2019

Reducing Tumour Volume Uncertainty for the Benefit of Radiation Therapy Cancer Patients

Lauren Rebecca Bell

UNIVERSITY
OF WOLLONGONG
AUSTRALIA

Reducing Tumour Volume Uncertainty for the Benefit of Radiation Therapy Cancer Patients

Lauren Rebecca Bell

This thesis is presented as part of the requirements for the conferral of the degree:

Doctor of Philosophy

The University of Wollongong
School of Physics

October, 2019

This research has been conducted with the support of an Australian Government Research Training Program Scholarship.

Declaration

I, *Lauren Rebecca Bell*, declare that this thesis is submitted in partial fulfilment of the requirements for the conferral of the degree *Doctor of Philosophy*, from the University of Wollongong, is wholly my own work unless otherwise referenced or acknowledged. This document has not been submitted for qualifications at any other academic institution.

Lauren Rebecca Bell

October 29, 2019

Abstract

The efficacy of radiotherapy is dependent on its precision and accuracy. Increasingly conformal, modulated radiation fields can be reproducibly delivered to small, complex volumes within the human body. However, treatment is not without uncertainty. This thesis focuses on limitations in radiotherapy accuracy due to uncertainty in delineation of the volume requiring treatment.

Delineation of the target volume occurs on computed tomography (CT) imaging (with or without aid from other imaging) and is manually drawn by the radiation oncologist. The inability to visualise individual cancer cells in situ with current imaging means that an oncologist's contour is highly subjective. This potential inconsistency in contours and hence, treatment plans for the same patient can obscure results of trial data, as well as result in suboptimal plans for the individual patient.

This thesis investigated the importance of inter-observer variation in radiotherapy. The variability in high risk clinical target volume (CTV_{HR}) delineation for high dose rate (HDR) cervical cancer brachytherapy was quantified and the dosimetric consequences of this variation was assessed. A wide range in variability when considering each individual contour was observed, with an average mean absolute surface distance (MASD) of (7.13 $\pm$ 6.37) mm and an average dice similarity coefficient (DSC) of (0.63 $\pm$ 0.15). The maximum dimensions of the CTV_{HR} contour demonstrated 0.75 cm - 1.82 cm variation and the mean volume varied by 28.5 cm^3. The average dosimetric uncertainty across a treatment course that arises as a result of this variability was shown to be 2.9 Gy - 3.4 Gy

$EQD2_{10}$ for an Australian dataset and 1.9 Gy - 2.8 Gy $EQD2_{10}$ for a gold standard European dataset. The European data demonstrated a correlation between lateral dimensions and volume with dose to 98% of the volume (D98). This was validated with the independent Australian dataset.

In an attempt to minimise inter-observer variation, an atlas-based auto-segmentation method was proposed and investigated for whole breast external beam radiotherapy. The utilisation of this approach in the radiotherapy treatment planning process would minimise inter-observer variation arising from manual contouring. The auto-segmented contours were accurate, demonstrating good similarity when compared to the gold standard contour (DSC $>$ 0.7 and MASD $<$ 1 cm).

Finally, it is acknowledged that it is unlikely that inter-observer variation will be entirely eliminated as long as subjective interpretation of images remains a requirement in radiotherapy. Hence, a delineation uncertainty margin was proposed and investigated. The coordinate system most appropriate for this margin (as assessed on whole breast radiotherapy volumes) was spherical coordinates, however cardinal expansions were also found to be appropriate for small breast volumes. Delineation uncertainty margin magnitudes ranged from 3.2 mm - 14.5 mm for spherically defined margins, 3.0 mm - 17.6 mm for polar margins and 1.6 mm - 15.8 mm for cartesian defined margins. Acceptable clinical target volume (CTV) observer coverage was determined with $>$ 95% of observer contours included in the delineation margin for all coordinate systems. The volume expanded by the margin was assessed to ensure that unnecessary healthy tissue was not included and malignant tissue was not missed. The delineation uncertainty margin produced volumes within accepted tolerances for the majority of cases.

In order to improve patient outcomes, sources of uncertainty must be minimised. Inter-observer variation is one of the largest sources of uncertainty in the radiotherapy process. This thesis presents a series of investigations that attempt to understand the impact this

uncertainty has on treatment, minimise the human intervention in the process such that consistency might be improved and manage its presence for the benefit of improved cancer patient treatment outcomes.

Acknowledgments

This research was supported by an Australian Government Research Training Program Scholarship, the South Western Sydney Local Health District Radiation Oncology Scholarship and the 2016 Endeavour Research Fellowship.

First and foremost I would like to express my deepest appreciation to my supervisors - my greatest allies in this endeavour. Professor Peter Metcalfe, your expertise and guidance has been invaluable, not just for this work, but for my career as a whole. Thank you to Associate Professor Lois Holloway for sharing your wealth of knowledge, experience and never-ending support. You are an inspiring mentor. To Doctor Elise Pogson, thank you for providing encouragement, support and practical assistance whenever I needed it! I am so appreciative to have had a team as amazing as you three supporting and motivating me through this doctorate.

Thank you to the staff of the Liverpool Cancer Care Centre and Ingham Institute Medical Physics group for your support and input throughout the years. In particular, thanks to Professor Shalini Vinod and Lucy Ohanessian who 'showed me the ropes' for brachytherapy planning. It was a pleasure working with you. To Doctor Jason Dowling from CSIRO, thank you for your assistance with image processing. This work has benefited greatly from your expertise. Thank you to Associate Professor Taran Paulsen Hellebust for your assistance and mentorship whilst I was in Norway. I'm especially grateful to you for welcoming me into your home and making me feel part of your family when I was so far away from my own. To the staff at Oslo University Hospital, I

greatly value your help with this work, as well as your friendship. Tusen takk! Thank you to my colleagues at the Illawarra Cancer Care Centre, particularly Professor Martin Carolan, for providing the time and support required for me to complete this thesis. Thank you to the Centre for Medical Radiation Physics for the opportunities to travel nationally and internationally to promote this work. To all the staff, students and friends whose paths have crossed with mine along the way, thank you for sharing this journey with me and enriching my experience.

To my family, your unwavering love and support has made this possible. I am proud to share this accomplishment with you.

Most of all, thank you Trent for your limitless encouragement, motivation and love.

Contents

List of Publications

Published

Bell, L. R., Dowling, J. A., Pogson, E. M., Metcalfe, P., Holloway, L. (2017), Atlas-based segmentation technique incorporating inter-observer delineation uncertainty for whole breast. *J Phys: Conf Ser*, 777: 012002.

Vinod, S. K., Lim, K., **Bell, L.**, Veera, J., Ohanessian. L., Juresic, E., Borok, N., Chan, P., Chee, R., Do, V., Govindarajulu, G., Sridharan, S., Johnson, C., Moses, D., Van Dyk, S., Holloway, L. (2017), High risk CTV delineation for cervix brachytherapy: Application of GEC ESTRO guidelines in Australia and New Zealand. *J Med Imaging Radiat Oncol*, 61: 133-140.

Bell, L. R., Pogson, E. M., Metcalfe, P. E. and Holloway, L. (2016), Defining and assessing an anisotropic delineation margin for modern radiotherapy. *Med. Phys.*, 43: 6644-6653.

Pogson, E. M., **Bell, L.**, Batumalai, V., Koh, E. S., Delaney, G., Metcalfe, P. and Holloway, L. (2014), A comparison of coordinate systems for use in determining a radiotherapy delineation margin for whole breast. *J Phys: Conf Ser*, 489: 012057.

In Preparation

Bell, L., Holloway, L. C., Bruheim, K., Petric, P., Kirisits, C., Tanderup, K., Potter, R., Vinod, S., Lim, K., Pogson, E. M., Metcalfe, P., Hellebust, T. P. (2019), Dose Planning Variations Related to Delineation Uncertainties in MRI-Guided Brachytherapy for Locally Advanced Cervical Cancer. **(In Preparation).**

Bell, L., Vinod, S., Ohanessian. L., Lim, K., Pogson, E. M., Metcalfe, P., Holloway, L. C. (2019), Validation of Dose Planning Variations Related to Delineation Uncertainties in MRI-Guided Brachytherapy for Locally Advanced Cervical Cancer for the Australian Context. **(In Preparation).**

List of Awards

Endeavour Research Fellowship

2016

Merit-based scholarship funded by the Australian Government to support citizens around the world to undertake study, research and professional development.

