2015

Hitting the target: analysis of delineation and dosimetric uncertainty in radiotherapy

Michael Geoffrey Jameson
University of Wollongong

UNIVERSITY OF WOLLONGONG

Centre for Medical Radiation Physics

Faculty of Engineering

Hitting the target: Analysis of delineation and dosimetric uncertainty in radiotherapy

Michael Geoffrey Jameson

B.Med.Rad.Phys.Adv(Hons)

A Dissertation Submitted in Fulfilment of

The Requirements for Award of the Degree of

Doctor of Philosophy From

University of Wollongong

December 2015

ABSTRACT

Through advances in radiation delivery systems and image guidance, the accuracy and precision of radiation therapy has improved in recent times. Some aspects with respect to the accuracy and precision with which treatments are prescribed and planned have also improved, however it has not been to the same extent. Radiotherapy has moved from 2D to 3D treatment planning and now incorporates multimodality imaging into the contouring process, but there is still variation in tumour delineation. This thesis is an investigation into the impact of contouring, planning, and organ motion variation on dosimetry and modelled outcome in a variety of disease scenarios. The effect of contouring uncertainty in the lung was investigated with a retrospective dataset of non-small cell lung cancer patients. Planning uncertainty due to planner experience was studied using a head and neck case and the influence of organ motion was considered in the post-prostatectomy setting. Finally, the techniques developed to analyse contouring variation were applied to a gynaecological clinical trial benchmarking dataset to incorporate contouring uncertainty into the trial sample size calculation.

For some treatment sites, the uncertainty in radiotherapy target delineation is greater than that of organ motion and setup error. As radiotherapy treatment techniques have become more conformal, the relative importance of contouring uncertainty has increased compared to other sources of error in the treatment chain. Understanding the impact of contouring variation on modelled outcome would aid in the development of contouring guidelines, adaptive radiotherapy

protocols, margin definition and clinical trial quality assurance. The impact of contouring variation on modelled outcome was assessed for a series of non-small cell lung cancer patients. The results of this work should inform the choice of metric used and ensure that future contouring studies are more consistent and comparable.

A significant advantage of IMRT over standard conformal techniques is the ability to highly conform the dose distribution around sensitive healthy tissues. This increased conformity comes at the expense of increased plan complexity and delivery time. In the context of clinical trials, variation in treatment planning approaches, and the experience of centres in IMRT planning, has been shown to result in significant variations in dosimetry. There are a variety of techniques available for producing an IMRT plan and planner experience may have an impact on the final plan quality. The influence of planner experience on IMRT plan quality was assessed through a head and neck case planning study. Treatment delivery time and monitor units ranged from 15-25 minutes and approximately 800-1200 MU with delivery time increasing with decreasing planner experience. The planner with the least experience had the poorest plan, as indicated by achieving the fewest PTV constraints of all planners studied.

It has been known for some time that the prostate bed can experience inter- and intra-fraction motion due to its proximity to the bladder and bowel, organs that are constantly filling and emptying. Endorectal balloons (ERBs) have been used in prostate radiotherapy as organ stabilising devices. In this work, ERBs in the

post prostatectomy setting were evaluated. The ERB significantly improved inter-fraction reproducibility for the rectum and the CTV. Concordance indices for non-ERB and ERB of 0.50 ± 0.12/0.71 ± 0.07 for the rectum and 0.72 ± 0.15/0.73 ± 0.11 for the CTV were demonstrated. However, the improved geometric stability with the ERB did not translate into a statistically significant benefit in inter-fraction dosimetric stability.

Protocol deviations in Randomised Controlled Trials have been found to result in a significant decrease in outcomes. In some cases, the magnitude of the detrimental effect can be larger than the anticipated benefits of the interventions involved. The accuracy of radiotherapy contouring is one of largest contributors to protocol deviations in radiotherapy trials. It is well recognised that robust methodology and quality assurance is required to ensure the validity of RCTs. This study aims to model the effect of contouring variation on tumour control probability (TCP) and consequently on clinical trial sample size. PORTEC3 is a phase III clinical trial comparing concurrent chemoradiation and adjuvant chemoradiotherapy with pelvic radiation alone in high risk advanced stage endometrial carcinoma. A benchmarking exercise was performed for the PORTEC3 RCT amongst Australian and New Zealand centres. The results of this benchmarking exercise were then used to assess the robustness of the sample size calculations. This work provides a framework to incorporate quantified uncertainties as part of routine benchmarking exercises in RCT sample size calculations to ensure robust results are obtained from RCTs.

LIST OF ABREVIATIONS

3D	three dimensional
3DCRT	three dimensional conformal radiotherapy
4D	four dimensional
4FLD	four field
ANZ	Australia and New Zealand
CBCT	cone beam computed tomography
CCORE	centre for collaborative outcomes research
CD	compact disc
CDK1	cyclin-dependent kinase 1
CERR	computational environment for radiotherapy research
CI	conformity index (dose)
CI	concordance index (contours)
CMRP	centre for medical and radiation physics
CN	conformity number
COM	centre of mass
CONSORT	consolidated standards of reporting trials
COV	centre of volume
CT	computed tomography
CTV	clinical target volume
DDR	DNA damage response
DICOM	digital imaging and communication in medicine
DNA	deoxyribonucleic acid
dPETCT	diagnostic positron emission tomography computed tomography
DSB	double strand break
DSC	dice similarity coefficient

DVH	dose volume histogram
DWI	diffusion weighted imaging
EORTC	european organisation for research and treatment of cancer
EPID	electronic portal imaging device
ERB	endo-rectal balloon
EUD	equivalent uniform dose
FDG	flurodeoxyglucose
FLT	fluorothymidine
FMISO	fluoromisonidazole
FSU	functional subunit
gEUD	generalised equivalent uniform dose
GS	gold standard
GTV	gross tumour volume
HI	homogeneity index
ICCC	Illawarra cancer therapy centre
ICRU	international commission on radiological units and measures
IGRT	image guided radiation therapy
Imax	maximum isodose
IMRT	intensity modulated radiation therapy
Irec	inferior rectum
ITV	internal target volume
kV	kilo voltage
LCTC	Liverpool cancer therapy centre
LINAC	linear accelerator
LQ	linear quadratic
MDC	monodansylcadaverine

MLC	Multi-leaf collimator
MLD	mean lung dose
MPM2	mitotic phosphoprotein 2
MRI	magnetic resonance imaging
NCI	national cancer institute
NSCLC	non-small cell lung cancer
NTCP	normal tissue complication probability
OAR	organ at risk
PET	positron emission tomography
PORTEC	post-operative radiation therapy in endometrial carcinoma
pPETCT	planning positron emission tomography computed tomography
PRV	planning risk volume
PTV	planning target volume
QA	quality assurance
QANTEC	quantitative analysis of normal tissue effects in the clinic
QART	quality assurance radiation therapy
RB	retinoblastoma protein
RCT	randomised controlled trial
RI	reference isodose
ROG	radiation oncology group
RTOG	radiation therapy oncology group
RVR	remaining volume at risk
SA-β-gal	senescence-associated β-galactosidase
SPECT	single photon emission tomography
SPSS	statistical package for the social sciences
Srec	superior rectum

STAPLE	simultaneous truth and performance level estimation
TCP	tumour control probability
TD_{50}	dose for 50% control
TPS	treatment planning system
TROG	Trans-Tasman radiation oncology group
TV	target volume
VMAT	volumetric modulated arc therapy
VRI	volume of reference isodose

LIST OF PUBLICATIONS

1. **Jameson MG**, Holloway LC, Vial PJ, Vinod SK, Metcalfe PE. A review of methods of analysis in contouring studies for radiation oncology. Journal of medical imaging and radiation oncology. 2010;54(5):401-10.

2. Batumalai V, Koh E, Delaney G, Holloway L, **Jameson MG**, Papadatos G, *et al*. Interobserver variability in clinical target volume delineation in tangential breast irradiation: a comparison between radiation oncologists and radiation therapists. Clinical Oncology. 2011;23(2):108-13.

3. Arumugam S, **Jameson MG**, Xing A, Holloway L. An accuracy assessment of different rigid body image registration methods and robotic couch positional corrections using a novel phantom. Medical physics. 2013;40(3):031701.

4. Arumugam S, Xing A, **Jameson MG**, Holloway L. An algorithm to calculate a collapsed arc dose matrix in volumetric modulated arc therapy. Medical physics. 2013;40(7):071724.

5. Batumalai V, **Jameson MG**, Forstner DF, Vial P, Holloway LC. How important is dosimetrist experience for intensity modulated radiation therapy? A comparative analysis of a head and neck case. Practical radiation oncology. 2013;3(3):e99-e106.

6. **Jameson MG**, De Leon J, Windsor AA, Cloak K, Keats S, Dowling JA, *et al*. Endorectal balloons in the post prostatectomy setting: Do gains in stability lead to more predictable dosimetry? Radiotherapy and Oncology. 2013;109(3):493-7.

7. **Jameson MG**, Kumar S, Vinod SK, Metcalfe PE, Holloway LC. Correlation of contouring variation with modeled outcome for conformal non-small cell lung cancer radiotherapy. Radiotherapy and Oncology. 2014;112(3):332-6.

8. Batumalai V, Quinn A, **Jameson MG**, Delaney G, Holloway L. Imaging dose in breast radiotherapy: does breast size affect the dose to the organs at risk and the risk of secondary cancer to the contralateral breast? Journal of Medical Radiation Sciences. 2015;62(1):32-9.

9. **Jameson MG**, Ohanessian L, Batumalai V, Patel V, Holloway LC. Comparison of Oncentra® Brachy IPSA and graphical optimisation techniques: a case study of HDR brachytherapy head and neck and prostate plans. Journal of Medical Radiation Sciences. 2015.

10. Leon JF, **Jameson MG**, Windsor A, Cloak K, Keats S, Vial P, *et al*. Superior target volume and organ stability with the use of endorectal balloons in post-prostatectomy radiotherapy. Journal of medical imaging and radiation oncology. 2015.

11. Pogson, E. M., Begg, J., **Jameson, MG**, Dempsey, C., Latty, D., Batumalai, V., at al. A phantom assessment of achievable contouring concordance across multiple treatment planning systems. *Radiotherapy and Oncology*. 2015.

12. **Jameson, MG.**, McNamara, J., Bailey M., Metcalfe PE., Holloway, LC., Results of the Australasian (TROG) radiotherapy benchmarking exercise in preparation for participation in the PORTEC-3 trial. Journal of medical imaging and radiation oncology (Accepted)

PUBLISHED ABSTRACTS

1. Holloway L, **Jameson M**, Prasad S, Forstner D. A modeling study: head and neck hypofractionation achievable with IMRT. Australasian Physical & Engineering Sciences in Medicine. 2008;31(4):509.

2. **Jameson M**, Metcalfe P, Vial P, Goozee G, Holloway L. The feasibility of utilising a siemens electronic portal imaging device for routine radiotherapy quality assurance. Australasian Physical & Engineering Sciences in Medicine. 2008;31(4):505.

3. Arumugam S, **Jameson M**, Holloway L. Development of a QA Phantom for online image registration and resultant couch shifts. 2010.

4. Holloway L, **Jameson M**, Batumalai V, Koh E, Papadatos G, Lonergan D, *et al*. Estimating a Delineation Uncertainty Margin to Account for Inter-observer Variability in Breast Cancer Radiotherapy. International Journal of Radiation Oncology* Biology* Physics. 2010;78(3):S741.

5. Arumugam S, Xing A, **Jameson M**, Holloway L, Goozee G. Development of a software of VMAT delivery using EPID. 2011.

6. Balumalai V, **Jameson M**, Blakeney S, Franji I, Andrew K, Nguyen C, *et al*. 113 oral How much experience is enough? a comparative analysis of IMRT plans. Radiotherapy and Oncology. 2011;99:S43-S4.

7. **Jameson M**, De Leon J, Windsor A, Cloak K, Holloway L, Vial P, *et al*. TH-C-BRA-04: Endorectal Balloons in Post-Prostatectomy: Do Gains in Stability Lead to More Predictable Dosimetry? Medical Physics. 2012;39(6):4000-.

8. Windsor A, De Leon J, **Jameson M**, Cloak K, Zammit A, Ko R, *et al*. Endorectal Balloons in Postprostatectomy Radiation Therapy—Improved Stability of Clinical Target Volumes and Reduction of Geographical Miss. International Journal of Radiation Oncology* Biology* Physics. 2012;84(3):S391-S2.

9. Ho GK, Kumar S, Arumugam S, **Jameson M**, Degruyter S, Phan PD, *et al*., editors. Non-small cell lung cancer (nsclc): changes in volume during radiotherapy and potential adaptive radiotherapy planning. journal of thoracic oncology; 2013: lippincott williams & wilkins 530 walnut st, Philadelphia, PA 19106-3621 USA.

10. Pogson E, McNamara J, **Jameson M**, McDowall R, Lim A, Dempsey C, *et al*. SU-E-J-213: An Evaluation of the Reproducibility of Radiotherapy Contouring Utilizing Multiple Institutions and Treatment Planning Systems. Medical Physics. 2013;40(6):200-.

11. Ohanessian L, Vinod S, Dinsdale G, Franji I, Scotti A, Lim K, *et al*., editors. Static field intensity modulated radiation therapy versus helical tomotherapy-retrospective study comparing the two treatment modalities in the planning of post operative endometrial patients. international journal of gynecological cancer; 2014: Lippincott Williams & Wilkins 530 Walnut St, Philadelphia, PA 19106-3621 USA.

STATEMENT OF AUTHORSHIP

This thesis is submitted to the University of Wollongong in fulfilment of the requirements for the Degree of Doctor of Philosophy. The work presented in this thesis is, to the best of my knowledge and belief, original except as acknowledged in the text. I hereby declare that I have not submitted this material, either in full or in part, for a degree at this or any other institution.

Michael Jameson

Signature:…… Date: ……

ACKNOWLEDGEMENTS

First and foremost thanks must go to my supervisor A/Prof Lois Holloway. Lois' enthusiasm and excitement for research is infectious and has made the completion of this work an enjoyable experience. Many problems have been identified and solved over early morning coffee and to say that this thesis would not have been completed without Lois' support, is an understatement. Thanks must also go to my other supervisor Prof Peter Metcalfe. Pete was my honours supervisor and since that time his influence on my development as a professional physicist has been substantial. I value greatly his wisdom and support and have benefited from his mentorship over the years.

Thank you to the physics staff of the Liverpool and Campbelltown cancer therapy centres. In particular, to the director of medical physics Gary Goozée. Gary was also my clinical supervisor over the course of my training, and therefore, often a sounding board for PhD issues. Thanks to Gary and the physics staff for supporting me in completing my PhD while working within the cancer therapy centre.

Thanks to Ingham Medical Physics group for providing a pleasant environment in which to work while finishing up this thesis. Thank you to the oncologists who provided their advice and time contouring for various parts of the work presented here, A/Prof Shalini Vinod, Drs Miriam Boxer and Denise Lonergan. Thanks to Shivani Kumar and Vikneswary Batumalai for their sound treatment planning and Sydney restaurant advice.

To Dr Jeremy de Leon who spent many weekends contouring for the rectal balloon work. Thanks also, to Mike Bailey and Jo McNamara who put up with my constant harassment, questions and even came into work on the weekend for me! I must also thank A/Prof Martin Carolan for supporting the collaboration with his department on the PORTEC3 work. Thanks to Pearly Khaw and Carien Creutzberg for their support and guidance. To Dr Farhannah Aly for reviewing this thesis and providing useful feedback.

To my family, thanks for always being there supporting me throughout my academic pursuits. To Alex for her constant support and dedication.

Finally, this thesis is dedicated to my grandfather Kevin Green. His effervescent curiosity and enthusiasm for all things, but in particular, technology never ceased to impress and inspire me.

TABLE OF CONTENTS

ABSTRACT 2

List of abreviations 5

List of publications 9

Published abstracts 10

Statement of authorship 11

Acknowledgements 12

TABLE OF CONTENTS 14

Chapter 1: Introduction 18

1.1 Account of scientific progress linking the publications in the thesis 20

1.2 Specific aims and objectives 24

1.2.1 The impact of contouring variation on modelled radiotherapy outcome 24

1.2.2 The influence of planner experience on IMRT plan quality 25

1.2.3 Investigation of organ motion, dosimetry, and margins in the presence of organ stabilising devices 26

1.2.4 Benchmarking and assessing the impact of contouring variation in radiotherapy clinical trials 28

1.3 The journey 30

Chapter 2**:** Literature review 31

2.1 Radiotherapy 31

2.1.1 Cancer and the role of radiotherapy 31

2.1.2 Trends in radiotherapy treatment delivery 32

2.2 Radiotherapy treatment planning 35
2.2.1 Imaging 35
2.2.2 Contouring 37
2.2.3 Assessing plan quality 41
2.3 Methods of analysis in contouring studies for Radiation Oncology 43
2.4 Motion in radiotherapy 45
2.4.1 Types of motion 45
2.4.2 Image Guided Radiation Therapy (IGRT) 46
2.4.3 Impact of organ motion on plan quality 47
2.4.4 Strategies to reduce motion 48
2.4.5 Strategies to account for motion 48
2.5 Radiobiological modelling 49
2.5.1 Mechanisms of radiation induced cell death 50
2.5.2 The Rs of radiobiology 54
2.5.3 The linear quadratic model 55
2.5.4 Equivalent Uniform Dose (EUD) 57
2.5.5 Tumour Control Probability (TCP) 57
2.5.6 Normal tissue complication probability (NTCP) 58
2.6 Radiotherapy clinical trials 61
2.6.1 Trial design and sample size calculations 62
2.6.2 Quality assurance in radiotherapy clinical trials 64
2.7 References 66

Chapter 3: A review of methods of analysis in contouring studies for radiation oncology 81

Chapter 4: Correlation of contouring variation with modelled outcome for conformal non-small cell lung cancer radiotherapy 93

Chapter 5: How important is dosimetrist experience for intensity modulated radiation therapy? A comparative analysis of a head and neck case 106

Chapter 6: Endorectal balloons in the post prostatectomy setting: Do gains in stability lead to more predictable dosimetry? 116

Chapter 7: Superior target volume and organ stability with the use of endorectal balloons in post-prostatectomy radiotherapy 127

Chapter 8: Results of the Australasian (TROG) radiotherapy benchmarking exercise in preparation for participation in the PORTEC-3 trial 137

Chapter 9: The impact of contouring uncertainty on radiotherapy clinical trial sample size; A novel methodology applied to the PORTEC-3 trial 146

Chapter 10: Discussion and conclusions 175

10.1 General Discussion 175

10.2 The impact of contouring variation on modelled radiotherapy outcome 177

10.3 The influence of planner experience on IMRT plan quality 179

10.4 Investigation of organ stability, dosimetry, and margins in the presence of organ stabilising devices 180

10.5 Benchmarking and assessing the impact of contouring variation in radiotherapy clinical trials 181

10.6 Future work 183
10.7 Summary 185
10.8 References 186
Appendix A: A phantom assessment of achievable contouring concordance across multiple treatment planning systems 188
Appendix B: Contouring Variability and its Effect on Radiobiology Parameters for Head and Neck Cancer 193

Chapter 1: Introduction

"It is best to prove things by actual experiment; then you know; whereas if you depend on guessing and supposing and conjectures, you never get educated."

(Mark Twain, 1906)

"*With 14 million new cases and 8 million related deaths in 2012 cancer is a major cause of morbidity and mortality worldwide*"[1]. The technology of radiotherapy planning and delivery is constantly evolving to meet the challenge of safely delivering a therapeutic dose to cancerous tissues. A prerequisite to delivering safe, precise radiotherapy is understanding the sources and impact of uncertainty in each step of the treatment chain[2]. When determining the benefit or otherwise of new technologies and techniques through clinical trials, rigorous methodology must be adopted to ensure protocol compliance[3].

The accuracy and precision of radiation therapy has improved in recent times through advances in radiation delivery systems and image guidance. Although, some aspects, with respect to the accuracy and precision with which treatments are prescribed and planned have improved, it has not been to the same extent. Radiotherapy has moved from 2D to 3D treatment planning and now incorporates multimodality imaging into the contouring process, but there is still variation in tumour delineation and inverse planning.[4]. Through the analysis of delineated 3D images, that is, contours, in radiotherapy planning it is

possible to investigate a number of aspects of the planning and delivery process (Figure 1). This thesis is by publications, and is an investigation into the impact of uncertainty in contouring, planning, and organ motion, on dosimetry and modelled outcome.

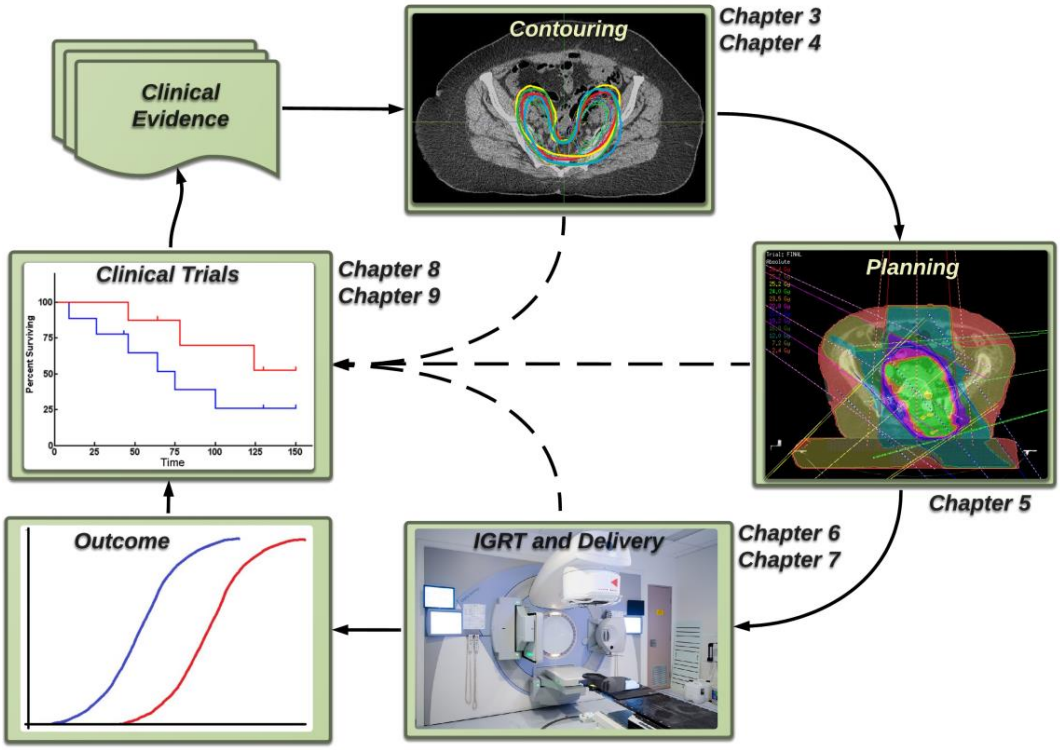


Figure1.1 Radiotherapy process diagram identifying which aspect of the treatment chain each chapter addresses

The analogy of William Tell shooting the apple from his son's head (Figure 1.2) is not new in radiation therapy[5]. But, it is of particular relevance to the work presented in this thesis; accurately defining and hitting the target while avoiding injury.

Figure 1.2 William Tell shooting at the apple, woodcut from *Ein Schönes Spiel…von Wilhelm Thellen*, by O. Schweitzer, 1698.

1.1 Account of scientific progress linking the publications in the thesis

The investigation into contouring variation in radiotherapy began with a review of the literature. Contouring variation has for a number of years been recognised as a major uncertainty in radiotherapy[6]. A clinically focused review had been published earlier[7] but there was no work summarising the methods of analysis of contouring variation. With the advent of 3D planning in external beam radiotherapy and brachytherapy, there had been a large increase in the number of contouring studies being published. Here, a contouring study is broadly classified as an investigation that analyses the variation between a number of delineations on medical images in order to elucidate some information about the planning and delivery process. Most commonly, the

information sought is about inter- or intra-observer contouring variation but may also include the impact of organ motion, image quality, contouring guidelines, clinical trial protocol evaluation, atlas development and training. The literature review, presented in chapter 3 identified a number of different methods of analysis. These techniques were explained with advantages and disadvantages in particular situations. What was obvious from the literature was that there was no consensus on the appropriate techniques to use, and that methods employed were dictated by bespoke software and expertise available to investigators rather than evidence.

The work presented in chapter 4 aimed to address the issue identified in the literature review, i.e. the lack of consensus in analysis technique in contouring studies. This was achieved by establishing which contouring variation metrics were most likely to impact on dosimetry and modelled outcome and therefore, be most relevant to reporting. The impact of contouring variation on dosimetry had been investigated previously[8-10], but this was the first study assessing the correlation between these two factors. The contouring variation metrics that were most significantly correlated with modelled outcome were identified for conformal lung cancer radiotherapy. This work presented a methodology that could be employed in other tumour sites and treatment techniques to ascertain the most relevant metrics of contouring variation to report. This work was repeated for head and neck cancer inversely planned radiotherapy, see

Appendix A. The best achievable contouring concordance between planning systems was investigated by our group using a phantom study, see Appendix B.

Similar to contouring uncertainty, inter-observer variation in radiotherapy planning has been identified as a confounding factor on radiotherapy trials[11]. The International Commission on Radiological Units (ICRU) has outlined procedures for prescribing and reporting in radiotherapy[12-14] that guide the planning process. In the context of inverse planning there may be inter-observer variation due to planner experience with respect to adjusting parameters to achieve the end result. A study assessing the impact of planner experience on dosimetry is presented in chapter 5. This work demonstrated that planner experience can influence both plan quality and delivery efficiency in the context of head and neck inverse planning.

The resulting dosimetric impact of day-to-day organ deformation and position can be similar to that of contouring variation. Involving both systematic and random uncertainties, see section 2.2.2.1. Therefore, similar analysis techniques to those used in chapter 4 can be employed to assess the dosimetric impact of organ motion using contoured daily cone beam computed tomography (CBCT) imaging. Endo-rectal balloons (ERBs) have been used extensively in prostate radiotherapy[15] to stabilise the prostate and minimise the amount of rectal wall in the high dose area. The use of ERBs in the post-prostatectomy setting had not been investigated to the same extent. Chapters 6 and 7 employ similar analysis

techniques to those used in chapter 4 to assess the impact of organ motion on dosimetry and margins with and without the ERB. This was the first study published investigating the day-to-day reproducibility of the prostate bed with the ERB *in situ*. The significance of the findings were that ERBs did reduce organ motion, particularly for the rectum. Further study is warranted to confirm whether this translates into better dosimetric reproducibility with a larger patient cohort.

Contouring and dosimetric uncertainty has been shown to be a major confounding factor in radiotherapy clinical trials[16]. Yet, there are a number of other uncertainties in clinical trials, that is, predicted treatment response, combined modality treatment effect, patient dropout etc. The difference in these uncertainties in the response rate and patient dropout are routinely accounted for in sample size calculations for clinical trials. Using the analysis techniques from previous chapters, it is possible to ascertain the uncertainty in modelled outcome due to contouring and planning uncertainty. The study presented in chapter 8 details the results of a benchmarking study for the PORTEC3 trial[17]. PORTEC3 is a phase III clinical trial comparing concurrent chemoradiation and adjuvant chemoradiotherapy with pelvic radiation alone in high risk advanced stage endometrial carcinoma. The benchmarking study quantified the contouring and planning variation amongst participating centres adhering to the same protocol in Australia and New Zealand. Chapter 9 utilised the contouring and modelled outcome analysis techniques mentioned above to

establish impact of contouring variation for the PORTEC3 trial. This work presented a novel technique for incorporating contouring uncertainty into the sample size calculation for a randomised controlled trial (RCT).

1.2 Specific aims and objectives

1.2.1 The impact of contouring variation on modelled radiotherapy outcome

For some treatment sites, the uncertainty in radiotherapy target delineation is greater than that of organ motion and setup error[7].

As radiotherapy treatment techniques have become more conformal, the relative importance of contouring uncertainty has increased compared to other sources of error in the treatment chain[7]. While many studies have analysed contouring uncertainty geometrically, few have considered the potential impact on dosimetry[18].

Given varying anatomy and treatment goals, certain clinical sites may be more susceptible to dosimetric impacts of contouring variation than others. Understanding the impact of contouring variation on modelled outcome would

aid in the development of contouring guidelines, adaptive radiotherapy protocols and clinical trial quality assurance.

Research question:

> *What is the relationship between contouring variation and predicted outcome using radiobiological modelling for non-small cell lung cancer (NSCLC)?*

Chapter 4 analyses the correlation between geometric contouring variation and tumour control probability (TCP) for a series of NSCLC patients.

1.2.2 The influence of planner experience on IMRT plan quality

Intensity modulated radiation therapy (IMRT) has become the standard of care for a number of treatment sites and is performed in over 90% of Australian centres[19]. The advantage of IMRT over standard conformal techniques is the ability to sculpt the dose distribution around sensitive healthy tissues[20]. This increased conformity comes at the expense of increased plan complexity and delivery time[21]. In the context of clinical trials, variation in treatment planning approaches, and the experience of centres in IMRT planning, has been shown to result in variation in treatment plans[22]. There are a variety of techniques

available for producing an IMRT plan and planner experience may have an impact on the final plan quality.

Research question:

What is the impact of planner experience on the quality of radiotherapy treatment plans in the head and neck region?

Chapter 5 presents an analysis of head and neck IMRT plans generated by six different planners of varying IMRT planning experience.

1.2.3 Investigation of organ motion, dosimetry, and margins in the presence of organ stabilising devices

Adjuvant radiotherapy delivered post radical prostatectomy results in longer time to biochemical failure and improved local control compared to surveillance[23]. There is also a survival benefit associated with adjuvant radiotherapy for patients <70 years old or who had positive surgical margins[23]. It should be noted, these results are derived from the pre-prostate speficic antigen era and are currently under investigation in a number phase three randomised trials[24, 25]. Owing to excellent target coverage and critical structure sparing, intensity modulated treatment techniques are the preferred method of treatment delivery in post prostatectomy radiotherapy.

The target volume in post prostatectomy radiotherapy is bounded by the bladder and rectum and therefore may experience deformation due to organ motion. Furthermore, bladder and rectal changes day to day can be significant[26].

Endo-rectal balloons have been used to stabilise anatomy extensively in intact prostate radiotherapy[27]. It remains to be demonstrated if endo-rectal balloons actually improve dosimetric reproducibility on a day-to-day basis.

Research questions:

Does the use of an endo-rectal balloon in situ improve dosimetric precision in post-prostatectomy radiotherapy?

Does the use of endo-rectal balloons reduce the required planning target volume (PTV) margin for organ motion in post prostatectomy patients?

Chapter 6 presents a geometric and dosimetric comparison of two cohorts of post-prostatectomy patients treated with and without an endo-rectal balloon *in situ*. Chapter 7 uses the same cohort of patients studied in chapter 6 but specifically analyses organ motion and the required PTV margin to account for it.

1.2.4 Benchmarking and assessing the impact of contouring variation in radiotherapy clinical trials

A randomised controlled trial (RCT) is the most effective means available to answer questions about treatment effectiveness when designed, conducted and reported appropriately[28]. It is well recognised that robust methodology and quality assurance (QA) is required to ensure the validity of RCTs[29].

Accurate delineation of target volumes and organs at risk for radiation therapy planning is required for high quality treatment as it has a direct flow-on effect for the rest of the radiotherapy chain. The ability of clinicians to contour according to protocol has been investigated for a number of RCTs[16, 30-33]. The accuracy and consistency of contouring in a RCT may be affected by heterogeneity within contributing institutions technology and experience[7].

Protocol deviations in RCTs have been found to result in a significant decrease in survival and local control[16]. In some cases the magnitude of the detrimental effect can be larger than the anticipated benefits of the interventions involved[16, 34]. Implementation of appropriate QA of radiotherapy measures for clinical trials has been found to result in fewer deviations from protocol[35].

The modelled impact of dosimetric uncertainty on sample size for RCTs showed that reduced uncertainty in dose resulted in a significant reduction in required patient numbers[36]. Dosimetric uncertainty is influenced by contouring variation and has been demonstrated to be significant for a number of clinical sites[37-39].

Research questions:

What is the magnitude of endometrial cancer contouring variation in Australia and New Zealand?

What is the impact of contouring variation on the statistical power of clinical trials and can it be accounted for by ensuring optimum patient trial recruitment numbers?

Chapter 8 presents the results of a benchmarking QA study performed in Australia and New Zealand for the PORTEC3 RCT[17]. Chapter 9 assesses the impact of contouring variation on clinical trial design using the benchmarking dataset from the PORTEC3 clinical trial.

1.3 The journey

The research presented in this thesis was undertaken in the School of Physics within the Faculty of Engineering and information Sciences. Expertise and laboratory support was provided within the Centre for Medical Radiation Physics, at the University of Wollongong. Treatment planning facilities and clinical research supervision were also provided by the Liverpool Cancer Therapy Centre (LCTC) and the Ingham Institute for Applied Medical Research where most of the day to day research was undertaken. The Illawarra Cancer Care Centre (ICCC) at Wollongong Hospital also provided data and clinical guidance for a portion of the research undertaken. The contouring, planning and organ motion studies were performed at LCTC. The clinical trial QA and statistical power studies were completed at ICCC.

Chapter 2: Literature review

2.1 Radiotherapy

2.1.1 Cancer and the role of radiotherapy

In Australia, excluding non-melanoma skin cancer, 123920 people were diagnosed with cancer in 2014[40]. Although the mortality rates from cancer are falling, in 2014 cancer related deaths still accounted for approximately 3/10 of all deaths in Australia[40]. The five year overall survival of cancer patients has improved in Australia from 46% in 1982-1986 to 67% in 2007-2011, however this has not been consistent across all tumour types[40]. In 2014 the most commonly diagnosed cancers in males were estimated to be prostate, bowel, skin (melanoma), lung and head and neck[40]. While in women the most commonly diagnosed cancer sites were estimated to be breast, bowel, skin (melanoma), lung, and uterine[40].

In 2013 the Collaboration for Cancer Outcomes Research and Evaluation (CCORE) provided a report to the Australian government department of health and aging reviewing optimal radiotherapy utilisation rates[41]. These rates estimate the number as a percentage of diagnosed cancer patients that would be treated with each resource as part of an optimal treatment regimen. The reported optimal rates of radiotherapy and brachytherapy were 48.3% and 3.3%, while chemotherapy was 8.9%[41]. Meaning, that in the Australian setting,

based on the best available data, radiotherapy is indicated for 48.3% of notifiable cancers, either delivered as a monotherapy or in combination with chemotherapy or brachytherapy[41].

