were calculated specifically for each patient in our study. As a result of the differing strategies for determining therapeutic dose, the study by Newhauser et al. (2009), which took therapeutic dose from Miralbell et al., reported that therapeutic dose was the greatest contributing factor to the risk of second cancers in the thyroid, and our study reports that stray dose is the greatest contributing factor. Nonetheless, the fact that our findings for dose from stray radiation agree within 1.3%, provides some external validation of our results.

Our study was also similar to that of Brodin et al. (2011) in that they, too, considered the contribution of stray and therapeutic dose in their estimates of second cancer risk in the thyroid. Additionally, they considered a cell sterilization factor for dose, as we did. However, they chose to follow a different dose-response model than we did to account for cell sterilization. Specifically, they chose a linear plateau-model while we selected a linear exponential model. Additionally, their study focused on different methods of radiation delivery than ours did. They studied the effects of rotational intensity modulated radiation therapy with photons (rotational IMRT), conventional 3D radiation therapy with photons (3D CRT), and intensity modulated radiation therapy with protons (IMPT). Because they took values of dose from stray radiation in IMPT from Newhauser et al. (2009), our results for dose from stray radiation agreed with theirs. However, they used a different model for predicting risk than we did. Specifically they estimated excess absolute risk following the method of Schneider et al. (2008) which takes into account attainted age, sex, and age at exposure while we estimated excess relative risk using BEIR VII (NRC 2006) which takes sex, age at exposure and organ specific factors into account. Thus, the main difference in these estimates is that Brodin et al.

(2011) used data from the Life Span Study (UNSCEAR 2000) and a weighted mean approach to estimate thyroid specific results, and at the dose levels of our project, any resulting differences in the respective estimations of risk were considered negligible.

In an attempt to compare one additional aspect of our study to previous works, namely the effect of using a cell sterilization factor we surveyed several works which address this topic (Brodin et al. 2011, Bhatti et al. 2010, Schneider et al. 2005, Ronckers et al. 2005, and Sigurdson et al. 2005). This revealed no overall consensus regarding which model is most the appropriate model to use for pediatric CSI patients. In particular, Brodin et al. (2011) used data from Schneider et al. (2008) to compare the linear, linear-exponential, and plateau dose-response models. Their finding was that the plateau dose-response model provided the best fit. Because, they did not provide all the parameters used in their analysis, we did not make a direct comparison with the model used in this work. For our remaining comparisons, we plotted dose response curves from Sigurdson et al. (2005), Ronckers et al. (2005), Schneider et al. (2006), and Bhatia et al. (2010) in Figure 31 below.