Nucletron Prize

2015

Joint best postgraduate presentation at the 10th Student Research Symposium of the NSW/ACT Branch of the Australasian College of Physical Scientists and Engineers in Medicine

Australian Postgraduate Award

2013-2016

South Western Sydney Local Health District Radiation Oncology Scholarship

2013-2016

Conference Presentations

Bell, L., Hellebust, T. P., Bruheim, K., Petric, P., Tanderup, K., Pogson, E., Metcalfe, P., Holloway, L. (2016, November). Correlating delineation uncertainty metrics with dosimetric outcome for cervical brachytherapy. Presentation delivered at Engineering and Physics Sciences in Medicine (EPSM), Sydney, Australia.

Bell, L., Dowling, J., Pogson, E., Metcalfe, P., Holloway, L. (2015, November). Atlas based auto-segmentation method incorporating delineation uncertainty for whole breast radiotherapy. Presentation delivered at the Engineering and Physical Sciences in Medicine Conference (EPSM), Wellington, New Zealand

Bell, L., Pogson, E., Metcalfe, P., Holloway, L. (2015, November). Assessing a delineation margin for modern radiotherapy. Presentation delivered at the Engineering and Physical Sciences in Medicine Conference (EPSM), Wellington, New Zealand

Bell, L., Pogson, E., Metcalfe, P., Holloway, L. (2015, October). Defining and assessing a delineation margin for modern radiotherapy. Presentation delivered at the Innovations in Cancer Treatment and Care Conference, Sydney, Australia

Bell, L., Pogson, E., Metcalfe, P., Holloway, L. (2013, November). Assessing whole breast inter-observer delineation uncertainty in spherical coordinates. Presentation delivered at the Engineering and Physical Sciences in Medicine Conference (EPSM), Perth, Australia.

Poster Presentations

Bell, L., Dowling, J., Pogson, E., Metcalfe, P., Holloway, L. (2016, May). An atlas based auto-contouring technique incorporating interobserver variation. E-Poster presented at European Society for Radiotherapy and Oncology (ESTRO), Turin, Italy.

Bell, L., Dowling, J., Pogson, E., Metcalfe, P., Holloway, L. (2016, January). An Atlas based auto-segmentaiton technique incorporating inter-observer delineation uncertainty. Poster presented at Mini Micro and Nano Dosimetry and Innovative Technologies in Radiation Oncology (MMND-ITRO), Hobart, Australia.

Bell, L., Pogson, E., Metcalfe, P., Holloway, L. (2014, September). Quantitative assessment of an interobserver delineation margin for breast radiotherapy treatment planning. Poster presented at the Combined Scientific Meeting (CSM), Melbourne, Australia.

Pogson, E., **Bell, L.**, Batumalai, V., Koh, E., Delaney, G., Metcalfe, P., Holloway, L. (2013, May). A comparison of coordinate systems for use in determining a radiotherapy delineation margin for whole breast. Poster presented at the International Conference on the Use of Computers in Radiation Therapy (ICCR), Melbourne, Australia.

List of Figures

List of Tables

List of Abbreviations

4D-CT	Four-Dimensional Computed Tomography
CERR	Computational Environment for Radiotherapy Research
CI	Concordance Index
COM	Centre of Mass
CT	Computed Tomography
CTV	Clinical Target Volume
CTV_{HR}	High-Risk Clinical Target Volume
CTV_{IR}	Intermediate-Risk Clinical Target Volume
D100	Dose to 100% of the volume
$D2cm^3$	Dose to 2 cm^3 of the Volume
D90	Dose to 90% of the Volume
D98	Dose to 98% of the Volume
DSC	Dice Similarity Coefficient
DVH	Dose Volume Histogram
EBRT	External Beam Radiotherapy
EIT	Extra-Included Tissue

EQD2	Equi-Effective Dose in 2 Gy Fractions
ESTRO	European Society for Radiotherapy and Oncology
EUD	Equivalent Uniform Dose
FIGO	International Federation of Gynaecology and Obstetrics
GEC	Groupe Europèen de Curiethèrapie
GTV	Gross Tumour Volume
HDR	High Dose Rate
ICRU	International Commission on Radiation Units and Measurements
IGABT	Image-Guided Adaptive Brachytherapy
IMRT	Intensity Modulated Radiotherpay
MD-X	Maximum Lateral Dimensions
MD-Y	Maximum Anterior/Posterior Dimensions
MD-Z	Maximum Superior/Inferior Dimensions
MDU	Margin for Delineation Uncertainty
MRI	Magnetic Resonance Imaging
MT	Missed Tissue
NSCLC	Non Small-Cell Lung Cancer
OAR	Organ at Risk
PET	Positron Emission Tomography
PTV	Planning Target Volume
rSD	Relative Standard Deviation
SBRT	Stereotactic Body Radiotherapy

SD	Standard Deviation
STAPLE	Simultaneous Truth and Performance Level Estimation
US	Ultrasound
VMAT	Volumetric Arc Therapy

Chapter 1

Introduction

The efficacy of radiotherapy is dependent on its precision and accuracy. Increasingly conformal, modulated radiation fields can be reproducibly delivered to small, complex volumes within the human body. The advancements in technology that have enabled this precision have resulted in improvements in patient outcomes. In order to continue to provide optimal treatment for patients, remaining uncertainties must be minimised. One of the largest sources of uncertainty affecting radiotherapy accuracy is variability in delineation of the volume requiring treatment.

Delineation of the target volume occurs on computed tomography (CT) imaging (with or without aid from other imaging) and is manually drawn by the radiation oncologist. Herein lies the problem. By defining a binary volume (malignant cells inside, healthy cells outside), differences in observer decisions are emphasised, as opposed to identifying a region of varying probability of malignant cells. The issue is complicated by the inability to visualise individual cancer cells in situ with current imaging. A radiotherapy plan optimised to an oncologist's contour will differ to that of another oncologist. This potential inconsistency in treatment plans for the same patient can obscure results of trial data, as well as result in suboptimal plans for the individual patient. Further complicating the issue is the lack of knowledge of a ground truth. Not only will different oncologists produce different contours, and hence treatment plans, we

do not, and cannot know with current technologies, which oncologist is correct at a cellular level without retrospective pathological studies.

Volumes are an essential component of the treatment planning process, mostly due to the facilitation of dose-volume histograms that are used in dose reporting. However, the aforementioned issues justify questioning whether they are an optimal tool for planning purposes. Perhaps in the future radiotherapy treatments will dosimetrically vary with the probability of microscopic cancer, as opposed to relying on dose gradients to encompass any cells that may have been missed in the original target volume. Nevertheless, the current process is such that a hard line is drawn and hence, attempts to understand, minimise and manage this uncertainty in treatment planning need to be investigated.

This thesis investigates the variation in observers when delineating a target volume. It investigates the impact on dosimetric outcomes and methods to minimise the human intervention in the process such that consistency might be improved; and defines and analyses an uncertainty margin to best manage this uncertainty.

1.1 Project Aims

This section outlines the scope of this project by stating the project aims with associated research questions and the thesis chapters that address each. The three main aims explored in this thesis centered on delineation uncertainty are outlined below.

Aim 1. Understand the impact of delineation uncertainty on radiotherapy efficacy

Research questions:

What is the magnitude of inter-observer variation for high dose rate cervical cancer brachytherapy in the Australian context?

How does inter-observer variation affect the dosimetric parameters

obtained during the radiotherapy treatment planning process?

Do common metrics used to assess contouring consistency correlate with dosimetric outcomes?