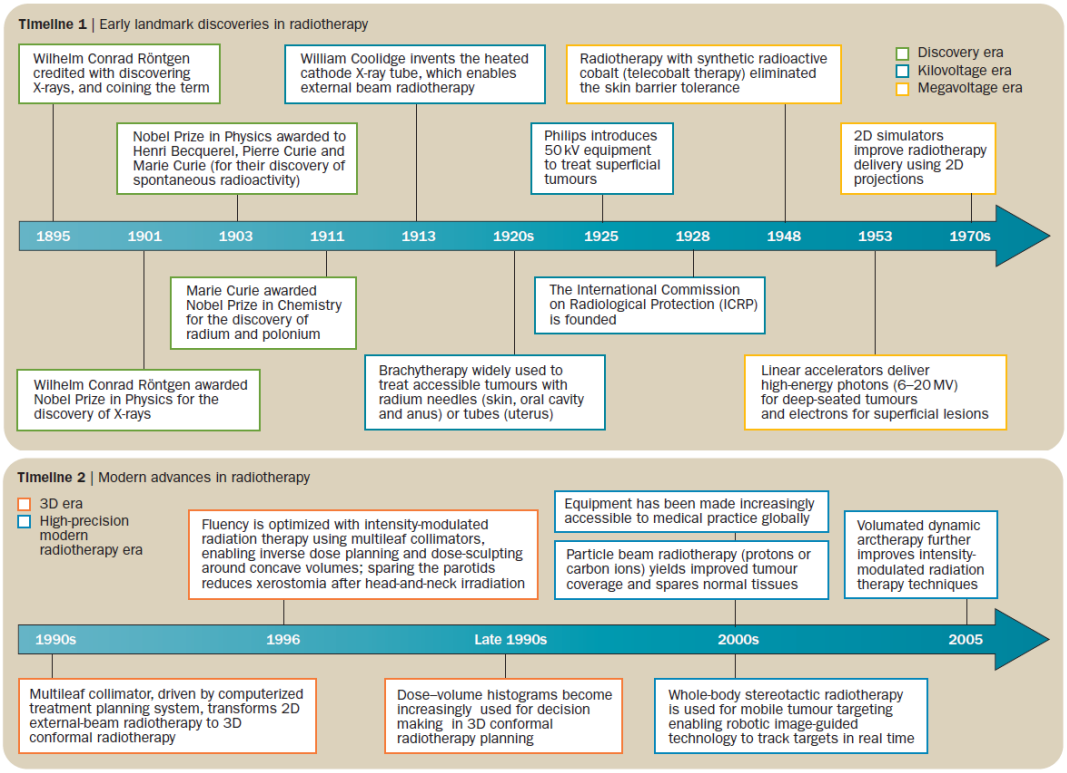


Figure 2.1 From Thariat *et al*[42]. Timeline of radiotherapy evolution from the discovery of X-rays by Röntgen to modern intensity modulated techniques.

2.1.2 Trends in radiotherapy treatment delivery

The discovery of X-rays by German physicist Wilhelm Röntgen in 1895 heralded the beginning of the use of radiation in medicine (Figure 2.1)[43]. Although it should be noted that the rays Röntgen named, and was the first to systematically describe, were also studied in the 1800s by various others[44]. The time between discovery and reported first use of X-rays for a medical

purpose was 60 days[45], although, there is some debate as to whether this is historically correct. The first verifiable reported use of X-rays medically was in Stockholm[46] to treat basal cell carcinoma and was reported at the 1899 Swedish Society of Medicine meeting.

The first 40 years of radiotherapy was dominated by the use of kilovoltage X-ray beams. These were categorised into soft, medium and hard X-rays by penetrative properties. The lower energy beams were used to treat a variety of skin cancers, dermatological and inflammatory conditions in the era before antibiotics and steroids[44, 47]. While the higher energy or harder beams were used to treat deep-seated tumours. There are a number of drawbacks to treating deep tumours with kilovoltage energies; the dose to skin and overlying tissues is quite high due to attenuation of the low energy beam, absorption in bone and long treatment times[48].

Teletherapy (external beam) devices using Radium were also manufactured in North America and Europe[44]. With the advent of nuclear reactors man-made isotopes became available in 1948, and Cobalt-60 was used as a Teletherapy source widely for 20-30 years[42]. Linear accelerators (linacs) were developed before and during the second world war and the first electron accelerator designed for medical use was installed in the Hammersmith hospital, London in 1953[48]. The first patient treated in North America with a 6 MV linac was at Stanford in 1956[49]. Both Co-60 and linac based mega-voltage therapies allowed

skin-sparing application of radiation dose to deep tumours in the pelvis and thorax for the first time.

The first computed tomography (CT) image of a patient was acquired in 1971[50] and in the 1980s was being implemented in radiotherapy departments[42]. This permitted more accurate definition of the tumour and healthy tissues. Dose distributions could now be sculpted in three dimensions (3D) using treatment planning systems (TPS) with beams eye view and linacs with multileaf collimators (MLCs). This so called 3D-conformal radiotherapy saw many tumour sites benefit from higher doses and improved organ at risk (OAR) sparing[42].

Intensity modulated radiation therapy (IMRT) was first proposed by Brahme[51] in 1988 and started entering clinics due to technology advances (e.g. MLC) in the late 1990s[52]. IMRT modulates the intensity of the radiation to enable precise shaping of the dose distribution to the target while avoiding healthy tissue[21]. There are a variety of different techniques for delivering IMRT including; beam compensators[53], and MLCs in both step and shoot[54] and sliding window mode[55]. Volumetric modulated arc therapy (VMAT), first proposed by Yu[56] (called Intensity Modulated Arc Therapy) in 1995, is a form of rotational IMRT and has become standard of care in many centres. This was later refined by Otto[57] to improve the optimisation technique required to generate a plan.

2.2 Radiotherapy treatment planning

Radiotherapy treatment planning is the process in which the tumour and healthy tissues are defined in 3D and used to generate a dose distribution that will guide the technical delivery of the beam by the treatment machine (Figure 2.2). With the increasing use of computers in radiotherapy, treatment planning has evolved from 2D radiographs to multiple 3D datasets that include both high resolution anatomical information and functional information. Computers have also had an influence on beam delivery, from standard open fields using beams-eye-view to intensity modulated beams that closely shape the dose to targets while avoiding healthy tissues.

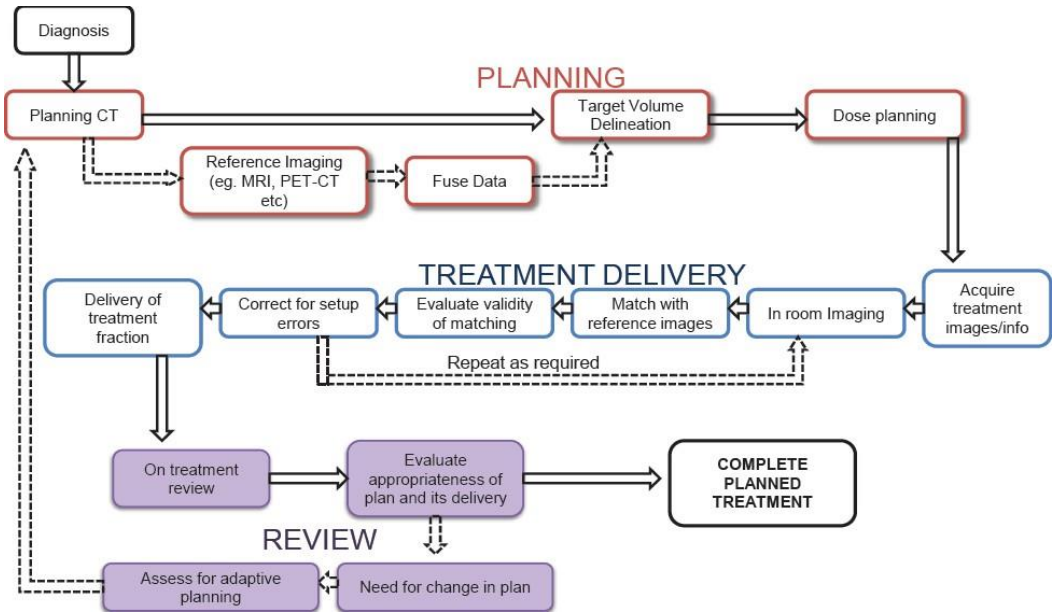


Figure 2.2 Radiotherapy process including treatment planning steps in red, from Gupta *et al*[58]

2.2.1 Imaging

Radiotherapy planning relies largely on 3D imaging. CT was the first imaging modality that allowed the visualisation of the tumour in relation to the

surrounding tissues in 3D. CT also provides electron density information, which is required for understanding radiation transport in tissue. However, soft tissue definition on CT images can be poor and lead to target delineation uncertainties[59]. Other imaging modalities can be registered with CT to better delineate the location of tumours and OARs. Magnetic resonance imaging can provide excellent anatomical soft tissue definition[59]. Other functional imaging techniques include positron emission tomography (PET) and single photon emission tomography (SPECT).

2.2.1.1 Positron Emission Tomography (PET)

PET imaging has been shown to improve target delineation in a number of treatment sites, particularly head and neck, and lungs[60]. Different molecular imaging agents enable the visualisation of different tumour characteristics including metabolism (FDG), hypoxia (FMISO), and proliferation (FLT). PET imaging can also be used for response assessment and may prove valuable in the setting of adaptive radiotherapy[61].

2.2.1.2 Magnetic Resonance Imaging (MRI)

MRI refers to the production of 2D and 3D images that correspond to the macroscopic density distribution of nuclear spins within the volume being imaged[62, 63]. MRI can provide excellent soft tissue definition in areas that CT does not, for example, defining the apex of the prostate. In recent times, there

has been accelerated research growth in the application of MRI to radiotherapy[64]. This is in part due to a lack of ionising radiation required for imaging as well as newer MRI scanners having wide bore designs that can accommodate patient immobilisation devices[65]. MRI can also be used to image functional characteristics of tumours and healthy tissues. Diffusion weighted imaging (DWI) makes use of the limited diffusion of water molecules to generate an image and therefore is a measure of cellularity. DWI may prove useful for treatment response assessment in a number of tumour sites[66]. Two factors limiting the uptake of MRI in radiotherapy is the lack of required electron density information for dose calculation[67] and geometric distortion[68].

2.2.2 Contouring

Contouring refers to the process of segmenting anatomical structures on digital images[69]. This is of particular importance in radiotherapy planning as the segmentation of tumour and healthy tissues is used to guide the treatment and identify areas to be avoided. Technically contouring can be performed on any image type but CT is typically used as it is required for dose calculation by the major treatment planning system vendors[69]. It is commonplace in radiotherapy planning for PET and MRI data to be registered to the planning CT to aid in defining primary tumours and involved nodal regions.

It is widely accepted that contouring is one of the largest sources of uncertainty in the radiotherapy treatment process[70]. Assessing this uncertainty is difficult due to the lack of a derivable ground truth from imaging data[69]. It is, in theory, possible to establish ground truth through invasive techniques (surgical intervention or biopsy) but in practice this is impractical and has additional uncertainty when registering specimens to imaging data[71]. Alternatively investigators assess variation from different physicians (inter-observer) or the same physician (intra-observer) and in a number of different clinical situations (see contouring studies section).

In radiotherapy planning for the individual patient, the dosimetric accuracy is closely related to quality of contouring.[69]. Manual delineation is currently the most widely used method of target and normal structure contouring, which is time consuming and subject to error for the reasons mentioned above. Reducing contouring time and achieving universally precise contours is the goal of automated contouring[69]. Indeed, automated contouring is required in adaptive radiotherapy where many datasets need to be delineated quickly, accurately and consistently[69]. Highly accurate automatic segmentations are currently achievable for some organs and image types but there are still a number of challenges to be faced including image artefact, patient specific features, organ motion and unpredictable shapes of abnormal tumour growth[69].

2.2.2.1 Volumes and margins

To ensure consistent definition of dose distributions in three-dimensional space the ICRU has proposed a set of principles. These principles, for prescribing, recording and reporting photon beam therapy have been published in a number of reports[12, 13, 72]. These reports describe a number of volumes to be used in defining radiotherapy treatments (Figure 1.3). The gross tumour volume (GTV) is the macroscopic extent of malignant growth as determined by palpation or imaging. The clinical target volume (CTV) is the volume which contains the GTV and any microscopic malignant disease. The planning target volume (PTV) is a volume which contains the CTV plus a margin to account for organ, tumour, and patient movement, and uncertainty in delineation and setup. The treated volume and irradiated volumes are defined as the volume of the prescription and tissue volume which receives a dose that is considered significant in relation to normal tissue tolerance. The internal target volume (ITV) is the volume that accounts for movement and deformation of the CTV due to physiological processes. The organs at risk (OARs) or critical normal structures are tissues that might influence treatment planning or prescription through potential morbidity if irradiated. The planning organ at risk volume (PRV) is the OAR plus a margin to account for uncertainties and variations in position and definition to avoid serious complication. The remaining volume at risk (RVR) is the imaged volume, excluding any contoured OARs and CTVs.

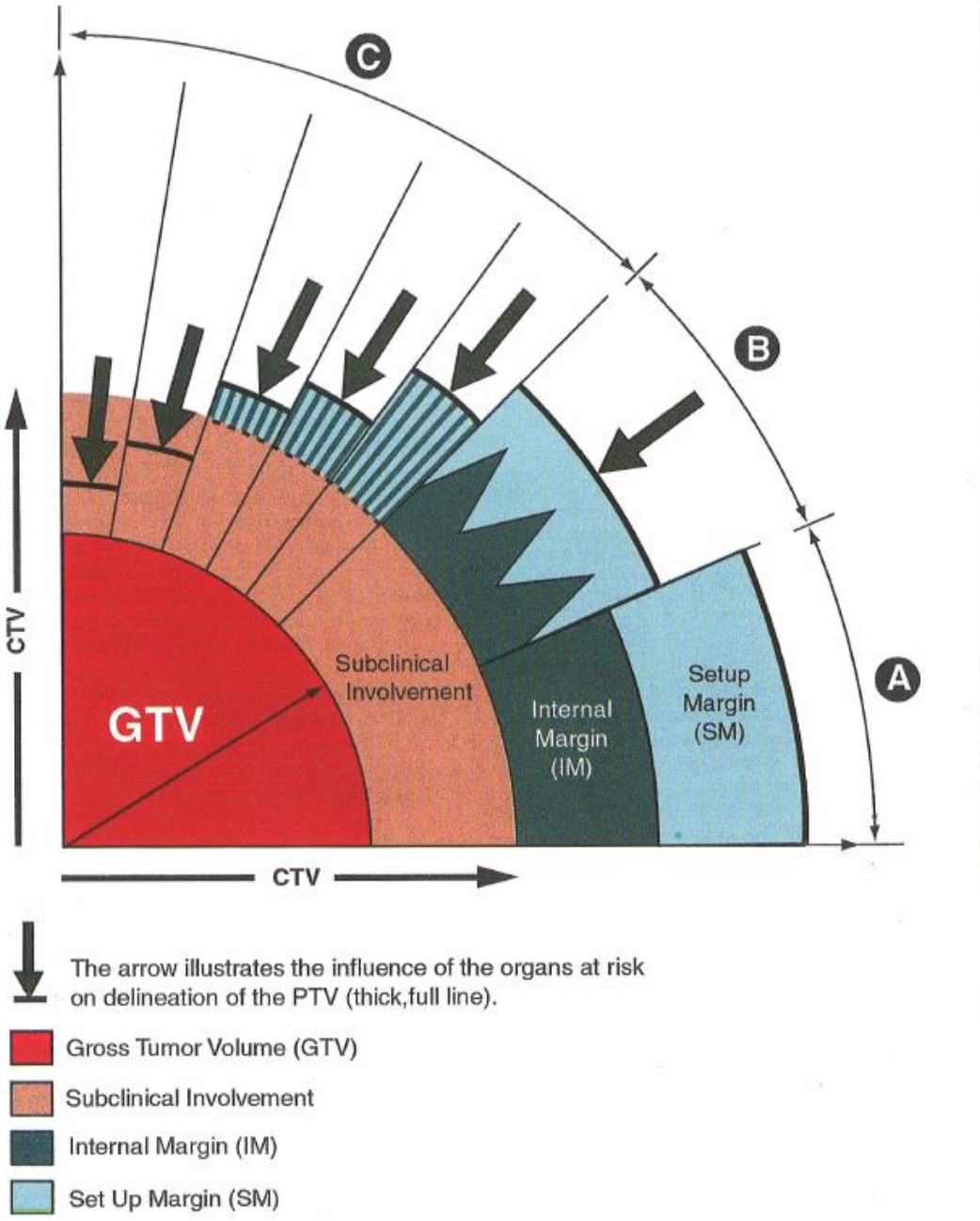


Figure 1.3 Diagram showing the relationship between different treatment volumes as defined by ICRU report 62[13]

In defining the PTV margin one must account for geometric uncertainties in the treatment planning and treatment process. These include tumour delineation (see Section 2.3), unknown extent of microscopic spread of malignant disease,

organ motion and patient setup[73, 74]. To calculate the required margin to account for these uncertainties, they are commonly classified as systematic or random[75]. If the mean irradiation geometry of the fractionated treatment differs from the geometry in the treatment plan this is considered a systematic error[75]. Variations in position around the mean from fraction to fraction are considered random errors[75]. Stroom and Heijmen[75] also note that the source of random and systematic errors may be the same. The impact of systematic errors is larger than that of random errors and thus in modern margin recipes these are given a larger weighting[73]. By far the most popular margin recipe is that of van Herk, Equation 1.

$$Margin = 2.5\Sigma + 0.7\sigma \quad (1)$$

Where Σ is the standard deviation of the systematic errors and σ is the standard deviation of the random errors.

2.2.3 Assessing plan quality

Geometric and dosimetric accuracy are closely related in radiotherapy, contouring dictates where the dose is to be delivered but the quantum of dose deposited is also important. It has been stated many times in the literature that deviations of 7-10% in delivered dose can be detected clinically[70, 76, 77]. In his 1984 paper on dosimetric precision Brahme states, "*If the normalized dose response gradient is higher than 3, as is frequently the case, the relative standard*

deviation of mean dose in the target volume should be less than 3 per cent to achieve an absolute standard deviation in tumour control probability of less than 10 per cent"[78]. Assessing plan quality consists of checking 1) that the plan matches the treatment intent (i.e. prescription) and 2) that the delivered dose matches the plan[21]. When evaluating whether a plan matches treatment intent ,the radiation oncologist and the planner can make use of the dose display, dose volume histograms (DVH) and some planning systems provide tools that allow for assessment of tumour control probability (TCP) and normal tissue control probability (NTCP) [79] (see Section 2.4). Dose conformity indices were introduced by the Radiotherapy Oncology Group (RTOG) as method of assessing how closely and uniformly the prescription isodose conformed to the target volume[80], Equations 2-4.

$$Quality\ of\ coverage = \frac{I_{min}}{RI} \tag{2}$$

$$Homogeneity\ index = \frac{I_{max}}{RI} \tag{3}$$

$$Conformity\ index = \frac{V_{RI}}{TV} \tag{4}$$

Where, I_{min} is the minimal isodose surrounding the target, RI is the reference isodose, I_{max} is the maximum isodose in the target, V_{RI} is the volume of the reference isodose and TV is the target volume. Since the RTOG recommendations there have been a variety of different techniques proposed to

assess the conformity of the prescription dose to the dose to normal tissues in a general way that enables comparison between studies[80-82].

2.3 Methods of analysis in contouring studies for Radiation Oncology

Chapter 3 of this thesis is a review article that was published in 2010 and was the first publication towards this PhD thesis. Chapter 3 provides a detailed overview of contouring studies and methods of analysis in radiation oncology. There have been four other reviews in this area[6, 7, 18, 83], all differ slightly in scope but nonetheless overlap the subject area covered in Chapter 2. Here, a brief summary of these reviews will be provided outlining common issues identified in the literature.

Weiss and Hess[7] published a review of the available literature in 2002. The aim was to evaluate impact of inter-observer variability in contouring on the global geometric accuracy in radiotherapy. From the literature 18 studies were identified and reviewed with respect to tumour site, number of patients and observers, volume of interest and key results. From these studies Weiss and Hess hypothesised the causes of contouring variation and gave a number of recommendations including the use of clear protocols, advanced imaging and peer review to reduce uncertainty.

Njeh's[6] commentary in 2008 aimed to bring attention to the issue of contouring variation as the "weakest link" in the overall radiotherapy treatment chain. In the article Njeh outlines the planning and delivery process and definitions of accuracy and precision. Some of the solutions for contouring variability identified by Njeh include the use of appropriate imaging for delineation, PETCT in head and neck cancer for example. Njeh also recommends continued education and peer review as possible solutions to contouring variation.

The articles by Hanna *et al*[83] and Fotina *et al*[18] both deal specifically with the metrics used to quantify contouring variation. Hanna *et al* performed a systematic review using PubMed using search terms relevant to contouring studies. Hanna *et al* identified 63 studies across a range of tumour sites, the most common of which was lung. Fotina *et al*, is not strictly a review but did perform a comprehensive literature search of overlap metrics for contour comparison. They then calculated this metrics for a series of 7 prostate and 8 lung cases that were contoured by 8 observers.

All of the articles save for the review by Njeh acknowledged the issue that there is no consistent method or form of reporting used for contouring variation studies, with respect to the number of patients and observers to the metrics of comparison used. Hanna *et al* recommended the use of an overlap metric, DICE similarity index for instance, in combination with volume and centre of mass.

Fotina *et al* agreed with Hanna *et al* but also recommended descriptive statistics and a statistical measure of agreement.

2.4 Motion in radiotherapy

Motion is a confounding factor in the delivery of effective radiotherapy. Motion in this context, refers to deviation of the target and normal structures from their planned position, with respect to the treatment coordinate system. This motion can occur over a range of time scales. There are a number of methods that have been proposed to account for motion, these depend on the type and magnitude of the motion and the treatment site in question (i.e. gating for lung, transponders for prostate).

2.4.1 Types of motion

The main sources of motion encountered in radiotherapy can be broadly classified as intra-fraction and inter-fraction motion. Intra-fraction motion is that which occurs during a treatment fraction. Inter-fraction motion is defined as motion that occurs between treatment fractions. Motion that operates on intra-fraction time scales includes: cardiac, respiration, organ filling, peristalsis, and patient movement. Furthermore, inter-fraction motion can arise from day-to-day differences in organ filling, treatment setup, and response of normal and tumour volume changes due to radiation[84]. For intra-fraction motion the

trajectory of the tumour or organ will vary depending on the location within the body and the fixation of the tumour with respect to its surroundings[85].

2.4.2 Image Guided Radiation Therapy (IGRT)

IGRT refers to the integration of imaging equipment within the treatment room to acquire images of the patient in the treatment position prior to or during radiotherapy[86, 87]. Accounting for, and minimising the impact of motion on radiotherapy treatment is the aim of IGRT. The technologies used to deliver IGRT are varied in sophistication and complexity, but all use imaging to align the patient to the planned position. The ideal properties of an IGRT system have been described by Mageras[88] and include: accuracy and precision, efficiency, integration, broad application, reduced radiation dose, real time data collection and cost effectiveness. Radiation based systems may use the mega-voltage treatment beam to generate an image using an electronic portal imaging device (EPID) or film. Further, a cone-beam CT (CBCT) may be used, kVCBCT consists of a kilo-voltage imaging source (usually orthogonal to the treatment beam) and a flat panel detector. The Tomotherapy system uses the treatment beam with a reduced energy to generate a mega-voltage fan-beam CT image. Other systems use multiple kilo-voltage sources stereoscopically to localise bony anatomy or markers. Non-radiation based IGRT systems may use optical cameras, electromagnetic tracking, ultrasound or MRI to discern patient anatomy or markers in order to align the patient with the treatment beam. All of these

technologies, implementations, commissioning procedures and limitations have been reviewed in detail by De Los Santos *et al*[89].

2.4.3 Impact of organ motion on plan quality

When treating a moving target, the delivered dose distribution may not match that of the treatment plan, which does not typically include uncertainty due to motion. The extent to which the delivered dose differs from that of the planned dose depends on how the motion interferes with the delivery, and is of particular importance in dynamic and modulated deliveries[85]. Motion may interfere constructively or destructively with MLC motion, gantry rotation, collimator rotation, or dose rate[85]. The frequency of the motion in question will also play a role. Inter-fraction motion will cause day-to-day differences from the planned dose which will average out over the course of treatment. Intra-fraction motion may cause differences from the planned dose, which is averaged out over that treatment session. It has been stated previously that these effects will not constitute a problem as, over many fractions the cumulative impact is to only slightly smear the dose distribution[90]. But, with the increasing use of hypofractionated treatment delivery the potential impact of organ motion on the delivered dose is demanding increased investigation[91].

2.4.4 Strategies to reduce motion

Depending on the treatment site a number of different motion reduction techniques have been reported. Thermoplastic masks are now commonplace fixation devices for the treatment of head and neck patients. These masks are placed in a water bath to soften and are then moulded to the patient, used daily for position they can reduce inter-fraction setup error[92]. Abdominal compression has demonstrated motion reduction for lung and liver treatments by reducing the amount by which the diaphragm can move freely[93]. A number of different products have been proposed for prostate radiotherapy. The Rectafix, is a plastic rod which is inserted into the rectum during simulation and treatment. The Rectafix increases the separation of the rectum and the prostate and reduces rectal motion. Endorectal balloons serve a similar purpose in that they are also inserted into the rectum during simulation and treatment to stabilise the rectum and move the posterior rectal wall away from the high dose region[15]. Hydrogel spacers, are injected under transrectal ultrasound guidance between the rectum and prostate and last for a number of months. The gel creates a space between the rectum and prostate and results in a reduction in rectal doses for the majority of the prostate patients[94].

2.4.5 Strategies to account for motion

Pre-treatment imaging can be used to reduce the impact of inter-fraction motion on the delivered dose. Margins can also be used to account for inter- and

intra- fraction motion however the aim of IGRT is to reduce treatment margins and thus the volume of normal tissue irradiated[85]. One of the simplest and earliest proposed methods of accounting for intra-fraction motion was to only turn the beam on when the target is inside the beam aperture, this is known as gating and was proposed in 1980s by a number of investigators[95-97]. For lung radiotherapy, this requires capturing tumour motion in the planning CT scan using 4D techniques[98]. Furthermore a respiratory signal needs to be collected during treatment in order to gate the beam, this signal may come from a bellows belt, fiducial markers, spriometry, or external surrogate[98]. Breath hold techniques have also proved useful in gated treatments, whereby the patient holds their breath at a desired point in the breathing cycle[98]. Currently the most advanced technique to account for motion of the target during treatment is realtime tracking[98]. The Cyberknife system uses fiducial markers and fluoroscopic techniques to track target motion and compensate with a robotic treatment unit[99]. Recently Keall *et al*[100] reported on the use of electromagnetic fiducials to guide dynamic MLC tracking of prostate radiotherapy.

2.5 Radiobiological modelling

Radiobiology underpins the discipline of radiation oncology. Classical radiobiology informed modern developments in fractionation, the linear quadratic model, and our understanding of the repair of radiation damage[101].

However, radiobiology also holds promise in elucidating methods of optimisation of biological and physical factors for personalised biologically based treatment planning[101]. Radiobiological modelling is a valuable tool in the assessment of complex radiotherapy treatment plans[102]. For example, the comparison of IMRT and conformal plans for prostate radiotherapy[103] or step-and-shoot IMRT verse Tomotherapy for head and neck cancer[104].

2.5.1 Mechanisms of radiation induced cell death

Radiotherapy exploits the ability of radiation to induce death in cells, of particularly interest is the death of tumour cells. There are a number of ways in which radiation can cause the death of a cell and, these are influenced by the DNA damage response (DDR) system[105], here death is classified as the inability of a cell to proliferate. How and when cells die is determined by the DDR, which can vary between different types of tumour and normal cells and within populations of tumour cells[105]. The characteristics of different types of cell death are outlined in Table 2.1. Apoptosis is a highly regulated form of cell death that is an essential and normal part of many physiological processes, which can be induced by irradiation[105]. Autophagy translated means 'self-eating' and refers to a process where cells consume their own cytoplasm. Autophagy has been observed post irradiation although it is not clear if it is the cell trying to survive or dying in this context[105]. Mitotic catastrophe is the process whereby a cell dies while it is dividing, usually due to entering into mitosis with some

accumulated DNA damage[105]. Necrosis occurs when conditions are incompatible with normal cellular processes, i.e. exposure to radiation[105]. When cells permanently lose the ability to divide they are classified as senescent[105], radiation induced DNA damage can cause senescence in cells[106].

Table 2.1 Characteristics of different types of cell death from Okada and Mak [107]

Type of cell death	Morphological changes			Biochemical features	Common detection methods
	Nucleus	Cell membrane	Cytoplasm		
Apoptosis	Chromatin condensation; nuclear fragmentation; DNA laddering	Blebbing	Fragmentation (formation of apoptotic bodies)	Caspase dependent	Electron microscopy; TUNEL staining; annexin staining; caspase-activity assays; DNA-fragmentation assays; detection of increased number of cells in subG1/G0; detection of changes in mitochondrial membrane potential
Autophagy	Partial chromatic condensation; no DNA laddering	Blebbing	Increased number of autophagic vesicles	Caspase-independent; increase lysosomal activity	Electron microscopy; protein-degradation assays; assays for marker protein translocation to autophagic membranes; MDC staining

Mitotic catastrophe	Multiple micronuclei; Nuclear fragmentaion	-	-	Caspase-independent (at early stage) abnormal CDK1/cyclin B activation	Electron microscopy; assays for mitotic markers (MPM2); TUNEL staining
Necrosis	Clumping and random degradation of nuclear DNA	Swelling; rupture	Increased vacuolation; mitochondrial swelling	-	Electron microscopy; nuclear staining (usually negative); detection of inflammation and damage in surrounding tissues
Senescence	Distinct hetrochromatic structure (senescence associated hetrochromatic foci)	-	Flattening and increased granularity	SA-β-gal activity	Electron microscopy; SA-β-gal staining; growth-arrest assays; assays for increased p53, INK4A and ARF levels (usually increased); assays for RB phosphorylation (usually hypophosphorykated); assays for metalloproteinase activity (usually upregulated)

CDK1, cycline-dependent kinase 1; MDC, monodansylcadaverine; MPM2, mitotic phosphoprotein 2; SA-β-gal, senescence-associated β-galactosidase; RB retinoblastoma protein.

2.5.2 The Rs of radiobiology

Fractionation in radiotherapy was a consequence of technological limitations of early X-ray equipment[108] but was later developed through experiments performed in France in the 1920s[109], the goal of these experiments was to sterilise rams using kV radiation. It was observed that skin damage could be reduced if the total dose was divided into multiple small fractions. It was after this that fractionation began to be used in radiotherapy, exploiting repair and repopulation to spare normal tissues and reoxygenation and redistribution to damage the tumour[86]. The Rs of radiobiology (Figure 1.4) are; repair, repopulation, redistribution, reoxygenation and radiosensitivity. Repair refers to the process by which the function of a cell is restored after acquiring some damage from irradiation. Radiation can cause single and double strand breaks to DNA, 1 Gy will cause about 1000 single strand breaks and 40 DSB[110]. Depending on the type of strand break the cell may employ excision repair, mismatch repair or recombination repair[86]. Repopulation refers to the process whereby surviving cells, after irradiation, begin to proliferate. Redistribution or reassortment of cells within the cell cycle is a regular occurrence in homeostasis. It is important in radiotherapy however as different phases of the cell cycle are more sensitive to radiation than others with M phase most sensitive and S most resistant[86]. Reoxygenation of cells is important in radiotherapy as oxygenated cells are more sensitive to radiation damage and

hypoxic cells are more resistant[86]. There are a multiplicity of factors that influence radiosensitivity of human tumours which are broadly classified into tumour (hypoxia, tumour kinetics and number of clonegens), host (defence, volume effect and genetic predisposition) and treatment (dose, type of radiation and fractionation) factors[86].

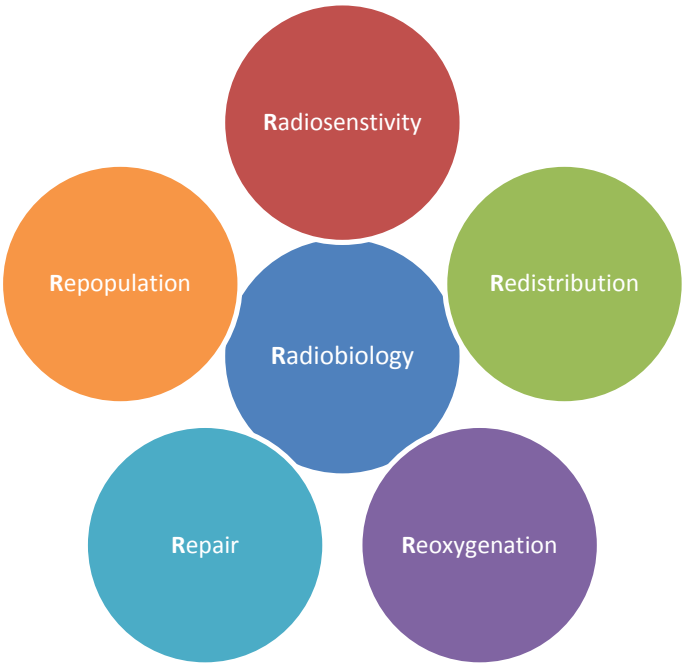


Figure 1.4 The R's of radiobiology

2.5.3 The linear quadratic model

The linear quadratic (LQ) model uses a second order polynomial with a zero constant term to fit cell survival data[111]. The formula for cell survival is then:

$$-\ln(S) = \alpha D + \beta D^2$$
$$p(survival) = e^{-\alpha D - \beta D^2} \quad (5)$$

Where D is the dose and α and β are constants. The LQ model is favoured over other power law models because it gives a more accurate description of cell survival for low doses[111]. The shape of the curve (Figure 2.3) for this model is determined by the ratio of α Gy^{-1} and β Gy^{-2}, α/β Gy can be seen in figure 10 as the point on the curve at which the damage from the linear and quadratic components is equal[111]. This model has been in wide spread use for a number of years owing to its ability to accurately predict radiation response both *in vitro* and *in vivo*[111].

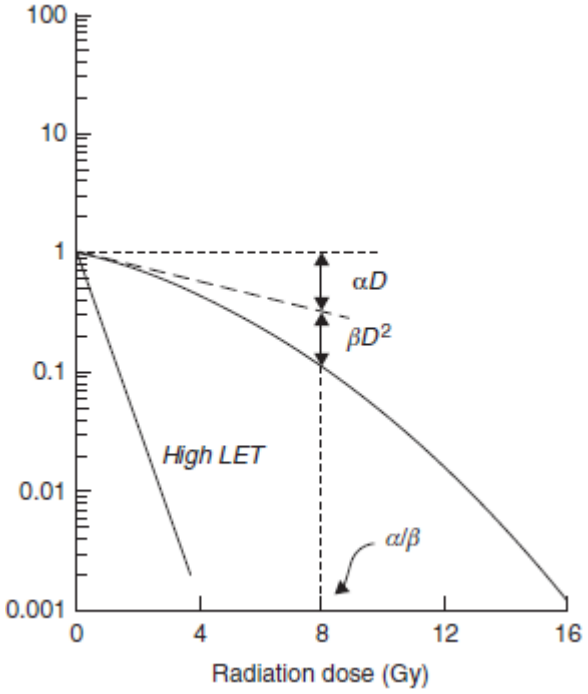


Figure 2.3 The linear Quadratic model
from Joiner, 2009

2.5.4 Equivalent Uniform Dose (EUD)

EUD was proposed by Niemierko in 1997 and is defined as the uniform dose that, if delivered over the same number of fractions as the non-uniform dose distribution of interest, yields the same radiobiological effect[112]. In 1999 Niemierko extended the notion of EUD to normal tissues with the generalised EUD[113]:

$$gEUD = \left(\sum_i v_i D_i^a\right)^{\frac{1}{a}} \quad (6)$$

Where v_i is the fractional organ volume receiving the dose D_i and a is a parameter describing the volume effect which is tissue specific.