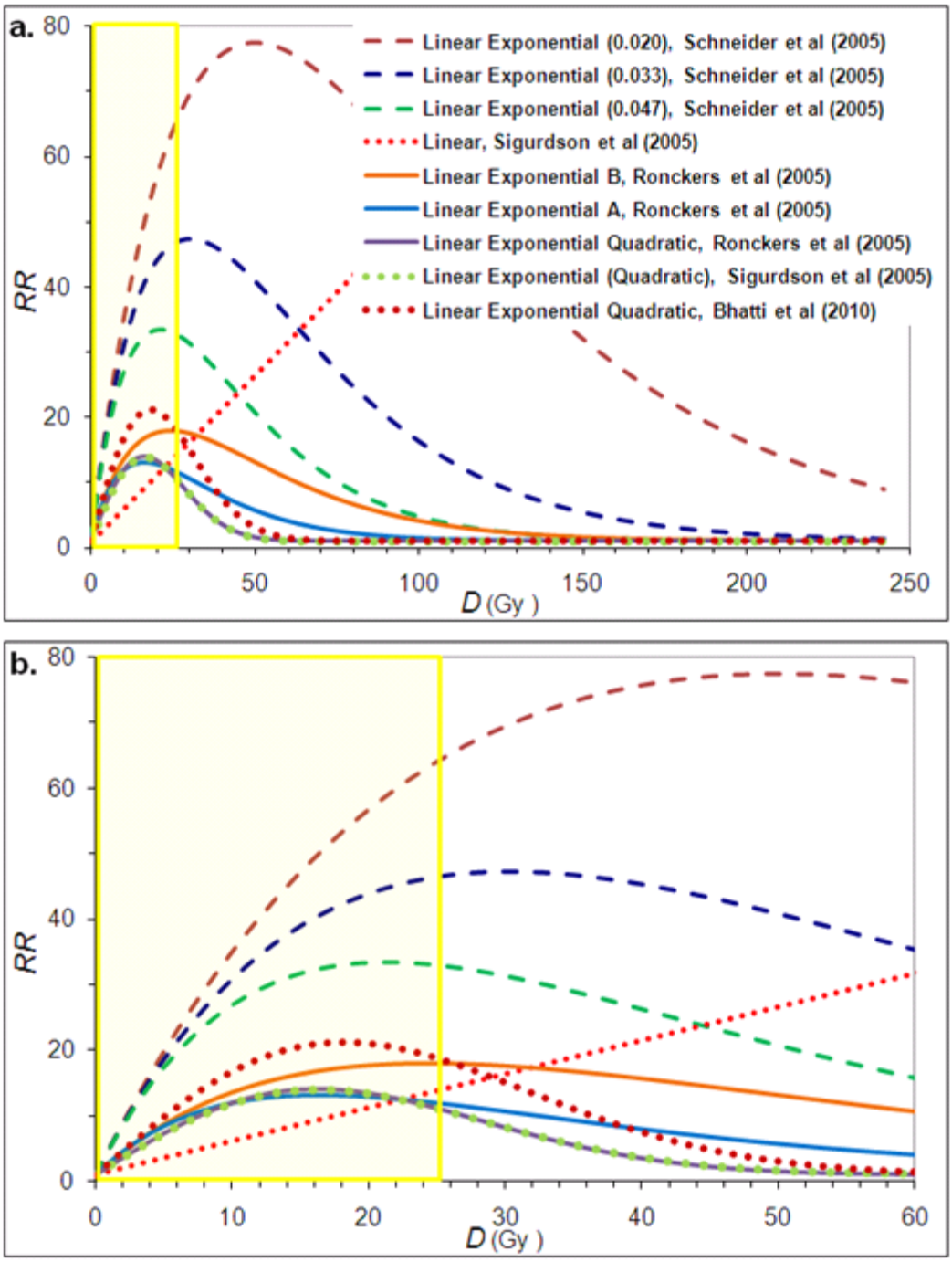

Figure 31. Comparison of dose-response models

Relative risk is plotted against absorbed dose for several dose-response models in the literature. The model used in this study is the linear exponential from Schneider et al. (2005) with a thyroid-specific cell sterilization parameter of 0.033 (dashed navy blue curve). In (a) the dose response models are shown for high and low dose regions. In (b) the dose response models are shown to the maximum therapeutic dose, including boost, that would be delivered for a population of pediatric patients being treated with CSI. The yellow boxes in both images represent the area which corresponds to the interval of therapeutic dose in this study, 0 Gy to 25 Gy.

The dose-response model that we used to estimate D_{CE} was based on the work of Schneider et al. (2006) in which they determined that a linear exponential model provided the best fit to their data. Also in their work, they provided a 95% confidence interval for their exponential term. We graphed these and used them as a lower and upper bound estimate for the linear exponential used in our work (Figure 31). Additional works which based their analysis on data from the Childhood Cancer Survivorship Study (CCSS) included that of Ronckers et al. (2005) which examined several curves. The best fit was achieved with a linear exponential quadratic model (Figure 31). In the work by Sigurdson et al. (2005), they report that a linear exponential curve affords the best fit (Figure 31), however, when graphed this curve overlies the linear exponential quadratic defined by Ronckers et al. (2005). In an update to the CCSS study by Bhatti et al. (2010), the dose-response curve from Sigurdson et al. (2005) was updated to a new linear exponential quadratic (Figure 31) which more closely resembles the a linear exponential curves from Schneider et al. (2005). In particular, it resembles the lower-bound estimate from Schneider et al. (2005). Thus, through comparison with these previous works, we found that use of the cell sterilization factor provided by Schneider et al. (2006) provides a comparatively conservative estimate of risk (Figure 31b). This is a particularly important finding when one considers that studies such as ours may be used in clinical decision making.

4.2 Major findings and implications

One of the major findings from this work is that for a sample of patients receiving CSI, D_{CE} made only a 2.5% difference in predicted estimates for *RRR*. However, on a patient-by-patient basis use of D_{CE} made as much as an 11% difference in estimates of *RRR* for an individual patient. Because the patient with an 11% difference, patient 1, was subjected to the same SOC and physician approval criteria as the other patients in the sample, this finding has two potential implications. The first implication is that predicted risk can be very patient specific, particularly in the case of proton CSI. Reasons for this include a potential for errors in range and patient set-up. This suggests that it may be more prudent to perform personalized risk estimates on a per-patient-basis, i.e., class estimates for risk, at least within our population could have resulted in an 11% error in risk. A second implication is that regions of higher dose may be more sensitive to the cell sterilization effect. In particular, there may be

implication with respect to regions that are subject to higher dose gradients or regions in which there is a higher potential for dose irregularities, i.e., in structures that move, such as the lung or heart.