Content Overview:

Chapter 3 quantifies the variability in high-risk clinical target volumes (CTV_{HR}) delineation for high dose rate (HDR) cervical cancer brachytherapy and assesses the dosimetric consequences of this variation on a gold standard target volume. Chapter 4 investigates the dose planning variation that arises from inter-observer variation when individual plans are optimised to individual contours of a Groupe Européen de Curiethérapie and the European Society for Radiotherapy and Oncology (GEC-ESTRO) inter-observer variation dataset. Chapter 5 validates the findings of Chapter 4 in the Australian context.

Aim 2. Minimise delineation uncertainty by automating the contouring process

Research questions:

Can an accurate automatic segmentation method be developed to reduce the magnitude of delineation uncertainty?

Content Overview:

Chapter 6 investigates an atlas-based auto-segmentation method that incorporates inter-observer variation. It is assessed on a whole breast external beam radiotherapy dataset.

Aim 3. Manage delineation uncertainty using an uncertainty margin

Research questions:

What is the best coordinate system to define a delineation uncertainty margin for whole breast external beam radiotherapy?

Can inter-observer variation be accounted for appropriately with an uncertainty margin?

Content Overview:

Chapter 7 presents an investigation into the most appropriate coordinate system to define a delineation uncertainty margin for whole breast radiotherapy. Chapter 8 expands on these findings to define and assess a margin for this cohort.

Chapter 2

Literature Review

2.1 Radiotherapy

Globally, cancer incidence and mortality is growing, with an estimated 18.1 million new cases and 9.6 million deaths in 2018 [1]. The 2018 Global Cancer Statistics report identifies the cumulative risk of developing cancer before the age of 75 as 21.4% and the risk of death as 17.7% [1]. Australia and New Zealand have the highest overall incidence rates globally with a cumulative risk in males of 49.1% and 33.3% in females [1]. The cumulative risk of death is comparatively low (10.8% in males and 8.1%) indicating the tremendous ongoing treatment burden for the Australian health care system [1].

The most common modalities available to treat cancer are radiotherapy, chemotherapy and surgery. The radiotherapy utilisation rate in Australia is 48.3% [2]. This means that approximately half of cancer patients would benefit from radiotherapy treatment at some point in their treatment journey, either in an adjuvant setting, or as the primary treatment modality. Radiotherapy utilises ionising radiation such as photons, electrons, protons or heavy ions, that damage and can destroy cancer cells by depositing energy within their vicinity. This energy deposition also damages healthy cells in the area being irradiated. However, healthy cells repair faster than cancer cells when exposed to radiation. By delivering radiation in small doses with short breaks in between, the differences in repair

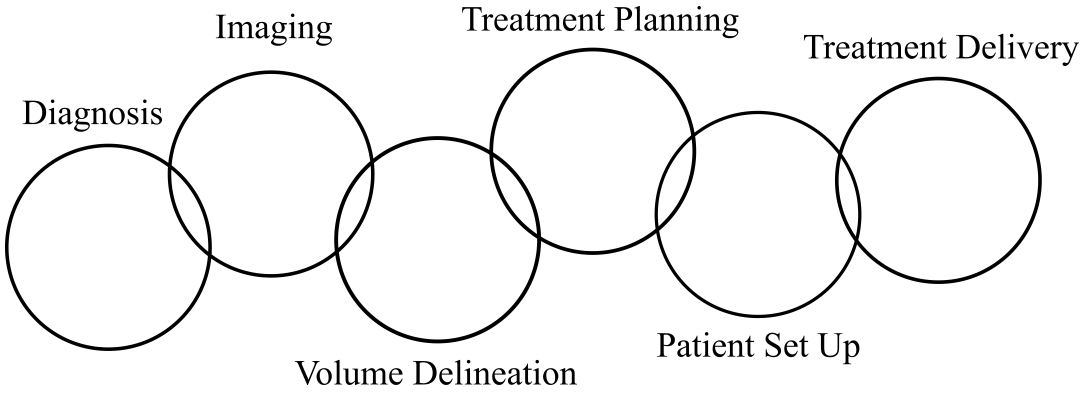


Figure 2.1: General radiotherapy process [3].

time can be fortuitously exploited. The cumulative damage to the cancer cells relative to the healthy cells increases with each fraction, forming the basis of therapy.

Radiotherapy is a multi-step, multi-discipline process. The general radiotherapy process is outlined in Figure 2.1.

Diagnosis is achieved via a combination of diagnostic imaging, biopsies, pathology and review of symptoms. Once a cancer diagnosis is made, the patient is referred to an oncologist who, as part of a team of specialists, will decide on the treatment course. If radiotherapy is considered, a radiation oncologist is referred to. One of the first things a radiation oncologist does is review existing diagnostic images and request further imaging for therapeutic purposes.

Imaging is an important part of the treatment chain as it facilitates tumour localisation and provides the tissue information required to calculate dose. CT is required in most treatment methods, however positron emission tomography (PET), magnetic resonance imaging (MRI) and ultrasound (US) provide beneficial supplementary information and are recommended in many instances.

Volume delineation is performed by the radiation oncologist and involves identifying the area which is intended to be treated on the CT dataset (or otherwise as appropriate). This often requires a volumetric delineation of the region. Target and organ at risk (OAR) delineation is essential for dose prescription, recording and reporting. International

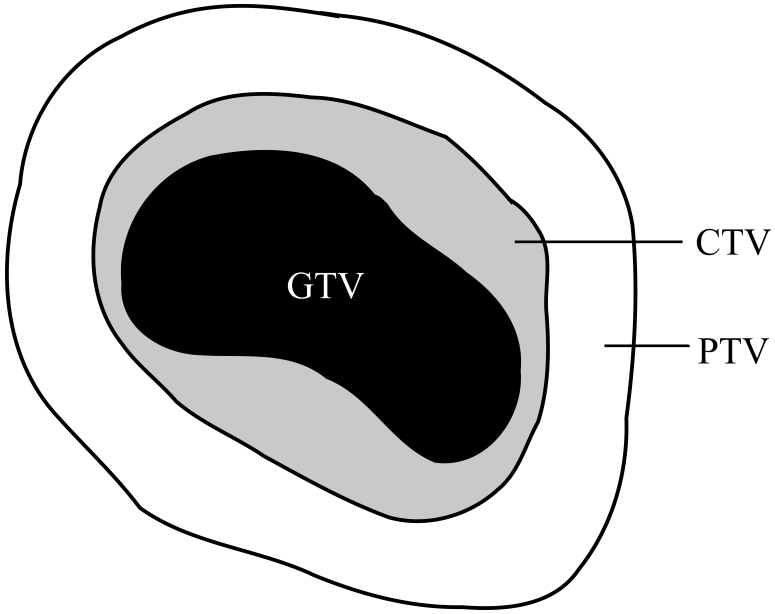


Figure 2.2: Schematic illustration of the different radiotherapy volumes. The clinical target volume (CTV) encompasses the gross tumour volume (GTV) and microscopic spread of the cancer. The planning target volume (PTV) is an extension of the CTV in the cardinal directions to account for uncertainties.

Commission on Radiation Units and Measurements (ICRU) Reports 50 and 62 identify specific regions required for treatment with varying concentrations of malignant cells [4, 5]. Figure 2.2 is a schematic illustration of the different volumes. The gross tumour volume (GTV) should include the malignant growth as is visible or palpable. The clinical target volume (CTV) includes the GTV and any microscopic spread of the malignant disease. The planning target volume (PTV) is an extension of the CTV to account for all treatment uncertainties. The OARs are also delineated, to which dose should be minimised.

The radiotherapy planning stage involves the development and application of a set of instructions that will be sent to the treatment machine to achieve a desired outcome. The amount of dose that should be delivered to the CTV and the dose constraints to OARs is specified. Treatment parameters such as fluence, beam energy, angle and aperture are optimised to achieve these goals.

Once the treatment is planned, the patient is positioned on the treatment couch in a

specific, reproducible way. Immobilisation equipment and techniques are used to minimise uncertainties. Immobilisation is essential in ensuring the planned treatment is delivered as intended.