2.5.5 Tumour Control Probability (TCP)

Radiation dose response curves are sigmoidal in shape with the likelihood of a radiation effect increasing with increasing dose. There are three standard approaches that have been commonly used to mathematically model dose-response; Poisson, logistic and probit[114, 115]. The only model with a radiobiological background is the Poisson model as it is based on the Poisson statistical model of cell kill[116]:

$$P(D) = 2^{-e\left[e\gamma\left(1-D/D_{50}\right)\right]} \quad (7)$$

Where D_{50} is the 50% response dose and γ is the maximum value of the normalised dose response gradient. The logistic model is widely used in biology applications, outside of radiation oncology, for estimating response probabilities[115]. One of the drawbacks of this model is that there is no simple mechanistic basis and, therefore, no biological interpretation of its parameters[115]. Despite this, the logistic model enjoys widespread use in radiobiology to describe dose response in empirical TCP models. For example, Källman *et al* [116] used:

$$P(D) = \frac{1}{\left[1 + \frac{D_{50}}{D}\right]^{4\gamma}} \tag{8}$$

The probit model has been used for its ease of computation when approximating the Poisson model[116, 117]. It is also useful for estimating the impact of dosimetric and biological uncertainties[116, 117]:

$$P(D) = \frac{1}{2}\left[1 - Erf\left[\sqrt{\pi\gamma}\left(1 - \frac{D}{D_{50}}\right)\right]\right] \tag{9}$$

2.5.6 Normal tissue complication probability (NTCP)

One of the most widely used NTCP models, particularly in north America, is the Lyman model[118]. This model calculates NTCP as function of uniformly irradiated dose in a fractional organ volume[119]:

$$NTCP = (2\pi)^{-1/2} \int_{-\infty}^{t} e^{\left(-x^2/2\right)} dx$$

Where;

$$t = \left(D - TD_{50}(v)\right) \Big/ \left(m \cdot TD_{50}(v)\right) \tag{10}$$

$$TD_{50}(v) = TD_{50}(1) \cdot v^{-n}$$

Where D is the dose to the irradiated volume fraction v, the m parameter determines the slope of the NTCP curve at 50% complications, $TD_{50}(v)$ is the dose that gives an NTCP of 50%. The Lyman model assumes a power law relationship between tolerance dose and irradiated volume fraction although there is no biological basis for this. Instead, it is a mathematically convenient technique that agrees with clinical data, here n is restricted to values 0-1[120]. In order to use this model with non-uniform dose distributions, histogram reduction techniques are used, the most common of which is the Kutcher-Burman[121] method.

The volume and structure of tissue irradiated is an important factor when considering clinical tolerance[122]. The concept of functional subunits (FSUs) was introduced by Withers *et al*[123] in 1988. FSUs are defined (with respect to tumour) as the largest tissue volume, or unit of cells, that can be regenerated from a single surviving clonogenic cell. Within an organ FSUs can be arranged in a parallel or serial architecture. In a parallel architecture it is thought that FSUs function independently[122], therefore, a threshold volume (i.e. the number of

irradiated FSUs) must be considered. The risk of complication in a parallel organ depends on the total dose and is less influenced by hot spots. Parallel organs include the kidney, liver and lung[122]. In serial organs the function of that organ is dependent on each individual FSU. Serial organs include the spinal cord, intestine and oesophagus[122]. As the function of the organ depends on the function of each FSU, hotspots are important in predicting clinical response[122].

Kallman *et al*[116] introduced the relative seriality or s-model in 1992 which was designed to describe the response of an organ with a mixture of serial and parallel FSUs:

$$NTCP = \left\{1 - \prod_{I} [1 - P(D_i)^s]^{v_i}\right\}^{1/s} \quad (11)$$

Where, v_i is the organ volume receiving a dose D_i and $P(D_i)$ is the complication. The parameter s describe the relative contributions of the serial and parallel tissue architectures with a value of one for completely serial and zero for completely parallel[114].

Emami *et al*[124] in 1991 published a paper that outlined normal tissue radiation tolerance doses according to how much of the organ is irradiated, 1/3, 2/3 or the whole volume. Due the paucity of available data for all relevant organs the expert panel took the approach of using consensus to determine the tolerance

doses. In the same issue of the journal Burman *et al* fit the Lyman model to the data presented by Emami *et al* to provide estimated NTCP. The Quantitative Analysis of Normal Tissue Effects in the Clinic (QUANTEC) report was published in the international journal or radiation oncology biology physics in 2010[125]. This was a series of reviews and vision papers that aimed to provide focused summaries of dose/volume/outcome data for a number of organs relevant to radiotherapy as the first significant update since the Emami data[125]. Some of the values reported in the special edition include lung whole organ V20≤30% to ensure no greater that 20% chance of symptomatic pneumonitis or rectal whole organ V75<15% for <15% chance of ≥ grade 2 late rectal toxicity[126]. Some limitations, areas for improvement and opportunities for future research were also identified[127-131].

2.6 Radiotherapy clinical trials

The first randomised, medical therapeutic clinical trial was run by Hill in 1946-48 and demonstrated that streptomycin was superior to bed rest alone for the treatment of tuberculosis[132], Sir Austin Bradford Hill said of the trial that it "*can be seen to have ushered in a new era of medicine*". The earliest trials in radiation oncology were conducted in Manchester, England, in 1948[133], and involved the investigation of breast cancer. There were a number of trials conducted in North America in the 1950s examining the role of radiotherapy in breast and lung cancer[133]. Two trials in the 1960s made an impact on patient

management, an early Hodgkin disease and locally advanced prostate cancer trial where radiotherapy was investigated as a primary curative treatment[133]. After a number of years of smaller national and cooperative groups in Europe and North America performing trials there was a need for a more unified multidisciplinary approach. This new approach saw the formation of the radiation oncology group (ROG)[134] within the European organisation for research and treatment of cancer (EORTC) and radiation therapy oncology group (RTOG)[133] within the National Cancer Institute (NCI). In Australia and New Zealand the Trans-Tasman radiation oncology group (TROG) was formed in 1989. All of these organisations recognise the important role of QA in ensuring the quality of radiotherapy trial data in terms of integrity, consistency, reliability and accuracy[134].

2.6.1 Trial design and sample size calculations

The results of clinical trials underpin the modern healthcare system, it is therefore desirable that they are designed and run punctiliously. The elements of good trial design include[132, 135, 136]:

I. clearly stated objectives, specification of eligibility
II. treatments and endpoints
III. determination of detectable treatment difference
IV. specification of treatment assignment
V. sample size assumptions
VI. reporting

There are a number of different types of clinical trial, which may be broadly classified into 'phases' based on the general intent of the trial. Phase I trials are designed to determine the maximum tolerated dose of a new agent[132]. Phase II trials are used to test new treatments that show promise for an anti-cancer effect[132]. The goal of phase III or randomised controlled trials is to compare treatment regimens. From the CONSORT 2010 statement[135] "*Randomised controlled trials, when appropriately designed, conducted, and reported, represent the gold standard in evaluating healthcare interventions*", the statement goes on to say "*however, randomised trials can yield biased results if they lack methodological rigour*".

One of the key areas in which RCTs can lack rigour is sample size calculation[137]. The four basic components of a sample size calculation for a comparative study are; Type I error (α), power, event rate in the control group, and a treatment effect[138]. A Type I error (α) is defined as a false positive (treatment A is found to be superior to treatment B when, in fact, it is not) and a Type II error (β) is a false negative (treatment A is found to be no better than treatment B when, in fact, it is). Conventionally α is set at 0.05 which equates to a 5% chance of making a false positive conclusion, and β is set to 0.20 or a 20% chance of a false negative conclusion[138]. The power is the probability of rejecting the false negative conclusion and is thus the 1- β, which would equal 0.80 or 80% for the previous example.

2.6.2 Quality assurance in radiotherapy clinical trials

Quality assurance in radiotherapy clinical trials has increased in recent times through cooperative trial research groups like TROG, RTOG, and EORTC. There are also efforts underway to harmonise quality assurance processes internationally to create a more homogeneous approach[35, 139]. It was demonstrated by Peters *et al*[140] that poor quality non-compliant head and neck radiotherapy was associated with a 20% reduction in overall survival. A conclusion that was only possible due to the availability of trial QA data for retrospective analysis. Furthermore it has been shown through secondary analysis that protocol deviations may predict poor outcomes[141].

2.6.2.1 Types of quality assurance strategies

Quality assurance requirements for sites wishing to participate in EORTC clinical trials have been classified into five different levels[142]. Level 1 consists of a facility questionnaire and an external reference dosimetry audit. Level 2 is a benchmarking or dummy run exercise. Level 3 involves performing case reviews or audits on a limited number of cases. Level 4 requires extensive case review or audit. Level 5 involves performing a complex dosimetry audit.

The facility questionnaire usually consists of a structured document that is filled in by a participating institution with information pertaining to; available technology, treatment techniques, staffing, and treatment workload[142].

Dosimetry audits have been carried out by various organisations for many years[143-145] as there are many local department factors that can influence calibration, including; staff skill level, available equipment, adherence to protocols and secondary standards laboratory used[142]. Benchmarking or dummy run exercises involve providing trial investigators with data from a typical case and asking them to 'treat' the case using the trial protocol[146, 147]. Benchmarking exercises can be performed at any time during trial recruitment but are ideally run before site activation. If there are large deviations from trial protocol the site can be notified and the benchmarking repeated. Benchmarking exercises can also be useful in drawing attention to shortcoming and ambiguities in the protocol[142]. Case review or audit involves planning data being sent to a centralised facility for review of compliance with trial protocol. There are a number of treatment planning items that can be verified using case audits including contouring of targets and OARs, dosimetry, imaging, and planning techniques[142]. Complex dosimetry checks are performed to ensure that departments can actually plan and deliver complicated radiotherapy treatments. These typically involve generating a plan on a physical phantom and then delivering that plan to the phantom and measuring the dose[148].

2.6.2.2 Impact of quality on clinical trial outcome

It has been reported that the quality of the radiotherapy delivered in a clinical trial can impact on the outcome of that trial [140, 146, 149]. Further, a decrease in variation in absorbed dose in a clinical trial can lead to a significant reduction in

the sample size required to answer the trial question[36]. In a meta-analysis of eight cooperative group trials Ohri *et al*[141] reported that protocol deviations were associated with increased risk of treatment failure and increased mortality. In a review of EORTC dummy run literature Fairchild *et al* [146] reported that if a centre had taken part in a credentialing exercise they were more likely have positive results in future individual case audits.

2.7 References

1. Cancer IAfRo. World cancer report 2014. Geneva: WHO. 2014.

2. Thwaites D, editor Accuracy required and achievable in radiotherapy dosimetry: have modern technology and techniques changed our views? Journal of Physics: Conference Series; 2013: IOP Publishing.

3. Weber DC, Hurkmans CW, Melidis C, Budach W, Langendijk JH, Peters LJ, et al. Outcome impact and cost-effectiveness of quality assurance for radiotherapy planned for the EORTC 22071–24071 prospective study for head and neck cancer. Radiotherapy and Oncology. 2014.

4. Van Dye J, Batista J, Bauman GS. Accuracy and uncertainty considerations in modern radiation oncology. The Modern Technology of Radiation Oncology. 2013;3:361-412.

5. Brahme A. The need for accurate target and Dose specifications in conventional and conformal radiation therapy: An introduction. Acta Oncologica. 1997;36(8):789-92.

6. Njeh C. Tumor delineation: The weakest link in the search for accuracy in radiotherapy. Journal of medical physics/Association of Medical Physicists of India. 2008;33(4):136.

7. Weiss E, Hess CF. The impact of gross tumor volume (GTV) and clinical target volume (CTV) definition on the total accuracy in radiotherapy theoretical

aspects and practical experiences. Strahlentherapie und Onkologie. 2003;179(1):21-30.

8. Spoelstra FOB, Senan S, Le Péchoux C, Ishikura S, Casas F, Ball D, et al. Variations in Target Volume Definition for Postoperative Radiotherapy in Stage III Non-Small-Cell Lung Cancer: Analysis of an International Contouring Study. International Journal of Radiation Oncology, Biology, Physics. 2009;76(4):1106-13.

9. Nelms BE, Tomé WA, Robinson G, Wheeler J. Variations in the contouring of organs at risk: test case from a patient with oropharyngeal cancer. International Journal of Radiation Oncology, Biology, Physics. 2010.

10. Foppiano F, Fiorino C, Frezza G, Greco C, Valdagni R. The impact of contouring uncertainty on rectal 3D dose–volume data: Results of a dummy run in a multicenter trial (AIROPROS01–02). International journal of radiation oncology, biology, physics. 2003;57(2):573-9.

11. Nelms BE, Robinson G, Markham J, Velasco K, Boyd S, Narayan S, et al. Variation in external beam treatment plan quality: an inter-institutional study of planners and planning systems. Practical radiation oncology. 2012;2(4):296-305.

12. ICRU Report 50: Prescribing, Recording and Reporting Photon Beam Therapy. ICRU Publ Bethesda MD. 1993.

13. ICRU Report 62. Prescribing, recording, and reporting photon beam therapy (Supplement to ICRU Report 50). Bethesda, MD: International Commission on Radiation Units and Measurements, 1999.

14. Hodapp N. [The ICRU Report 83: prescribing, recording and reporting photon-beam intensity-modulated radiation therapy (IMRT)]. Strahlentherapie und Onkologie. 2012(188):97-9.

15. van Lin ENT, van der Vight LP, Witjes JA, Huisman HJ, Leer JW, Visser AG. The effect of an endorectal balloon and off-line correction on the interfraction systematic and random prostate position variations: a comparative study. International Journal of Radiation Oncology* Biology* Physics. 2005;61(1):278-88.

16. Peters LJ, O'Sullivan B, Giralt J, Fitzgerald TJ, Trotti A, Bernier J, et al. Critical impact of radiotherapy protocol compliance and quality in the treatment

of advanced head and neck cancer: Results from TROG 02.02. Journal of Clinical Oncology. 2010;28(18):2996.

17. Creutzberg C. Randomized Trial of Radiation Therapy With or Without Chemotherapy for Endometrial Cancer (PORTEC-3).

18. Fotina I, Lütgendorf-Caucig C, Stock M, Pötter R, Georg D. Critical discussion of evaluation parameters for inter-observer variability in target definition for radiation therapy. Strahlentherapie und Onkologie. 2012;188(2):160-7.

19. Barber J, Vial PJ. A survey of modulated therapy use in Australia & New Zealand. EPSM; Wellington2015.

20. Webb S. Intensity-modulated radiation therapy: CRC Press; 2001.

21. Ezzell GA, Galvin JM, Low D, Palta JR, Rosen I, Sharpe MB, et al. Guidance document on delivery, treatment planning, and clinical implementation of IMRT: report of the IMRT Subcommittee of the AAPM Radiation Therapy Committee. Medical physics. 2003;30(8):2089-115.

22. Bohsung J, Gillis S, Arrans R, Bakai A, De Wagter C, Knöös T, et al. IMRT treatment planning—A comparative inter-system and inter-centre planning exercise of the ESTRO QUASIMODO group. Radiotherapy and oncology. 2005;76(3):354-61.

23. Bolla M, van Poppel H, Tombal B, Vekemans K, Da Pozzo L, de Reijke TM, et al. Postoperative radiotherapy after radical prostatectomy for high-risk prostate cancer: long-term results of a randomised controlled trial (EORTC trial 22911). The Lancet. 2012.

24. Kneebone A. Radiotherapy - Adjuvant Versus Early Salvage (RAVES).

25. Parker C. Radiation Therapy and Androgen Deprivation Therapy in Treating Patients Who Have Undergone Surgery for Prostate Cancer (RADICALS).

26. Haworth A, Paneghel A, Herschtal A, Duchesne G, Williams S, Tai K, et al. Verification of target position in the post-prostatectomy cancer patient using cone beam CT. Journal of medical imaging and radiation oncology. 2009;53(2):212-20.

27. Smeenk RJ, Louwe RJW, Langen KM, Shah AP, Kupelian PA, van Lin ENJ, et al. An Endorectal Balloon Reduces Intrafraction Prostate Motion During Radiotherapy. International Journal of Radiation Oncology* Biology* Physics. 2011;83(2):661-9.

28. Plint AC, Moher D, Morrison A, Schulz K, Altman DG, Hill C, et al. Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. Medical journal of Australia. 2006;185(5):263.

29. Schulz K, Altman D, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMC medicine. 2010;8(1):18.

30. Dieckmann K, Pötter R, Wagner W, Prott FJ, Hörnig-Franz I, Rath B, et al. Up-front centralized data review and individualized treatment proposals in a multicenter pediatric Hodgkin's disease trial with 71 participating hospitals: the experience of the German-Austrian pediatric multicenter trial DAL-HD-90. Radiotherapy and Oncology. 2002;62(2):191-200.

31. Dusserre A, Garavaglia G, Giraud JY, Bolla M. Quality assurance of the EORTC radiotherapy trial 22863 for prostatic cancer: the dummy run. Radiotherapy and Oncology. 1995;36(3):229-34.

32. Li XA, Tai A, Arthur DW, Buchholz TA, Macdonald S, Marks LB, et al. Variability of target and normal structure delineation for breast cancer radiotherapy: an RTOG Multi-Institutional and Multiobserver Study. International Journal of Radiation Oncology* Biology* Physics. 2009;73(3):944-51.

33. Valley JF, Bernier J, Tercier PA, Fogliata-Cozzi A, Rosset A, Garavaglia G, et al. Quality assurance of the EORTC radiotherapy trial 22931 for head and neck carcinomas: the dummy run. Radiotherapy and Oncology. 1998;47(1):37-44.

34. Fairchild A, Straube W, Laurie F, Followill D. Does quality of radiation therapy predict outcomes of multicenter cooperative group trials? A literature review. International Journal of Radiation Oncology* Biology* Physics. 2013;87(2):246-60.

35. Bekelman JE, Deye JA, Vikram B, Bentzen SM, Bruner D, Curran Jr WJ, et al. Redesigning radiotherapy quality assurance: opportunities to develop an efficient, evidence-based system to support clinical trials—report of the National Cancer Institute Work Group on Radiotherapy Quality Assurance.

International Journal of Radiation Oncology* Biology* Physics. 2012;83(3):782-90.

36. Pettersen MN, Aird E, Olsen DR. Quality assurance of dosimetry and the impact on sample size in randomized clinical trials. Radiotherapy and Oncology. 2008;86(2):195-9.

37. Barghi A, Johnson C, Warner A, Bauman G, Battista J, Rodrigues G. Impact of Contouring Variability on Dose-Volume Metrics used in Treatment Plan Optimization of Prostate IMRT. Cureus. 2013;5(11).

38. Jameson MG, Kumar S, Vinod SK, Metcalfe PE, Holloway LC. Correlation of contouring variation with modeled outcome for conformal non-small cell lung cancer radiotherapy. Radiotherapy and Oncology. 2014.

39. Stanley J, Dunscombe P, Lau H, Burns P, Lim G, Liu H-W, et al. The Effect of Contouring Variability on Dosimetric Parameters for Brain Metastases Treated With Stereotactic Radiosurgery. International Journal of Radiation Oncology* Biology* Physics. 2013;87(5):924-31.

40. Cancer in Australia: an overview, 2012. Australian Institute of Health and Welfare & Australasian Association of Cancer Registries 2012, Contract No.: 70.

41. Barton M, Jacob S, Shafiq J, Wong K, Delaney G, Hanna T. Review of optimal radiotherapy utilisation rates. Ingham Institute for Applied Medical Research, Liverpool. 2013.

42. Thariat J, Hannoun-Levi J-M, Myint AS, Vuong T, Gérard J-P. Past, present, and future of radiotherapy for the benefit of patients. Nature reviews Clinical oncology. 2013;10(1):52-60.

43. Roentgen W. On a new kind of ray (first report). Münchener medizinische Wochenschrift (1950). 1959;101:1237.

44. Lederman M. The early history of radiotherapy: 1895–1939. International Journal of Radiation Oncology* Biology* Physics. 1981;7(5):639-48.

45. Grubbe EH. Priority in the therapeutic use of X-rays. Radiology. 1933;21(2):156-62.

46. Thorp N. Proton therapy for childhood malignancies: including a historical background summary. 2011.

47. Fletcher GH. Regaud lecture perspectives on the history of radiotherapy. Radiotherapy and Oncology. 1988;12(4):253-71.

48. Bernier J, Hall EJ, Giaccia A. Radiation oncology: a century of achievements. Nature Reviews Cancer. 2004;4(9):737-47.

49. Thwaites DI, Tuohy JB. Back to the future: the history and development of the clinical linear accelerator. Phys Med Biol. 2006;51(13):R343.

50. Hounsfield GN. Computed medical imaging. Journal of computer assisted tomography. 1980;4(5):665-74.

51. Brahme A. Optimization of stationary and moving beam radiation therapy techniques. Radiotherapy and Oncology. 1988;12(2):129-40.

52. Webb S. Contemporary IMRT: developing physics and clinical implementation: CRC Press; 2004.

53. Chang SX, Cullip TJ, Deschesne KM, Miller EP, Rosenman JG. Compensators: An alternative IMRT delivery technique. Journal of Applied Clinical Medical Physics. 2004;5(3).

54. Shepard D, Earl M, Li X, Naqvi S, Yu C. Direct aperture optimization: a turnkey solution for step-and-shoot IMRT. Medical physics. 2002;29(6):1007-18.

55. LoSasso T, Chui C-S, Ling CC. Physical and dosimetric aspects of a multileaf collimation system used in the dynamic mode for implementing intensity modulated radiotherapy. Medical physics. 1998;25(10):1919-27.

56. Yu CX. Intensity-modulated arc therapy with dynamic multileaf collimation: an alternative to tomotherapy. Phys Med Biol. 1995;40(9):1435.

57. Otto K. Volumetric modulated arc therapy: IMRT in a single gantry arc. Medical physics. 2008;35(1):310-7.

58. Gupta T, Narayan CA. Image-guided radiation therapy: Physician's perspectives. Journal of medical physics/Association of Medical Physicists of India. 2012;37(4):174.

59. Hunter KU, Eisbruch A. Advances in imaging: target delineation. The Cancer Journal. 2011;17(3):151-4.

60. Konert T, Vogel W, MacManus MP, Nestle U, Belderbos J, Grégoire V, et al. PET/CT imaging for target volume delineation in curative intent radiotherapy of non-small cell lung cancer: IAEA consensus report 2014. Radiotherapy and Oncology. 2015.

61. Min M, Lin P, Lee MT, Shon IH, Lin M, Forstner D, et al. Prognostic role of metabolic parameters of 18F-FDG PET-CT scan performed during radiation therapy in locally advanced head and neck squamous cell carcinoma. European journal of nuclear medicine and molecular imaging. 2015:1-11.

62. Lauterbur PC. Image formation by induced local interactions: examples employing nuclear magnetic resonance. Nature. 1973;242(5394):190-1.

63. Garroway AN, Grannell PK, Mansfield P. Image formation in NMR by a selective irradiative process. Journal of Physics C: Solid State Physics. 1974;7(24):L457.

64. Metcalfe P, Liney G, Holloway L, Walker A, Barton M, Delaney G, et al. The potential for an enhanced role for MRI in radiation-therapy treatment planning. Technology in cancer research & treatment. 2013;12(5):429-46.

65. Liney GP, Moerland MA, editors. Magnetic resonance imaging acquisition techniques for radiotherapy planning. Seminars in radiation oncology; 2014: Elsevier.

66. Moffat BA, Chenevert TL, Lawrence TS, Meyer CR, Johnson TD, Dong Q, et al. Functional diffusion map: a noninvasive MRI biomarker for early stratification of clinical brain tumor response. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(15):5524-9.

67. Dowling JA, Lambert J, Parker J, Salvado O, Fripp J, Capp A, et al. An atlas-based electron density mapping method for magnetic resonance imaging (MRI)-alone treatment planning and adaptive MRI-based prostate radiation therapy. International Journal of Radiation Oncology* Biology* Physics. 2012;83(1):e5-e11.

68. Walker A, Liney G, Metcalfe P, Holloway L. MRI distortion: considerations for MRI based radiotherapy treatment planning. Australasian Physical & Engineering Sciences in Medicine. 2014;37:103-13.

69. Sharp G, Fritscher KD, Pekar V, Peroni M, Shusharina N, Veeraraghavan H, et al. Vision 20/20: perspectives on automated image segmentation for radiotherapy. Medical physics. 2014;41(5):050902.

70. Van Dyk J, Batista J, Bauman GS. Accuracy and uncertainty considerations in modern radiation oncology. The Modern Technology of Radiation Oncology. 2013;3:361-412.

71. Van Baardwijk A, Bosmans G, Boersma L, Buijsen J, Wanders S, Hochstenbag M, et al. Pet-ct–based auto-contouring in non–small-cell lung cancer correlates with pathology and reduces interobserver variability in the delineation of the primary tumor and involved nodal volumes. International Journal of Radiation Oncology* Biology* Physics. 2007;68(3):771-8.

72. Units ICoR, (ICRU) M. Prescribing, Recording, and Reporting Photon-Beam Intensity-Modulated Radiation Therapy (IMRT): ICRU 83: Oxford University Press; 2010.

73. Van Herk M, editor Errors and margins in radiotherapy. Seminars in radiation oncology; 2004: Elsevier.

74. Hamilton C, Ebert M. Volumetric uncertainty in radiotherapy. Clinical Oncology. 2005;17(6):456-64.

75. Stroom JC, Heijmen BJ. Geometrical uncertainties, radiotherapy planning margins, and the ICRU-62 report. Radiotherapy and oncology. 2002;64(1):75-83.

76. Wambersie A, Dutreix J, Dutreix A. Dosimetric precision required in radiotherapy. Consequences of the choice and performance required of detectors. Journal belge de radiologie. 1969;52(2):94.

77. Flamant R, Malaise E, Dutreix J, Hayem M, Pierquin B, Tubiana M, et al. A clinical therapeutic study of irradiating tonsil neoplasms with 20 MeV photon or electron beans. European journal of cancer. 1967;3(3):169.

78. Brahme A. Dosimetric precision requirements in radiation therapy. Acta Oncologica. 1984;23(5):379-91.

79. Fraass B, Doppke K, Hunt M, Kutcher G, Starkschall G, Stern R, et al. American Association of Physicists in Medicine Radiation Therapy Committee Task Group 53: quality assurance for clinical radiotherapy treatment planning. Medical physics. 1998;25(10):1773-829.

80. Feuvret L, Noël G, Mazeron J-J, Bey P. Conformity index: a review. International Journal of Radiation Oncology* Biology* Physics. 2006;64(2):333-42.

81. Leung LHT, Kan MWK, Cheng ACK, Wong WKH, Yau CC. A new dose–volume-based Plan Quality Index for IMRT plan comparison. Radiotherapy and Oncology. 2007;85(3):407-17.

82. Park Y-K, Park S, Wu H-G, Kim S. A new plan quality index for dose painting radiotherapy. Journal of Applied Clinical Medical Physics. 2014;15(4).

83. Holloway L, Jameson M, Batumalai V, Koh E, Papadatos G, Lonergan D, et al. Estimating a Delineation Uncertainty Margin to Account for Inter-observer Variability in Breast Cancer Radiotherapy. International Journal of Radiation Oncology* Biology* Physics. 2010;78(3):S741.

84. Sharpe MB, Craig T, Moseley DJ. Image guidance: treatment target localization systems. 2007.

85. Korreman SS. Motion in radiotherapy: photon therapy. Phys Med Biol. 2012;57(23):R161.

86. Marcu L, Bezak E, Allen BJ. Biomedical Physics in Radiotherapy for Cancer: Csiro publishing; 2012.

87. Jaffray DA. Image-guided radiotherapy: from current concept to future perspectives. Nature Reviews Clinical Oncology. 2012;9(12):688-99.

88. Mageras GS, editor Introduction management of target localization uncertainties in external-beam therapy. Seminars in radiation oncology; 2005: WB Saunders.

89. De Los Santos J, Popple R, Agazaryan N, Bayouth JE, Bissonnette J-P, Bucci MK, et al. Image guided radiation therapy (IGRT) technologies for radiation therapy localization and delivery. International Journal of Radiation Oncology* Biology* Physics. 2013;87(1):33-45.

90. Bortfeld T, Jiang SB, Rietzel E, editors. Effects of motion on the total dose distribution. Seminars in radiation oncology; 2004: Elsevier.

91. Colvill E, Poulsen PR, Booth J, O'Brien R, Ng J, Keall P. DMLC tracking and gating can improve dose coverage for prostate VMAT. Medical physics. 2014;41(9):091705.

92. Fuss M, Salter BJ, Cheek D, Sadeghi A, Hevezi JM, Herman TS. Repositioning accuracy of a commercially available thermoplastic mask system. Radiotherapy and oncology. 2004;71(3):339-45.

93. Heinzerling JH, Anderson JF, Papiez L, Boike T, Chien S, Zhang G, et al. Four-dimensional computed tomography scan analysis of tumor and organ motion at varying levels of abdominal compression during stereotactic treatment of lung and liver. International Journal of Radiation Oncology* Biology* Physics. 2008;70(5):1571-8.

94. Song DY, Herfarth KK, Uhl M, Eble MJ, Pinkawa M, Van Triest B, et al. A multi-institutional clinical trial of rectal dose reduction via injected polyethylene-glycol hydrogel during intensity modulated radiation therapy for prostate cancer: analysis of dosimetric outcomes. International Journal of Radiation Oncology* Biology* Physics. 2013;87(1):81-7.

95. Mah K, Henkelman RM, editors. Time varying dose due to respiratory motion during radiation therapy of the thorax. Proceedings of the 8th International Conference on the Use of Computers in Radiation Therapy; 1984.

96. Peltola SMS, editor Gated radiotherapy to compensate for patient breathing. Proceedings of the Eleventh Varian Users Meeting; 1986.

97. Rekonen A, Toivonen J, editors. Breathing gated radiation therapy: possibilities and need. Proc XIV Int Conf on Medical and Biological Engineering and VII Int Conf on Medical Physics; 1885.

98. Keall PJ, Mageras GS, Balter JM, Emery RS, Forster KM, Jiang SB, et al. The management of respiratory motion in radiation oncology report of AAPM Task Group 76a). Medical physics. 2006;33(10):3874-900.

99. Hoogeman M, Prévost J-B, Nuyttens J, Pöll J, Levendag P, Heijmen B. Clinical accuracy of the respiratory tumor tracking system of the cyberknife: assessment by analysis of log files. International Journal of Radiation Oncology* Biology* Physics. 2009;74(1):297-303.

100. Keall PJ, Colvill E, O'Brien R, Ng JA, Poulsen PR, Eade T, et al. The first clinical implementation of electromagnetic transponder-guided MLC tracking. Medical physics. 2014;41(2):020702.

101. Joiner MC, van der Kogel AJ, Steel GG. Introduction: the significance of radiobiology and radiotherapy for cancer treatment. Basic Clinical Radiobiology Fourth Edition, Hodder Arnold Publication, London. 2009:1-10.

102. Dale RG, Jones B. Radiobiological modelling in radiation oncology: British Inst of Radiology; 2007.

103. Luxton G, Hancock SL, Boyer AL. Dosimetry and radiobiologic model comparison of IMRT and 3D conformal radiotherapy in treatment of carcinoma of the prostate. International Journal of Radiation Oncology* Biology* Physics. 2004;59(1):267-84.

104. van Vulpen M, Field C, Raaijmakers CP, Parliament MB, Terhaard CH, MacKenzie MA, et al. Comparing step-and-shoot IMRT with dynamic helical tomotherapy IMRT plans for head-and-neck cancer. International Journal of Radiation Oncology* Biology* Physics. 2005;62(5):1535-9.

105. Wouters BG. Cell death after irradiation: how, when and why cells die. Basic Clinical Radiobiology 4th ed London: Hodder Education. 2009:27-40.

106. Campisi J, di Fagagna FdA. Cellular senescence: when bad things happen to good cells. Nature reviews Molecular cell biology. 2007;8(9):729-40.

107. Okada H, Mak TW. Pathways of apoptotic and non-apoptotic death in tumour cells. Nature Reviews Cancer. 2004;4(8):592-603.

108. Jameson MG, Holloway LC, Vial PJ, Vinod SK, Metcalfe PE. A review of methods of analysis in contouring studies for radiation oncology. Journal of medical imaging and radiation oncology. 2010;54(5):401-10.

109. Metcalfe P, Kron T, Hoban P. The physics of radiotherapy x-rays and electrons: Medical Physics Publ.; 2012.

110. Olive PL. The role of DNA single-and double-strand breaks in cell killing by ionizing radiation. Radiation research. 1998;150(5s):S42-S51.

111. Joiner MC. Quantifying cell kill and cell survival. Basic clinical radiobiology. 2009;4:41-55.

112. Niemierko A. Reporting and analyzing dose distributions: a concept of equivalent uniform dose. Medical physics. 1997;24(1):103-10.

113. Niemierko A. A generalized concept of equivalent uniform dose (EUD). Med Phys. 1999;26(6):1100.

114. Li XA, Alber M, Deasy JO, Jackson A, Jee K-WK, Marks LB, et al. The use and QA of biologically related models for treatment planning: Short report of the TG-166 of the therapy physics committee of the AAPMa). Medical Physics. 2012;39(3):1386-409.

115. Bentzen SM. Dose–response relationships in radiotherapy. Basic Clinical Radiobiology, 4th ed London: Hodder Arnold. 2009:56-66.

116. Källman P, Ågren A, Brahme A. Tumour and normal tissue responses to fractionated non-uniform dose delivery. International journal of radiation biology. 1992;62(2):249-62.

117. BENTZEN M, SL TUCKER S. Quantifying the position and steepness of radiation dose-response curves. International journal of radiation biology. 1997;71(5):531-42.

118. Lyman JT. Complication probability as assessed from dose-volume histograms. Radiation Research. 1985;104(2s):S13-S9.

119. Kong F-MS, Pan C, Eisbruch A, Haken RKT, editors. Physical Models and Simpler Dosimetric Descriptors of Radiation Late Toxicity. Seminars in Radiation Oncology; 2007: Elsevier.

120. Yorke ED, editor Modeling the effects of inhomogeneous dose distributions in normal tissues. Seminars in radiation oncology; 2001: Elsevier.

121. Kutcher G, Burman C, Brewster L, Goitein M, Mohan R. Histogram reduction method for calculating complication probabilities for three-dimensional treatment planning evaluations. International Journal of Radiation Oncology* Biology* Physics. 1991;21(1):137-46.