An additional finding of this work was that as the value of $\overline{RRR}$ approaches one, i.e., when $\overline{RRR} > 0.8$, calculations of sample size become increasingly important and sensitive to factors such as the experimental standard deviations of the respective populations. In other words, when there is a less than 20% difference in the estimates for risk from proton vs. photon CSI, a larger population of patients is needed in order to confidently determine if a difference in risk exists. While this is an expected result, we were surprised at the magnitude of the difference in sample size in relation to shifts in standard deviation. More specifically, we found a 6 patient difference for a change in experimental standard deviation of 0.001 (Figure 27c) which can have profound implications with respect to resources and the required time for studies in which eligible patient pool is limited, as is the case in pediatric populations.

Finally, we found that application of a cell sterilization factor, D_{CE}, increased the experimental standard deviation of our population. As a result, when D_{CE} was applied to our population, our estimate of $\overline{RRR}$ was more susceptible to Type I and Type II errors and required larger sample sizes (Figures 26, 28). Thus, while the exact numerical impact of a cell sterilization factor may differ with the specific dose-response model that is selected (Figure 31), our study shows that when a cell sterilization factor is taken into account, it can impact sample size. Thus, without careful consideration a higher probability of statistical error can be introduced (Schafer and Gilbert 2006). Moreover, when we consider the dose range of this study, the actual dose-response model becomes less important, because the models differ most in the high dose region. It is the low dose region which applies most to pediatric populations and accordingly has the greatest potential to effect clinical decision making.

4.3 Study strengths

This project is one of the first population-based virtual studies of comparative risk that departs from the traditional use of D_{MA}. To our knowledge it is the only study that explores the relationship between cell sterilization and statistical significance in an epidemiological study of small sample groups, e.g., groups which include pediatric cancers and patients treated with advanced radiotherapy. Additionally, no prior study has applied similar concepts to a

population in which both therapeutic and stray dose were determined on a patient-specific basis. Thus, the major strength of this work is that it revealed the potential impact of cell sterilization factors on sample size in small-scale radio-epidemiological studies. Additionally, it provides the first general analysis of what sample sizes are needed to meet the dual requirements of equipoise and descriptive statistics.

4.4 Study limitations

There were several limitations to our study. For clarity we have grouped them into two general categories: those related to the design of the virtual clinical trial and those which are model specific. With respect to the design of the virtual clinical trial, we have defined three potential limitations. In the first of these, we note that we used a mean radiation weighting factor for neutrons which was generalized for our patient population. Ideally the neutron weighting factor would be determined on a patient-specific basis; however, given the time constraints for this work and the fact that field energies varied little between patients, this was not a serious limitation. A second limitation, related to trial design, was that we considered only one organ. In particular, we performed our risk analysis on the thyroid which is small, of fairly constant density, and did not display large gradient of dose in our treatment plans. As a result, our study had little dosimetric variation which resulted in small standard deviations in dose i.e., the experimental standard deviation for H_{total} was less than 8.5% for proton CSI and less than 5.5% for photon CSI. While this limited the applicability of our risk analysis to other organs with greater tissue heterogeneity, it also provided a solid first step in our investigation of the effect of a factor for cell sterilization because the thyroid is one of the organs with the greatest known effect (Bhatti et al. 2011). Thus, by focusing on the thyroid we were able to limit the potential for competing variables in our study. Finally, we did not include results from intensity modulated photon therapy, IMRT. In this respect, use of IMRT may have introduced a CSI technique for photons with tissue sparing that was more comparable to that of CSI with passively scattered protons. This may have increased our values of *RRR* which may have impacted our findings regarding sample size when D_{CE} vs. D_{MA} were applied. However, because IMRT is not the SOC for CSI, we are confident that our use of the FIF technique for photon CSI may ultimately yield the most clinically relevant results.

With respect to the model specific limitations of our study, we have also defined three potential limiting factors, the first of which deals with our choice for the thyroid specific cell