Radiation treatment modalities fall into two broad categories: external beam radiotherapy (EBRT) and brachytherapy. EBRT is delivered via a linear accelerator which produces a beam of radiation that can be centred on the treatment target. A combination of different beam angles and shapes can be used to conform dose to the target volume within the patient in three dimensions. Brachytherapy utilises radioactive sources by placing them in close proximity of the treatment volume. For treatment sites other than the skin, this is achieved via interstitial needles or intra-cavitary techniques. Sources can be high or low dose rate, requiring short treatment times or permanent insertion respectively.

2.2 Types of Uncertainties in Radiotherapy

Every step of the radiotherapy process has the potential to introduce uncertainty into the accuracy of the delivered dose. An uncertainty in radiotherapy is defined as any deviation from the planned irradiation geometry during the radiotherapy process. The introduction of uncertainty may result in an under-dosing of the target volume, and an over-dose of the organs at risk [3]. If the uncertainty is large enough or target volumes small enough, geometric miss may occur.

Uncertainties can have systematic and/or random components. Systematic uncertainties are consistent, repeatable uncertainties that occur throughout a fractionated treatment course. The combined standard deviation (SD) of all systematic uncertainties is represented by 'Σ' [6]. Random uncertainties are unpredictable fraction-to-fraction variations. The combined SD of all random errors are represented by 'σ' [6].

Sources of uncertainties in radiotherapy can be broadly classified into set up uncertainties, motion uncertainties and delineation uncertainties. Motion uncertainties encompass any

movement during treatment that deviates from the patient anatomy at the time of treatment planning imaging. This encompasses motion of the patient after they have been set up for treatment (e.g. moving their head), as well as motion during the treatment delivery (e.g. lung motion due to breathing).

Set up uncertainties occur when there is a discrepancy between intended and actual treatment position at the time of treatment set up. This includes a shift or rotation of the patient in a particular direction, an incorrect field shape or size and incorrect isocentre positioning [7]. Set up uncertainties can be inter-fraction, meaning the uncertainty can vary between fractions; or intra-fraction meaning during fractions, as exampled by a couch positional difference between fields. Anatomical changes such as systematic weight loss, or variable bladder volume can be classified as set up uncertainty.

Delineation uncertainties represent the uncertainty in defining the 'true' treatment volume on the planning images. This uncertainty is difficult to quantify as the true treatment volume can only be determined with biopsies and pathology since individual microscopic cancer cells aren't visible on current therapeutic imaging. Delineation uncertainty results in intra-observer variation and inter-observer variation. Intra-observer variation occurs when a single observer produces different contours when delineating multiple times (presented with the same information). Inter-observer variation is the variation in contours produced by different observers. Inter-observer variation is the focus of this thesis. The origin of this variation lies in the individual, human interpretation of the data presented. In practice, surrogates such as CT images are used for delineation however this enables differences in image interpretation due to reasons including observer experience, imaging modality, windowing and levelling and room lighting.

2.3 Managing Uncertainties in Radiotherapy

It is unrealistic to expect that all uncertainties can be eliminated with current radiotherapy technology, however the presence of uncertainty and its impact can be minimised in a number of ways. Patient immobilisation equipment such as breast boards, head and neck masks and vacuum bags can serve to increase positional accuracy and constancy throughout treatment. Image-guided radiotherapy includes the use of imaging prior, during and after treatment to ensure the anatomy at the time of treatment matches the anatomy that the treatment plan was generated on. Conventional on-treatment imaging utilises a kilovoltage source and detector system mounted to the linear accelerator, enabling bony anatomy and some soft tissue to be visualised in a planar or volumetric acquisition. This imaging is primarily focused on maximising set up accuracy and in some cases motion management. Pre- and post- treatment imaging can be more sophisticated due to the increased time available for acquisition and scheduling flexibility. This can include PET, MRI or high quality CT imaging for target localisation improvements or for characterisation of the tumours range of motion.

Finally, quality assurance programs are put in place to ensure that all equipment in the radiotherapy chain behaves in an accurate and reproducible way. For example, the impact of superior on-treatment imaging can be negated if the couch shifts as determined by the software, aren't physically achievable.

The implementation of these methods to account for radiotherapy treatment uncertainties has served to reduce their magnitude. To account for residual set-up and organ positional uncertainties, ICRU reports 50 and 62 state that a margin should be generated around the CTV [4, 5]. To determine a margin, the magnitude of the combined uncertainty impacting the dose distribution must be determined. A common method utilises the central limit theorem to estimate the combined contribution of each source of error at the multiple time points that they may manifest [6]. The theorem assumes that the sum of many distributions with arbitrary shape will approach a normal Gaussian distribution. As such,

the SD from the mean of the measured errors is used to represent a population of errors [8].

ICRU Report 62 suggests that the root sum squares of the SD in each source of uncertainty should be used as a means of combining them (Equation 2.1) [5]. Linear additions of each source of uncertainty are not recommended.

$$\sigma_{total} = \sqrt{\sigma_{setup}^2 + \sigma_{motion}^2} \tag{2.1}$$

There are no recommendations in the ICRU reports on how to combine systematic and random components to achieve a singular SD for each source of uncertainty. Systematic uncertainties are known to affect the dose distribution quantitatively more than random uncertainties [9, 10]. Furthermore, systematic uncertainties are shown to shift the distribution as opposed to random uncertainties which blur the dose distribution [11]. This inherent difference between the two types implies that independent margins for systematic and random uncertainties should be used, and that a weighted sum should be used when combining these margins.

Several weighted combinations of systematic and random uncertainties have been published in regards to defining an uncertainty margin. These algorithms will be referred to as 'margin recipes'. Recipes accounting for just random uncertainties, as well as empirical recipes accounting for the microscopic expansion of the GTV to the CTV have been published. Table 2.1 presents the main margin recipes that exist in current literature.

There are two main derivations of the CTV-PTV margin in the literature [9, 14].

The margin method proposed by van Herk et al assessed the impact of uncertainties on the delivered dose and used this to determine the smallest acceptable margin around the CTV to account for it [14]. The margin aimed to ensure that 90% of the population received a minimum dose to the CTV of 95% the nominal dose. The systematic and random

Table 2.1: Margin recipes to expand radiotherapy volumes to account for uncertainties (CTV to PTV). Adapted and updated from Table 2 in van Herk et al [6].

Authors	Year	Recipe
Bel et al [12]	1996	0.7σ
Antolak and Rosen [13]	1999	1.65σ
Stroom et al [9]	1999	$2\Sigma + 0.7\sigma$
van Herk et al [14]	2000	$2.5\Sigma + 0.7\sigma$
		$2.5\Sigma + 1.64(\sigma - \sigma_p)$
McKenzie et al [15]	2000	$2.5\Sigma + \beta(\sigma - \sigma_p)$
Parker et al [16]	2002	$\Sigma + \sqrt{(\sigma^2 + \Sigma^2)}$
van Herk et al [10]	2002	$2.5\Sigma + 0.7\sigma - 3mm$
		$\sqrt{(2.7\Sigma)^2 + (1.6\sigma)^2 - 2.8mm}$
Topolnjak et al [17]	2008	$2.5\Sigma + 0.3\sigma$
McKenzie et al [18]	2002	1.3Σ
		$1.3\Sigma + 0.5\sigma$ (for blurring)
		$1.3\Sigma - 0.5\sigma$
Stroom et al [8]	2014	$2.1\Sigma + 0.8\sigma$ for clinical plans
		$2.5\Sigma + 0.9\sigma$ ideal scenario

components of uncertainties were treated separately.

The change in dose distribution for an individual patient was determined by first convolving the planned dose distribution with the random errors to estimate the cumulative blurred dose distribution arising from multiple treatment fractions. The blurred dose distribution was shifted by an amount and direction corresponding to a given systematic error. This was repeated for all systematic errors to obtain the dose distribution that would be delivered when accounting for all uncertainties.