122. Dörr W, Van der Kogel AJ. The volume effect in radiotherapy. Basic clinical radiobiology, 4th Aufl Hodder Arnold, London. 2009:191-206.

123. Withers HR, Taylor JM, Maciejewski B. Treatment volume and tissue tolerance. International Journal of Radiation Oncology* Biology* Physics. 1988;14(4):751-9.

124. Batumalai V, Koh E, Delaney G, Holloway L, Jameson M, Papadatos G, et al. Interobserver variability in clinical target volume delineation in tangential breast irradiation: a comparison between radiation oncologists and radiation therapists. Clinical Oncology. 2011;23(2):108-13.

125. Marks LB, Ten Haken RK, Martel MK. Guest editor's introduction to QUANTEC: a users guide. International Journal of Radiation Oncology* Biology* Physics. 2010;76(3):S1-S2.

126. Marks LB, Yorke ED, Jackson A, Ten Haken RK, Constine LS, Eisbruch A, et al. Use of normal tissue complication probability models in the clinic. International Journal of Radiation Oncology* Biology* Physics. 2010;76(3):S10-S9.

127. Jaffray DA, Lindsay PE, Brock KK, Deasy JO, Tomé W. Accurate accumulation of dose for improved understanding of radiation effects in normal tissue. International Journal of Radiation Oncology* Biology* Physics. 2010;76(3):S135-S9.

128. Jeraj R, Cao Y, Ten Haken RK, Hahn C, Marks L. Imaging for assessment of radiation-induced normal tissue effects. International Journal of Radiation Oncology* Biology* Physics. 2010;76(3):S140-S4.

129. Bentzen SM, Parliament M, Deasy JO, Dicker A, Curran WJ, Williams JP, et al. Biomarkers and surrogate endpoints for normal-tissue effects of radiation therapy: the importance of dose–volume effects. International Journal of Radiation Oncology* Biology* Physics. 2010;76(3):S145-S50.

130. Deasy JO, Bentzen SM, Jackson A, Ten Haken RK, Yorke ED, Constine LS, et al. Improving normal tissue complication probability models: the need to adopt a "data-pooling" culture. International Journal of Radiation Oncology* Biology* Physics. 2010;76(3):S151-S4.

131. Jackson A, Marks LB, Bentzen SM, Eisbruch A, Yorke ED, Ten Haken RK, et al. The lessons of QUANTEC: recommendations for reporting and gathering data on dose–volume dependencies of treatment outcome. International Journal of Radiation Oncology* Biology* Physics. 2010;76(3):S155-S60.

132. Green S, Benedetti J, Smith A, Crowley J. Clinical trials in oncology: CRC press; 2012.

133. Cox JD. Brief history of comparative clinical trials in radiation oncology: perspectives from the silver anniversary of the Radiation Therapy Oncology Group. Radiology. 1994;192(1):25-32.

134. Grégoire V, Bartelink H, Bernier J, Bolla M, Bosset J-F, Collette L, et al. EORTC Radiation Oncology Group: 50 years of continuous accomplishments. European Journal of Cancer Supplements. 2012;10(1):150-9.

135. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMC medicine. 2010;8(1):18.

136. Moher D, Schulz KF, Altman DG, Group C. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. The Lancet. 2001;357(9263):1191-4.

137. Freiman JA, Chalmers TC, Smith Jr H, Kuebler RR. The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial: survey of 71 negative trials. New England Journal of Medicine. 1978;299(13):690-4.

138. Schulz KF, Grimes DA. Sample size calculations in randomised trials: mandatory and mystical. The Lancet. 2005;365(9467):1348-53.

139. Melidis C, Bosch WR, Izewska J, Fidarova E, Zubizarreta E, Ishikura S, et al. Radiation therapy quality assurance in clinical trials–Global harmonisation group. Radiotherapy and oncology: journal of the European Society for Therapeutic Radiology and Oncology. 2014;111(3):327.

140. Peters LJ, O'Sullivan B, Giralt J, Fitzgerald TJ, Trotti A, Bernier J, et al. Critical impact of radiotherapy protocol compliance and quality in the treatment of advanced head and neck cancer: results from TROG 02.02. Journal of Clinical Oncology. 2010;28(18):2996-3001.

141. Ohri N, Shen X, Dicker AP, Doyle LA, Harrison AS, Showalter TN. Radiotherapy protocol deviations and clinical outcomes: a meta-analysis of cooperative group clinical trials. Journal of the National Cancer Institute. 2013:djt001.

142. Weber DC, Poortmans PM, Hurkmans CW, Aird E, Gulyban A, Fairchild A. Quality assurance for prospective EORTC radiation oncology trials: the challenges of advanced technology in a multicenter international setting. Radiotherapy and Oncology. 2011;100(1):150-6.

143. Izewska J, Andreo P. The IAEA/WHO TLD postal programme for radiotherapy hospitals. Radiotherapy and oncology. 2000;54(1):65-72.

144. Ebert M, Harrison K, Cornes D, Howlett S, Joseph D, Kron T, et al. Comprehensive Australasian multicentre dosimetric intercomparison: issues, logistics and recommendations. Journal of medical imaging and radiation oncology. 2009;53(1):119-31.

145. Thwaites D, Williams J, Aird E, Klevenhagen S, Williams PC. A dosimetric intercomparison of megavoltage photon beams in UK radiotherapy centres. Phys Med Biol. 1992;37(2):445.

146. Fairchild A, Collette L, Hurkmans C, Baumert B, Weber D, Gulyban A, et al. Do results of the EORTC dummy run predict quality of radiotherapy delivered within multicentre clinical trials? European Journal of Cancer. 2012;48(17):3232-9.

147. Weber DC, Tomsej M, Melidis C, Hurkmans CW. QA makes a clinical trial stronger: evidence-based medicine in radiation therapy. Radiotherapy and Oncology. 2012;105(1):4-8.

148. Ibbott GS, Haworth A, Followill DS. Quality assurance for clinical trials. Frontiers in oncology. 2013;3.

149. Abrams RA, Winter KA, Regine WF, Safran H, Hoffman JP, Lustig R, et al. Failure to adhere to protocol specified radiation therapy guidelines was associated with decreased survival in RTOG 9704—a phase III trial of adjuvant chemotherapy and chemoradiotherapy for patients with resected adenocarcinoma of the pancreas. International Journal of Radiation Oncology* Biology* Physics. 2012;82(2):809-16.

Chapter 3: A review of methods of analysis in contouring studies for radiation oncology

Statement of joint authorship

Title

A review of methods of analysis in contouring studies for radiation oncology

Authors

Michael G Jameson, Lois C Holloway, Philip J Vial, Shalini K Vinod and Peter E Metcalfe

Michael G Jameson (candidate)

Developed literature survey, analysed and interpreted literature and wrote manuscript

Lois C Holloway

Contributed to literature survey and manuscript

Philip J Vial

Contributed to literature survey and manuscript

Shalini K Vinod

Contributed to literature survey and manuscript

Peter E Metcalfe

Contributed to literature survey and manuscript

Chapter 4: Correlation of contouring variation with modelled outcome for conformal non-small cell lung cancer radiotherapy

Statement of joint authorship

Title

Correlation of contouring variation with modelled outcome for conformal non-small cell lung cancer radiotherapy

Authors

Michael G Jameson, Shivani Kumar, Shalini K Vinod, Peter E Metcalfe and Lois C Holloway

Michael G Jameson (candidate)

Developed experimental design and scientific method, collected data, analysed and interpreted data, wrote manuscript

Shivani Kumar

Contributed to scientific method and manuscript

Shalini K Vinod

Contributed to scientific method and manuscript

Peter E Metcalfe

Contributed to scientific method and manuscript

Lois C Holloway

Contributed to experimental design and scientific method, contributed to data analysis and interpretation, contributed to manuscript

Radiotherapy and Oncology 112 (2014) 332–336

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Contouring

Correlation of contouring variation with modeled outcome for conformal non-small cell lung cancer radiotherapy

Michael G. Jameson [a,b,c,*], Shivani Kumar [a,c,d], Shalini K. Vinod [a,c,d], Peter E. Metcalfe [b], Lois C. Holloway [a,b,c,d,e]

[a] Liverpool and Macarthur Cancer Therapy Centres; [b] Centre for Medical Radiation Physics, University of Wollongong; [c] Ingham Institute of Applied Medical Research; [d] South Western Sydney Clinical School, University of New South Wales; and [e] Institute of Medical Physics, School of Physics, University of Sydney, Australia

ARTICLE INFO

Article history:
Received 4 June 2013
Received in revised form 21 March 2014
Accepted 27 March 2014
Available online 19 May 2014

Keywords:
TCP
EUD
Contouring variation
Lung cancer

ABSTRACT

Background and purpose: Contouring variation is a well know uncertainty in modern radiotherapy. This study investigates the relationship between contouring variation, tumor control probability (TCP) and equivalent uniform dose (EUD) for conformal non-small cell lung cancer (NSCLC) radiotherapy.
Material and methods: Seven patients were retrospectively recruited to the study and multiple PTV contours were generated based on CT and PET imaging by three observers. Plans were created for each PTV volume. Volumes were analyzed geometrically using volume, location, dimension and conformity index (CI). Radiobiological plan analysis consisted of two TCP models and EUD. Spearman's correlation coefficient (ρ) was used to quantify the association between geometric variation and radiobiological metrics.
Results: The variation in CI and TCP for the study was 0.66–0.90% and 0.19–0.68%. Changes in lateral dimension and volume were significantly correlated with TCP and EUD with an average ρ of −0.49 and 0.43 ($p < 0.01$) respectively.
Conclusions: TCP and geometric contour variation show significant correlation. This correlation was most significant for changes in lateral dimensions of PTV volumes. This association may be used in the assessment of contouring protocol violations in multicenter clinical trials and aid in the design of future contouring studies.
 Radiotherapy and Oncology 112 (2014) 332–336

Variability in contouring remains one of the largest uncertainties in modern radiotherapy, despite continued research and investigation over many years [1]. Methods of assessment of contouring variation and possible solutions through education, automation and intervention have been investigated extensively and reviewed [1–3]. Variation in contouring has been attributed to; observer experience [4], imaging modality [5,6], the use of guidelines [7], and patient and tumor factors such as site, stage, age and size [8].

Spoelstra et al. [9] investigated the impact of contouring variation on dosimetry, showing that the introduction of a contouring protocol significantly reduced the risk of radiation-induced lung toxicity. Other studies have shown that contouring variation can have a significant impact on dosimetry for head and neck, breast and post-prostatectomy patients [10,11]. Vinod et al. [12] utilized tumor control probability (TCP) in assessing the impact of incorporating F-18 fluorodeoxyglucose positron emission tomography (FDG-PET) on non-small cell lung cancer (NSCLC) planning. They demonstrated an impact of contouring variation on TCP ranging from 1% to 24%.

Understanding the impact of contouring variation on TCP would aid in development of contouring guidelines and adaptive radiotherapy (ART) protocols. It would also assist in the design of clinical trial protocols where contouring variation is a known confounding factor [13,14]. At present, there is no consistent or widely accepted method of contour comparison [1,2]. Fotina et al. [2] investigated the relationships between different comparison metrics in an effort to recommend a minimum parameter set for "full description" of contouring variation. Ideally the choice of metric should correlate with clinical outcome, as different tumor sites and planning techniques will differ in sensitivity to delineation variability. The present study was designed to investigate the relationship between contouring variation and outcome surrogates specifically TCP, equivalent uniform dose (EUD) and mean lung dose using a series of NSCLC patient datasets. The aim was to recommend geometric parameters for the assessment of contouring variation, that relate to clinical outcome.

* Corresponding author. Address: Cancer Therapy Centre, Liverpool Hospital, Locked Bag 7103, Liverpool, NSW 1871, Australia.
E-mail address: michael.jameson@sswahs.nsw.gov.au (M.G. Jameson).

http://dx.doi.org/10.1016/j.radonc.2014.03.019
0167-8140/

Materials and methods

Patient datasets

Seven previously treated NSCLC patients were selected for the study (Supplementary Table 1), five from a prior investigation [12] and two additional. Free breathing CT scans were acquired with contrast using a Siemens Somatom™ (Siemens Medical Solutions, Erlangen, Germany) system with 2 mm slice thickness and a 512 × 512 matrix. Diagnostic PET CT (dPETCT) scans were acquired using a Phillips Gemini™ (Philips Healthcare, Best, Netherlands) PETCT system. Radiotherapy planning PETCT (pPETCT) scans were also acquired between 17 and 91 days after the dPETCT using the same system with a flat bed insert. Patients were injected with 400 MBq of FDG, rested 60 min for uptake in supine position before image acquisition.

Delineation methods

Three radiation oncologists (observers A, B and C) experienced in treating lung cancer contoured gross tumor volumes (GTVs) based on the planning CT, dPETCT and pPETCT for each patient. All contouring was completed using either Focal™ (Elekta Oncology Systems, Stockholm, Sweden) or Pinnacle^3™ (Philips Healthcare, Best, Netherlands) treatment planning systems (TPSs). The lung and mediastinal windows were initially set to W850 L-750 and W400 L20 but were allowed to be adjusted. A standard uptake value (SUV) of 2.5 was used to visualize the primary on the dPETCT and pPETCT. Nodes were contoured separately if not contiguous with the primary otherwise, a single GTV was delineated. A uniform expansion of 8 mm clipped to bone was used to define the clinical target volume (CTV) from GTV. The CTV to planning target volume (PTV) margin was 15 mm craniocaudally and 10 mm mediolaterally and was not allowed to be adjusted once created.

Reference "gold standard" volume

To assess geometric and dosimetric variation a 'gold standard' (GS) PTV was used. Assuming that the true tumor existed within the observer contours a volume was created in the computerized environment for radiation research (CERR) [15] with the simultaneous truth and performance level estimation (STAPLE) algorithm using all observer contours [16].

Treatment planning

For treatment planning the dPETCT and pPETCT were rigidly registered to the planning CT by a senior dosimetrist and contours copied to the planning CT. Ten conformal radiotherapy (CRT) treatment plans were generated per patient, one for each PTV (CT, dPETCT and pPETCT) for each observer (A, B and C) and the STAPLE volume using the Xio™ (Elekta Oncology Systems, Stockholm, Sweden) TPS to give a total of 70 plans. All plans were generated for a Siemens Oncor Impression linear accelerator with 1 cm leaves (Siemens Medical Solutions, Erlangen, Germany) using a 2.5 mm calculation grid. The originally treated CRT plan (3–5 conformal beams) (Table 1) was projected on each PTV and modified to meet International Commission on Radiation Units and Measurements (ICRU) objectives for coverage and dose homogeneity by adjusting beam apertures and weights [17]. Further planning goals limited the volume of lung (excluding PTV) receiving 20 Gy to less than 35% ($V_{20} \leqslant 35\%$) and the maximum spinal cord dose to 45 Gy.

Geometric analysis

All CT, structure and dose data were imported into CERR [15]. An in-house developed MATLAB (The Mathworks Inc, Natick, MA) application was used to calculate volume, center of mass (COM) location and maximal dimensions [X(med/lat), Y(ant/post) and Z(sup/inf)] for each volume. The conformity index (CI) [1] was calculated for each observer volume and the GS volume:

$$\mathrm{CI} = \frac{\mathrm{GS} \cap V_i}{\mathrm{GS} \cup V_i} \quad (1)$$

where GS = gold standard and V_i = observer contour. A CI equal to unity indicates perfect agreement while a CI of zero reflects no overlap. The contours for each patient were treated as a single group, representing the range of variation for analysis.

Radiobiological analysis and mean lung dose

Assuming the GS volume represented the true PTV, GS dose volume histograms (DVHs) were exported for plans generated based on the observer volumes. The impact of observer variation on normal lung tissue was quantified with mean lung dose (MLD), as a predictor for radiation pneumonitis [18]. The in-house developed Comp Plan software [19] was used to calculate MLD and the following radiobiological metrics (see Supplementary Material for details).

Statistics

Descriptive statistics were used to assess geometric and radiobiological variation. Spearman's nonparametric rank-correlation coefficient (ρ) was used to quantify the association between geometric variation (volume, COM, dimension and CI) and radiobiological metrics (TCP, EUD and MLD) with a p value less than 0.05 considered significant. SPSS® (SPSS Inc, Chicago, IL) software was used for all statistical analysis.

In our study we have used 63 separate contours to investigate the relationship between contouring variation and TCP/EUD, which gives the power to detect $\rho = 0.4$, $\alpha = 0.05$ (two sided) and $1\text{-}\beta = 0.9$ [20] (see Supplementary Material for sample size calculation).

Results

Fig. 1 shows the GS volume and observer contours in the axial plane for cases 6 and 7. Cases 4 and 7 exhibited the largest variation in volume with mean contoured volumes of 867 ± 168 cm^3 and 899 ± 419 cm^3 (Table 1). The largest radiobiological variation was also noted in cases 4 and 7. Both had involved mediastinal nodes included in the GS volume which were not contoured by all observers. As such the nodes were out-of-field for some plans resulting in mean GS EUDs of 53.6 ± 8.6 Gy and 43.6 ± 21.2 Gy.

While there was relatively large geometric variation in PTV contours for case 6 there is little variation in TCP and EUD. As the tumor location was in the RLL the beam arrangement consisted of anterior and posterior obliques. The majority of the contouring variation was in the ant/post direction while the lateral variation was minor, resulting in very little contouring variation in the beams eye view.

Fig. 2 demonstrates the relationship between PTV volume and MLD. While it is intuitive that larger PTV volumes will result in higher MLD it was not strictly the case. The figure clearly shows that variations in PTV volume of up to 500 cm^3 can have very little impact on MLD. This is because the MLD is also dependent on PTV location, beam arrangement, lung size and technique.

Supplementary Fig. 1 shows the average difference in TCP when analyzed by observer and imaging modality was generally less than −0.1%, with standard deviation ranging 0.1–0.35%. The average difference in EUD for the PTV ranged from less than −1 Gy to −6 Gy. The average difference in MLD was less than −2 Gy for all

Table 1
Variation in TCPlogit, TCPpossion, EUD and contouring metrics from GS reference volume between observers for all patients.

Patient	Stats	TCP_{logit}	EUD (Gy)	$TCP_{possion}$	ΔCOM (cm)	CI	Dimension			Volume (cm^3)
							X (cm)	*Y* (cm)	*Z* (cm)	
1	Mean	0.40	45.62	0.15	0.27	0.85	11.72	11.32	10.34	701.11
	SD	0.00	0.19	0.00	0.18	0.06	0.43	0.83	0.33	99.17
2	Mean	0.63	60.48	0.43	0.92	0.76	8.46	10.21	8.41	349.98
	SD	0.01	1.02	0.02	0.47	0.09	1.68	1.54	1.17	92.44
3	Mean	0.62	59.67	0.41	0.19	0.89	11.5	14.38	13.29	1019.14
	SD	0.01	0.41	0.01	0.13	0.10	0.65	0.34	0.37	129.6
4	Mean	0.57	53.56	0.34	0.88	0.79	14.2	11.11	14.22	867.39
	SD	0.07	8.60	0.10	0.71	0.08	0.43	0.96	5.38	168.36
5	Mean	0.64	60.79	0.44	0.21	0.90	10.49	11.21	10.28	622.31
	SD	0.00	0.09	0.00	0.11	0.03	0.29	0.71	0.43	38.64
6	Mean	0.68	64.25	0.50	0.31	0.85	7.35	8.06	6.38	227.28
	SD	0.04	2.68	0.06	0.47	0.22	1.19	1.44	0.52	70.5
7	Mean	0.44	43.60	0.24	0.96	0.67	12.12	11.66	14.4	899.29
	SD	0.22	21.17	0.16	0.49	0.20	2.76	1.49	2.55	418.7

Abbreviations: COV = center of volume; CI = conformity index; EUD = equivalent uniform dose; TCP = tumor control probability; SD = standard deviation.

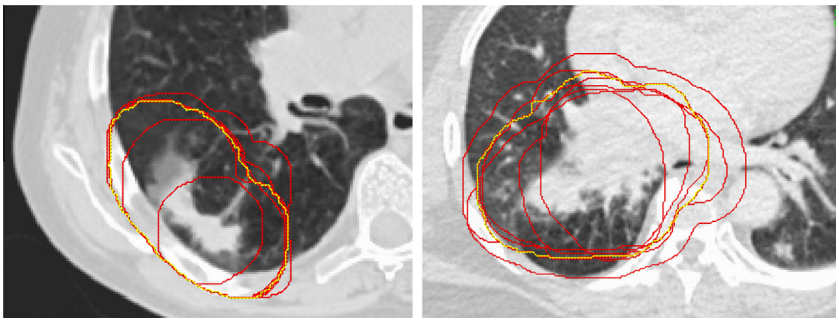


Fig. 1. Coronal and axial CT images of observer (red lines) and STAPLE reference (gold lines) PTV contours for case #6 (left) and case #7 (right). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

observers and imaging modalities except for observer X which was 1 Gy, standard deviation ranged approximately 3.5–13 Gy.

Spearman's ρ was calculated to investigate the relationship between TCP_{logit}, $TCP_{possion}$, EUD and geometric variation for all cases (Table 2). X and Z dimension demonstrated the highest correlation with average Spearman coefficients of 0.49 ($p < 0.01$) and 0.48 ($p < 0.01$). ΔCOM and CI showed the weakest correlation with average Spearman coefficients of −0.25 and 0.31 ($p < 0.05$).

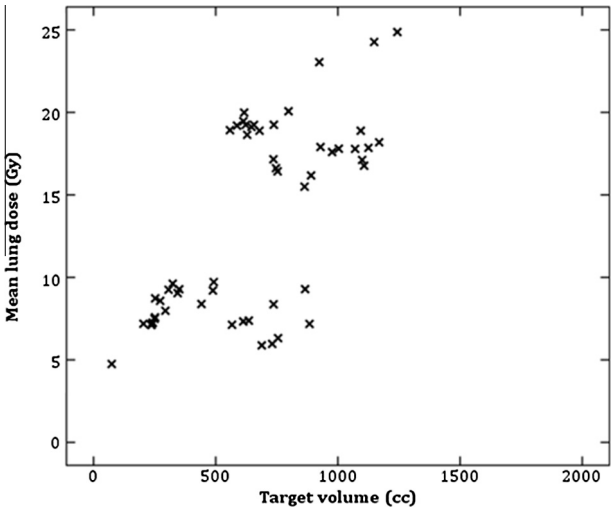


Fig. 2. Impact of variation in target volume on mean lung dose.

To investigate dependencies in the data Spearman's ρ was also analyzed with respect to the imaging modality and observer, these results can be seen in Supplementary Table 2. What this shows is that when the data are broken down by modality and observer the metrics with the highest correlation are still XDim, ZDim and Volume as in Table 2. The correlation values actually increase for some metrics while maintaining statistical significance. This is largely due to the fact that observer Z was a significant outlier when contouring on the PET scans.

Discussion

While there have been studies quantifying the magnitude of geometric contouring variation in NSCLC, investigations focusing on the dosimetric impact and correlation remain sparse. Understanding the association between contouring variation and local control for specific tumor sites and treatment techniques is highly desirable. Here we present the first study comparing the correlation of contouring metrics with modeled outcome, demonstrating the link between geometric contouring variation and TCP for NSCLC conformal radiotherapy.

The choice of reference volume to compare all other contours to is contentious [21] and a source of uncertainty for this study. STAPLE was chosen as it reflects most probable volume based on the observer volumes, other studies have used the most experienced observer, consensus or average [1]. The rank from most to least correlated geometric factor was in descending order X-Dim, Z-Dim, volume, CI, ΔCOM and Y-Dim. The inter-model (TCP, EUD) differences were small (<2%) and similar in magnitude. This

Table 2
Spearman's ρ for TCP_{logit}, EUD and $TCP_{possion}$ and geometric variation for all patients.

Model	ΔCOM		Dimension						Volume		CI	
			X		Y		Z					
	ρ	sig.	ρ	sig.	ρ	sig.	ρ	sig.	ρ	sig.	ρ	sig.
TCP_{logit}	−0.25	0.05	−0.48	0.00	−0.22	0.08	−0.46	0.00	−0.42	0.00	0.32	0.01
EUD	−0.25	0.05	−0.51	0.00	−0.24	0.06	−0.51	0.00	−0.45	0.00	0.31	0.02
$TCP_{possion}$	−0.24	0.06	−0.48	0.00	−0.24	0.06	−0.46	0.00	−0.42	0.00	0.31	0.02

Abbreviations: COM = center of mass; CI = conformity index; EUD = equivalent uniform dose; TCP = tumor control probability.

suggests that for NSCLC the relationship between geometric variation and radiobiology is not sensitive to the radiobiology model or parameters used. This may not be the case for other sites where the tumor response is less well known. This will likely vary based on clinical site and technique.

One of the limitations of the current study is the number of patients analyzed. Seven patients were assessed and may not be representative of the NSCLC population. There were three observers that contoured each case three times, giving nine distinct PTV volumes per case. It is possible that consistent over or under contouring by one observer may bias the results of the study. Also, it has been demonstrated that different planning systems create margins and interpret volumes of contours differently [22]. The impact of this was minimized for this study by creating all margins and plans in Xio. The conformal planning technique used may be less sensitive to contouring variation than IMRT and VMAT for example. Techniques with tighter margins would be more sensitive to variation in delineation when assessed using this method. For each treatment plan the DVH of the gold standard reference volume is analyzed, therefore, increasing the margins increases the chance of irradiating the reference volume to a sufficient dose for tumor control, conversely smaller margins decrease this chance. The trade off in these two scenarios is the amount of healthy lung irradiated. Lütgendorf-Caucig et al. [23] evaluated COM displacements to estimate a PTV margin to ensure minimum CTV coverage of 95% for 90% of patients using the margin recipe of van Herk et al. [24]. Adopting this approach for inter-observer and intra-imaging contouring variation a margin of 11 mm would be required based on the data in this study.

The geometric parameters most closely related to outcome for NSCLC were X-Dim, ΔCOM and CI. In a review of methods of analysis in contouring studies it was found that X-Dim was employed in only 1/10 of the lung studies reviewed, while ΔCOM and CI were used in 4/10 and 2/10 [1], and volume was the metric of choice in 8/10 studies. This highlights that the choice of metric for assessment of contouring variation is not driven by relevance to clinical outcome but likely by the tools available to investigators. The choice of metric will differ based on treatment site and technique. Therefore, it is recommended that prior to conducting a contouring study the most relevant metrics should be determined for the given treatment site and planning technique. This knowledge combined with the minimum required set of geometric descriptors [2] will ensure results of future contouring studies are consistent and comparable.

Conclusion

TCP and geometric contour variation demonstrate a significant relationship for conformal NSCLC radiotherapy with changes in medial–lateral dimension showing the strongest correlation. Care should be taken in the choice of GS reference volume. The technique presented here may be used in the assessment of contouring protocol violations in multicenter clinical trials and in particular in the choice of metrics used for analysis.

Conflict of interest statement

None.

Acknowledgement

The authors gratefully acknowledge Dr Wei Xuan for statistical advice.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.radonc.2014.03.019.

References

[1] Jameson MG, Holloway LC, Vial PJ, Vinod SK, Metcalfe PE. A review of methods of analysis in contouring studies for radiation oncology. J Med Imaging Radiat Oncol 2010;54:401–10.
[2] Fotina I, Lütgendorf-Caucig C, Stock M, Pötter R, Georg D. Critical discussion of evaluation parameters for inter-observer variability in target definition for radiation therapy. Strahlenther Onkol 2012;188:160–7.
[3] Weiss E, Hess CF. The impact of gross tumor volume (GTV) and clinical target volume (CTV) definition on the total accuracy in radiotherapy theoretical aspects and practical experiences. Strahlenther Onkol 2003;179:21–30.
[4] Peters LJ, O'Sullivan B, Giralt J, Fitzgerald TJ, Trotti A, Bernier J, et al. Critical impact of radiotherapy protocol compliance and quality in the treatment of advanced head and neck cancer: results from TROG 02.02. J Clin Oncol 2010;28:2996–3001.
[5] Breen SL, Publicover J, De Silva S, Pond G, Brock K, O'Sullivan B, et al. Intraobserver and interobserver variability in GTV delineation on FDG-PET-CT images of head and neck cancers. Int J Radiat Oncol Biol Phys 2007;68:763–70.
[6] Steenbakkers RJHM, Duppen JC, Fitton I, Deurloo KEI, Zijp LJ, Comans EFI, et al. Reduction of observer variation using matched CT-PET for lung cancer delineation: a three-dimensional analysis.[see comment]. Int J Radiat Oncol Biol Phys 2006;64:435–48.
[7] Li XA, Tai A, Arthur DW, Buchholz TA, Macdonald S, Marks LB, et al. Variability of target and normal structure delineation for breast cancer radiotherapy: an RTOG Multi-Institutional and Multiobserver Study. Int J Radiat Oncol Biol Phys 2009;73:944.
[8] Gao Z, Wilkins D, Eapen L, Morash C, Wassef Y, Gerig L. A study of prostate delineation referenced against a gold standard created from the visible human data. Radiother Oncol 2007;85:239–46.
[9] Spoelstra FOB, Senan S, Le Péchoux C, Ishikura S, Casas F, Ball D, et al. Variations in target volume definition for postoperative radiotherapy in stage iii non-small-cell lung cancer: analysis of an international contouring study. Int J Radiat Oncol Biol Phys 2009;76:1106–13.
[10] Mitchell DM, Perry L, Smith S, Elliott T, Wylie JP, Cowan RA, et al. Assessing the effect of a contouring protocol on postprostatectomy radiotherapy clinical target volumes and interphysician variation. Int J Radiat Oncol Biol Phys 2009;75:990–3.
[11] Nelms BE, Tomé WA, Robinson G, Wheeler J. Variations in the contouring of organs at risk: test case from a patient with oropharyngeal cancer. Int J Radiat Oncol Biol Phys 2012;82(1):368–78.
[12] Vinod SK, Kumar S, Holloway LC, Shafiq J. Dosimetric implications of the addition of 18 fluorodeoxyglucose-positron emission tomography in CT-based radiotherapy planning for non-small-cell lung cancer. J Med Imaging Radiat Oncol 2012;54:152–60.
[13] Petersen RP, Truong PT, Kader HA, Berthelet E, Lee JC, Hilts ML, et al. Target volume delineation for partial breast radiotherapy planning: clinical characteristics associated with low interobserver concordance. Int J Radiat Oncol Biol Phys 2007;69:41–8.
[14] Poortmans PMP, Venselaar JLM, Struikmans H, Hurkmans CW, Davis JB, Huyskens D, et al. The potential impact of treatment variations on the results of radiotherapy of the internal mammary lymph node chain: a

quality-assurance report on the dummy run of EORTC Phase III randomized trial 22922/10925 in Stage I-III breast cancer. Int J Radiat Oncol Biol Phys 2001;49:1399–408.
[15] Deasy JO, Blanco AI, Clark VH. CERR: a computational environment for radiotherapy research. Med Phys 2003;30:979.
[16] Warfield SK, Zou KH, Wells WM. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Trans Med Imaging 2004;23:903–21.
[17] ICRU Report 62. Prescribing, recording, and reporting photon beam therapy (Supplement to ICRU Report 50). Bethesda, MD: International Commission on Radiation Units and Measurements; 1999.
[18] Marks LB, Bentzen SM, Deasy JO, Kong FMS, Bradley JD, Vogelius IS, et al. Radiation dose volume effects in the lung. Int J Radiat Oncol Biol Phys 2010;76:S70–6.
[19] Holloway LC, Miller JA, Kumar S, Whelan BM, Vinod SK. Comp Plan: a computer program to generate dose and radiobiological metrics from dose–volume histogram files. Med Dosim 2012;37(3):305–9.
[20] Machin D, Campbell MJ, Tan S-B, Tan S-H. Sample size tables for clinical studies. Wiley-Blackwell; 2011.
[21] Vorwerk H, Beckmann G, Bremer M, Degen M, Dietl B, Fietkau R, et al. The delineation of target volumes for radiotherapy of lung cancer patients. Radiother Oncol 2009;91:455–60.
[22] Pooler AM, Mayles HM, Naismith OF, Sage JP, Dearnaley DP. Evaluation of margining algorithms in commercial treatment planning systems. Radiother Oncol 2008;86:43–7.
[23] Lütgendorf-Caucig C, Fotina I, Stock M, Pötter R, Goldner G, Georg D. Feasibility of CBCT-based target and normal structure delineation in prostate cancer radiotherapy: multi-observer and image multi-modality study. Radiother Oncol 2011;98:154–61.
[24] van Herk M, Remeijer P, Rasch C, Lebesque JV. The probability of correct target dosage: dose-population histograms for deriving treatment margins in radiotherapy. Int J Radiat Oncol Biol Phys 2000;47:1121.

Supplementary Table 1 Patient characteristics

Patient	TNM	Histology	Prescribed dose (Gy/#)	Primary location
1	T3N0M0	Undifferentiated NSCLC	45/25(PCT)	LUL
2	T2N0M0	SCC	60/30	LLL
3	T3N0M0	SCC	60/30	Left hilum
4	T1N3M0	Large cell carcinoma	60/30	Left hilum
5	T4N3M0	Large cell carcinoma	60/30	LUL
6	T1aN0M0	Large cell carcinoma	60/30	RLL
7	T2N2M0	Adenocarcinoma	60/30	RLL

Abbreviations: NSCLC = non small cell lung cancer; SCC = squamous cell carcinoma; PCT = Pancoast tumor; LL = left lung; LLL = left lower lobe; LUL = left upper lobe; RLL = right lower lobe

Radiobiology calculations

Equivalent uniform dose (EUD) [1]:

$$EUD = \left(\sum_i (v_i D_i^a)\right)^{1/a} \quad (2)$$

where v_i = normalized volume for the voxel being considered, D_i = the dose to the voxel being considered and a is a parameter related to the structure being considered (a = -1 for PTVs [2]) that drives the model.