sterilization factor alpha, α. In our work we used estimates for α which were determined by Schneider et al. (2006) with data from Hodgkin's patients. As noted in our uncertainty analysis (section 2.10.2), there are some conflicting reports in the literature regarding whether or not cell-sterilization occurs similarly for Hodgkin's patients vs. non-Hodgkin's patients (Sigurdson et al. 2005; Ronckers et al. 2006). Moreover the region of low doses, such as was used in our study, is the region in which the differences are greatest between the dose-response curves suggested by Sigurdson et al., Ronckers et al., and Schneider et al. (Figure 31). As a result, we agree that this could be a very notable limitation in our work. However, in light of our results for uncertainty in *RRR,* i.e., the two major contributors to relative uncertainty in *RRR* are uncertainty in absorbed dose for proton CSI, on the order of 2800%, vs. relative uncertainty in the cell sterilization factor, on the order of 50%. We remain confident that uncertainty in the thyroid-specific cell sterilization factor does not invalidate our results and, indeed, was the impetus for much of the uncertainty analysis in this work. In a second limitation stemming from the dose-response model, we note that the model used in this work, the BEIR VII model for *ERR/H* does not take dose fractionation into account. Moreover, the BEIR VII data is based on A-bomb survivor data and because risk is not linear with dose, linear adaptations of the BEIR VII data to our region of low dose, is not optimal. However, until more data is available, use of the BEIR VII model represents a reasonable choice. Finally, a third limitation, which is associated with the dose-response model, is that while radiation therapy (RT) accounts for roughly 8% of second cancers in adults other factors such as lifestyle, environmental factors and genetic susceptibility may account for more than 90% of the remaining occurrences of second cancers (Berrington de Gonzalez et al. 2011). This is a limitation because current models do not describe how RT interacts with the aforementioned other factors. However, this is an active area of research (Morton and Chanock 2011), so when results become available, the results from our study can be adjusted in a straightforward manner. In this respect, this limitation may be overcome in a future work.

4.5 Future work

Given the biological and statistical complexity of this subject, future studies are needed to determine the effect of D_{CE} vs. D_{MA} in other organs. This includes organs with differing tissue heterogeneity, organs which move, and organs which receive partial in-field irradiation. Additionally, in an effort to address the uncertainty associated with the selection of appropriate

dose-response models, efforts to retrospectively compare the results of this study with actual patient outcomes would be greatly beneficial. Such work may be particularly challenging in the case of proton CSI given the relative rarity of pediatric cancer and the historically limited availability of proton therapy. However, this constraint is fading. There are approximately 28 proton therapy centers in operation world-wide (Particle Therapy Co-Operative Group, 2010); thus, there is a growing potential for creating a database of treatment outcomes which would facilitate retrospective comparisons.

Chapter Five

5 Conclusions

This work was motivated by the unique challenges associated with epidemiological studies of late effects from contemporary radiation therapy. First, the risk of developing a radiation-related second cancer increases with the time since exposure, and the latency for solid tumors is 5 years or more. Thus, by the time a study can be completed, the contemporary treatment is no longer contemporary. Second, while second cancers account for 18% of cancer diagnoses (Howlander et al. 2011), it is still difficult to accrue sufficient numbers of patients for a population based outcome study. This challenge is compounded when rare sub-groups, such as patients who received pediatric radiation are considered. Third, the traditional approach of estimating second cancers from radiation organ dose considers only mean organ dose. This approach overlooks the effect of cell sterilization which at high doses makes second cancer formation unlikely. In our work, we used a prospective method to compare and predict the risk of radiation related second cancers for different contemporary radiation therapy techniques, i.e., a virtual clinical trial. We tested this method for a recurring challenge in radio-epidemiological studies, small sample sizes. As a result, we considered (1) a rare pediatric cancer requiring craniospinal irradiation and (2) a rare but well understood second cancer, thyroid cancer. We considered the effect cell sterilization on risk prediction and, consequently, sample size in a micro-clinical trial. As a result, we found that in a trial of radiation therapy techniques, when the difference in the predicted risk of second cancers is less than 10%, inclusion of cell sterilization in the dose model results in a need for larger sample sizes if a study is designed to achieve 80% statistical power.

This work demonstrated that it is possible to prospectively evaluate contemporary radiation therapy techniques for radiation-related second cancers using a virtual clinical trial approach. This work provided specific results which indicate that these types of studies can achieve statistical significance, even with small sample sizes. Specifically, the required sample size (N) for our specific study was unity, regardless of whether cell sterilization was included in the dose model. Reasons for this include relatively large differences in absorbed dose in the thyroid from proton vs. photon CSI ($\overline{RRR}_{\mathrm{MA}}$ = 0.052, $\overline{RRR}_{\mathrm{CE}}$ = 0.087) and small variations in absorbed dose across the patient sample due to small variations in patient anatomy relative to the treatment field ($\sigma_{RRR_{\mathrm{MA}}}$= 0.014, $\sigma_{RRR_{\mathrm{CE}}}$= 0.021). Additionally, this work found that for