Once the change in dose distribution due to uncertainties for a population of patients was known, the margin could be determined. For the systematic component, an ellipsoid of vector radius $\alpha\Sigma$ was used to describe the possible shifts of the CTV, where Σ is a vector describing the systematic component of the uncertainties and α is the factor that ensures 90% confidence that the ellipsoid encompasses the systematic uncertainties in three dimensions.

After correcting the planned dose distribution for systematic uncertainties, the difference between 95% isodose (for example) of the planned dose distribution and the 95% isodose of the blurred dose distribution was the additional margin required to account for random uncertainties. This could be expressed as $\beta\sigma$-$\beta\sigma_\rho$ where $\beta\sigma$ is the distance between a

reference 50% and the 95% isodose surface for the cumulative dose distribution, and the $\beta\sigma_p$ is the same distance but for the planned dose distribution. σ_p is the SD describing the width of the penumbra. The 50% isodose surface was taken as a reference point as its position is constant, even with blurring.

The margins for systematic uncertainties, and random uncertainties were linearly added. The margin recipe that met the specified criteria of 95% dose to the CTV for 90% of the population is seen in Equation 2.2.

$$2.5\Sigma + 0.7\sigma \tag{2.2}$$

Limitations with this derivation include the fact that idealised dose distributions and dose conformation were assumed, the errors were assumed to have isotropic effects in all directions, and no rotational errors and shape variations were considered.

A second type of margin derivation method, proposed by Stroom et al, utilises 'coverage probabilities' [9]. The CTV for a particular patient was converted into a 3D binary image, with values of one inside the CTV and zero outside. The binary CTV was convolved with the normal distribution of systematic uncertainties to obtain a coverage probability matrix. This is the probability of each point to be covered by the CTV. The 2.5% iso-probability curve corresponding to the volume bounded by voxels having a coverage probability of 2.5%, was taken as the expansion of the CTV for systematic uncertainties (PTV_1). To account for random uncertainties, a smaller expansion was needed. PTV_1 was convolved with the normal distribution of random uncertainties to obtain the final coverage probability matrix. The 25% iso-probability volume was taken to account for random deviations.

Based on the condition that on average, 99% of the CTV receives greater than 95% of the planned dose, the margin recipe seen in Equation 2.3 was suggested.

$$2\Sigma + 0.7\sigma \tag{2.3}$$

Limitations with this recipe lie in the fact that only set up errors were considered and all errors were mutually independent and Gaussian in nature.

It is interesting to note that even with two completely different methods of defining a margin recipe and with different, though admittedly similar criteria, Equation 2.3 and Equation 2.2 are near identical, with just a small numerical difference between the Σ scaling coefficient.

2.4 Inter-observer Variation - The Weak Link

Radiotherapy has undergone substantial technological advancement in the last 50 years resulting in significant improvement in local control rates [19]. Treatments have progressed from large, open fields with limited beam angles, to highly conformal volumetric treatments [19]. Paramount to this technology and technique progression is the requirement to correctly identify the region to be treated [3, 20]. The ability to accurately deliver conformal treatments, often consisting of high doses in single or few fractions, depends on delineation accuracy to ensure geometrical miss of the tumour and over-irradiation of normal tissue is avoided.

The quality of radiotherapy is only as strong as its weakest link. A large scale clinical trial showed that radiotherapy of head and neck cancer could have markedly inferior outcome if treatment protocols were not adhered to [21]. Specifically, poor radiotherapy was found to result in an approximate 20% decrease on 2-year overall survival for advanced head and neck cancer. Target volume definition inaccuracies were identified as having the most serious implications, exceeding the impact of dose distribution inadequacies and incorrect dose prescriptions.

The largest factor contributing to uncertainty in delineation, is inter-observer variation [3].

Inter-observer variation occurs due to the difficulty in identifying the true contour based on imaging. For the specific case of the CTV, microscopic disease cannot be visualised with current therapeutic imaging capabilities and spatial resolution limitations. Without this knowledge of a true volume, the CTV is defined based on an individual clinician's judgement on what tissue may contain malignant cells at one particular time-point. The factors influencing this judgment include observer experience and training [21–24], the imaging modalities available [25–28], availability and adherence to contouring protocols [21, 29] and patient/tumour specific factors [30, 31].

2.5 Quantifying Inter-Observer Variation

Metrics used to quantify delineation uncertainty are varied amongst the literature, with no consensus between studies as to which metric(s) should be used [32–34]. Metrics to describe a radiotherapy target delineation include volume, dimensions, shape and centre of volume. Most studies utilise a combination of metrics due to the limitations in each [34]. It appears that in different anatomical locations, some metrics may be more appropriate than others [35, 36].

2.5.1 Volume

Volume is a metric that describes the amount of tissue encompassed by a delineation. It is typically determined by identifying the number of voxels contained within the contour and multiplying by the voxel size. However, different software may calculate volume slightly differently. Volume comparison is excellent for providing a comparison of the size of delineations, however it lacks spatial and shape information. Volume is the most commonly reported metric in delineation studies [33, 34].

2.5.2 Centre of Mass

The centre of mass (COM) is a single point identifying the centre of a delineation in three dimensions. This metric gives an indication of the positional similarities between delineations, but no shape or size information is obtained. COM is often reported in delineation uncertainties studies, however it is rarely the only reported metric [34].

2.5.3 Dimensions

A delineation can be described by how much it extends in specific dimensions. Delineation uncertainty amongst a cohort of delineations can be determined by calculating the SD in extension along a specific axis. Dimensions can be determined by defining a cuboid around the delineation such that the surfaces of the cuboid abut the surfaces of the delineation [33]. The known dimensions of the cuboid therefore indicate the dimensions of the delineation in those directions. Alternatively, a vector emanating from the COM of the delineation can identify the dimensions of the shape by finding the distance at which the vector intersects the delineation. Depending on the number of axes reported, dimension metrics can describe the shape and size of a delineation. Dimensions can sometimes be misleading if there are surface variations that occur between the axes investigated.

2.5.4 Volume Overlap Metrics

Volume overlap metrics are measures of the overlap between two volumes. They attempt to overcome the lack of spatial information in a volume metric, however in a relative capacity only. Generally, a value of 1 represents 100% overlap of the two delineations and a value of 0 indicates zero overlap. Volume overlap metrics commonly utilise the union and/or intersection volumes of two delineations. Two of the most common metrics in the literature are the Jaccard Index and the Dice Similarity Coefficient (DSC). The

Jaccard Index is the ratio of the volume of overlap to the encompassing volume of two delineations as seen in Equation 2.4.

$$JaccardIndex = \frac{|V_1 \cap V_2|}{|V_1 \cup V_2|} \tag{2.4}$$

The Jaccard index has been used extensively in the literature but is referred to by a number of names including 'conformity index' [37], 'concordance index' (CI) [31] and 'ratio of common to encompassing volume' [25, 29, 38]. This lack of consistent terminology adds to the difficulty in comparing delineation uncertainty studies. It was first proposed as a tool for botanic comparisons and was termed the 'coefficient de communaute' [39, 40].

The DSC was first proposed in an ecology journal (called the 'coincidence index') as a metric that quantifies the association between two groups independent of which group is used as the base [41]. It is similar to the Jaccard index however places greater weight on the intersecting volume of delineations.

$$DSC = \frac{2|V_1 \cap V_2|}{|V_1| + |V_2|} \tag{2.5}$$

The above metrics can be used to characterise a cohort of delineations; however it is most meaningful to compare each delineation to a reference delineation. This is especially relevant for overlap metrics. The reference delineation should be a ground truth delineation which ideally encompasses 100% of the malignant cells and 0% normal tissue. However, in the absence of patient specific pathology slides, the absolute accuracy of the target volume is unable to be determined. Therefore, the reference delineation must be a best estimation for the ground truth. Methods to determine a reference delineation include a mathematical estimate of the true contour [42] or manually generated by an individual or panel of experts [33]. One study by Gao et al utilised images from the Visible Human Project to use as a gold standard [43].