TCP based on the logit model (TCP_{logit})[3]

$$TCP_{logit} = \prod \left[\frac{1}{1+\left(\frac{D_{50}}{D_i}\right)^{4\gamma 50}}\right]^{v_i} \quad (3)$$

where D_{50} = 51.24 Gy is the dose for 50% control, γ_{50} = 0.83 is the slope of the dose response curve

[4]

TCP based on Possion statistics ($TCP_{possion}$) [3]

$$TCP = \left(\frac{1}{2}\right)^{\sum v_i exp\left[e\gamma_{50}\left(1-{}^{D_i}/_{D_{50}}\right)\right]} \quad (4)$$

where D_{50} = 64 Gy and γ_{50} = 1.3 [5]

Sample size calculation

In order to estimate the number of observations needed for the study, a sample size estimate is required. The method described in Machin's text book [6] was used, where for two normally distributed variables it can be shown that:

$$u_\rho = \frac{1}{2}\log\left[\frac{1+\rho}{1-\rho}\right] + \frac{\rho}{2(N-1)} \quad (1)$$

In the above equation N is the sample size and ρ is the predicted correlation for significance level α and power of 1-β, then:

$$N = \frac{\left(z_{1-\alpha} + z_{1-\beta}\right)^2}{{u_\rho}^2} \quad (2)$$

In order to calculate a value or N to substitute into equation (1) we can calculate an initial u_ρ denoted u_ρ^0 using:

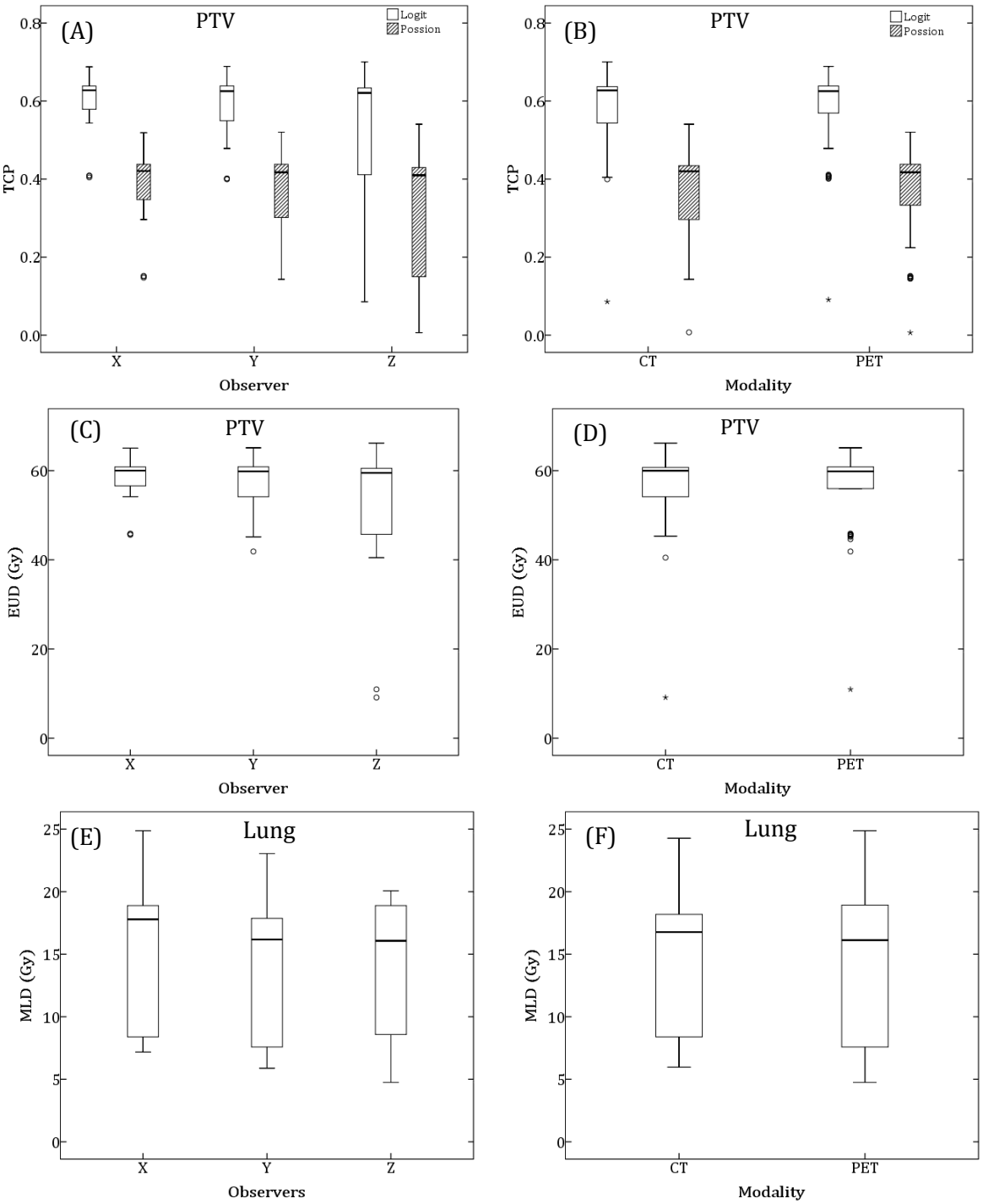
$$u_\rho^0 = \frac{1}{2}\log\left[\frac{1+\rho}{1-\rho}\right] \quad (3)$$

Machin [6] has tabulated results for which this process was repeated until the two consecutive values of N were within unity (Table 1). From this table it can be seen that for ρ = 0.5, α = 0.05 (two sided) and 1-β = 0.8 a sample size of 29 is needed. However, this is based on linear correlation (i.e. Pearson) and in this study the non- parametric Spearman's correlation was used. There is not a straight forward method for calculating a sample size for Spearman's ρ however Siegal states that Spearman is about 91% as efficient as Pearson [7] therefore the sample size required for Spearman is 32.

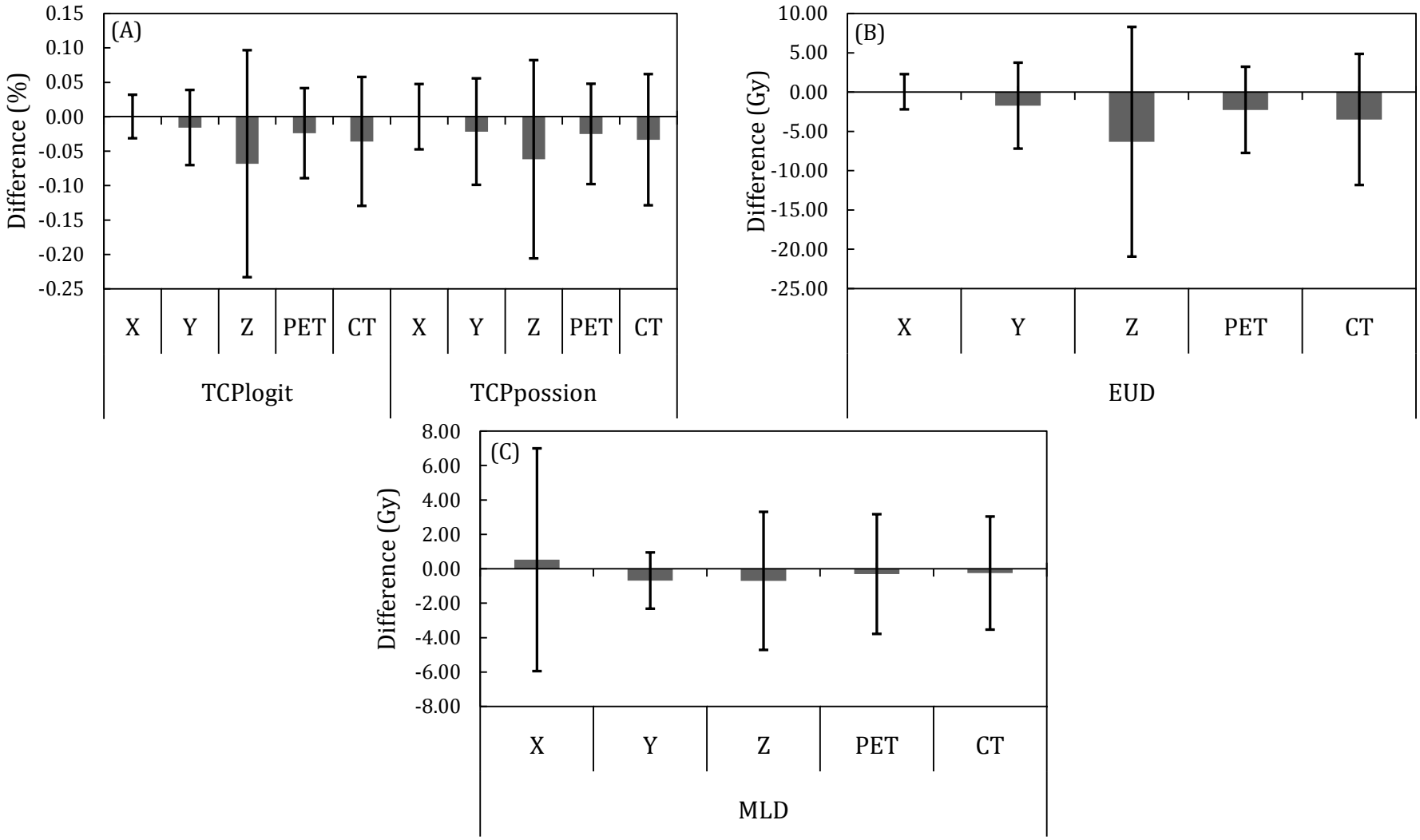
In our study we have used 63 separate contours to investigate the relationship between contouring variation and TCP/EUD, which using the above table gives us the power to detect ρ = 0.4, α = 0.05 (two sided) and 1-β = 0.9.

Supplementary Table 2 Sample sizes for detecting a statistically significant correlation coefficient. Table adapted from [1].

ρ	α		Power 1-β	
	One-sided	Two-sided	0.8	0.9
0.1	0.025	0.05	782	1046
	0.05	0.10	617	853
	0.10	0.20	450	655
0.2	0.025	0.05	193	258
	0.05	0.10	153	211
	0.10	0.20	112	162
0.3	0.025	0.05	84	112
	0.05	0.10	67	92
	0.10	0.20	50	71
0.4	0.025	0.05	46	61
	0.05	0.10	37	50
	0.10	0.20	28	39
0.5	0.025	0.05	29	37
	0.05	0.10	23	31
	0.10	0.20	18	24
0.6	0.025	0.05	19	25
	0.05	0.10	16	21
	0.10	0.20	12	16
0.7	0.025	0.05	13	17
	0.05	0.10	11	14
	0.10	0.20	9	12
0.8	0.025	0.05	10	12
	0.05	0.10	8	10
	0.10	0.20	7	8
0.9	0.025	0.05	7	8
	0.05	0.10	6	7
	0.10	0.20	5	6



Supplementary Figure 1 TCP and EUD results for PTV (A-D) and V20, EUD and MLD results for lung (E & F) grouped with respect to imaging modality and observer



Supplementary Figure 2 Difference in TCP (A) and EUD (B) between STAPLE plan and observer plans for PTV and MLD (C) between the STAPLE plan and observer plans for lung tissue (error bars indicate ± σ).

References

[1] Niemierko A. Reporting and analyzing dose distributions: a concept of equivalent uniform dose. Med Phys. 1997;24:103.
[2] Wu Q, Mohan R, Niemierko A, Schmidt-Ullrich R. Optimization of intensity-modulated radiotherapy plans based on the equivalent uniform dose. Int J Radiat Oncol Biol Phys. 2002;52:224-35.
[3] Källman P, Ågren A, Brahme A. Tumour and normal tissue responses to fractionated non-uniform dose delivery. Int J Radiat Biol. 1992;62:249-62.
[4] Okunieff P, Morgan D, Niemierko A, Suit HD. Radiation dose-response of human tumors. Int J Radiat Oncol Biol Phys. 1995;32:1227.
[5] Martel MK, Ten Haken RK, Hazuka MB, Kessler ML, Strawderman M, Turrisi AT, et al. Estimation of tumor control probability model parameters from 3-D dose distributions of non-small cell lung cancer patients. Lung Cancer. 1999;24:31-7.
[6] Machin D, Campbell MJ, Tan S-B, Tan S-H. Sample size tables for clinical studies: Wiley-Blackwell; 2011.
[7] Siegel S, Castellan NJ. Non parametric statistics for the behavioral sciences: McGraw-Hill International; 1988.

Chapter 5: How important is dosimetrist experience for intensity modulated radiation therapy? A comparative analysis of a head and neck case

Statement of joint authorship

Title

How important is dosimetrist experience for intensity modulated radiation therapy? A comparative analysis of a head and neck case

Authors

Vikneswary Batumalai**, Michael G Jameson**, Dion F Forstner, Philip J Vial and Lois C Holloway

Vikneswary Batumalai

Developed experimental design and scientific method, collected data, analysed and interpreted data, wrote manuscript

Michael G Jameson (candidate)

Contributed to experimental design and scientific method, contributed to analyses and interpretation of data, contributed to manuscript

Dion F Forstner

Contributed to scientific method and manuscript

Philip J Vial

Contributed to scientific method and manuscript

Mark Sidhom

Contributed to scientific method and manuscript

Lois C Holloway

Contributed to scientific method and manuscript

Chapter 6: Endorectal balloons in the post prostatectomy setting: Do gains in stability lead to more predictable dosimetry?

Statement of joint authorship

Title

Endorectal balloons in the post prostatectomy setting: Do gains in stability lead to more predictable dosimetry?

Authors

Michael G Jameson, Jeremiah De Leon, Apsara A Windsor, Kirrily Cloak, Sarah Keats, Jason A Dowling, Shekhar S Chandra, Philip J Vial, Mark Sidhom, Lois C Holloway and Peter E Metcalfe

Michael G Jameson (candidate)

Developed experimental design and scientific method, collected data, analysed and interpreted data, wrote manuscript

Jeremiah De Leon

Contributed to experimental design and scientific method, contributed to analyses and interpretation of data, contributed to manuscript

Apsara A Windsor

Contributed to scientific method and manuscript

Kirrily Cloak

Contributed to scientific method and manuscript

Sarah Keats

Contributed to scientific method and manuscript

Jason A Dowling

Contributed to data analysis and manuscript

Shekhar S Chandra

Contributed to data analysis and manuscript

Philip J Vial

Contributed to scientific method and manuscript

Mark Sidhom

Contributed to scientific method and manuscript

Lois C Holloway

Contributed to experimental design and scientific method, contributed to analyses and interpretation of data, contributed to manuscript

Peter E Metcalfe

Contributed to scientific method and manuscript

Radiotherapy and Oncology 109 (2013) 493–497

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Endorectal ballon spacing

Endorectal balloons in the post prostatectomy setting: Do gains in stability lead to more predictable dosimetry?

Michael G. Jameson [a,b,c,*], Jeremiah De Leon [a], Apsara A. Windsor [a,d,h], Kirrily Cloak [a], Sarah Keats [a], Jason A Dowling [e], Shekhar S Chandra [e], Philip Vial [a,c,f], Mark Sidhom [a], Lois Holloway [a,b,c,f,g] Peter Metcalfe [b,c]

[a] Liverpool and Macarthur Cancer Therapy Centres; [b] Centre for Medical Radiation Physics, University of Wollongong; [c] Ingham Institute; [d] Collaboration for Cancer Outcomes Research and Evaluation, Liverpool Hospital; [e] Australian e-Health Research Centre, CSIRO Computational Informatics; [f] Institute of Medical Physics, School of Physics, University of Sydney; [g] SWSCS, School of Medicine, University of New South Wales; and [h] Central Coast Cancer Centre, Gosford, Australia

ARTICLE INFO

Article history:
Received 21 December 2012
Received in revised form 14 August 2013
Accepted 15 August 2013
Available online 14 September 2013

Keywords:
Endorectal balloon
EUD
Inter-fraction motion
IMRT
Post prostatectomy

ABSTRACT

Purpose: To perform a comparative study assessing potential benefits of endorectal-balloons (ERB) in post-prostatectomy patients.
Method and materials: Ten retrospective post-prostatectomy patients treated without ERB and ten prospective patients treated with the ERB *in situ* were recruited. All patients received IMRT and IGRT using kilovoltage cone-beam computed tomography (kVCBCT). kVCBCT datasets were registered to the planning dataset, recontoured and the original plan recalculated on the kVCBCTs to recreate anatomical conditions during treatment. The imaging, structure and dose data were imported into in-house software for the assessment of geometric variation and cumulative equivalent uniform dose (EUD) in the two groups.
Results: The difference in location (ΔCOV) for the bladder between planning and each CBCT was similar for each group. The range of mean ΔCOV for the rectum was 0.15–0.58 cm and 0.15–0.59 cm for the non-ERB and ERB groups. For superior-CTV and inferior-CTV the difference between planned and delivered $D_{95\%}$ (mean ± SD) for the non-ERB group was 2.1 ± 6.0 Gy and −0.04 ± 0.20 Gy. While for the ERB group the difference in $D_{95\%}$ was 8.7 ± 12.6 Gy and 0.003 ± 0.104 Gy.
Conclusions: The use of ERBs in the post-prostatectomy setting did improve geometric reproducibility of the target and surrounding normal tissues, however no improvement in dosimetric stability was observed for the margins employed.
 Radiotherapy and Oncology 109 (2013) 493–497

Adjuvant radiation therapy delivered post radical prostatectomy (PP) results in longer time to biochemical failure and improved local control compared to watchful waiting [1]. Long-term follow-up of the EORTC 22911 trial concluded that there is a survival benefit associated with adjuvant radiation following surgery for patients <70 years or having positive margins [1]. Intensity modulated radiation therapy (IMRT) is becoming the standard of care for this patient cohort as it provides excellent target coverage and normal tissue sparing. However, precise daily localization and immobilization is required for accurate delivery of highly conformal treatments.

Variation in prostate bed location day-to-day has been investigated previously [2–4]. Rectal volumes can vary significantly throughout treatment from −40 to +60% compared to planning [3]. Similarly bladder variation can be major, varying up to 200 cm^3 [3]. These day-to-day fluctuations can have a substantial dosimetric effect on both the prostate bed and the normal tissues [2,3]. Daily imaging with registration to bony anatomy may not be sufficient in accounting for prostate bed motion. Relative to bony anatomy, prostate bed displacement exceeded 5 mm in 21% of treatments in the cranio–caudal (C–C) direction and 9% in the anterior–posterior (A–P) for 20 patients [4].

Endorectal balloons (ERBs) have been used to stabilize the internal anatomy for whole prostate treatment [5]. The ERB sits within the rectum immobilizing the prostate and pushing parts of the rectum out of the high dose region. ERBs reduced planned anal wall and rectal wall doses when compared to no ERB [6]. While ERBs have been used in the PP setting [7], no study has investigated inter-fraction geometric and dosimetric stability PP. Two PP patient cohorts treated with or without ERB were used to investigate organ motion and deformation over the treatment course. Further, the delivered equivalent uniform dose for targets and normal tissues was compared to planned equivalent uniform dose (EUD) for both cohorts.

* Corresponding author. Address: Cancer Therapy Centre, Liverpool Hospital, Locked Bag 7103, Liverpool, NSW 1871, Australia.
E-mail address: michael.jameson@sswahs.nsw.gov.au (M.G. Jameson).

0167-8140/$ - see front matter © 2013 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.radonc.2013.08.024

Methods

Patient data

Subsequent to ethics approval ten patients most recently treated with adjuvant radiotherapy following radical prostatectomy were retrospectively selected as the non-ERB group for this pilot study. A further ten patients were recruited prospectively after radical prostatectomy to the ERB group (Supplementary Table 1). The original planning CT was used for patients in the non-ERB group. Two planning scans were obtained for the ERB group, one without the ERB and one with the ERB *in situ*. QLRAD (QLRAD B.V., Dalfsen, The Netherlands) ERBs were used in this study, details of which can be found elsewhere [5]. Images were acquired with a Siemens Somatom™ (Siemens Medical Solutions, Erlangen, Germany) system using 2 mm slice thickness and a 512 × 512 matrix. All patients were instructed to drink 500 mL of water prior to simulation and treatment. Additional fiber supplements were initiated one week prior to simulation and continued throughout treatment. The image guided radiotherapy (IGRT) regimen consisted of daily cone–beam CT (CBCT) fractions 1–5 and weekly thereafter. All CBCTs were performed on a Synergy accelerator equipped with XVI (Elekta Oncology Systems, Stockholm, Sweden) with 120 kVp, 1056 mAs using a 20 cm FOV and a bowtie filter.

Delineation

All planning and daily CBCT images were imported into Focal (Elekta Oncology Systems, Stockholm, Sweden) for registration and contouring. CBCT images were rigidly registered to planning scans based on bony anatomy per standard IGRT practice on treatment. Target volumes and normal tissues were contoured on the planning scan and all CBCTs for each patient (Fig. 1). The PP clinical target volume (CTV) was delineated according to national consensus guidelines [8]. Briefly, the inferior border of the CTV was approximately 5 mm below the vesicourethral anastimosis covering all surgical clips. The inferior CTV (infCTV) was bounded by the pubis symphysis, levator ani, obturator internus and anterior rectal wall. Above the superior edge of the symphysis pubis the superior CTV (supCTV) was extended cranially to include all of the seminal vesical bed and the distal portion of the vas deferens. A uniform CTV to planning target volume (PTV) margin of 10 mm was used. The anterior border of the supCTV encompasses the posterior 15 mm of the bladder. Similarly the whole rectum was contoured in three sections: superior rectum from the inferior aspect of sacroiliac joint to superior border of pubic symphasis, inferior rectum extended to a line drawn from the coccyx to the inferior border of the pubic symphasis and the anus which extended another 4 cm. The inner wall of the rectum and anus was contoured by contracting the outer wall contour by 5 mm [6].

Planning

All patients were planned with step and shoot IMRT, consisting of seven or nine non-opposing, coplanar fields using Xio® (Elekta Oncology Systems, Stockholm, Sweden). The prescribed dose was 70 Gy and 64.4 Gy to the inferior and superior CTVs respectively, delivered in 35 fractions. International Commission on Radiation Units and Measurements (ICRU) objectives for coverage and dose homogeneity were adhered to [9]. Additional planning goals limited the volume of rectum receiving 40 Gy to less than 60% ($V_{40} < 60\%$) and $V_{60} < 40\%$. Further objectives limited $V_{50} < 50\%$ for bladder and $V_{45} < 60\%$ for femoral heads [10]. After registration and contouring, each patient's original plan was recalculated on all CBCTs. Due to the unstable CT numbers of the non-scatter corrected CBCT images all relative electron densities were forced to unity, except for ERBs which were set to air [11]. For consistency in dosimetric comparison the planning CT was also density forced with the same values used for the CBCTs.

Geometric analysis

An in-house developed MATLAB (The Mathworks Inc., Natick, MA) application in conjunction with the CERR platform [12] calculated volume and center of volume (COV) location. The conformity index (CI) [13] was calculated for each CBCT volume and planned volume, for the CTV, rectum and bladder volumes and subvolumes:

$$\mathrm{CI} = \frac{V_{\mathrm{planned}} \cap V_{\mathrm{CBCT}i}}{V_{\mathrm{planned}} \cup V_{\mathrm{CBCT}i}} \quad (1)$$

where V_{planned} and $V_{\mathrm{CBCT}i}$ are the planning and ith CBCT contour volumes respectively. A CI equal to unity indicates perfect agreement while a CI of zero reflects no overlap. The mean absolute surface distance (MASD) between planning and CBCT volumes was computed as the average of all Euclidean surface distances per vertex from the planning scan to each CBCT [14].

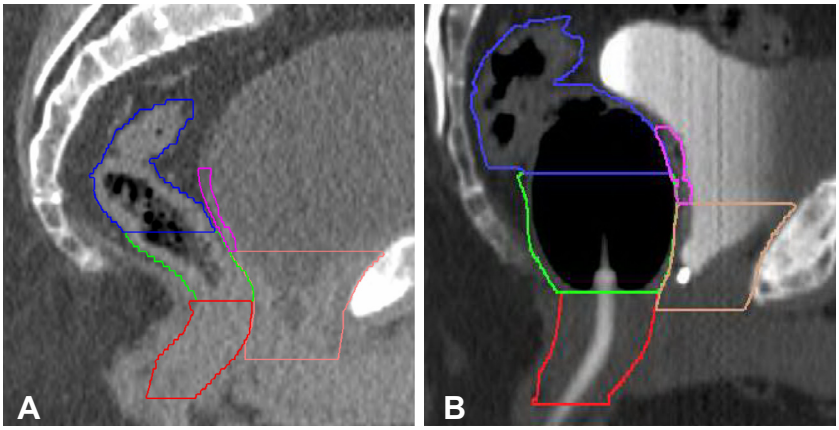


Fig. 1. Sagital views of targets and normal structures delineated for (A) non-ERB and (B) ERB groups. Blue: superior rectum; green: inferior rectum; red: anus; pink: superior CTV and beige: inferior CTV.

Dosimetric analysis

EUD has been shown to be a useful tool when comparing complex radiotherapy plans [15]. The form suggested by Niermierko [16] was used to evaluate the dose distributions for CTV, rectum and bladder volumes and sub-volumes:

$$\mathrm{EUD} = \left(\sum_i (v_i D_i^a)\right)^{\frac{1}{a}} \quad (2)$$

where v_i is the normalized volume for the voxel being considered, D_i is the dose to the voxel being considered and a is a parameter related to the structure being considered that drives the model. For the current study $a = 11.9$ for the rectum [17], $a = 8$ for the bladder [18] and $a = -1$ for the CTVs [18]. In-house software [19] was used to calculate EUD and dose volume parameters V_{50} for bladder, V_{40} for rectum and D_{95} for CTV.

Statistics

The SPSS® (SPSS Inc., Chicago, IL) package was used for all statistical analysis. The two-sample Kolmogorov–Smirnov (K–S) test was used to compare results between the non-ERB and ERB groups [20]. Where a p value <0.05 would indicate that the geometric and dosimetric distributions of the groups were statistically different.

Results

Patient data

For the non-ERB group 91 CBCTs were suitable for contouring while 71 were suitable for planning. From the 10 patients recruited to the ERB group, one patient (#10) had a superior non-ERB plan and was treated without ERB and one patient had to cease ERB use due to hemorrhoids. This left 70 CBCTs suitable for contouring and 69 were able to be planned. Image artifacts and clipping were the leading reasons why some CBCT datasets could not be contoured or planned.

Geometric variation

The mean ± SD CI for bladder was 0.54 ± 0.21 and 0.54 ± 0.20 for the non-ERB and ERB groups, respectively (Supplementary Fig. 1). Non-ERB/ERB CI values for the whole rectum, inferior rectum and superior rectum were 0.50 ± 0.12/0.71 ± 0.07 ($p < 0.01$), 0.51 ± 0.12/0.78 ± 0.08 ($p < 0.01$) and 0.42 ± 0.13/0.59 ± 0.11 ($p < 0.01$). Similarly for the CTV, infCTV and supCTV the CI values were 0.72 ± 0.15/0.73 ± 0.11 ($p < 0.05$), 0.87 ± 0.07/0.88 ± 0.05 ($p < 0.01$) and 0.54 ± 0.22/0.56 ± 0.15 ($p = 0.1$) for non-ERB/ERB.

The difference in location (ΔCOV) for the bladder between planning and each CBCT was similar for each group. The range of mean ΔCOV for Whole rectum was 0.15–0.58 and 0.15–0.59 cm for the non-ERB and ERB groups. However, the average change in Whole rectum volume from planning for the non-ERB and ERB groups ranged from 84–224% and 98–120%. The infCTV ΔCOV for non-ERB and ERB ranged from 0.01 to 0.05 and 0.02 to 0.14. The supCTV range of average change in volume was 75–126% and 55–141% for non-ERB and ERB respectively. The MASD is displayed in Supplementary Fig. 2 for non-ERB patient#8 and ERB patient#1 as representative cases.

Inter-fraction dosimetric stability

Fig. 2 shows variation in V_{40} from planning value over the course of treatment for Superior rectum and Inferior rectum. For supCTV and infCTV the difference (mean ± SD) in $D_{95\%}$ for the non-ERB group was 2.1 ± 6.0 Gy and −0.04 ± 0.20 Gy. While for the ERB group the difference in $D_{95\%}$ was 8.7 ± 12.6 Gy and 0.003 ± 0.104 Gy from the supCTV and infCTV. The difference in bladder V_{50} from planned was 20.9 ± 23.5% for the non-ERB group and 12.6 ± 22.1% for the ERB group. Fig. 3 depicts the difference between the planned and delivered EUD.

Discussion

This study represents the first comparison of inter-fraction organ deformation and dosimetric stability for patients treated with

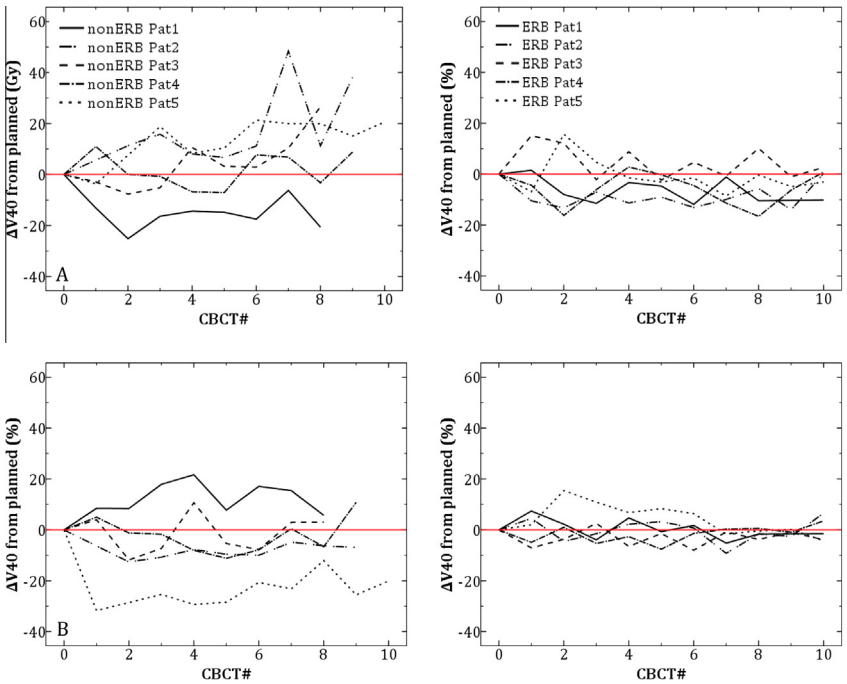


Fig. 2. Plots of the difference in V_{40} for superior (A) and inferior (B) rectum with no ERB (left column) and ERB (right column) red line indicates ideal situation of no variation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

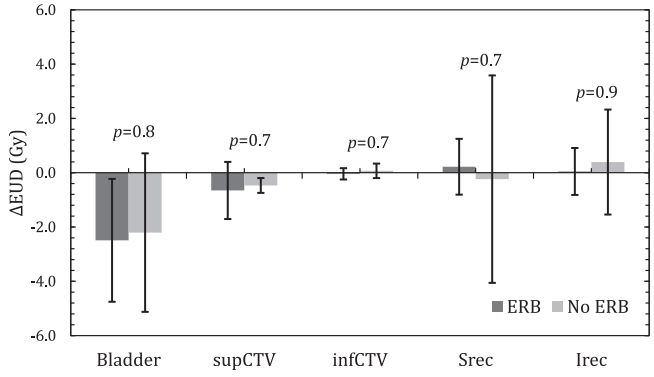


Fig. 3. Difference between planned EUD and average delivered EUD for a number of organs and sub-organs for each group, error bars indicate one standard deviation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

or without ERB in the post-prostatectomy setting. The ERB improved the CI of the Whole rectum, Superior rectum and Inferior rectum by 21%, 27% and 17% ($p < 0.01$). The improved geometric stability with the ERB did not translate into a statistically significant benefit in inter-fraction dosimetric stability based on ΔEUD. The reduced dosimetric stability seen for the bladder and supCTV is likely due to bladder filling and slight differences in ERB insertion depth between fractions.

The range of non-ERB organ motion was similar to those reported in other studies [21]. A study on the use of gold seed fiducials in PP radiotherapy reported slightly larger inter-fraction infCTV motion of 0.03 ± 0.09 cm, 0.4 ± 2.4 mm and −0.11 ± 0.21 cm in the LR, SI and AP directions [22]. Compared to ΔCOV in this study of 0.01–0.05 cm, these differences are likely due to the different methods of motion measurement used.

While the affects of the ERB on dosimetric stability were insignificant on average for the ERB cohort, there were indications that the ERB may be beneficial for some patients. However, it is difficult to know whether these patients would have been stable if treated without ERB. For example, patients in each group had single fraction variations between planned and delivered dosimetry that was deemed clinically significant by the treating physician. As reported elsewhere, the cumulative effect of per treatment differences is reduced significantly with fractionation [23].

Limitations in this study include the inherent uncertainty in contouring anatomical structures, which was minimized by having a single observer contour all structures in each group. The same window and level settings were used for all patients when contouring ERB on the CBCT scans. This minimized variation in defining the ERB lumen/tissue boundary. The dosimetric advantages of ERBs are affected by the differences in the original plan dosimetry, as well as inter-fraction and intra-fraction variations. In a planning comparison study Smeenk et al. reported a significant improvement in rectum V_{40} for the ERB versus non ERB group [6]. A recent study demonstrated reduced intra-fraction target motion with ERB for definitive prostate [24]. The choice of CTV to PTV margins influences the dosimetry analysis since larger margins will result in the CTV dosimetry being less sensitive to geometric instability [25]. Further investigation with a larger study is required to establish accurate margins and confirm the overall advantages of ERB in PP radiotherapy.

Conclusion

The use of ERBs in the post prostatectomy setting improved geometric reproducibility of target volumes and surrounding normal tissues. Improvements in dosimetric stability were inconclusive. A larger study incorporating inter- and intra-fraction motion is required to appreciate the potential benefit of ERBs in the post prostatectomy setting.

Conflict of interest

None.

Acknowledgements

NL-TEC and QL-RAD Inc® are gratefully acknowledged for providing the ERBs used in this study.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.radonc.2013.08.024.