proton CSI there is potential for increased variation in patient specific risk estimates, up to a factor of 27, when error in patient set-up is considered. Thus, in that case or others, when there is increased variation in patient specific risk estimates (σ_{RRR}) or when the difference in the predicted risk of second cancers is less than 10%, using alternate concepts for dose, i.e., use of cell sterilization rather than mean organ dose can greatly impact estimations for sample size. Furthermore, this study revealed that accurate estimation sample size may be vitally important to avoid statistical error. However, using alternate concepts for dose requires careful study design because the outcome is very sensitive to the factors used in the cell sterilization model. Ignoring or selecting wrong factors can result in under-sampling, which can have profound implications on the quality or even the validity of study results. Thus, we conclude that patient specific studies are the most appropriate types of studies for clinical decision making as statistical averaging may overestimate the errors for some individuals. This is especially true for proton therapy where range and set-up error dramatically impact organ dose. Moreover, it is especially important that the details of the treatment plans and their uncertainties be given careful consideration when comparing different treatment techniques because the details can change the outcome of the study.

Bibliography

Agosteo, S., C. Birattari, M. Caravaggio, M. Silari, G. Tosi (1998). "Secondary neutron and photon dose in proton therapy." Radiother Oncol **48**(3): 293-305.

Berrington de Gonzalez, A., R. E. Curtis, S. F. Kry, E. Gilbert, S. Lamart, C. D. Berg, M. Stovall, E. Ron, (2011). "Proportion of second cancers attributable to radiotherapy treatment in adults: a cohort study in the US SEER cancer registries." Lancet Oncol **12**(4): 353-360.

Bhatti, P., L. H. Veiga, C. M. Ronckers, A. J. Sigurdson, M. Stovall, S. A. Smith, R. Weathers, W. Leisenring, A. C. Mertens, S. Hammond, D. L. Friedman, J. P. Neglia, A. T. Meadows, S. S. Donaldson, C. A. Sklar, L. L. Robison, P. D. Inskip, (2010). "Risk of second primary thyroid cancer after radiotherapy for a childhood cancer in a large cohort study: an update from the childhood cancer survivor study." Radiat Res **174**(6): 741-752.

Boice, J. D. (1981). "Cancer following medical irradiation." Cancer **47**(5 Suppl): 1081-1090.

Brant, R. (2011). "Web-based Sample Size / Power Calculations." University of Calgary: Accessed June 2011.Brodin, N. P., P. M. Rosenschold, et al. (2011). "Radiobiological risk estimates of adverse events and secondary cancer for proton and photon radiation therapy of pediatric medulloblastoma." Acta Oncol **50**(6): 806-816.

N. P. Brodin, P. M. Rosenschold, M. C. Aznar, A. Kiil-Berthelsen, I. R. Vogelius, P. Nilsson, B. Lannering, T. Bjork-Eriksson, (2011)."Radiobiological risk estimates of adverse events and secondary cancer for proton and photon radiation therapy of pediatric medulloblastoma." Acta Oncol **50** (6): 806-816.

Dattalo, P. (2009). "A review of software for sample size determination." Eval Health Prof **32**(3): 229-248.

Dupont, W. D. and W. D. Plummer, Jr. (1990). "Power and sample size calculations. A review and computer program." Control Clin Trials **11**(2): 116-128.

Fontenot, J., P. Taddei, Y. Zheng, D. Mirkovic, T. Jordan, W. Newhauser, (2008). "Equivalent dose and effective dose from stray radiation during passively scattered proton radiotherapy for prostate cancer." Phys Med Biol **53**(6): 1677-1688.

Friedman, D. L., J. Whitton, W. Leisenring, A. C. Mertens, S. Hammond, M. Stovall, S. S. Donaldson, A. T. Meadows, L. L. Robison, J. P. Neglia, (2010). "Subsequent neoplasms in 5-year survivors of childhood cancer: the Childhood Cancer Survivor Study." J Natl Cancer Inst **102**(14): 1083-1095.

Giebeler, A. (2009). "CSI plan edit procedure." Internal Report MDACC-10-0063.
Hall, E. J. (2006). "Intensity-modulated radiation therapy, protons, and the risk of second cancers." Int J Radiat Oncol Biol Phys **65**(1): 1-7.

Hall, E. J. (2009). "Is there a place for quantitative risk assessment?" J Radiol Prot **29**(2A): A171-184.

Howell, R. M., A. Giebeler, W. Koontz-Raisig, A. Mahajan, C. J. Etzel, A. M. D'Amelio, S. Randeniya, W. D. Newhauser (in preparation). "Comparisson of passively scattered proton and photon field-in-field craniospinal irradiation for medulloblastoma."