2.6 Impact of Imaging on Inter-Observer Variation

CT images are the current standard of care for external beam radiotherapy due to the provision of electron densities that are required for treatment planning. However, a number of other auxiliary imaging modalities and techniques are often utilised to improve visualisation of the regions of interest. Most commonly used in clinical treatment planning workflow are MRI and PET. The use of contrast enhancing agents can also improve the ability to visualise regions of interest. Image resolution has been shown to impact inter-observer agreement, with high resolution images producing the greatest agreement in delineations [44]. Each imaging modality and type has its own advantages and disadvantages which can be exploited to improve delineation uncertainty.

MRI has the advantage of improved soft tissue contrast compared to CT. Delineations drawn on MRIs differ in size and shape than their equivalents on CT for prostate, brain, bone metastases and breast cancers [38, 45–47]. However, despite improved tissue contrast and the associated hypothesis that delineation uncertainty may be reduced, this isn't always the case. Breast contours defined on both MRI and on CT were shown to have comparable delineation uncertainty [45, 47] and the combination of MRI to CT for brain delineation did not reduce inter-observer variation [46]. Conversely, head and neck and prostate cancer MRI contours had less inter-observer variation than CT [48, 49] and delineations of bone metastases were most consistent on MRI [47]. The difference in inter-observer variation reduction may be attributed to the fact that breast and brain target volumes are already fairly well visualised on CT, especially with contrast agents, therefore the improved soft tissue image quality of MRI is hypothesised to have a negligible impact on the uncertainty in delineation. For head and neck tumours as well as prostate where the boundaries are less obvious in CT, MRI enables better visualisation and therefore better consistency in delineation. Weltens et al doesn't discredit the use of MRI for brain, despite no change in delineation uncertainty. The utilisation of both

modalities in conjunction with each other enables an informed delineation decision, especially since the ground truth is not known [46].

In most cases PET helps reduce inter-observer variation in lung when acquired in conjunction with a CT (PET-CT) [25, 39, 50, 51]. This is due to the additional biological information of metabolically active areas in PET images compared to conventional CT. However, delineation uncertainty is still substantial for PET-CT lung delineations [25, 51, 52]. Additionally, one study found that the utilisation of PET-CT images for NSCLC did not improve the concordance index of observer volumes, though the authors acknowledge that the study was underpowered [53]. PET based tumour volumes have been shown to be strongly affected by the choice of threshold level in the specific uptake value [50]. For head and neck tumours, one study found no change in inter-observer variation between PET-CT and CT defined contours [54].

The use of multi-modality imaging is identified as being the most appropriate approach to target volume delineation [55]. Although multi-modality contouring may improve delineation uncertainty, it has been shown that delineation uncertainty still persists. Even when the structure being contoured is well visualised as in the Visible Human Projects images, delineation uncertainty remains [43]. Bernier, Hall and Giaccia identified molecular biology and genetics as being the next step in improving target definition and radiotherapy quality in general [19].

2.7 Impact of Anatomical Location on Inter-Observer Variation

Inter-observer variation is substantial in a number of anatomical sites including soft tissue sarcoma, pituitary adenoma and brain tumours [20, 30, 34, 46, 56]. Some anatomical sites are more conducive to inter-observer variation than others with high contrast tumours more easily visualised than soft tissue organs. However, inter-observer variation is still

substantial due to requirement of human interpretation of images. A brief summary of the literature for some key anatomical sites is presented below.

2.7.1 Lung

The inter-observer variation in thoracic tumours is generally large, with substantial variations in measured volume [24, 57, 58], CI [24], overlap metrics [58, 59] and dimensions [58]. The most commonly assessed volume in the literature is the GTV for non small-cell lung cancer (NSCLC), demonstrating large inter-observer variation [24, 31, 57–60]. Delineation uncertainty in lymph node contouring was also found to be substantial [61, 62]. Inter-observer variation is present in NSCLC GTVs regardless of whether radiation oncologists or radiologists are contouring, though radiologists were found to systematically delineate smaller volumes than radiation oncologists [24]. One study found small overall inter-observer variation in peripheral lung tumours for stereotactic body radiotherapy (SBRT), with the variation largest in the cranial-caudal direction [31]. This differs from the large variation observed in other studies due to the increased visualisation of the tumours due to their location away from the chest wall.

In theory, four-dimensional CT (4D-CT) might provide better image quality for lung tumours than conventional CT, however this was not found to impact inter-observer variation, with one study identifying large deviations [59]. Of the 4D-CT reconstruction methods, volumes defined on 'mid-position' reconstructions (reconstruction of every part of the anatomy in its average position), have the smallest inter-observer variation [63].

2.7.2 Cervix

Inter-observer variation has been observed in all cervical brachytherapy MRI volumes (intermediate risk CTV (CTV_{IR}), CTV_{HR} and GTV) defined on MR images [64]. The

CTV_{HR} was found by Petric et al to have the smallest variability amongst observers and Dimopoulous et al found the mean volume differences were not significantly different for this volume [64, 65]. CTV_{IR} was shown to have significant variation, specifically in the posterolateral directions [64, 65]. Variation in the CTV_{HR} was largest in the cranial-caudal direction [64]. Volume and overlap metrics were most commonly assessed in studies assessing inter-observer variation in cervix radiotherapy [64–66]. Inter-observer variation was present for CT defined contours, however contours were more consistent, and larger, than when defined on MR images [66].

Contours defined for external beam radiotherapy also exhibit substantial variation when assessing CTV volumes, overlap metrics and COM [20, 67].

2.7.3 Breast

Whole breast PTVs were found to have substantial inter-observer variation as reported most commonly by variation in volume [22, 29, 68, 69] and overlap metrics [22, 69, 70]. Although these metrics provides an overall indication of how consistent delineations are, they do not provide any information about inter-observer variation in specific anatomical regions. Inter-observer variation was found to vary locally on the target volume [29, 69]. Hurkmans et al found the largest variations in the posterior, cranial and medial directions of whole breast target volumes [29]. Batumalai et al found that despite high CI for whole breast target volumes, the medial and posterior edges of whole breast PTVs displayed marked variation between observers [22].

Another commonly contoured volume in breast radiotherapy is glandular breast tissue. In general, glandular breast tissue is consistently contoured by observers [37, 45, 71]. However, similar to whole breast volumes, regions of highest inter-observer variation are the cranial and posterior directions [37].

In post-operative breast target volumes, inter-observer variation was considerable [71–75]. Inter-observer variation was found to remain substantial, even after contouring

guidelines implementation [75]. Region specific analysis of surface variations indicate that inter-observer variation in these volumes is spatially dependent [75]. The presence of gold fiducial markers or surgical clips on inter-observer variation for lumpectomy cavities was shown to improve variation [73, 75]. In the pre-operative setting, inter-observer variation is not substantial [74, 76].

2.7.4 Prostate

Large variability has been observed in the literature for prostate delineation with one study finding a range in CTV volume from 39.9 to 180.5 cm^3 [77]. Volume and volume overlap metrics were the most common metrics for assessing inter-observer variation [77–80]. Dimension metrics were also utilised to provide some spatial information [23, 81, 82]. The apex of the prostate exhibited the largest variations for the majority of studies investigating inter-observer variation for this site [79, 80, 83, 84]. More generally, large variability was in the craniocaudal length [23, 82] with one study suggesting that contouring consistency may be improved here if delineation is performed on all three planes due to CT pixel anisotropy in the sagittal and coronal planes [23]. The seminal vesicles also exhibited poor consistency in observer delineation [80, 83, 84]. The central portion of the prostate exhibited the highest consistency [83].

Prostate contours defined on CT exhibited greater variation amongst observers as assessed by volume ratio than contours defined on MRI [78]. The utilisation of both MRI and CT images when contouring reduced the volume of prostate PTVs as well as the variability amongst observers when compared to CT alone [80]. Although MRI appears to be the imaging modality of choice for minimised delineation uncertainty, the exact sequence should be optimised as Nyhom et al found substantial variability in observers for different MRI sequences [81].