References

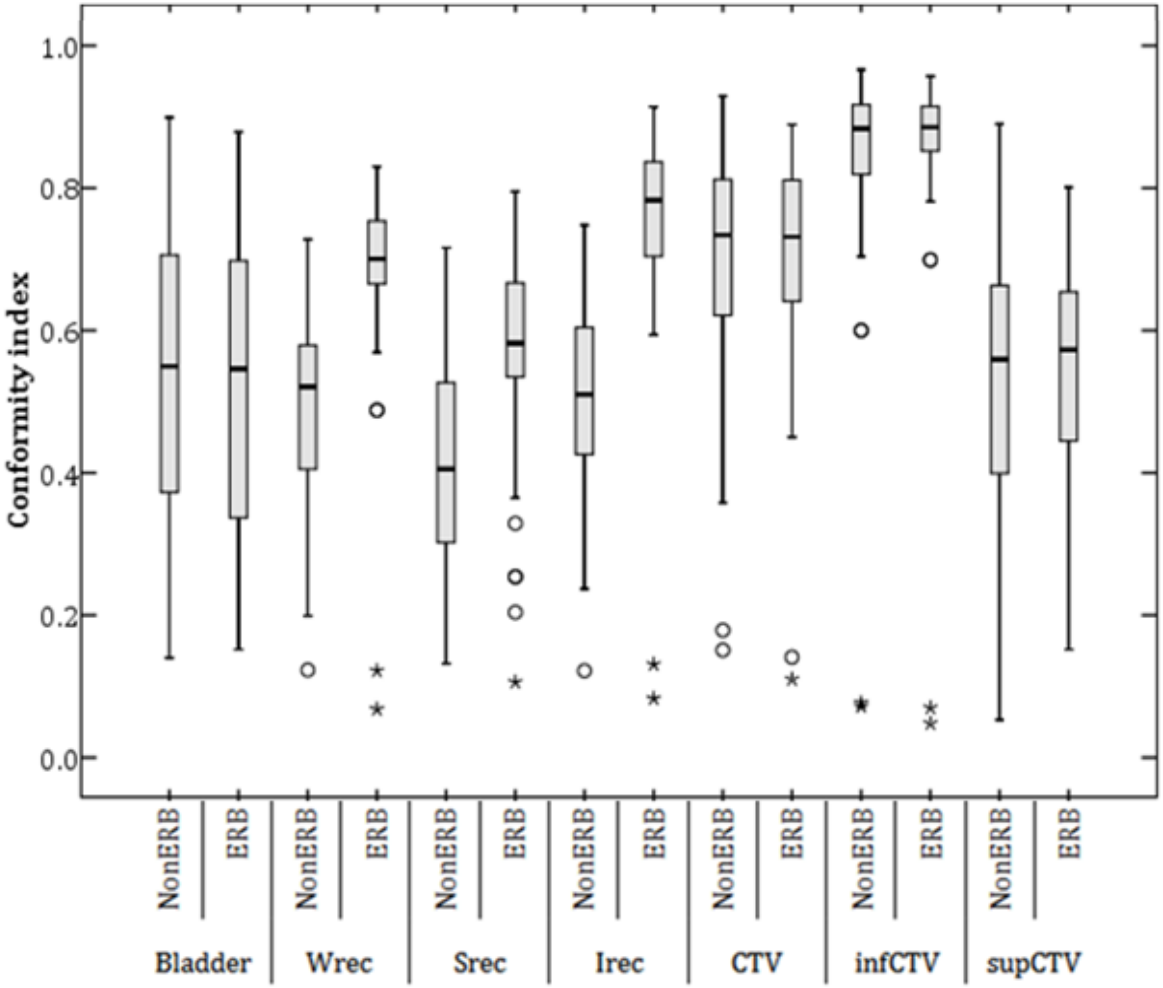
[1] Bolla M, van Poppel H, Tombal B, et al. Postoperative radiotherapy after radical prostatectomy for high-risk prostate cancer: long-term results of a randomised controlled trial (EORTC trial 22911). Lancet 2012;380:2018–27.
[2] Fiorino C, Foppiano F, Franzone P, et al. Rectal and bladder motion during conformal radiotherapy after radical prostatectomy. Radiother Oncol 2005;74:187–95.
[3] Haworth A, Paneghel A, Herschtal A, et al. Verification of target position in the post-prostatectomy cancer patient using cone beam CT. J Med Imaging Radiat Oncol 2009;53:212–20.
[4] Klayton T, Price R, Buyyounouski MK, et al. Prostate bed motion during intensity-modulated radiotherapy treatment. Int J Radiat Oncol Biol Phys 2012;84:130–36.
[5] Smeenk RJ, Teh BS, Butler EB, van Lin ENJ, Kaanders JHAM. Is there a role for endorectal balloons in prostate radiotherapy? A systematic review. Radiother Oncol 2010;95:277–82.
[6] Smeenk RJ, van Lin ENJ, van Kollenburg P, et al. Endorectal balloon reduces anorectal doses in post-prostatectomy intensity-modulated radiotherapy. Radiother Oncol 2011;101:465–70.
[7] Teh BS, Mai W-Y, Augspurger ME, et al. Intensity modulated radiation therapy (IMRT) following prostatectomy: more favorable acute genitourinary toxicity profile compared to primary IMRT for prostate cancer. Int J Radiat Oncol Biol Phys 2001;49:465–72.
[8] Sidhom MA, Kneebone AB, Lehman M, et al. Post-prostatectomy radiation therapy: consensus guidelines of the Australian and New Zealand Radiation Oncology Genito-urinary group. Radiother Oncol 2008;88:10–9.
[9] Blodgett TM, Meltzer CC, Townsend DW. PET/CT: form and function. Radiology 2007;242:360–85 [Review] [267 refs].
[10] Trans Tasman Radiation Oncology Group: RAVES – radiotherapy adjuvant versus early salvage a Phase III multi-centre randomised trial comparing adjuvant radiotherapy (RT) with early salvage RT in patients with positive margins or extraprostatic disease following radical prostatectomy (TROG 08.03) – Protocol. 2008.
[11] Jain P, Marchant T, Green M, et al. Inter-fraction motion and dosimetric consequences during breast intensity-modulated radiotherapy (IMRT). Radiother Oncol 2009;90:93–8.
[12] Deasy JO, Blanco AI, Clark VH. CERR: a computational environment for radiotherapy research. Med Phys 2003;30:979.
[13] Jameson MG, Holloway LC, Vial PJ, Vinod SK, Metcalfe PE. A review of methods of analysis in contouring studies for radiation oncology. J Med Imaging Radiat Oncol 2010;54:401–10.
[14] Dowling J, Fripp J, Chandra S, et al. Fast automatic multi-atlas segmentation of the prostate from 3D MR images. Prostate Cancer Imaging: Image Anal Image-Guided Interventions 2011:10–21.
[15] O'Daniel JC, Garden AS, Schwartz DL, et al. Parotid gland dose in intensity-modulated radiotherapy for head and neck cancer: is what you plan what you get? Int J Radiat Oncol Biol Phys 2007;69:1290–6.
[16] Niemierko A. Reporting and analyzing dose distributions: a concept of equivalent uniform dose. Med Phys 1997;24:103.
[17] Michalski JM, Gay H, Jackson A, Tucker SL, Deasy JO. Radiation dose–volume effects in radiation-induced rectal injury. Int J Radiat Oncol Biol Phys 2010;76:S123–9.
[18] Wu Q, Mohan R, Niemierko A, Schmidt-Ullrich R. Optimization of intensity-modulated radiotherapy plans based on the equivalent uniform dose. Int J Radiat Oncol Biol Phys 2002;52:224–35.
[19] Holloway LC, Miller JA, Kumar S, Whelan BM, Vinod SK. Comp Plan: a computer program to generate dose and radiobiological metrics from dose–volume histogram files. Med Dosim 2012;37:305–9.
[20] Wang KK-H, Vapiwala N, Deville C, et al. A study to quantify the effectiveness of daily endorectal balloon for prostate intrafraction motion management. Int J Radiat Oncol Biol Phys 2012;83:1055–63.

[21] Showalter TN, Nawaz AO, Xiao Y, Galvin JM, Valicenti RK. A cone beam CT-based study for clinical target definition using pelvic anatomy during postprostatectomy radiotherapy. Int J Radiat Oncol Biol Phys 2008;70:431.
[22] Schiffner DC, Gottschalk AR, Lometti M, et al. Daily electronic portal imaging of implanted gold seed fiducials in patients undergoing radiotherapy after radical prostatectomy. Int J Radiat Oncol Biol Phys 2007;67:610–9.
[23] Langen KM, Chauhan B, Siebers JV, Moore J, Kupelian PA. The dosimetric effect of intrafraction prostate motion on step-and-shoot intensity-modulated radiation therapy plans: magnitude, correlation with motion parameters, and comparison with helical tomotherapy plans. Int J Radiat Oncol Biol Phys 2012;84:1220–25.
[24] Smeenk RJ, Louwe RJW, Langen KM, et al. An endorectal balloon reduces intrafraction prostate motion during radiotherapy. Int J Radiat Oncol Biol Phys 2011;83:661–9.
[25] Jones BL, Gan G, Diot Q, et al. Dosimetric and deformation effects of image-guided interventions during stereotactic body radiation therapy of the prostate using an endorectal balloon. Med Phys 2012;39:3080.

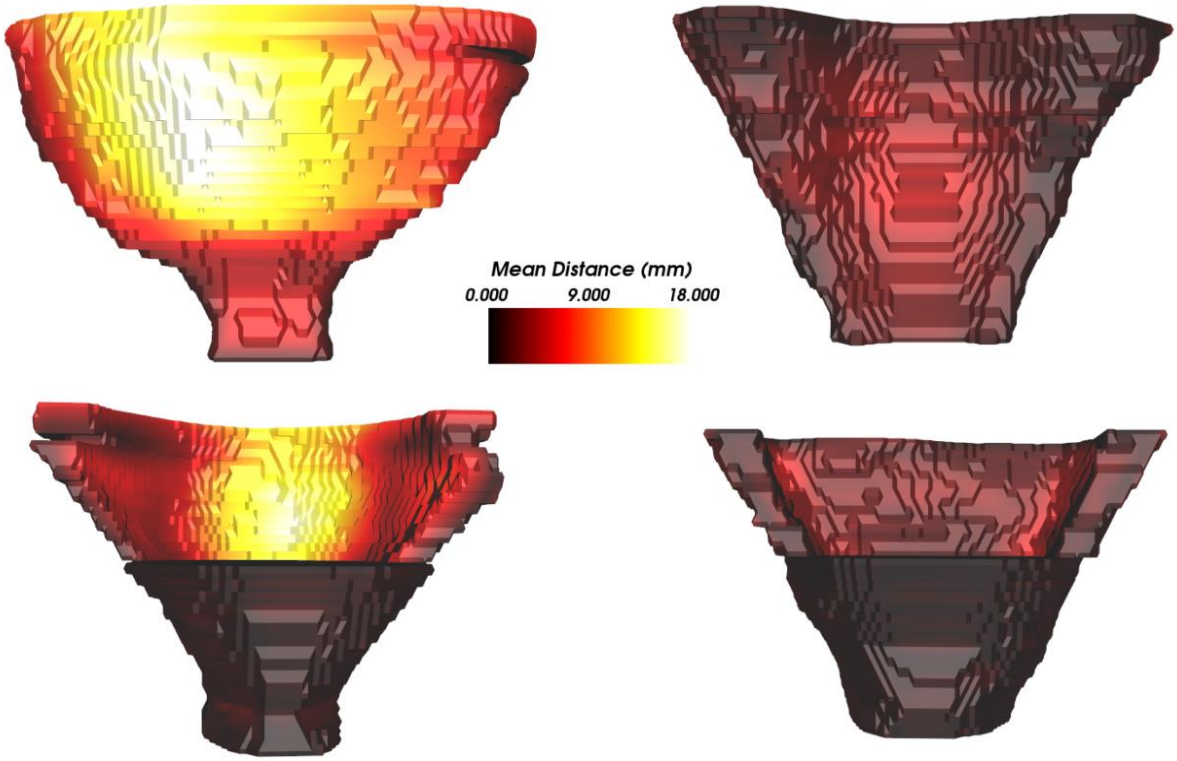
Supplementary Table 1 Summary of patient characteristics for each group

	Non-ERB group	ERB group
Age (years)		
Median	64	64
Range	52-74	52-72
Time from surgery to RT (months)		
Median	7.9	24.4
Range	2.6-99.4	3-75.7
Pre-RT PSA (ng/mL)		
Median	0.12	0.07
Range	0.03-0.26	0.04-0.21
Gleason Score (n)		
6	0	1
7	6	6
8	2	1
9	2	2
Pathology (n)		
pT1	1	0
pT2	3	6
pT3	6	4
Extracapsular extension (n)		
Yes	8	4
No	2	6
Surgical margin		
Positive	8	5
Negative	2	5

Abbreviations: ERB = Endorectal balloon; RT = radiation therapy; PSA = prostate specific antigen



Supplementary Fig 1. Distribution of conformity index results for various organs and sub organs for non-ERB and ERB groups.



Supplementary Fig 2. Posterior (top row) and anterior (bottom row) views of 3D CTV for no ERB (left column, case#8) and ERB (right column, case#1) with mean Euclidean distance from planning scan to CBCT represented by color.

Chapter 7: Superior target volume and organ stability with the use of endorectal balloons in post-prostatectomy radiotherapy

Statement of joint authorship

Title

Superior target volume and organ stability with the use of endorectal balloons in post-prostatectomy radiotherapy

Authors

Jeremiah F de Leon, **Michael G Jameson**, Apsara Windsor, Kirrily Cloak, Sarah Keats, Philip Vial, Lois Holloway, Peter Metcalfe and Mark Sidhom

Jeremiah F de Leon

Developed experimental design and scientific method, collected data, analysed and interpreted data, wrote manuscript

Michael G Jameson (candidate)

Contributed to experimental design and scientific method, contributed to data analysis and interpretation, contributed to manuscript

Apsara Windsor

Contributed to scientific method and manuscript

Kirrily Cloak

Contributed to scientific method and manuscript

Philip Vial

Contributed to scientific method and manuscript

Lois C Holloway

Contributed to scientific method and manuscript

Peter E Metcalfe

Contributed to scientific method and manuscript

Mark Sidhom

Contributed to experimental design, scientific method and manuscript

Chapter 8: Results of the Australasian (TROG) radiotherapy benchmarking exercise in preparation for participation in the PORTEC-3 trial

Statement of joint authorship

Title

Results of the Australasian (TROG) radiotherapy benchmarking exercise in preparation for participation in the PORTEC-3 trial

Authors

Michael G Jameson, Joanne McNamara, Mike Bailey, Peter E Metcalfe, Lois C Holloway, Kerwyn Foo, Viet Do, Linda Mileshkin, Carien Creutzburg and Pearly Khaw

Michael G Jameson (candidate)

Developed experimental design and scientific method, analysed and interpreted data, wrote manuscript

Joanne McNamara

Contributed to experimental design and scientific method, contributed to data analysis and interpretation, contributed to manuscript

Mike Bailey

Contributed to experimental design and scientific method, contributed to data analysis and interpretation, contributed to manuscript

Peter E Metcalfe

Contributed to scientific method and manuscript

Lois C Holloway

Contributed to scientific method and manuscript

Viet Do

Contributed to data analysis and manuscript

Linda Mileshkin

Contributed to scientific method and manuscript

Carien Creutzburg

Contributed to scientific method and manuscript

Pearly Khaw

Contributed to experimental design and scientific method, contributed to data analysis and interpretation, contributed to manuscript

sdjf

Chapter 9: The impact of contouring uncertainty on radiotherapy clinical trial sample size; A novel methodology applied to the PORTEC-3 trial

Statement of joint authorship

Title

The impact of contouring uncertainty on radiotherapy clinical trial sample size; A novel methodology applied to the PORTEC-3 trial

Authors

Michael G Jameson, Joanne McNamara, Mike Bailey, Kerwyn Foo, Anthony Yeo, Linda Mileshkin, Pearly Khaw, Carien Creutzburg, Peter E Metcalfe and Lois C Holloway

Michael G Jameson (candidate)

Developed experimental design and scientific method, collected data, analysed and interpreted data, wrote manuscript

Joanne McNamara

Contributed to experimental design and scientific method, contributed to data analysis and interpretation, contributed to manuscript

Mike Bailey

Contributed to experimental design and scientific method, contributed to data analysis and interpretation, contributed to manuscript

Anthony Yeo

Contributed to scientific method and manuscript

Linda Mileshkin

Contributed to scientific method and manuscript

Pearly Khaw

Contributed to scientific method and manuscript

Carien Creutzburg

Contributed to scientific method and manuscript

Peter E Metcalfe

Contributed to scientific method and manuscript

Lois C Holloway

Contributed to experimental design and scientific method, contributed to data analysis and interpretation, contributed to manuscript

***** The work presented in this chapter is currently under embargo by the PORTEC3 TMC, it will be submitted for publication once the trial has reported *****

The impact of contouring uncertainty on radiotherapy clinical trial sample size; A novel methodology applied to the PORTEC-3 trial

MG. Jameson[1,2,3]; J. McNamara[1,4]; M. Bailey[1,4]; K. Foo[5,6]; AET. Yeo[3]; L. Mileshkin[8]; P. Khaw[8]; C. Creutzberg[9]; PE. Metcalfe[1,3]; L. Holloway[1,2,3]

[1] Centre for Medical Radiation Physics, University of Wollongong, NSW, Australia
[2] Liverpool & Macarthur Cancer Therapy Centres, NSW, Australia
[3] Ingham Institute for Applied Medical Research, NSW, Australia
[4] Illawarra Cancer Care Centre, Wollongong, NSW, Australia
[5] Royal Prince Alfred Hospital, Sydney, NSW, Australia
[6] Sydney Medical School, University of Sydney, Sydney, NSW, Australia
[7] School of Physics, University of Sydney, Sydney, NSW, Australia
[8] Peter MaCallum Cancer Centre, VIC, Australia
[9] Department of Clinical Oncology, Leiden University Medical Centre, Leiden, The Netherlands

Keywords: RCT, sample size calculation, TCP, contouring variation, inter/intra-observer

Running head: Contouring Impact on Sample Size

Corresponding author:
Michael Jameson
Cancer Therapy Centre & Ingham Institute, Liverpool Hospital
Locked Bag 7103, LIVERPOOL NSW 1871
Australia
Phone: +402852750
Email: michael.jameson@sswahs.nsw.gov.au

Total number of pages: 27
Number of words in manuscript: 5104
Number of words in abstract: 355
Total number of tables: 3
Total number of figures: 6 (including supplementary)

Abstract

Background and purpose: Phase three clinical trials are powered based on an estimate of the treatment benefit differential between the standard and experimental arms. The accuracy of radiotherapy contouring may impact on the ability to distinguish between treatment arms. This study aims to model the effect of contouring variation on tumour control probability (TCP) and consequently on clinical trial sample size.

Material and methods: All Australasian observers participating in the PORTEC-3 trial were sent a de-identified CT of a female pelvis on which to contour relevant target structures and normal tissues. Each observer's contours were analysed using in-house code in conjunction with CERR in Matlab®. A "gold standard" consensus target was created by the trial review committee. Geometric analysis consisted of volume, centre of mass (COM), and DICE similarity coefficient with the "gold standard" consensus as a reference. Four-field-box, conformal and intensity modulated treatment plans were generated for each observer set of contours. A standard radiobiological model was used to estimate TCP for each plan calculated onto the "gold standard" contours. The uncertainty in trial sample size was calculated using standard statistical methods.

Results: The variation range in CTV volume, COM, and DICE similarity coefficient across observers was 293 cm^3, 0.29 – 2.7 cm, and 0.49 – 0.98 in relation to the "gold standard" respectively. The mean (± σ) variation in TCP compared to the "gold standard" was -0.29 ± 0.45%, 0.66 ± 0.52%, and 0.18 ± 0.63% for the four field, conformal, and IMRT plans respectively. A 0.29% decrease in TCP lead to a required increase of 3 (642 to 645) patients to maintain the same statistical power. For the worst case of a 1.63% decrease seen in one of the four field plans an extra 19 (642 to 661) patients would be required.

Conclusions: The variation seen in contour definition resulted in a sample size uncertainty of 1.4-2.4%. Radiotherapy clinical trials usually include quality assurance (QA) to ensure contouring variation is limited to an acceptable level. The method reported here could be applied to the results of such QA to improve or verify the accuracy of sample size and power calculations for future RT trials.

Introduction

The randomised controlled trial (RCT) is the most effective means available to answer questions about treatment effectiveness when designed, conducted and reported appropriately [1]. It is well recognised that robust methodology and quality assurance (QA) is required to ensure the validity of RCTs [2]. There are two types of error that trial designers go to great lengths to avoid: Type I and II errors, these are described in detail by Bentzen [3]. Briefly, Type I errors are false positives (treatment A is found to be superior to treatment B when, in fact, it is not) and Type II errors are false negatives (treatment A is found to be no better than treatment B when, in fact, it is), see Table 1. In a retrospective review of clinical trial benchmarking and case review initiatives, Fairchild *et al* demonstrated that QA measures should ensure optimal radiotherapy delivery [4]. There are currently efforts underway to harmonise QA initiatives amongst cooperative groups, one such endeavour is the homogeny of clinical trial groups QA standards [5].

Consistency of contouring according to protocol has been investigated for a number of RCTs [6-10]. The accuracy and consistency of contouring in a RCT may be affected by heterogeneity within contributing institutions technology and experience [11].

Poortmans *et al* [12] calculated the effect of variation in planning on the projected survival for the EORTC 22922/10925 advanced breast cancer RCT. A reduction in projected overall survival from 5% to 3.8% at ten years was estimated due to suboptimal dose distributions collected using a benchmarking case. Pettersen *et al* [13] modelled the impact of dosimetric uncertainty on sample size for RCTs and showed that reduced uncertainty in dose resulted in a significant reduction in required patient numbers. Dosimetric uncertainty is influenced by contouring variation and has been demonstrated to be significant for a number of clinical sites [14-16]. Thus contouring variation may impact on clinical trial outcomes and should be considered in trial design.

Table 1 Description of type I and II errors in clinical trials

		Reality	
		Treatment A ≠ B	Treatment A = B
Research	$p < 0.5$	Correct result	Error (Type I) False positive
	$p > 0.5$	Error (Type II) False negative	Correct decision

PORTEC-3 is a recently closed RCT comparing concurrent chemo-radiation and adjuvant chemotherapy verse pelvic radiation alone in high risk and advanced stage endometrial carcinoma [17]. The radiotherapy component of this RCT required investigators to delineate a number of target structures in the pelvis that were not typically contoured in Australasian centres at the time of recruitment commencement.

Therefore a decision was taken by the Trans-Tasman Oncology Group (TROG) to perform a bench marking exercise to assess contouring consistency amongst Australasian clinicians.

To date there has been no attempt in the literature to incorporate the findings of a benchmarking exercise into the sample size calculation of RCTs to account for the variation in delineation or planning. This paper presents novel methodology for undertaking this, utilising data from the PORTEC-3 benchmarking exercise. The results of the benchmarking study and associated methodology have been presented. Incorporation of the contouring variation observed in the PORTEC-3 benchmarking study into the RCT sample size calculation is presented in the current study.

Methods

The proposed methodology is described in Fig. 1. and consists of three main stages. The first is the assessment of contouring variation using data from the PORTEC-3 benchmarking exercise. The second involves analysing the impact of the contouring variation on dosimetry, this required the generation of treatment plans for the benchmarking contours. Unlike a benchmarking study the treatment plans (four field box, 4FLD box; conformal, 3DCRT; intensity modulated, IMRT) were generated by two investigators (MJ and JM), see section 2.2. Dosimetric variation was assessed using physical dose volume histograms (DVH) to calculate tumour control probability (TCP). The third stage consisted of incorporating the modelled variation into the RCT sample size calculation as uncertainty in the survival rates of the standard and experimental arms.

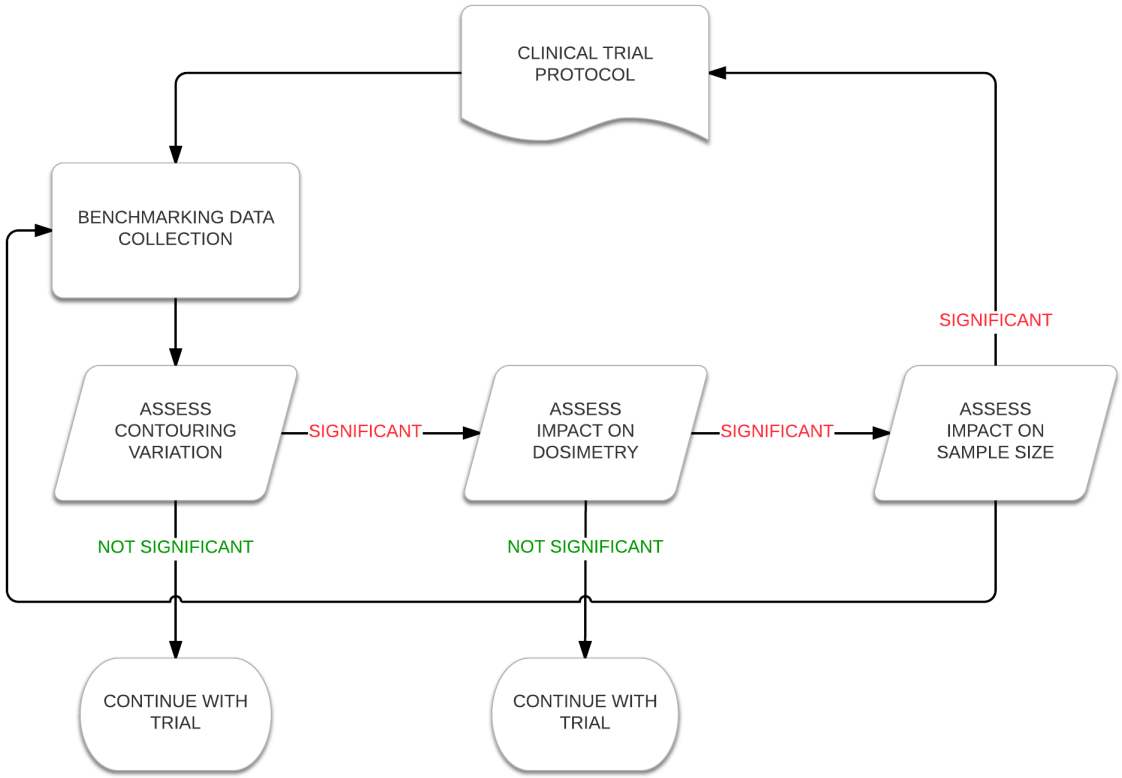


Fig. 1. Flow chart with decision points describing the proposed methodology for assessing the impact of contouring variation in RCT design (modified from Nelms [18, 19]).

1.1.1 Target delineation

Participating observers were asked to contour a test case according to trial protocol, where multiple observers from single institutions would be contributing patients, each individual observer took part. Contouring consisted of the CTV including the upper 50% of the vagina, the vaginal tissues superior to the vaginal marker, the paravaginal / parametrial soft tissues, and the distal common, external, and internal iliac lymph node regions. Inclusion of the sub-aortic pre-sacral nodes was recommended for tumours with involvement of the cervix. A margin of 7-10 mm was to be used from CTV to PTV with a margin of 12 mm in the upper vaginal region to account for bladder and rectal filling. “gold standard” reference volumes were created for comparison of observer

contours, these consisted of consensus target and organ at risk (OAR) volumes delineated by the local trial coordinators.

1.1.2 Treatment planning

Treatment plans were then generated using the Pinnacle3® v9.0 (Philips Medical Systems, Nederland B.V. Best, The Netherlands) treatment planning system (TPS). The Adaptive Convolve algorithm was employed with a dose grid of 3 mm^3. The sensitivity of planning technique to contouring variation was assessed by generating three different plans for all benchmarking and "gold standard" contours; 1) a 4FLD box using 10 MV photons with AP, PA and lateral beams, 2) a 3DCRT plan using seven 10 MV photon beams and 3) a 10 MV IMRT plan using "gold standard" OAR volumes in the optimization process as normal tissues were not delineated by all participating institutions.

1.1.3 Geometric variation

All available observer's contours were collated onto a single CT with the "gold standard" contours. This was loaded into CERR [20] and an in-house [15] developed MATLAB® (The Mathworks Inc, Natick, MA 2009) script was used to analyse each observer's target volumes. The volume and centre of mass (COM) of each observer's CTV contour was assessed. To quantify the variation with respect to the "gold standard" target the DICE similarity coefficient (DSC) was used [21]. For each observers' target volume, A, and the gold target volume, B, the DSC is defined as:

$$DSC = \frac{2(A \cap B)}{(A + B)} \quad (1)$$

A DSC equal to zero indicates that the two volumes do not overlap at all, while a DSC equal to one indicates perfect overlap.

1.1.4 Tumour Control Probability (TCP)

To assess the impact of contouring variation on TCP the "gold standard" target volume was assumed to be the true target. Treatment plans were developed as described in section 2.2 for each observer's target. The TCP calculated for the "gold standard" target volume. CTV DVHs were used to calculate TCP using the Comp Plan program [22].

$$TCP_{logit} = \prod \left[\frac{1}{1 + \left(\frac{TCD_{50}}{D_i}\right)^{4\gamma_{50}}} \right]^{v_i} \quad (2)$$

The logit model (equation 2) was used, where the dose to achieve 50% control is TCD_{50}, the slope of dose response curve is γ_{50}, the normalized volume is v_i and the dose to the voxel being considered is D_i. Parameters were chosen as the mean values from multiple institution adjuvant radiotherapy cohorts as reported by Okunieff *et al* [23] with TCD_{50} of 30.80 Gy, and a γ_{50} of 0.40 %/%. To assess any bias introduced by choice of model and parameters these calculations were repeated with a number of published models and data, see supplementary Fig. 1.

1.1.5 Statistical Considerations

The baseline number of required patients was calculated assuming an exponential survival curve. With a false positive error rate of 5% (α = 0.05), a power of 80% (1 – β = 0.80) assuming equal patient numbers in each arm (p = 0.50). An accrual period of 5 years and follow up of 2.5 years was used to detect a 10% difference in 5-year overall survival (OS) with the standard arm having OS of 65-75% [17]. The minimum number of patients required in the PORTEC-3 protocol was 655 with a target of 670 and the final number included at close was 686.

The variation in TCP due to contouring uncertainty obtained from the PORTEC-3 benchmarking exercise was then incorporated into the power calculation as uncertainties in the OS rates for each arm. These uncertainties were applied to both arms equally as patients were randomized.

The sample size calculation was based on a parallel fixed sample size clinical trial with survival as the main endpoint [17]. First the number of events (i.e. deaths) to be observed is calculated:

$$D = \frac{\left(Z_{1-\alpha/2}+Z_{1-\beta}\right)^2}{p(1-p)\ln(\delta)^2} \qquad (3)$$

Where D is the number of deaths, p is the allocation ratio (i.e. 0.5 for a 1:1 allocation), δ is the hazard ratio $\delta = log(R_N)/log(R_S)$ where, R_N and R_S are the survival rates in the new and standard arms. The $Z_{1-\alpha/2}$ and $Z_{1-\beta}$ values represent area under the normal distribution related to the significance level and statistical power respectively. Once the number of deaths (equation 4) required is known the number of patients, N, can be estimated.

$$D = ra\left(1 - \bar{R}^{-(f+a/2)/t}\right) \quad (4)$$

Where r is the accrual rate which is assumed to be known, f is the length of follow up from the end of accrual, t is the time at which $\bar{R}$ is estimated, a the accrual duration is then chosen to give the number of deaths required by Equation (4). Hence the number of patients $N = ra$. To assess the impact of choice of sample size calculation technique a number of difference methods were evaluated and compared in Supplementary Fig. 2.

Results

At the time of analysis, of the 31 datasets distributed to Australasian centres one dataset was missing, two were in non DICOM format, five were corrupt and not able to be imported into Pinnacle3®, and five of the datasets were exact copies of other submissions from the same institution (presumably reviewed by the contributing observer). This left 18 distinct datasets available for analysis as part of this study. The contours analysed as part of this study are displayed in Fig. 2.

1.1.6 Geometric variation

Variation in volume of the contoured target is illustrated in Fig. 3 A). The whiskers represent the minimum and maximum values while the box shows the 2nd and 3rd quartiles. The mean contoured CTV volume was 398.9 cm^3 (range: 228.4 – 521.4 cm^3). The distance between COM (ΔCOM) of the "gold standard" and each contoured target is shown in Fig. 3 A). Most ΔCOM were less than 2.0 cm with a mean of 1.4 cm (range: 0.3 – 2.7 cm). The mean DSC for the CTV was 0.73 (range: 0.49 – 0.98). The observer CTV volume with the highest CI also had the lowest ΔCOM.

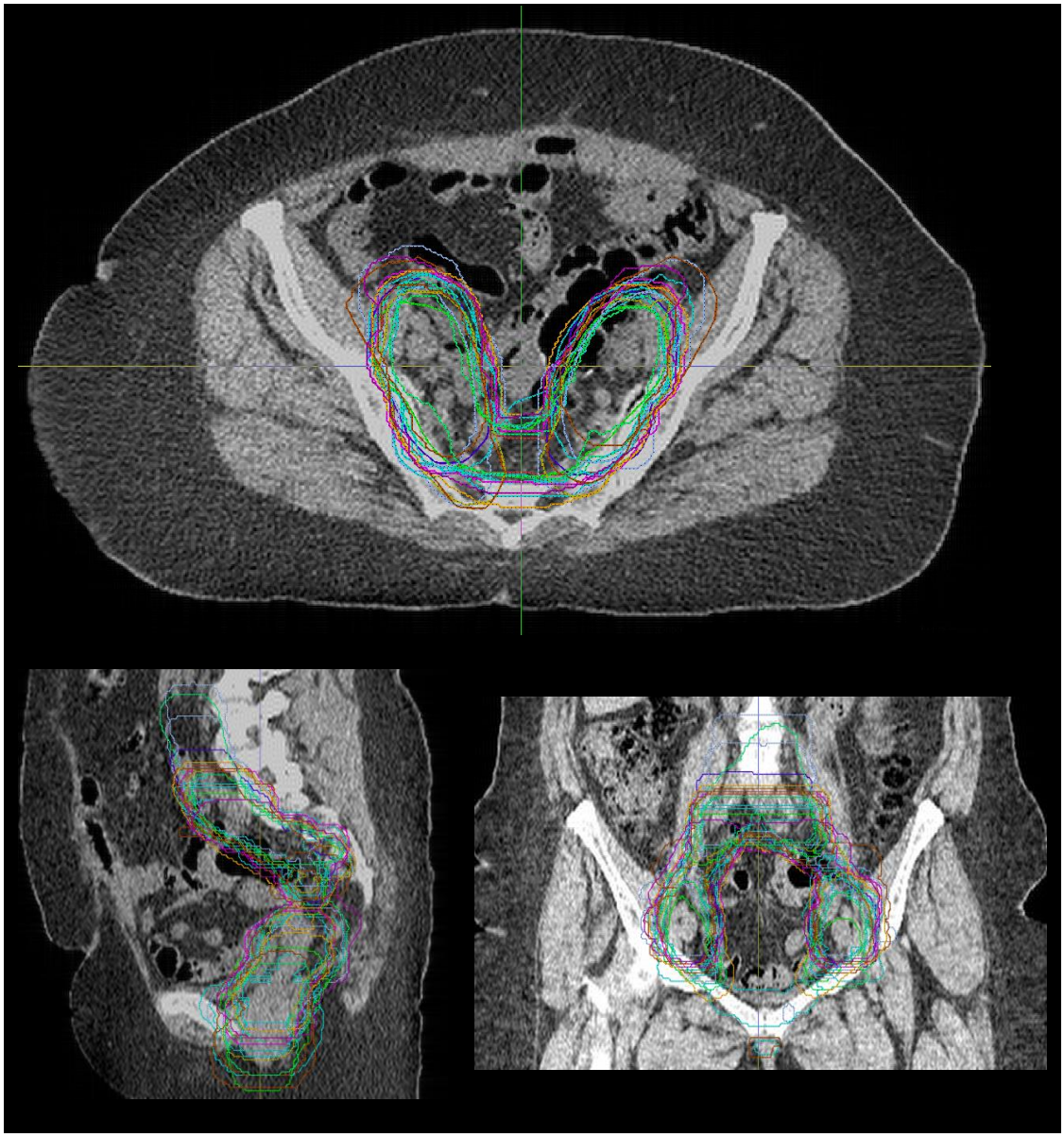


Fig. 2. Axial, sagittal and coronal slice through the pelvis showing variation in PTV definition

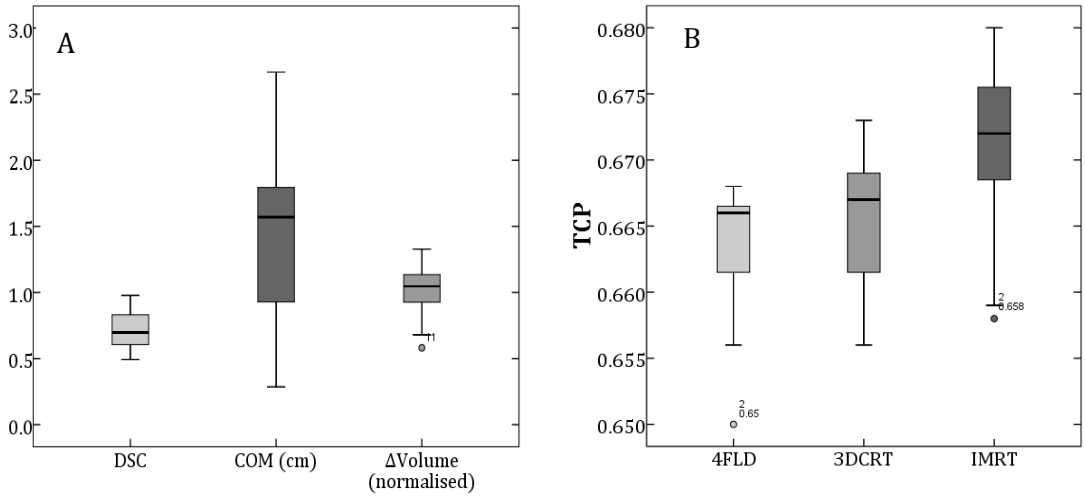


Fig. 3. Box and whisker plots representing the variation in A) variation in DSC and COM and normalized volume relative to the "gold standard" reference volume and B) the variation in TCP as calculated on the "gold standard" reference for each planning technique

1.1.7 Tumour Control Probability

The variation in TCP for each participating centre per planning technique is presented in Fig.3 B) and Table 2. There was a significant difference in mean TCP variation between IMRT 67.1% (σ = 0.6%) and 4FLD 66.3% (σ = 0.5%; $p < 0.01$), and 3DCRT 66.6% (σ = 0.5%) and IMRT ($p < 0.01$), but there was no significant difference between 3DCRT and 4FLD ($p = 0.25$). The TCP of the "gold standard" target volume and plan was 66.9% for the IMRT and 66.7% for the 4FLD box while the 3DCRT plan was 65.9%. In comparison to the 4FLD box plan, the IMRT and 3DCRT plans resulted in reduced small bowel dose.