Howell, R. M., N. E. Hertel, Z. Wang, J. Hutchinson, G. D. Fullerton, (2006). "Calculation of effective dose from measurements of secondary neutron spectra and scattered photon dose from dynamic MLC IMRT for 6 MV, 15 MV, and 18 MV beam energies." Med Phys **33**(2): 360-368.

Howell, R. M., S. B. Scarboro, S. F. Kry, D. Z. Yaldo, (2010a). "Accuracy of out-of-field dose calculations by a commercial treatment planning system." Phys Med Biol **55**(23): 6999-7008.

Howell, R. M., S. B. Scarboro, P. J. Taddei, S. Krishnan, S. F. Kry, and W. D. Newhauser, (2010b). "Methodology for determining doses to in-field, out-of-field and partially in-field organs for late effects studies in photon radiotherapy." Phys Med Biol **55**(23): 7009-7023.

Howlader, N., A. M. Noone, M. Krapcho, N Neyman, R. Aminou, W. Waldron, S.F. Altekruse, C.L. Kosary, J. Ruhl, Z Tatalovich, H. Cho, A. Mariotto, M.P. Eisner, D.R. Lewis, H.S. Chen, E.J. Feuer, K.A. Cronin, and B.K. Edwards, (2011). SEER Cancer Statistics Review, 1975-2008. Bethesda, MD, National Cancer Institute**:** http://seer.cancer.gov/csr/1975_2008/, based on November 2010 SEER data submission, posted to the SEER web site, 2011.

ICRP (1990). Recommendations of the International Commission on Radiological Protection: ICRP Publication 60. Ann ICRP. **21:** 1-3.

ICRP (2003). Relative biological effectiveness (RBE), quality factor (Q), and radiation weighting factor (wR): ICRP Publication 92. Ann ICRP. **33:** 4.

ICRP (2007). The 2007 recommendations of the international commission on radiological protection: ICRP Publication 103. Ann ICRP. **37:** 2-4.

ICRU (2007). Prescribing, Recording, and Reporting Proton-Beam Therapy, International Commission on Radiation Units and Measurements: Report 78. Oxford: Oxford University Press.

Inskip, P. D. and R. E. Curtis (2007). "New malignancies following childhood cancer in the United States, 1973-2002." Int J Cancer **121**(10): 2233-2240.

Kry, S. F. and M. Salehpour (2006). "In response to Dr. Schneider, Refers to article: Calculated risk of fatal secondary malignancies from intensity-modulated radiotherapy: In regard to Kry et al. (Int J Radiat Oncol Biol Phys 2005;62:1195-1203) " Int J Radiat Oncol Biol Phys **62**: 1290-1291.

Kry, S. F., M. Salehpour, D. S. Followill, M. Stovall, D. A. Kuban, R. A. White, Rosen, II, (2005). "Out-of-field photon and neutron dose equivalents from step-and-shoot intensity-modulated radiation therapy." Int J Radiat Oncol Biol Phys **62**(4): 1204-1216.

Linet, M. S., L. A. Ries, M. A. Smith, R. E. Tarone, S. S. Devesa, (1999). "Cancer surveillance series: recent trends in childhood cancer incidence and mortality in the United States." J Natl Cancer Inst **91**(12): 1051-1058.

Lunsford, T. K. and B. R. Lunsford (1995). "The Research Sample, Part I: Sampling." J Prosthetics Orthotics **7**(3): 105-112.

Mertens, A. C., Y. Yasui, J. P. Neglia, J. D. Potter, M. E. Nesbit, Jr., K. Ruccione, W. A. Smithson, L. L. Robison, (2001). "Late mortality experience in five-year survivors of childhood and adolescent cancer: the Childhood Cancer Survivor Study." J Clin Oncol **19**(13): 3163-3172.

Miralbell, R., A. Lomax, L. Cella, U. Schneider (2002). "Potential reduction of the incidence of radiation-induced second cancers by using proton beams in the treatment of pediatric tumors." Int J Radiat Oncol Biol Phys **54**(3): 824-829.

Morton, L. M. and S. J. Chanock (2011). "A step toward slaying the hydra of second cancers." Nat Med **17**(8): 924-925.

Neglia, J. P., D. L. Friedman, et al. (2001). "Second malignant neoplasms in five-year survivors of childhood cancer: childhood cancer survivor study." J Natl Cancer Inst **93**(8): 618-629.

Newhauser, W., Y. Zheng, P. Taddei, D. Mirkovic, J. Fontenot, A. Giebeler, R. Zhang, U. Titt, R. Mohan, (2008). "Monte Carlo proton radiation therapy planning calculations." Transactions of the American Nuclear Society **99**: 63.