2.8 Understanding Inter-Observer Variation - Dosimetric Impact

This review so far has identified that inter-observer variation is substantial in a number of anatomical locations and imaging modalities. However, the impact on dosimetric outcomes is a more relevant clinical endpoint especially since systematic uncertainties like inter-observer variation are known to shift the dose distribution [6].

There are two main ways to assess the dosimetric impact of inter-observer variation. Both methods require a cohort of patient images with multiple observers delineating the volume of interest, and a reference delineation. The first method involves optimising a plan for each of the observer delineations. The dose volume histogram (DVH) metrics can be determined for each of the plans on a common reference delineation. The variability in DVH metrics are indicative of the variation in planning that arises due to inter-observer variation. The second method requires a single plan optimised to the reference volume. The variability in DVH metrics as assessed on each of the individual delineations can be used to quantify the dosimetric impact of inter-observer variation.

Many studies have investigated the dosimetric impact of inter-observer variation. In the vast majority of studies, substantial variability in dosimetric parameters are observed due to inter-observer variation. For breast delineations, the dosimetric impact is consistent for the target volumes, however exhibiting substantial variability for the OARs [85]. Inter-observer variation in lung delineations can lead to unacceptable normal tissue doses including cardiac dose and mean lung dose [58, 86, 87]. For head and neck radiotherapy, substantial differences in dose volume histogram parameters were observed due to inter-observer variation for parotid and brachial plexus contours [88, 89]. Delineation uncertainty in paranasal sinus tumours have been shown to have considerable impact on delivered dose [90]. Inter-observer variability in brain delineations resulted in unacceptable plans for 30% of delineations [91]. Also, inter-observer contouring variability was shown to strongly impact DVH parameters for

penile bulb contouring [92].

Delineations of the left anterior descending coronary artery have exhibited large inter-observer variation, resulting in large dosimetric variation [93, 94]. The use of a heart reference dataset reduced the dosimetric variation in this volume, however in a limited capacity compared to the whole heart, left main artery and right coronary artery which were all reduced to less than 1 Gy dosimetric variability [94].

The dosimetric variability in rectum PTVs was observed to be as large as 10% for three-dimensional conformal radiotherapy plans [95]. Although reduced when contouring guidelines are implemented, this variability still persists, reducing to approximately 5% variability in PTV doses. The same trend was observed for volumetric arc therapy (VMAT) plans, however greater variability was observed. This suggests an increasing lack of robustness of plans to inter-observer variation as techniques become more conformal.

Conflicting results are observed for oropharynx cancer. Feng et al found substantial inter-observer variation in the OAR delineations, but noted minimal dosimetric difference (< 0.9 Gy) indicating that intensity modulated radiotherapy (IMRT) planning for this treatment site is robust to inter-observer variation [96]. However, Nelms et al found that OAR delineations exhibited substantial inter-observer variation, resulting in substantial dosimetric variation [97]. The conflicting findings may be explained by the fact that both studies were relatively underpowered, with the former using 3 observers and 10 patients and the later only 1 patient and 32 observers.

A handful of studies have shown that despite substantial inter-observer variation, the resulting dosimetric impact is minimal. For breast seroma volumes, PTV margins adequately accounted for delineation uncertainties, because minimal dosimetric impact was observed as assessed by the fractional volume contained by the 95% isodose surface and equivalent uniform dose (EUD) [98]. Inter-observer variation did not result in clinically significant outcome differences for prostate [84] and gastric cancer PTV

dosimetric parameters varied minimally [99]. For cervical cancer brachytherapy OARs, the dosimetric variability was reasonably consistent, however this was primarily due to the consistency in OAR contouring for this particular study [100]. Other studies have found up to 2-3 Gy variability in dose over the course of an entire treatment [101] and up to 10% variation [102] for OAR parameters. Sigmoid delineations exhibit the greatest variability, most likely due to the varying interpretation of sigmoid anatomy [103]. Cervical cancer brachytherapy CTV_{HR} can exhibit up to 5 Gy variability across a treatment course [101].

2.9 Minimising Inter-Observer Variation

Attempts to minimise inter-observer variation have been investigated in a number of studies. Vinod et al reviewed the interventions that have been employed to reduce inter-observer variability in volume definition in radiation oncology [34]. Non-imaging interventions can be categorised according to guides and protocols, teaching and contouring aids. The use of delineation aids such as contrast, radio-opaque wires, fiducial markers and palpation also aid in reducing contour inaccuracy [22, 29, 104]. Working collaboratively with other colleagues has also been demonstrated to reduce inter-observer variation [105].

2.9.1 Guidelines or Protocols

A number of studies have suggested that clearly defining protocols will reduce variation [57, 86, 87, 93, 106–109]. Statistically significant reductions in inter-observer variation have been shown in lung, breast and OAR delineation when observers followed a contouring protocol. Sources for contouring protocols in studies that assess these interventions often include trial protocols [86, 107]. Other sources of protocols come from advisory groups such as the ESTRO advisory committee on radiation oncology

practice, who have defined contouring guidelines for non-small cell lung cancer [110].

For NSCLC CTVs, the utilisation of a phase III trial contouring protocol resulted in a statistically significant reduction in delineation variation (assessing dimensions, specifically in the cranial-caudal dimensions) as well as a reduction in the resulting dosimetric variation (including mean lung dose, as well as heart, spinal cord and oesophagus DVH parameters) [86]. Another study investigating NSCLC delineation variability noted that there was statistically significant variation between observers delineating the GTV. After a 3 year period and a contouring protocol revision, no statistically significant variations were observed between the same observers [57]. Statistically significant variations in seroma GTVs for partial breast radiotherapy patients were also eliminated in a cohort of observers who were 'trained' according to a contouring protocol compared to observers who were 'untrained' [106]. A comparison of inter-observer variation in breast for radiation oncologists and radiation therapists was performed by Batumalai et al [22]. The inter-observer variation amongst the radiation therapists was smaller than that of the radiation oncologists. The authors hypothesised this to be due to the adherence of the contouring protocol. The oncologists were hypothesised to have weighted their contouring decisions on previous experience higher than on the protocol.

The utilisation of contouring protocols may result in a reduction of volumes, as was the case for seroma volumes in partial breast radiotherapy [106]; or an increase in volumes as was the case for prostatectomy CTVs [107]. In most cases however, observers following a delineation protocol produced more consistent volumes. Inter-observer variation was not reduced with the addition of a contouring protocol for breast [75] and lung cancer radiotherapy patients [60]. Despite the breast seromas showing a reduced variation, the left anterior descending coronary artery did not improve with guidelines [93].

Lack of protocol compliance for delineation uncertainty was shown to be the biggest

contributor to poorer patient outcomes in Peters et al [21]. Therefore, there is evidence that introducing and adhering to a protocol for delineation would reduce inter-observer variation and improve patient outcomes.

These studies suggest that guidelines and protocols improve consistency in delineation. The accuracy of the contours after the intervention of contouring protocol is often not assessed, therefore no comment can be made about the improvement in accuracy with contouring guidelines.

2.9.2 Teaching

Types of teaching interventions in the literature include didactic lectures, practical workshops or one-on-one teaching [78, 87, 95, 109, 111–116]. The impact on inter-observer delineation uncertainty as a result of teaching is variable with studies conflicting on whether a reduction is observed.

Prostate CTV delineations on CT were not improved despite MRI and CT prostate anatomy lectures and practical workshops. Participants felts their contouring skills and confidence had improved, however no significant difference in contour consistency was observed based on the intervention [112]. Similarly, for NSCLC, no significant differences were observed before and after an education course for radiation oncology residents [87]. The authors of this study speculate that the high quality of the contours prior to education may explain this anomalous finding. For radiation therapists delineating a range of volumes including prostate, bladder and rectum, large delineation variations were observed before an education session, and there was little improvement when repeated after the session [113].