Table 2 Difference in TCP from baseline value calculated for "gold standard" reference volume specific plan

	4FLD	*3DCRT*	*IMRT*
Mean	*-0.29%*	*0.66%*	*0.18%*
σ	*0.45%*	*0.52%*	*0.63%*
Max	*-1.63%*	*-0.34%*	*1.18%*

1.1.8 Sample Size

Table 3 summarises the effect of the TCP variation on the sample size calculation. Using the original survival estimates for the standard and experimental arm the sample size estimate was 642 patients. If the systematic and random TCP variation from the "gold standard" is included for the 4FLD, 3DCRT, and IMRT planning techniques the sample size estimate is 645 (σ = 9), 633 (σ = 13) and 639 (σ = 16) respectively.

Table 3 Sample size calculations for reference and incorporating contouring variation assessed in the PORTEC-3 benchmarking study

	Control rate (%)	Experimental rate (%)	Sample size estimate
Planned	65	75	642
4FLD	64.71 (σ = 0.45)	74.71 (σ = 0.45)	645 (σ = 9)
3DCRT	65.66 (σ = 0.52)	75.66 (σ = 0.52)	633 (σ = 13)
IMRT	65.18 (σ = 0.63)	75.18 (σ = 0.63)	639 (σ = 16)

The relationship between control in the standard (with corresponding assumptions from section 2.5) and the required sample size is displayed in Fig. 4 with the baseline sample size marked in red and the mean (solid) ± σ (dashed) for 4FLD marked in green.

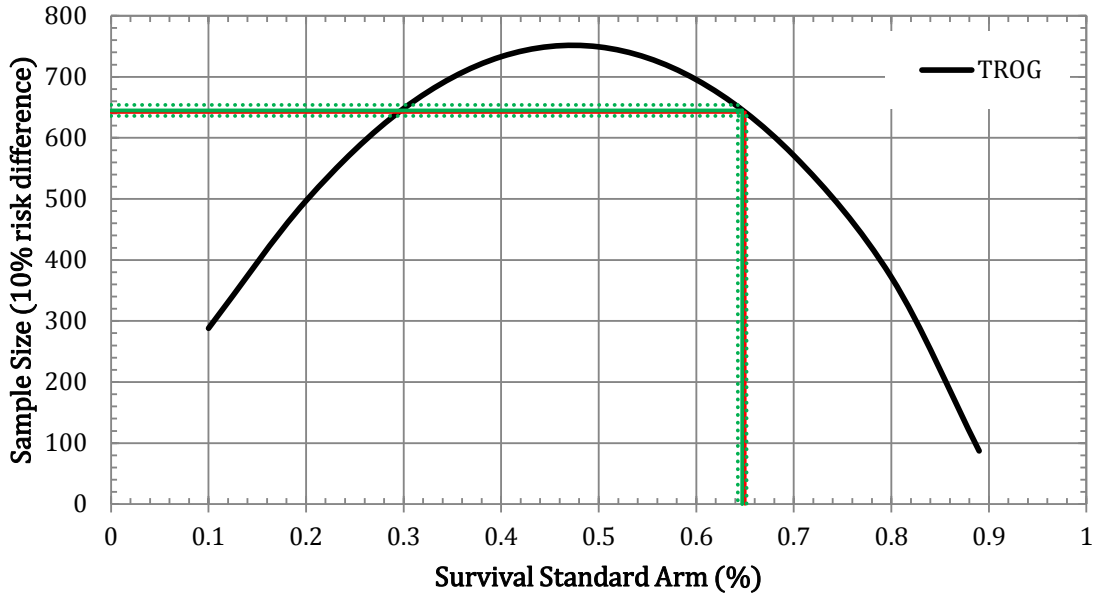


Fig. 4. Plot showing relationship between control rate and sample size. Obtained by varying the survival rate and keeping the risk differential at 10%. Red line indicates baseline sample size and green represents mean (solid) ± σ (dashed) sample size for 4FLD technique taking contouring variation into account

1.1.9 Discussion

Clinical trials should be conducted with robust methodology and QA. In trials that include radiotherapy it has been shown that as the heterogeneity of the radiotherapy in a trial goes up, so too does the number of patients needed to detect a significant treatment differential [3]. Thus, it is in the best interest of the trial investigators to minimize the heterogeneity in delivered radiotherapy. One of the largest contributing factors to treatment heterogeneity in radiotherapy is contouring variation [9].

There are a number approaches to minimizing contouring variation in RCTs involving radiotherapy, these QA measures have been reviewed by Webber *et al* [24]. Level 1 is site credentialing [25] whereby a site that is contributing patients to a trial(s) is credentialed by an outside body. This usually involves a facility questionnaire and a dosimetry audit. Level 2 is the bench marking exercise [12] in which an example case is sent to institutions that are interested in contributing patients to a trial. These institutions generally contour and/or plan this case according to trial protocol and the results are made available to the institution. Levels 3 and 4 consist of individual case review [26], where either limited selected RCT cases (level 3) or an extensive number (level 4) are reviewed by the RCT committee or other QA group. The results of the independent case review are then provided to the participating institutions. Level 5 is a complex dosimetry check consisting of generating a protocol specific plan on a physical phantom, irradiating the phantom and having results reviewed by an independent team. Many trials use a combination of these approaches to assess institutions when contributing patients to a RCT.

In the current study the results of the ANZGOG/TROG initiated PORTEC-3 benchmarking study have been incorporated into a RCT sample size calculation. For the reference conditions used, the number of patients required was not significantly affected by the variation in contouring observed. This can be attributed to the contouring variation observed not being substantial given the very large target volume. Additionally, due the location of the target volume (medial) the dosimetry was relatively insensitive to contouring variation. Observer's tended to over rather than under contour, resulting in adequate coverage of the "gold standard" reference CTV, however potential over exposure of OARs (e.g. small bowel and bladder). The largest

variation in TCP was observed for the IMRT technique, this was as expected as the more conformal the dose distribution the greater the sensitivity to contouring variation.

The amount of contouring variation may have been larger than if the benchmarking study was run later in the trial. As observers contribute patients to a trial over time they become more familiar with the protocol which may reduce the amount of contouring variation [4]. As this was an ANZGOG/TROG initiated benchmarking study (i.e. not general QA) all of the observers were from Australian and New Zealand centres and therefore may not be representative of the wider international group contributing patients to the trial. The benchmarking study only consisted of one patient data set and contouring variation may be influenced by patient specific parameters. For example, unusual anatomy and poor image quality due to patient size. Also, the benchmarking patient was stage IIA grade 3 and the PORTEC-3 trial allows for a variety of high risk and advanced stage stratifications [17].

There are a number of uncertainties and potential bias associated with this type of analysis. These relate to radiobiological modelling of TCP with respect to model/parameter choice, sample size calculation methods employed, and the assumptions on which they are based. These have been assessed in the supplementary material section and potential impacts stated above. Due to the limited availability of TCP model parameters for some tumour sites and differences in underlying statistical assumptions used, uncertainty and bias analysis should be performed for each trial protocol when the methodology proposed in this work is employed Fig. 1.

The sample size calculation method used was based on the TROG statistical guidelines and was compared to other techniques (see supplementary Fig. 2.). For the assumptions made (0.65 control rate with 10% risk difference, see 2.5 *Statistical Considerations*) the spread among the different calculation techniques was approximately 40 patients. Although this variance seems large the gradient of each of the techniques in supplementary Fig. 2. is approximately equal at a control rate of 0.65. Therefore, the reported difference in sample size from baseline will be equal for each technique. Ideally statistical design of clinical trials would model the incorporation of uncertainties involved in the parameters used. The final number of patients included in PORTEC-3 was 686, 16 more than the planned target of 670 with a minimum requirement of 655. The minimum number required differs from the 642 calculated in above, this is likely due the use of a different method of calculation (see Supplementary Fig.2.).

In this study local control (TCP) in the adjuvant setting was modelled as a surrogate for overall survival as used in a sample size calculation. Although there is no data on the link between TCP and overall survival for endometrial carcinoma one of the aims in controlling local disease is the prevention of metastatic spread. In a retrospective analysis of high risk patients (stage IC, grade 3) registered but not eligible for the original PORTEC-1 trial, local relapse rates for adjuvant RT alone were 13%, while the rates of distant metastases and overall survival were 31% and 74% at 5 years [27]. In the combined modality setting Greven *et al* reported the results of adjuvant radiotherapy combined with cisplatin/paclitaxel chemotherapy [28]. The four-year recurrence rates were 2%, and 19% for pelvic regional and distant disease. Furthermore, overall survival and disease free survival at four years was 85% and 81% respectively.

For the purpose of the sample size calculations, TCP uncertainty due to contouring variation was assumed to be equal in both arms however only the experimental arm received chemotherapy. Therefore the effect on TCP of contouring variation may be differential as chemotherapy will shift the dose response curve [29, 30]. Additionally, TCP uncertainty due to contouring variation may mask the benefit of combined modality regimens and can impact on overall survival [9]. This differential effect depending on treatment arm may change the assumed risk difference (10% for PORTEC-3) between the two arms and hence may have a large impact on sample size. For example, a 1% decrease in risk differential (from 10% to 9%) equates to an additional 160 (80 in each arm) patients required to maintain $1 - \beta = 0.80$.

There will likely also be contouring variation in any previous studies on which the trial in question was based thus, one might argue that the impact of contouring variation is already taken into account in the randomization process. Nevertheless, the contouring variation or the impact of this contouring variation is likely to vary between the previous studies and the study in question. Typically the number of clinicians contributing to RCTs is larger than the pilot studies on which they are based increasing the probability of inter-observer variation due lack of familiarisation with technique, and small patient numbers treated at contributing sites [9]. Moreover, radiotherapy treatment and planning technology changes over time. This change in technology might be explicit due to new techniques (e.g 3DCRT to IMRT), images used for contouring (e.g. CT to MRI based planning), or less obvious due to changes in planning tools and image quality [31].

The current study employed two planners to complete the treatment plans using contours from the contributing observers, this was to enable the comparison of 4FLD,

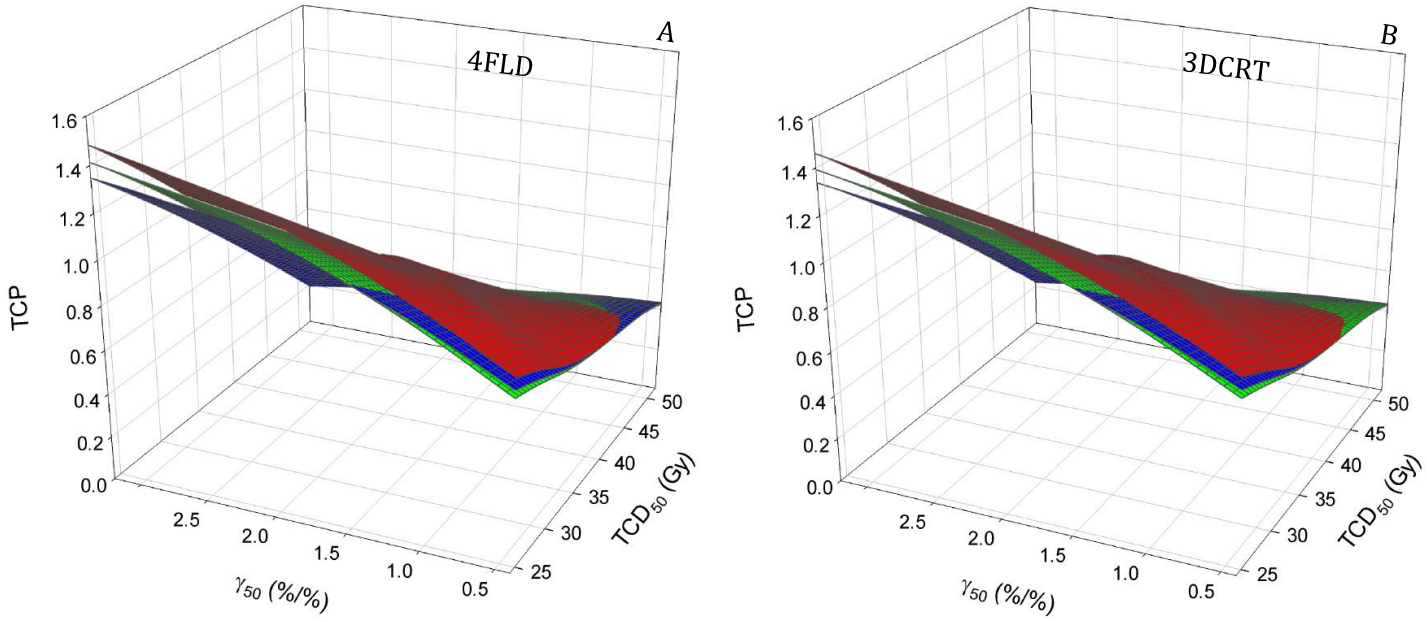
3DCRT and IMRT techniques. In reality, a large clinical trial would contain planning variation, possibly increasing the impact on sample size.

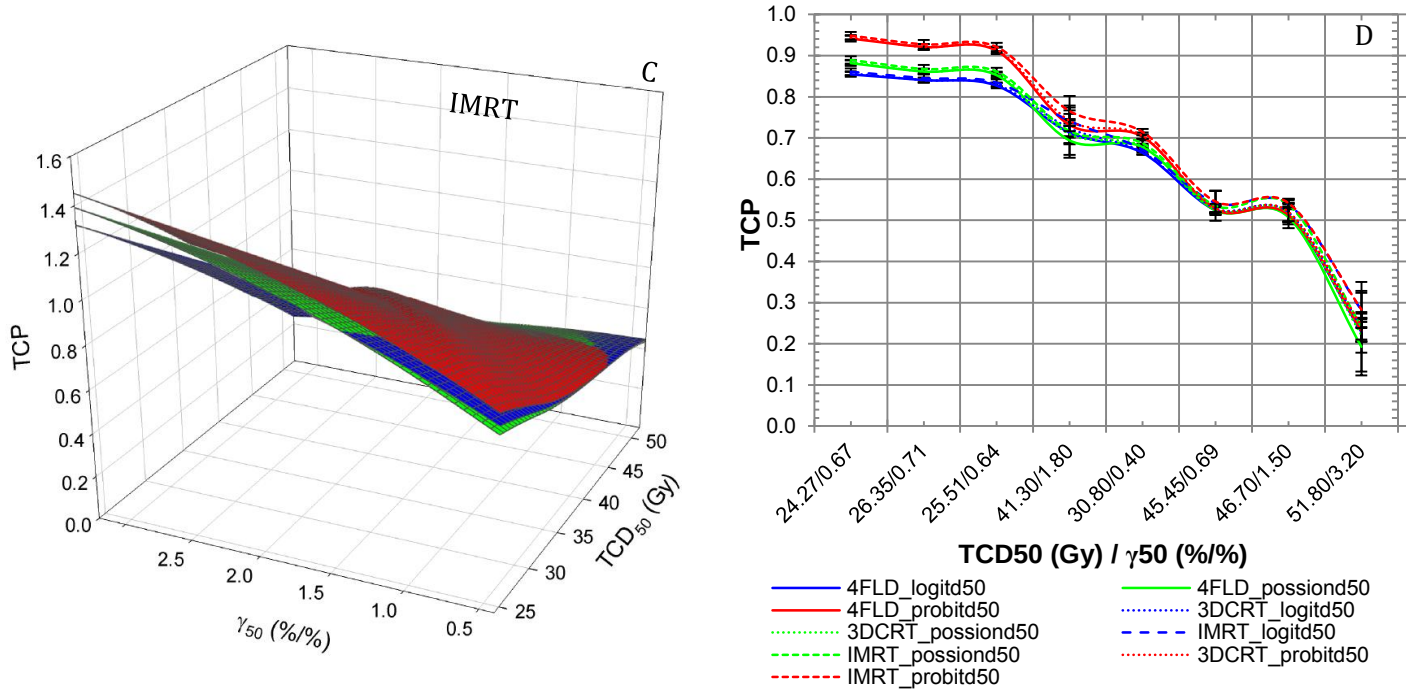
In a modelling study, Pettersen *et al* demonstrated that as the uncertainty in delivered dose increases, the required sample size to answer a clinical question to a given power increases [13]. Poortmans *et al* reported the results of a benchmarking study from the EORTC 22922/10925 protocol and claimed that the dosimetric variation observed may lead to a falsely non-significant result, fortunately this was not the case [32, 33]. Both of these studies advocate for rigorous QA and dosimetry credentialing of centres before contributing to RCTs. Previous work has shown that geometric contouring variation is significantly correlated with variation in TCP [15]. While a number studies have assessed contouring variation with benchmarking datasets [4] this is the first to assess the impact of that variation on the sample size calculation.

Conclusion

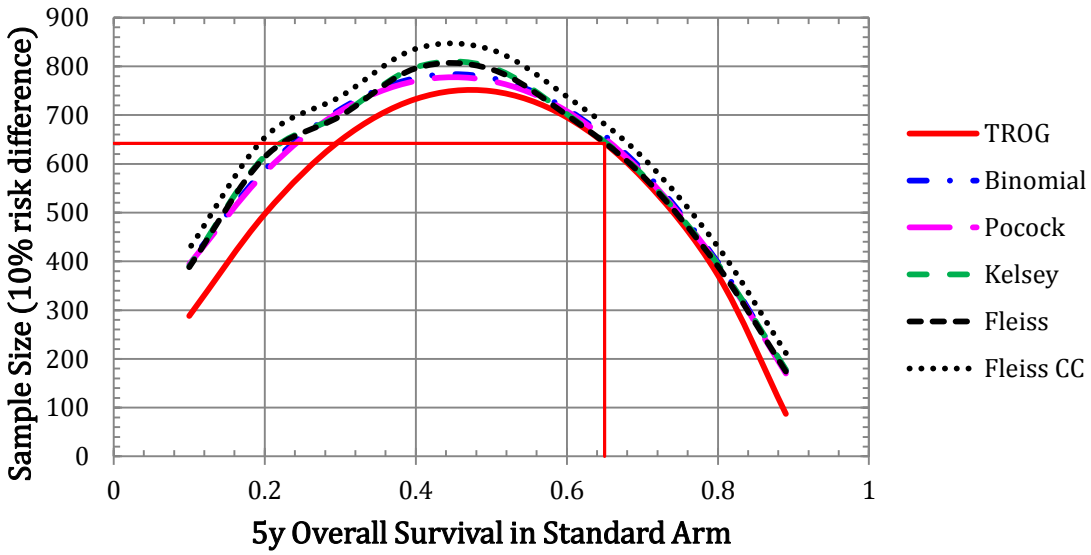
A methodology for the incorporation of contouring uncertainty available through preclinical trial QA to assess necessary sample size has been proposed and tested using data available from the PORTEC-3 ANZGOG/TROG benchmarking exercise. It was demonstrated that contouring variation can result in an increase in required sample size. The impact of contouring variation on sample size varies with respect to the sample size calculation method and the treatment technique. Consequently, this type of assessment should be performed in the initial protocol development stage of radiotherapy RCTs. It is of particular importance in combined modality trials where the impact of the contouring variation may differ depending on the arm of the trial.

Supplementary Material





Supplementary Fig. 5. Radiobiological model and parameter uncertainty analysis. Figures A), B) and C) show 3D mesh of TCP values for corresponding TCD50 and gamma50 parameters. As can be seen from the figures the probit model consistently returns higher TCP values except for larger TCD50 where the logit is higher for IMRT and 4FLD and the possion for 3DCRT. These were calculated using the model and parameter values listed in D) [22, 34].



Supplementary Fig. 6. Sample size calculated using the TROG guidelines [35], Binomial [36], Pocock [37], Kelsy, Fleiss and Fleiss with continuity correction [38]. Assuming a 10% risk differential between the standard and experimental arms. Red vertical and horizontal lines indicate standard arm survival and corresponding sample size estimate from PORTEC3 protocol.

References

1. Plint AC, Moher D, Morrison A, Schulz K, Altman DG, Hill C, et al. Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. Medical journal of Australia. 2006;185(5):263.

2. Schulz K, Altman D, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMC medicine. 2010;8(1):18.

3. Bentzen SM. Towards evidence based radiation oncology: improving the design, analysis, and reporting of clinical outcome studies in radiotherapy. Radiotherapy and Oncology. 1998;46(1):5-18.

4. Fairchild A, Collette L, Hurkmans C, Baumert B, Weber D, Gulyban A, et al. Do results of the EORTC dummy run predict quality of radiotherapy delivered within multicentre clinical trials? European Journal of Cancer. 2012;48(17):3232-9.

5. Melidis C, Bosch WR, Izewska J, Fidarova E, Zubizarreta E, Ulin K, et al. Global Harmonization of Quality Assurance Naming Conventions in Radiation Therapy Clinical Trials. International Journal of Radiation Oncology* Biology* Physics. 2014;90(5):1242-9.

6. Dieckmann K, Pötter R, Wagner W, Prott FJ, Hörnig-Franz I, Rath B, et al. Up-front centralized data review and individualized treatment proposals in a multicenter pediatric Hodgkin's disease trial with 71 participating hospitals: the experience of the German-Austrian pediatric multicenter trial DAL-HD-90. Radiotherapy and Oncology. 2002;62(2):191-200.

7. Dusserre A, Garavaglia G, Giraud JY, Bolla M. Quality assurance of the EORTC radiotherapy trial 22863 for prostatic cancer: the dummy run. Radiotherapy and Oncology. 1995;36(3):229-34.

8. Li XA, Tai A, Arthur DW, Buchholz TA, Macdonald S, Marks LB, et al. Variability of target and normal structure delineation for breast cancer radiotherapy: an RTOG Multi-Institutional and Multiobserver Study. International Journal of Radiation Oncology* Biology* Physics. 2009;73(3):944-51.

9. Peters LJ, O'Sullivan B, Giralt J, Fitzgerald TJ, Trotti A, Bernier J, et al. Critical impact of radiotherapy protocol compliance and quality in the treatment of advanced head and neck cancer: Results from TROG 02.02. Journal of Clinical Oncology. 2010;28(18):2996.

10. Valley JF, Bernier J, Tercier PA, Fogliata-Cozzi A, Rosset A, Garavaglia G, et al. Quality assurance of the EORTC radiotherapy trial 22931 for head and neck carcinomas: the dummy run. Radiotherapy and Oncology. 1998;47(1):37-44.

11. Jameson MG, Holloway LC, Vial PJ, Vinod SK, Metcalfe PE. A review of methods of analysis in contouring studies for radiation oncology. Journal of Medical Imaging and Radiation Oncology. 2010;54(5):401-10.

12. Poortmans PMP, Venselaar JLM, Struikmans H, Hurkmans CW, Davis JB, Huyskens D, et al. The potential impact of treatment variations on the results of radiotherapy of the internal mammary lymph node chain: a quality-assurance report on the dummy run of EORTC Phase III randomized trial 22922/10925 in

Stage I-III breast cancer1. International Journal of Radiation Oncology* Biology* Physics. 2001;49(5):1399-408.

13. Pettersen MN, Aird E, Olsen DR. Quality assurance of dosimetry and the impact on sample size in randomized clinical trials. Radiotherapy and Oncology. 2008;86(2):195-9.

14. Barghi A, Johnson C, Warner A, Bauman G, Battista J, Rodrigues G. Impact of Contouring Variability on Dose-Volume Metrics used in Treatment Plan Optimization of Prostate IMRT. Cureus. 2013;5(11).

15. Jameson MG, Kumar S, Vinod SK, Metcalfe PE, Holloway LC. Correlation of contouring variation with modeled outcome for conformal non-small cell lung cancer radiotherapy. Radiotherapy and Oncology. 2014.

16. Stanley J, Dunscombe P, Lau H, Burns P, Lim G, Liu H-W, et al. The Effect of Contouring Variability on Dosimetric Parameters for Brain Metastases Treated With Stereotactic Radiosurgery. International Journal of Radiation Oncology* Biology* Physics. 2013;87(5):924-31.

17. Creutzberg C. Randomized Trial of Radiation Therapy With or Without Chemotherapy for Endometrial Cancer (PORTEC-3).

18. Nelms BE, editor The Most Important Medical Devices. AAMD Annual Meeting; 2011; St. Louis, MO, USA.

19. Nelms BE, Tomé WA, Robinson G, Wheeler J. Variations in the contouring of organs at risk: test case from a patient with oropharyngeal cancer. International Journal of Radiation Oncology* Biology* Physics. 2012;82(1):368-78.

20. Deasy JO, Blanco AI, Clark VH. CERR: a computational environment for radiotherapy research. Medical Physics. 2003;30:979.

21. Dice LR. Measures of the amount of ecologic association between species. Ecology. 1945;26(3):297-302.

22. Holloway LC, Miller J-A, Kumar S, Whelan BM, Vinod SK. Comp Plan: A computer program to generate dose and radiobiological metrics from dose-volume histogram files. Medical Dosimetry. 2012;37(3):305-9.

23. Okunieff P, Morgan D, Niemierko A, Suit HD. Radiation dose-responce of human tumours. International Journal of Radiation Oncology* Biology* Physics. 1995;32(4):1227-37.

24. Weber DC, Poortmans PM, Hurkmans CW, Aird E, Gulyban A, Fairchild A. Quality assurance for prospective EORTC radiation oncology trials: the challenges of advanced technology in a multicenter international setting. Radiotherapy and Oncology. 2011;100(1):150-6.

25. Palta JR, Deye JA, Ibbott GS, Purdy JA, Urie MM. Credentialing of institutions for IMRT in clinical trials. International Journal of Radiation Oncology* Biology* Physics. 2004;59(4):1257-9.

26. Matzinger O, Poortmans P, Giraud JY, Maingon P, Budiharto T, van den Bergh A, et al. Quality assurance in the 22991 EORTC ROG trial in localized prostate cancer: Dummy run and individual case review. Radiotherapy and Oncology. 2009;90(3):285-90.

27. Creutzberg CL, van Putten WL, Wárlám-Rodenhuis CC, van den Bergh AC, De Winter KA, Koper PC, et al. Outcome of high-risk stage IC, grade 3, compared with stage I endometrial carcinoma patients: the Postoperative Radiation Therapy in Endometrial Carcinoma Trial. Journal of Clinical Oncology. 2004;22(7):1234-41.

28. Greven K, Winter K, Underhill K, Fontenesci J, Cooper J, Burke T. Final analysis of RTOG 9708: adjuvant postoperative irradiation combined with cisplatin/paclitaxel chemotherapy following surgery for patients with high-risk endometrial cancer. Gynecologic oncology. 2006;103(1):155-9.

29. Joiner MC, van der Kogel A. Basic Clinical Radiobiology Fourth Edition: CRC Press; 2009.

30. Jones B, Dale R. The potential for mathematical modelling in the assessment of the radiation dose equivalent of cytotoxic chemotherapy given concomitantly with radiotherapy. 2014.

31. Jameson M, Bailey M, Foo K, Yeo AET, Holloway L, Metcalfe P. Methodology for assessment of impact of contouring variation on clinical trial design. EPSM-ABEC; Darwin, NT, Australia2011.

32. Poortmans P, Kirkove C, Budach V, Maingon P, Valli M, Collette S, et al. Irradiation of the internal mammary and medial supraclavicular lymph nodes in stage I to III breast cancer: 10 years results of the EORTC radiation oncology and

breast cancer groups phase III trial 22922/10925. Eur J Cancer. 2013;49(Suppl 3):S1-19.

33. Poortmans PM, Venselaar JL, Struikmans H, Hurkmans CW, Davis JB, Huyskens D, et al. The potential impact of treatment variations on the results of radiotherapy of the internal mammary lymph node chain: a quality-assurance report on the dummy run of EORTC Phase III randomized trial 22922/10925 in Stage I–III breast cancer. International Journal of Radiation Oncology* Biology* Physics. 2001;49(5):1399-408.

34. Okunieff P, Morgan D, Niemierko A, Suit HD. Radiation dose-response of human tumors. International Journal of Radiation Oncology* Biology* Physics. 1995;32(4):1227-37.

35. TROG. Statistical Guidlines TPP E5. Trans Tasman Radiation Oncology Group Limited 2009.

36. Brant R. Inference for Proportions: Comparing Two Independent Samples [10/01/2015]. Available from: http://stat.ubc.ca/~rollin/stats/ssize/b2.html.

37. Pocock SJ. Clinical trials: a practical approach: Wiley Chichester; 1983.

38. Dean A, Sullivan K, Soe M. OpenEpi: open source epidemiologic statistics for public health, version 3.03 [updated 22/09/201410/01/2015]. Available from: www.OpenEpi.com.

Chapter 10: Discussion and conclusions

10.1 General discussion

Despite much advancement in the techniques and technology associated with radiotherapy treatment planning and delivery, there are a number of aspects that still present challenges and require further research. Accurate and precise delineation of targets and normal tissues is still one of the largest uncertainties in the radiotherapy planning chain[1]. One of the shortcomings of the contouring study literature is the lack of consensus on metrics of comparison of contours[2], the result of which is the inability to compare results between studies. Chapter 4 addresses this problem by presenting a framework for establishing the most significant metrics of variation for particular treatment sites and techniques. This, combined with a minimum set of metrics[2] and appropriate statistical presentation should allow for the inter-comparison of contouring studies in the future.

There has been considerable effort devoted to the development of robust automatic treatment planning techniques but these are yet to become widespread[3, 4]. The results of chapter 5 demonstrate that plan quality decreased with decreasing planner experience, and, the efficiency of plan delivery also increased with increasing planner experience. These results could

be seen as an example as to why automated inverse planning techniques warrant further investigation in head and neck radiotherapy.

A number of techniques to account for organ motion in the delivery of radiotherapy have been proposed but not widely adopted[5]. This lack of adoption may be due to the difficulty is assessing the effectiveness or otherwise of these devices. The analysis techniques used in chapter 4 were suited perfectly to investigating the organ motion problem on delineated daily CBCT imaging and could be applied to a number of situations to assess the effectiveness of IGRT approaches.

Clinical trials are regarded as the gold standard when it comes to making informed decisions about health care interventions[6]. However, it is also recognised that clinical trials need to be robustly designed and implemented to ensure results are unbiased[6]. Quality assurance in radiotherapy clinical trials is the key tool in ensuring that the results of a trial are valid and widely applicable. By assessing the impact of contouring variation on modelled outcome using the same analysis techniques described in chapter 4 it is possible to make clinical trials more robust to radiotherapy planning uncertainty, thus increasing the effectiveness and impact of these trials which form the cornerstone of radiotherapy practice.

This thesis represents a body of work investigating areas of uncertainty in radiotherapy planning, delivery and clinical trials. The main themes of the conducted research were outlined in Chapter 1 and include:

I. The impact of contouring variation on modelled radiotherapy outcome
II. The influence of planner experience on IMRT plan quality
III. Investigation of organ stability, dosimetry, and margins in the presence of organ stabilising devices
IV. Benchmarking and assessing the impact of contouring variation in radiotherapy clinical trials

Like William Tell, guiding the arrow to hit the apple, the delivery of safe and effective radiotherapy needs to be both accurate and precise. Accuracy and precision are inextricably linked in the aim of radiotherapy; to maximise the probability of cure without injury. The work presented here seeks to address the issue of precision. The issue of accuracy can only be addressed once radiotherapy contouring, planning and treatment are precise.

10.2 The impact of contouring variation on modelled radiotherapy outcome

Chapter 4 investigated the relationship between geometric contouring variation and outcome surrogates in the form of tumour control probability (TCP),

equivalent uniform dose (EUD) and mean lung dose, for a series of non-small cell lung cancer (NSCLC) patients. With a view to recommend relevant geometric parameters for the assessment of contouring variation that relate to modelled clinical outcome. Seven patients were included in the study and contouring was performed by three observers on CT and PET imaging datasets. Geometric variation was assessed and compared to resulting variation in TCP, EUD and mean lung dose.

Statistically significant relationships were observed for most geometric parameters with the strongest correlation pertaining to medial-lateral dimension of the target volume, centre of mass, and concordance index. In Chapter 3 it was found that medial-lateral dimension was employed in only 1/10 of the lung studies reviewed, while centre of mass and concordance index were used in 4/10 and 2/10, and volume was the metric of choice in 8/10 studies. This highlights that the choice of metric for assessment of contouring variation is not driven by relevance to clinical outcome but likely by the tools available to investigators. The results of this work should inform the choice of metric used and ensure future contouring studies are more consistent and comparable.

Most investigations of contouring variation require some sort of reference volume to compare to. This is often called the 'gold standard' or 'reference' volume. As the true extent of the tumour is not known these gold standard

volumes can take a number of forms; mathematical averages, probabilistic (e.g. STAPLE), consensus, and most experienced observer, have all been used in the literature. This is an area that warrants further investigation as the choice of gold standard can have a strong influence on the analysis of results. Understanding the impact of the choice of gold standard on the typical contouring variation metrics is of interest. This, in conjunction with the impact on dosimetry, may also serve to provide some reference values for typical contouring variation metrics for future investigators.