Newhauser, W. D., J. D. Fontenot, A. Mahajan, D. Kornguth, M. Stovall, Y. Zheng, P. J. Taddei, D. Mirkovic, R. Mohan, J. D. Cox, S. Woo, (2009a). "The risk of developing a second cancer after receiving craniospinal proton irradiation." Phys Med Biol **54**(8): 2277-2291.

Newhauser, W. D., J. D. Fontenot, P. J. Taddei, D. Mirkovic, A. Giebeler, R. Zhang, A. Mahajan, D. Kornguth, M. Stovall, P. Yepes, S. Woo, R. Mohan, (2009b). "Contemporary Proton Therapy Systems Adequately Protect Patients from Exposure to Stray Radiation." AIP Conf Proc **1099**(1): 450-455.

NIH (2003). (National Institutes of Health) Report of the NCI-CDC Working Group to Revise the 1985 NIH Radioepidemiological Tables: NIH Publication 03-5387. Bethesda, MD.

NRC (2006). (National Council of the National Academies) Health Risks from Exposure to Low Levels of Ionizing Radiation: BEIR VII - Phase 2. Washington, DC.

Pelowitz, D. B. (ed.2005). MCNPXTM User's Manual, Version 2.5.0. Report LA-CP-05-0369. Los Alamos, New Mexico, Los Alamos National Laboratory.

Ries, L. A. G., D. Melbert, M. Krapcho, A. Mariotto, B.A. Miller, E.J. Feuer, L. Clegg, M.J. Horner, N. Howlader, M.P. Eisner, M. Reichman, B.K. Edwards, (2007). SEER Cancer Statistics Review, 1975-2004. Bethesda, MD, National Cancer Institute**:** http://seer.cancer.gov/csr/1975_2004/, based on November 2006 SEER data submission, posted to the SEER web site, 2007.

Ries, L. A. G., M. A. Smith, J.G. Gurney, M Linet, T. Tamra, J.L. Young, G.R Bunin, (1999). Cancer Incidence and Survival among Children and Adolescents: United States SEER Program 1975-1995, National Cancer Institute, SEER Program. NIH Pub. No. 99-4649. Bethesda, MD.

Ron, E., J. H. Lubin, R. E. Shore, K. Mabuchi, B. Modan, L. M. Pottern, A. B. Schneider, M. A. Tucker, J. D. Boice, Jr., (1995). "Thyroid cancer after exposure to external radiation: a pooled analysis of seven studies." Radiat Res **141**(3): 259-277.

Ronckers, C. M., A. J. Sigurdson, M. Stovall, S. A. Smith, A. C. Mertens, Y. Liu, S. Hammond, C. E. Land, J. P. Neglia, S. S. Donaldson, A. T. Meadows, C. A. Sklar, L. L. Robison, P. D. Inskip, (2006). "Thyroid cancer in childhood cancer survivors: a detailed evaluation of radiation dose response and its modifiers." Radiat Res **166**(4): 618-628.

Ruben, J. D., S. Davis, C. Evans, P. Jones, F. Gagliardi, M. Haynes, A. Hunter, (2008). "The effect of intensity-modulated radiotherapy on radiation-induced second malignancies." Int J Radiat Oncol Biol Phys **70**(5): 1530-1536.

Schafer, W. D. and E. S. Gilbert (2006). "Some statistical implications of dose uncertainty in radiation dose-response analysis." Radiat Res **166**: 303-312.

Schneider, U. and B. Kaser-Hotz (2005). "A simple dose-response relationship for modeling secondary cancer incidence after radiotherapy." Z Med Phys **15**(1): 31-37.

Schneider, U., A. Lomax, J. Besserer, P. Pemler, N. Lombriser, B. Kaser-Hotz, (2007). "The impact of dose escalation on secondary cancer risk after radiotherapy of prostate cancer." Int J Radiat Oncol Biol Phys **68**(3): 892-897.

Schneider, U., A. Lomax, P. Pemler, J. Besserer, D. Ross, N. Lombriser, B. Kaser-Hotz, (2006). "The impact of IMRT and proton radiotherapy on secondary cancer incidence." Strahlenther Onkol **182**(11): 647-652.

Schneider, U., D. Zwahlen, D. Ross, B. Kaser-Hotz, (2005). "Estimation of radiation-induced cancer from three-dimensional dose distributions: Concept of organ equivalent dose." Int J Radiat Oncol Biol Phys **61**(5): 1510-1515.

Schneider U., Walsh L., (2008). "Cancer risk estimates from the combined Japanese A-bomb and Hodgkin cohorts for doses relevant to radiotherapy." Radiat Environ Biophys **47**:253 – 63.