10.3 The influence of planner experience on IMRT plan quality

In Chapter 5, the impact of varying degrees of radiotherapy planner experience on plan quality was presented. Six planners generated IMRT treatment plans for a $T_2N_3M_0$ tonsilar carcinoma case according to department protocol. Plans were compared visually by an experienced radiation oncologist and also using a number of dose-volume constraints and conformity indices. Delivery efficiency and dose accuracy were also compared. Only 3/6 of the planners were able to meet the dose objectives for the PTV. All planners could meet the constraints for the brainstem, spinal cord, mandible and oral cavity, with the exception of one planner whom failed to meet the mandible constraint. No planners achieved the required dose volume constraints for the right parotid or larynx but these structures overlapped with the PTV. Interestingly, the radiation oncologist, on

slice by slice review, deemed all plans of a clinically acceptable quality. Treatment delivery time and monitor units ranged from 15-25 minutes and just under 800 to over 1200 MU with delivery time increasing with decreasing planner experience. The planner with the least experience had the poorest plan, as indicated by meeting the fewest PTV constraints.

10.4 Investigation of organ stability, dosimetry, and margins in the presence of organ stabilising devices

An investigation into the use of ERBs in the post prostatectomy setting is presented in Chapters 6 and 7. It has been known for some time that the prostate bed can experience inter- and intra-fraction motion due its proximity to the bladder and bowel, organs that are constantly filling and emptying[7]. This study was completed in two parts. The first of which, Chapter 6, addressed the question of whether the addition of an ERB *in situ* improved dosimetric inter-fraction reproducibility with the same treatment margins. The second, Chapter 7, investigated whether the organ motion component of the PTV margin could be reduced when an ERB is used. For both of these studies 20 patients were included in the investigation, 10 retrospective patients treated with standard practice and 10 prospective patients treated with an ERB *in situ*. The treatment consisted of IMRT with a prescribed dose of 70 Gy to the inferior CTV and 64.4 Gy to the superior CTV.

The ERB significantly improved inter-fraction reproducibility for the rectum and the CTV. Concordance indices for non-ERB and ERB of 0.50 ± 0.12/0.71 ± 0.07 for the rectum and 0.72 ± 0.15/0.73 ± 0.11 for the CTV. However, the improved geometric stability with the ERB did not translate into a statistically significant benefit in inter-fraction dosimetric stability based on a change in equivalent uniform dose (ΔEUD). A reduced dosimetric stability for the bladder and supCTV was found and is likely due to bladder filling and slight differences in ERB insertion depth between fractions. One of the positive aspects of using the ERB was that it reduced the impact of bladder filling on CTV stability. The results of Chapter 6 agree with previous investigations in that a differential PTV margin is warranted given the relative difference in stability between the superior and inferior CTV.

10.5 Benchmarking and assessing the impact of contouring variation in radiotherapy clinical trials

Radiotherapy clinical trial quality assurance has become a focus in recent times, due in part to some sobering secondary analyses[8, 9] of large, well-funded and run cooperative group run trials. Chapters 8 and 9 illustrate the implementation of a dummy run to assess any possible protocol non-compliance and a modelling study incorporating the results of the dummy run into the trial design. While it is well understood that uncertainty due to contouring variation is larger than that of setup error for some tumour sites[1], hence the routine use

of dummy run exercises, there has been no effort to account for this in the design of clinical trials.

The range of variation in volume was 228.5-497.8 cm^3 for CTV contouring. Uncertainty was largest in the z (superior / inferior) direction where investigators did not adhere to protocol contouring guidelines. For the benchmarking study the dose from the investigator submitted plans were analysed against a set of gold standard contours. Dosimetric variation in Chapter 8 was not substantial, although it should be noted that the four field (4FLD) box planning technique used is relatively insensitive to contouring variation within the borders of the "box" dose distribution. In Chapter 9, to remove variation due to planning and focus on contouring all planning was performed centrally using a class solution planning technique. The IMRT planning technique demonstrated the largest variation in TCP with a range of 0.65-0.68, due to the conformity of the dose distribution with the shape of the contour. This TCP variation did not have a large impact on the required sample size, only requiring an extra 19 patients for the worst case. However this work provides a framework to incorporate uncertainties quantified as part of routine dummy run exercises to ensure robust results are obtained from RCTs.

10.6 Future work

From the topics presented in this thesis there are a number of issues that justify further investigation, including:

I. The introduction of guidelines and minimum reporting requirements for the conduct of contouring studies in radiation oncology
II. Applying the technique presented in Chapter 4 to other treatment sites to determine the appropriate metrics of contour variation to report.
III. The intra-fraction stability of the post prostatectomy target volume with ERB *in situ*
IV. The development of automated methods of performing radiotherapy clinical trial quality assurance
V. The inclusion of planning and delivery uncertainties in prospective radiotherapy clinical trials

As outlined in Chapter 3 there is no consistent evidence based method of contour comparison within the literature. Contouring uncertainty has become increasingly important as the accuracy of dose calculation and radiation delivery has improved. Without a consistent method of reporting variation in contouring for tumour sites, it will continue to be problematic to combine the results of these studies in meta-analyses. There has been a push from some publishers to include a minimum set of information when reporting planning studies so that other investigators can repeat experiments and compare hypotheses, the same should apply to contouring studies[10]. A review or

recommendation publication detailing appropriate methodology and reporting for contouring studies would facilitate the comparison of results from different studies and allow for the appraisal of the quality of individual studies in a uniform way. Applying the work presented in Chapter 4 to other treatment sites and techniques will inform the appropriate contouring variation metrics to use.

Further investigation is warranted by extending the studies presented in Chapters 6 and 7, including a larger number of patients and analysing intra-fraction motion. The pilot study presented in Chapter 6 could be used to inform the sample size calculation for a larger prospective trial. As part of that trial the impact of the ERB on intra-fraction motion should also be considered. Although SBRT has not been used in the post-prostatectomy setting to date, a thorough understanding of intra-fraction motion is needed perform SBRT safely.

Radiotherapy clinical trials are expensive, time consuming and complicated to run[6]. It is therefore important that the methods used when conducting a trial are robust and the protocol strictly adhered to. The technique presented in Chapter 9 could be used in prospective clinical trials to ensure there is adequate statistical power in the design of the trial to account for treatment uncertainties. Expert review is the current method of individual case review in radiotherapy trials. This is very expensive and can be a limiting factor to recruitment in some instances. Moreover, for trials involving adaptive radiotherapy where a patient

may have many treatment plans that require review, manual expert review will not be viable.

10.7 Summary

This work assessed the impact of organ motion, planner experience and contouring variation on plan quality. Furthermore, the same techniques were applied to a clinical trial dummy run, the results of which were used to assess the impact of contouring variation on the statistical power of an RCT. A method to ascertain the most relevant metrics of use when assessing contouring variation was presented. The choice of such metrics will be site and planning technique specific. Planner experience was demonstrated to have an impact on the quality of radiotherapy planning for head and neck IMRT. A lack of planning experience also resulted in IMRT plans that were less efficient to deliver. It was shown that the ERBs reduce the amount of inter-fraction motion for post prostatectomy radiotherapy. However, a larger prospective trial is required to confirm these results and the dosimetric impact of the ERB. A technique for the incorporation of planning uncertainty into clinical trial sample size calculations was proposed. This process, combined with rigorous QA, provides a simple means to use data from dummy run exercises to ensure the robustness of the trial sample size calculations.

This dissertation represents a series of studies into the impact on radiotherapy quality and clinical trial power of organ motion, planner experience and contouring variation. The key priorities for continuing this work are:

- Developing standardised practices when performing and reporting contouring studies in radiation oncology
- Using the results of benchmarking procedures to ensure the robustness of sample size calculations in RCTs

It is hoped the work reported in this thesis will contribute to the way in which clinician defined contour uncertainties, organ motion uncertainties and there impact on dose targeting and hence tumour control are assessed and reported in the future.

10.8 References

1. Weiss E, Hess CF. The impact of gross tumor volume (GTV) and clinical target volume (CTV) definition on the total accuracy in radiotherapy theoretical aspects and practical experiences. Strahlentherapie und Onkologie. 2003;179(1):21-30.

2. Fotina I, Lütgendorf-Caucig C, Stock M, Pötter R, Georg D. Critical discussion of evaluation parameters for inter-observer variability in target definition for radiation therapy. Strahlentherapie und Onkologie. 2012;188(2):160-7.

3. Njeh C. Tumor delineation: The weakest link in the search for accuracy in radiotherapy. Journal of medical physics/Association of Medical Physicists of India. 2008;33(4):136.

4. Nelms BE, Robinson G, Markham J, Velasco K, Boyd S, Narayan S, et al. Variation in external beam treatment plan quality: an inter-institutional study of planners and planning systems. Practical radiation oncology. 2012;2(4):296-305.

5. Keall PJ, Mageras GS, Balter JM, Emery RS, Forster KM, Jiang SB, et al. The management of respiratory motion in radiation oncology report of AAPM Task Group 76a). Medical physics. 2006;33(10):3874-900.

6. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMC medicine. 2010;8(1):18.

7. Ost P, De Meerleer G, De Gersem W, Impens A, De Neve W. Analysis of prostate bed motion using daily cone-beam computed tomography during postprostatectomy radiotherapy. International Journal of Radiation Oncology* Biology* Physics. 2011;79(1):188-94.

8. Peters LJ, O'Sullivan B, Giralt J, Fitzgerald TJ, Trotti A, Bernier J, et al. Critical impact of radiotherapy protocol compliance and quality in the treatment of advanced head and neck cancer: results from TROG 02.02. Journal of Clinical Oncology. 2010;28(18):2996-3001.

9. Ohri N, Shen X, Dicker AP, Doyle LA, Harrison AS, Showalter TN. Radiotherapy protocol deviations and clinical outcomes: a meta-analysis of cooperative group clinical trials. Journal of the National Cancer Institute. 2013:djt001.

10. Yartsev S, Muren LP, Thwaites DI. Treatment planning studies in radiotherapy. Radiotherapy and Oncology. 2013;3(109):342-3.

Appendix A: A phantom assessment of achievable contouring concordance across multiple treatment planning systems

Radiotherapy and Oncology xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

A phantom assessment of achievable contouring concordance across multiple treatment planning systems

Elise M. Pogson [a,b,*], Jarrad Begg [b], Michael G. Jameson [a,b], Claire Dempsey [c,d], Drew Latty [e], Vikneswary Batumalai [b,i], Andrew Lim [f], Kankean Kandasamy [g], Peter E. Metcalfe [a,b], Lois C. Holloway [a,b,h,i]

[a] Centre for Medical Radiation Physics, University of Wollongong; [b] Liverpool and Macarthur Cancer Therapy Centres and the Ingham Institute, Liverpool Hospital; [c] Department of Radiation Oncology, Calvary Mater Newcastle Hospital; [d] School of Health Sciences, University of Newcastle; [e] The Crown Princess Mary Cancer Centre, Westmead; [f] Peter MacCallum Cancer Centre, Melbourne; [g] Prince of Wales Hospital, Randwick; [h] Institute of Medical Physics, University of Sydney; and [i] South Western Clinical School, University of New South Wales, Australia

ARTICLE INFO

Article history:
Received 17 December 2014
Received in revised form 1 September 2015
Accepted 18 September 2015
Available online xxxx

Keywords:
Radiotherapy
Contouring
Delineation
Inter-observer
DICE
Hausdorff

ABSTRACT

In this paper, the highest level of inter- and intra-observer conformity achievable with different treatment planning systems (TPSs), contouring tools, shapes, and sites have been established for metrics including the Dice similarity coefficient (DICE) and Hausdorff Distance. High conformity values, e.g. $DICE_{Breast\_Shape} = 0.99 \pm 0.01$, were achieved. Decreasing image resolution decreased contouring conformity.

Delineation of radiotherapy structures has direct clinical consequences. Contouring of nodal CTV sub-volumes in particular, is critical [1]. Even moderate geometrical differences in small neck Planning Target Volumes (PTVs) can impact on the target dose (up to 11 Gy reductions in D99 for DICE above 0.8) [2]. For non-small lung cancer variation a concordance index (CI) has been demonstrated to result in variation in Tumour Control Probability (TCP) [3], highlighting the correlation between contour variation and TCP. However, there are no reported contour variation metric baseline values considering uncertainties in the process such as different treatment planning systems (TPSs), importing and exporting processes, contour shapes, volumes and image resolution. Knowledge of these baseline values is important for clinical trials which commonly occur across multiple centres and TPSs. Current literature does not give clear guidelines for reporting contouring variability in inter-observer studies [4] with variation in methodology and metrics only enabling comparison between inter-observer studies in a limited fashion [5]. As such, calculating multiple metrics including a combination of descriptive statistics, overlap measures and statistical measures of agreement is recommended for multiple observer studies [6].

The number of studies reporting on auto-segmentation [7,8], and the inter- [9,10] and intra- [11] observer conformity of volumes is growing. Inadequate definition of the Gross Tumour Volume (GTV) or Clinical Target Volume (CTV) leads to systematic uncertainty which may result in geometric miss of the tumour throughout the course of patient radiation therapy [5]. As such there has been an increasing trend to assess, and reduce, the variability of these target volumes. This study determined the highest concordance metrics achievable, and how these metrics (details given in Supplementary Table 1) may vary in a best case phantom scenario considering: multiple sites, variation between TPSs, shapes, volume, tools utilized and adherence to auto-threshold settings within the protocol.

Methods

A Quasar Body phantom (Modus Medical Devices Incorporated, Ontario Canada) was used to provide an initial CT dataset. The Quasar phantom was scanned on a Brilliance Big Bore CT (Phillips Healthcare, The Netherlands) using a helical abdomen scanning sequence: 1 mm slice spacing, 2 mm slice thickness, standard

* Corresponding author at: School of Physics, Engineering and Information Sciences, University of Wollongong, Wollongong, NSW 2522, Australia.
E-mail address: elisep@uow.edu.au (E.M. Pogson).

http://dx.doi.org/10.1016/j.radonc.2015.09.022
0167-8140/

resolution (512 × 512) and field of view of 350 mm. This phantom had three inserts containing structures providing a range of surface contours and edges. In this study the 20-degree air wedge contained in the first insert (referred to as the triangular prism) and the entire empty third insert (an 8 cm diameter cylinder with semi-conic top) were used for contouring.

The Quasar phantom CT dataset was imported into MATLAB R2012a (Mathworks Incorporated, Natick USA). Uniform rectangular prisms and a patient breast volume (203 cm^3) were inserted into the CT dataset using 'Computational Environment for Radiotherapy Research' (CERR) [12,13] and MATLAB. High intensities were utilized to obtain optimal image contrast. The Quasar phantom with inserted shapes is displayed, with inter-observer contours, in Supplementary Fig. 1.

A contouring protocol set image window levels to Window/Level = 400/800 HU and described allowable techniques/tools. All eight rectangular prisms were auto-contoured using auto-threshold at recommended threshold values or other automated tools (e.g. Oncentra's magic-wand tool). Rectangular prisms 1, 4 and 8 (Supplementary Fig. 1) were manually contoured. Bounding boxes in auto-contouring and zoom functions were allowed. The breast contour was manually delineated; allowing interpolation between slices and/or copy to next slice. The triangular prism and cylinder were both delineated using automated tools (such as auto-threshold) and manually. All eight observers were blind to others contours. The TPSs used for contouring were; Eclipse Planning System 11.0.64 (Varian Medical Systems, Palo Alto Canada): two sites, Oncentra (Elekta, Stockholm Sweden): two sites, Pinnacle^3 9.0 (Philips, Netherlands): two sites, and FocalSim 4.80.01 (Elekta, Stockholm Sweden): two sites. These contours were then exported and collated in CERR.

The same original 512 × 512 data-set was contoured five times by four observers, with a minimal 24 h time lapse between contouring. Pairwise analysis of the Jaccard Index (JI) also known as conformity index or concordance index (CI) [6,14] (CI_{pairs} the average of all possible pairs of the JI which equates to CI_{gen} when mutual variability between all observers is the same [15]), Volume Overlap Index (VOI) and Hausdorff Distances (HDs) were calculated for each observer and averaged. This was performed for all manually contoured structures.

Different studies have different image resolutions. As such the Quasar phantom was resampled and contoured by five different observers, to show the expected inter-observer effects for differing sample/dataset pixel size and slice thickness. The resampling was performed in MATLAB with the overall volume maintained. Slice thickness was also set to the spacing of 2 mm, 4 mm and 8 mm keeping the resolution at 512 × 512 px (1.463 px/mm) and saved as DICOM. The resampled DICOM data were of the following resolutions; 512 × 512 px^2 (1.463 px/mm – a typical high resolution CT), 350 × 350 px^2 (1.000 px/mm), 245 × 245 px^2 (0.700 px/mm), 175 × 175 px^2 (0.500 px/mm), 88 × 88 px^2 (0.250 px/mm), and 44 × 44 px^2 (0.125 px/mm).

To allow comparison between observers, simultaneous truth and performance level estimation (STAPLE) volumes were generated as consensus gold standard reference volumes in CERR, using a 90% confidence interval with observers weighted equally. CERR was utilized to calculate the generalized kappa statistic as well as the DICE, and JI in three dimensions for all observers comparing to the gold standard STAPLE volume (Supplementary Table 1). The maximal HD, average Hausdorff Distance, CI_{pairs} and VOI was calculated in a pairwise analysis over all volumes in MilxView (Australian e-Health Research Centre (AEHRC), Australia) [16,17] (Supplementary Table 2).

The JI [18–20], DICE [4], Hausdorff distance [21] and Kappa (κ) statistic [22,23] outlined in Supplementary Table 1, are metrics commonly used to establish inter-observer variation [6]. JI and DICE values from CERR were verified in 3D Slicer [24–26] and MILXview and were consistent to within 2 significant figures.

Results

Eight auto-contoured, inter-observer rectangular prism contours from different TPSs were all within two pixels of the true volume on every slice, for every point within the contour (Fig. 1(a)). The maximum HD of these contours compared to the STAPLE ranged from 1 pixel width/height (0.68 mm) or 2 pixels added in quadrature (0.97 mm), with a maximum of 3 pixels (2.04 mm) for the auto-contoured rectangular prisms (Fig. 1(c)). As the STAPLE for square 5 is different to the true volume there are larger HDs and discrepancies for this volume. A pairwise HD measure, rather than to the STAPLE, is less sensitive to such errors and is used in all following analysis. Fig. 1(b) displays each inter-observer's DICE compared to the STAPLE. Inter- and intra- observer contour variation as measured by maximum HD relative to the STAPLE volumes was less than 7 mm for all volumes at normal resolution (1.463 px/mm). Kappa statistics comparing multiple shapes from the Quasar phantom show near perfect agreement for most shapes despite asymmetry from the breast contour (Supplementary Fig. 2).

Auto-contoured rectangular prisms were less conformal (kappa in the range of 0.61–0.80) than manually delineated shapes (kappa in the range of 0.81–1), (Supplementary Fig. 2), with other shapes having no difference. The contouring tool used did not show any observable effect in contour conformity. Average manual and auto-threshold DICE were in agreement (within the 95%

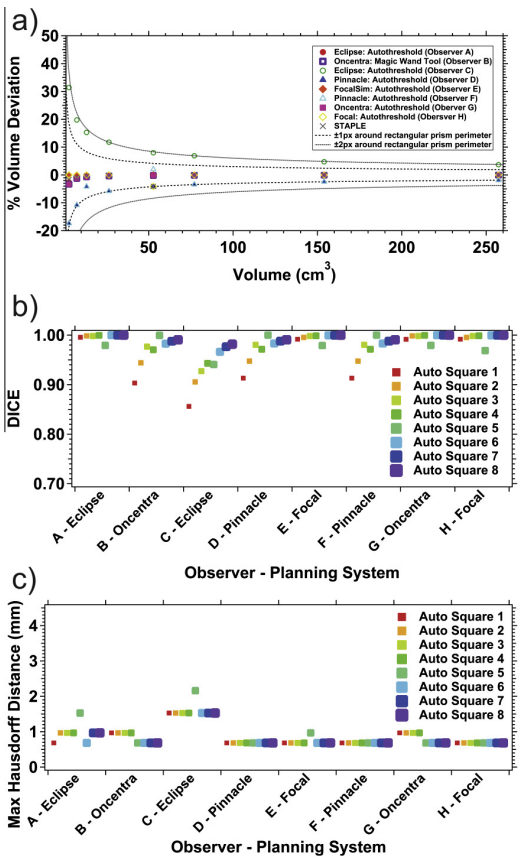


Fig. 1. Auto-contoured squares; a) Percentage deviation of volume from the true volume. Majority of contours are within 1 px and the rest within 2 px, b) DICE c) maximum HD from the STAPLE volume.

Please cite this article in press as: Pogson EM et al. A phantom assessment of achievable contouring concordance across multiple treatment planning systems. Radiother Oncol (2015), http://dx.doi.org/10.1016/j.radonc.2015.09.022

confidence limit) for all shapes. The JI, average DICE and kappa for the manually delineated shapes are summarized in Supplementary Table 2.

Inter-observer generalized kappa statistics for differing shapes is shown in Fig. 2(a). Decreasing image resolution reduces concordance, especially for smaller structure volumes e.g. triangular prism (47 cm^3). This is evident in the average DICE compared to the STAPLE volume in each image (Fig. 2(b)) and the average maximal HDs (Fig. 2(c)). The breast contour and some rectangular prisms with an image resolution of 0.250 px/mm and 0.125 px/mm were excluded as the outline was not visible at recommended window levels due to resampling.

As resolution decreases below 0.250 px/mm, the relative inter-observer DICE also decreases for manual contours, despite Fig. 2(b) showing good concordance compared to the STAPLE generated on each individual resolution dataset. Supplementary Fig. 3, displays the relative DICE of contours with lowering resolution compared to the highest resolution image (1.49).

Varying the slice thickness from 1 mm to 2 mm, 4 mm and 8 mm had no significant effect on inter-observer conformity.

Discussion

This investigation has demonstrated that despite the use of multiple treatment planning systems, it is possible to achieve close to perfect conformity between observers with a high contrast dataset. Conformity is reduced with reduction in image resolution and volume of the structure considered.

The relative deviation, as shown in Fig. 1, increases for smaller volumes i.e. up to ~30% for a small volume of 2.7 cm^3. Additional differences may occur during import and export through multiple TPSs, as the same volume exported from multiple TPSs has been shown to vary from 2% to −4% for small volumes (less than 250 cc) [27]. The HDs, as shown in Fig. 2(c), are increasing due to lengthening pixel sizes. This was similar to results shown in another study [28].

Inter-observer variation is shown to increase with lower resolution. Intra-observer variation is either in agreement or smaller than inter-observer variation similarly to previously reported clinical findings [5]. Disagreement between the same TPS is evident for contours generated using auto-threshold tools in the same TPS by different observers, (Fig. 1(c)). Hounsfield Units (HUs) used for Auto-thresholding were requested, and showed significantly different HUs had been used. This ambiguity is likely due to conversion between TPSs. We recommend that the conversion between multiple TPSs for inter-observer studies be performed and sent out with the study dataset in future studies. The highest achievable values are dependent upon image resolution, contour volume, number of observers, image contrast, window level and adherence to the protocol.

Previously reported values in breast radiotherapy CTV inter-observer studies include a JI of; 0.81 for radiation oncologist breast contouring [9], 0.84 for radiation therapist breast contouring [9], 0.87 for glandular breast volumes [14], 0.56 for partial breast volumes [14] and 0.82 for glioblastoma GTV's (Gross Tumour Volumes) [29]. An inter-observer breast contour generalized kappa of 0.97 ($p < 0.05$), maximal HD of 3.42 mm, average JI of 0.98 ± 0.01 and average DICE of 0.99 ± 0.01 was found in this study. This demonstrates the highest achievable values for future expert clinician contours compared to a STAPLE volume, for an acceptable number of observers (five or more, with a recommendation to have as large a number of expert observers as possible for small volumes [28]) and a standard CT image resolution (512 × 512). The gold standard STAPLE volume has been generated by the contours assessed here, whilst this has minimal effect, in an ideal study the aim would be to have a separate group of contours to generate a gold standard STAPLE and compare to this. To avoid this metrics such as CI_{pairs} or VOI may be utilized instead. Complexity of shape showed no observable effect in conformity, as the complicated breast contour achieved a higher average DICE, average JI and Kappa than the cylinder and rectangular prism, of similar volumes. However an assessment of more complicated irregular shapes than rounded breast contours still needs to be undertaken.

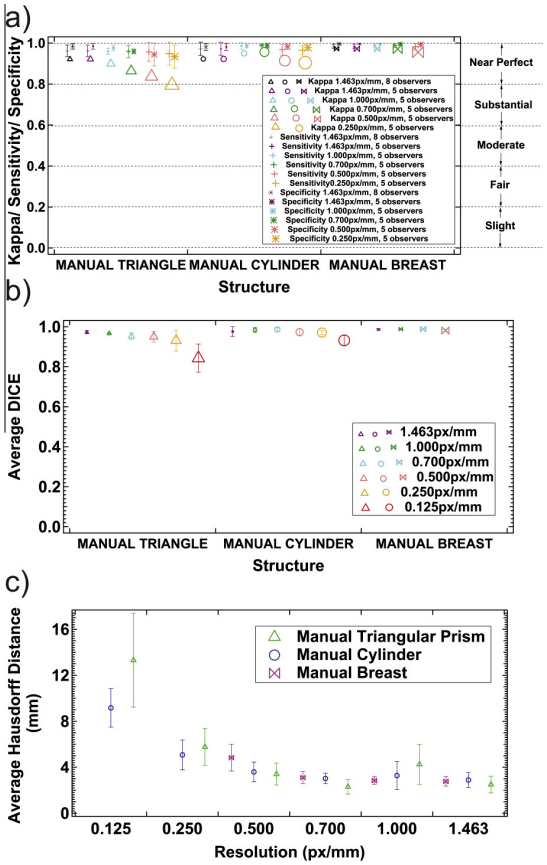


Fig. 2. Manually delineated Inter-observer a) STAPLE parameters with differing image resolution; Kappa, Specificity, Sensitivity and Volume, b) 5 observer average DICE and c) 5 observer average Hausdorff Distances. Error bars represent 1SD. The STAPLE in the resampled images have lower specificity and sensitivity with lowering resolution. The 95% confidence intervals also increase, for small volumes, with worsening resolution (as the amount of data is reduced).

In summary, multi-observer results from multiple TPSs, differing TPS tools, image resolution, image slice thickness, contour shapes and volumes has been established for average DICE, average JI, CI_{pairs}, VOI, kappa, average HD and maximum HD. Values obtained in this phantom study suggest that multiple sites and systems do not have significant impact on concordance metrics for these particular volumes. Values presented here may provide an upper bound as to what is achievable in future studies. Alternatively if images are of significantly different image resolution, extremely small volumes (such as a head and neck study), of more irregular shape, or with less observers, future studies might consider including another object/dataset to determine their highest achievable kappa, average DICE or average JI under these circumstances. This could be undertaken on a study by study basis.

Disclaimer

The views expressed in this article are my own and not an official position of the institution or funding support.

Please cite this article in press as: Pogson EM et al. A phantom assessment of achievable contouring concordance across multiple treatment planning systems. Radiother Oncol (2015), http://dx.doi.org/10.1016/j.radonc.2015.09.022

Sources of support

Cancer Australia and National Breast Cancer Foundation grant project number 1033237.

Conflict of interest declaration

The authors report a grant from Cancer Australia and National Breast Cancer Foundation, during the conduct of the study.

Acknowledgements

This work was supported by a Cancer Australia and National Breast Cancer Foundation grant project number 1033237.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.radonc.2015.09.022.

References

[1] Valentini V, Boldrini L, Damiani A, Muren LP. Recommendations on how to establish evidence from auto-segmentation software in radiotherapy. Radiother Oncol 2014;3:317–20.
[2] Voet PWJ, Dirkx MLP, Teguh DN, et al. Does atlas-based autosegmentation of neck levels require subsequent manual contour editing to avoid risk of severe target underdosage? A dosimetric analysis. Radiother Oncol 2011;98:373–7.
[3] Jameson MG, Kumar S, Vinod SK, Metcalfe PE, Holloway LC. Correlation of contouring variation with modeled outcome for conformal non-small cell lung cancer radiotherapy. Radiother Oncol 2014;112:332–6.
[4] Yang J, Beadle BM, Garden AS, et al. Auto-segmentation of low-risk clinical target volume for head and neck radiation therapy. Pract Radiat Oncol 2014;4. e31-7.
[5] Weiss E, Hess CF. The impact of gross tumor volume (GTV) and clinical target volume (CTV) definition on the total accuracy in radiotherapy. Strahlenther und Onkol 2003;179:21–30.
[6] Fotina I, Lütgendorf-Caucig C, Stock M, Pötter R, Georg D. Critical discussion of evaluation parameters for inter-observer variability in target definition for radiation therapy. Strahlenther Onkol 2012;188:160–7.
[7] Zhou W, Xie Y. Interactive contour delineation and refinement in treatment planning of image-guided radiation therapy. JACMP 2014;15:4499–522.
[8] Simmat I, Georg P, Georg D, et al. Assessment of accuracy and efficiency of atlas-based autosegmentation for prostate radiotherapy in a variety of clinical conditions. Strahlenther und Onkol 2012;188:807–15.
[9] Holloway LC, Jameson MG, Batumalai V, et al. Estimating a delineation uncertainty margin to account for inter-observer variability in breast cancer radiotherapy. Int J Radiat Oncol Biol Phys 2010;78:S741.
[10] Yamazaki H, Shiomi H, Tsubokura T, et al. Quantitative assessment of inter-observer variability in target volume delineation on stereotactic radiotherapy treatment for pituitary adenoma and meningioma near optic tract. Radiat Oncol 2011;6.
[11] Lütgendorf-Caucig C, Fotina I, Stock M, et al. Feasibility of CBCT-based target and normal structure delineation in prostate cancer radiotherapy: multi-observer and image multi-modality study. Radiother Oncol 2011;98:154–61.
[12] Apte A, Khullar D, Alaly J and Deasy JO. The Computational Environment for Radiotherapy Research (CERR). 2010; http://www.cerr.info/about.php.
[13] Deasy JO, Blanco AI, Clark VH. CERR: a computational environment for radiotherapy research. Med Phys 2003;30:979–85.
[14] Struikmans H, Wárlám-Rodenhuis C, Stam T, et al. Interobserver variability of clinical target volume delineation of glandular breast tissue and of boost volume in tangential breast irradiation. Radiother Oncol 2005;76:293–9.
[15] Kouwenhoven E, Giezen M, Struikmans H. Measuring the similarity of target volume delineations independent of the number of observers. Phys Med Biol 2009;54:2863.
[16] Dowling JA. Opportunities for image analysis in radiation oncology. Australas Phys Eng Sci Med 2014;37:275–7.
[17] Dowling JA, Fripp J, Chandra S. Fast automatic multi-atlas segmentation of the prostate from 3D MR images. In: Prostate Cancer Imaging. Image Analysis and Image-Guided Interventions. Berlin: Springer; 2011. p. 10–21.
[18] Petersen RP, Truong PT, Kader HA, et al. Target volume delineation for partial breast radiotherapy planning: clinical characteristics associated with low interobserver concordance. Int J Radiat Oncol Biol Phys 2007;69:41–8.
[19] Fein DA, McGee KP, Schultheiss TE, Fowble BL, Hanks GE. Intra- and interfractional reproducibility of tangential breast fields: a prospective on-line portal imaging study. Int J Radiat Oncol Biol Phys 1996;34:733–40.
[20] Feuvret L, Noël G, Mazeron J-J, Bey P. Conformity index: a review. Int J Radiat Oncol Biol Phys 2006;64:333–42.
[21] Jameson MG, Holloway LC, Vial PJ, Vinod SK, Metcalfe PE. A review of methods of analysis in contouring studies for radiation oncology. J Med Imag Radiat Oncol 2010;54:401–10.
[22] Ebert M, McDermott L, Haworth A, van der Wath E, Hooton B. Tools to analyse and display variations in anatomical delineation. Australas Phys Eng Sci Med 2012;35:159–64.
[23] Lim K, Small Jr W, Portelance L, et al. Consensus guidelines for delineation of clinical target volume for intensity-modulated pelvic radiotherapy for the definitive treatment of cervix cancer. Int J Radiat Oncol Biol Phys 2011;79:348–55.
[24] Fedorov A, Sonka M, Buatti J, et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network. Magn Reson Imaging 2012;30:1323–41.
[25] Fedorov A, Sonka M, Buatti J, et al. 3D Slicer. 2014; http://www.slicer.org/.
[26] Pinter C, Lasso A, Wang A, Jaffray D, Fichtinger G. SlicerRT: radiation therapy research toolkit for 3D Slicer. Med Phys 2012;39:6332–8.
[27] Ebert MA, Haworth A, Kearvell R, et al. Comparison of DVH data from multiple radiotherapy treatment planning systems. Phys Med Biol 2010;55:N337.
[28] Commowick O, Warfield SK. Estimation of inferential uncertainty in assessing expert segmentation performance from STAPLE. IEEE Trans Med Imaging 2010;29:771–80.
[29] Ryuji M, Toshinori H, Ryo T, Hideo N, Yasuyuki Y. Double reading for gross tumor volume assessment in radiotherapy planning. J Solid Tumors 2012;2:38.

Please cite this article in press as: Pogson EM et al. A phantom assessment of achievable contouring concordance across multiple treatment planning systems. Radiother Oncol (2015), http://dx.doi.org/10.1016/j.radonc.2015.09.022

Appendix B: Contouring Variability and its Effect on Radiobiology Parameters for Head and Neck Cancer

Yvette Griffiths, MSc, Sydney University, School of Physics
Supervisors: Dr Lois Holloway, Michael Jameson

Abstract

Inter and Intraobserver variation in delineation (or contouring) of tumour and normal structures is a widely recognised issue in radiotherapy. Many studies have quantified this variation and investigated ways to reduce it. If a contour is inaccurately delineated, the tumour may be underdosed or normal tissues overdosed. Currently there are studies that have shown a clinical impact from inter/intra observer variation through the use of radiobiological models for both tumour and normal tissues.

The aim of this project is to investigate a potential correlation between geometrical variations in contouring and radiobiologically modelled clinical outcome.

Multiple contours were generated mathematically and by observers on head and neck cancer CT datasets. An IMRT dose distribution was generated based on each contour. Then the contours were analysed for geometric variation and modeled clinical outcome. The contouring variation and modelled clinical outcome was correlated.

The results showed a 13.86% length variation in the x direction, as a percentage of the mean. The change in predicted clinical outcome was 56.68% (of the mean). A trend in correlation was seen between the length of the x, y and z dimensions and modelled clinical outcome. A trend was also seen between volume change and predicted clinical outcome.

The correlation trends found in this study could potentially be used for predicting the effect that contouring change clinical outcome. To achieve more conclusive results, a larger future study would be required, in order to develop guidelines to predict the effect of inaccurate structure delineation.