Sigurdson, A. J., C. M. Ronckers, A. C. Mertens, M. Stovall, S. A. Smith, Y. Liu, R. L. Berkow, S. Hammond, J. P. Neglia, A. T. Meadows, C. A. Sklar, L. L. Robison, P. D. Inskip, (2005). "Primary thyroid cancer after a first tumour in childhood (the Childhood Cancer Survivor Study): a nested case-control study." Lancet **365**(9476): 2014-2023.

Sisterson, J. (2005). "Ion beam therapy in 2004." Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms **241**: 713-716.

Stovall, M., R. Weathers, C. Kasper, S. A. Smith, L. Travis, E. Ron, R. Kleinerman, (2006). "Dose reconstruction for therapeutic and diagnostic radiation exposures: use in epidemiological studies." Radiat Res **166**(1 Pt 2): 141-157.

Taddei, P. J., J. D. Fontenot, Y. Zheng, D. Mirkovic, A. K. Lee, U. Titt, W. D. Newhauser, (2008). "Reducing stray radiation dose to patients receiving passively scattered proton radiotherapy for prostate cancer." Phys Med Biol **53**(8): 2131-2147.

Taddei, P. J., A. Mahajan, D. Mirkovic, R. Zhang, A. Giebeler, D. Kornguth, M. Harvey, S. Woo, W. D. Newhauser, (2010). "Predicted risks of second malignant neoplasm incidence and mortality due to secondary neutrons in a girl and boy receiving proton craniospinal irradiation." Phys Med Biol **55**(23): 7067-7080.

Taddei, P. J., D. Mirkovic, J. D. Fontenot, A. Giebeler, Y. Zheng, D. Kornguth, R. Mohan, W. D. Newhauser, (2009). "Stray radiation dose and second cancer risk for a pediatric patient receiving craniospinal irradiation with proton beams." Phys Med Biol **54**(8): 2259-2275.

Taylor, R. E. (1996). "Cancer in children: radiotherapeutic approaches." Br Med Bull **52**(4): 873-886.

UNSCEAR. (2000). Report to the General Assembly: Sources and effects of ionizing radiation. Annex I - Epidemiological evaluation of radiation-induced cancer.

Waters, L., J. Hendricks, G. McKinney, (2002). Monte Carlo N-Particle transport code system for multiparticle and high energy applications. Los Alamos, NM, Los Alamos National Laboratory.

Wrixon, A. D. (2008). "New recommendations from the International Commission on Radiological Protection--a review." Phys Med Biol **53**(8): R41-60.

Yan X, T. U., Koehler A M and Newhauser W D (2002). "Measurement of neutron dose equivalent to proton therapy patients outside of the proton radiation field." Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **476**: 429-434.

S. S. Yom, E. K. Frija, A. Mahajan, E. Chang, K. Klein, A. Shiu, J. Ohrt, S. Woo, (2007). "Field-in-field technique with intrafractionally modulated junction shifts for craniospinal irradiation," Int J Radiat Oncol Biol Phys **69** (4): 1193-1198.

Zheng, Y., W. Newhauser, J. Fontenot, P. Taddei, R. Mohan, (2007). "Monte Carlo study of neutron dose equivalent during passive scattering proton therapy." Phys Med Biol **52**(15): 4481-4496.

Vita

Annelise Giebeler was born in Redlands, California to Elizabeth Ann Shetler and James Frederick Giebeler. She started her undergraduate studies at Crafton Hills College in Yucaipa, California; San Bernardino Valley College in San Bernardino, California; and Riverside Community College in Riverside, California and completed them at California State University San Bernardino, in San Bernardino, California where she received the degrees of Bachelor of Science for mathematics and Bachelor of Science for for physics in March of 2000. After graduation, she worked as a substitute teacher for the San Bernardino City School District and attended classes at the Loma Linda University School of Medicine in Loma Linda, California. Then, from November, 2001, through July, 2005, she worked as a dosimetrist at Loma Linda University's James M. Slater Proton Treatment and Research Center. Next, from August, 2005, through September, 2006, she worked at The University of Texas M.D. Anderson Cancer Center's Proton Treatment Center in Houston, Texas, in preparation for the Center's first patient treatment in May of 2006. In January of 2007 she matriculated at The University of Texas Health Science Center at Houston Graduate School of Biomedical Sciences as a graduate research assistant in the Medical Physics program, and in September of 2008 she was awarded a predoctoral fellowship in The University of Texas M. D. Anderson Cancer Prevention Research Training Program. In December of 2010 she received the degree of Masters of Science in Medical Physics from The University of Texas Health Science Center at Houston Graduate School of Biomedical Sciences.

Permanent address:
1223 Rose St.
Mentone, CA 92359