Conversely, Tai et al found that as little as a single ten-minute one-to-one training session can statistically improve inter-observer variation for cervical oesophageal target volume delineation [111]. Improved consistency in both CT and MRI based prostate planning was observed after didactic lectures, practical sessions and with the assistance

of a prostate atlas [78]. A practical workshop on early stage rectal cancer delineation, coupled with a delineation atlas, resulted in decreased CTV volumes, reduced caudal dimensions and reduced surface dimensions in the anterior, lateral, posterior and sphincter region [115]. Furthermore, the introduction of teaching and contouring guidelines also showed improved dosimetric consistency for rectal radiotherapy [95].

Just as there is discordance about whether inter-observer variation improves with teaching, there is discordance about whether teaching improves delineation accuracy. Some studies observed an improvement in delineation accuracy as compared to the 'expert' contours after education, though these results were not significant [87, 113]. Conversely, head and neck squamous cell carcinoma CTVs as delineated by radiation oncology residents were found to remain sub-optimal despite didactic lectures and hand on practical sessions [116]. It is difficult to meaningfully assess accuracy in delineation without a pathological ground truth.

2.9.3 Automatic Segmentation

Manual delineation of target volumes in radiotherapy is both resource and time intensive. Automatic delineation techniques (auto-segmentation) have been shown to achieve fast, and consistent segmentation, reducing the presence of inter-observer variation. Auto-segmentation can be achieved via a number of methods that are summarised in Table 2.2.

At the time of completing the investigations for this thesis, atlas-based segmentation was the most advanced and frequently used method in medical image segmentation [125]. Hence, Chapter 6 and the following literature review is based on this method. As outlined in Table 2.2, it is acknowledged that the field has progressed, therefore future work arising from this thesis would investigate machine-learning based auto-segmentation.

Atlas-based segmentation involves inter-subject registration of retrospective patient

Table 2.2: Summary of automatic segmentation methods [119, 120].

Type	Method	Limitations
Threshold-Based	Regions of uniform brightness are segmented by applying a threshold to separate the object and background.	The object and background must have uniform brightness of distinct grey levels.
Edge-Based	Discontinuities in grey levels or colour are identified, enabling segmentation based on the boundaries of image features.	Noise and indistinct discontinuities in grey levels can result in inaccurate segmentation.
Region-Based	Pixels with similar properties (normally grey levels) are grouped together.	Requires optimal selection of properties on which to cluster pixels otherwise regions may be over/ under segmented
Textural Feature-Based	Regions of mutually related elements (texture) are segmented. Tone (pixel intensity properties) and structure (spatial relationship of pixels) are used to define texture.	Requires optimal selection of texture elements with which to cluster pixels otherwise regions may be over/under segmented
Model-Based	The structure of organs is modelled probabilistically as they have a repetitive geometry. This model can be used as a constraint in conjunction with other segmentation methods.	Requires manual interaction to spatially place an initial model on an image and choose appropriate parameters.
Atlas-Based	Information on anatomy, shape, size etc of organs for segmentation are compiled in the form of an atlas. This atlas is either used as a reference or co-registered with an image to enable segmentation of a new dataset.	Require expert knowledge in building the atlas/ database and may be limited in segmenting complex structures with variable shape.
Machine Learning-Based	Artificial neural networks process feature extraction of images quickly and iteratively to enable fast and accurate segmentation of data. The neural network learns for previous data and applies this to a new dataset.	Depends on previous data therefore limited by the limitations in data eg image quality. Require large datasets to begin, however can continually learn.

datasets containing delineations. The result of these registrations is an 'average atlas' image that is representative of that cohort. Auto-segmentation is achieved by deformedly registering the atlas to a new dataset and subsequently propagating the delineations. Atlas-based delineation in the context of inter-observer variation has been investigated in breast, endometrial, lung and head and neck radiotherapy [52, 69, 117, 118]. All studies have demonstrated improved consistency in automatic delineations as compared to manual contours, as well as notable time savings. In the specific case of endometrial cancer, a commercial based system for atlas-based delineation produced more consistent target volume delineations, with overlap metrics increasingly significantly compared to the manual delineations. Breast delineations were shown to have a decreased mean surface to surface distance when auto-contoured and required minimal editing by the observers [69]. Head and neck contours were not only shown to be more consistent, but by comparing to a gold standard contour, auto-contours were shown to be more accurate when using a commercial auto-contouring system [117].

Other methods of automatic delineation include automatic thresholding of specific uptake values in PET-CT datasets [52], as well as using a 'gold standard' breast dataset deformedly registered to the dataset of interest, enabling observers to study the expert consensus contour before undertaking their own delineation [121]. The Radiation Therapy Oncology Group contouring atlas for anorectal CTV is another example of this method [122]. Note that in this specific case, the use of the 'atlas' term does not refer to the definition outlined above.

Ultimately, auto-segmentation techniques require an oncologist to review, and in some cases edit the contours, thereby re-introducing inter-observer variation. Nevertheless, these techniques are efficient tools that provide a good starting point for defining target volumes in radiotherapy [123].

2.10 Managing Inter-Observer Variation

Accounting for inter-observer variation is increasingly paramount. As techniques and technologies evolve to reduce set up and motion uncertainties, uncertainty margins are reducing as well. Delineation uncertainty is less affected by treatment improvements as it originates in the planning stage of radiotherapy and is limited by imaging capabilities. Where in the past, larger radiotherapy margins, although not designed for delineation uncertainty, likely mitigated it to some extent; in recent practice it is inadequately accounted for. This is especially important for stereotactic techniques where millimetre margins are being proposed [124]. The risk of geometric miss due to inter-observer variation is arguably greater than ever.

As outlined earlier in this review, uncertainties in radiotherapy are typically accounted for by applying an uncertainty margin around the CTV. Although ICRU reports 50 and 62 state that a margin should account for all geometric uncertainties, in practice, radiotherapy margins often do not include delineation uncertainty [4, 5]. Studies investigating margin methodologies for radiotherapy, as summarised in Table 2.1, often do not include a delineation uncertainty component. The few studies that do include this error in their derivation often assume idealised dose distributions, spherical symmetry of the error or no shape variations of the target volume [10, 14]. If delineation uncertainty were to be included in the conventional margin application in the three cardinal dimensions, it is likely that the error would be inadequately accounted for. This is due to the spatially varying nature of delineation uncertainties.

The analysis of radiotherapy uncertainties at other points than the cardinal dimensions has been investigated for the prostate, bladder, cervix and breast [38, 64, 75, 125]. The cranial-caudal dimension exhibits delineation variability in lung, bladder, Hodgkin lymphomas and soft-tissue sarcomas, possibly due to the partial voluming effect of slice-based imaging [30, 31, 125, 126]. Breast target volumes have been shown to have the greatest inter-observer variability at the posterior and medial borders [22] whereas

prostate volumes vary most anterior-superior to the apex [38, 127]. Most of these studies have concluded that anisotropic margins are necessary to adequately account for spatially varying uncertainties. Three-dimensional, anisotropic extensions of the CTV to account for spatially varying uncertainties have been proposed [127–129]. Austin-Seymour implemented such a margin for both lung and head and neck sites however the margin did not consider inter-observer variation in its generation [129].

Chapter 3

High-Risk CTV Delineation for Cervix Brachytherapy: Application of GEC-ESTRO Guidelines in Australia and New Zealand

This chapter quantifies the variability in CTV_{HR} delineation for HDR cervical cancer brachytherapy on an Australian dataset and assesses the dosimetric consequences of this variation on a gold standard target volume. The findings in this chapter contribute to answering the first aim of this thesis: to understand the impact of delineation uncertainty on radiotherapy efficacy (Chapter 1.1).

The content in this chapter has been published. The candidate's contribution to this publication is in the form of data collection, data analysis and manuscript contribution.

Vinod, S. K., Lim, K., Bell, L., Veera, J., Ohanessian. L., Juresic, E., Borok, N., Chan, P., Chee, R., Do, V., Govindarajulu, G., Sridharan, S., Johnson, C., Moses, D., Van Dyk, S., Holloway, L. (2017), Highrisk CTV delineation for cervix brachytherapy: Application of GECESTRO guidelines in Australia and New Zealand. J Med Imaging Radiat Oncol, 61: 133-140. doi:10.1111/1754-9485.12509

3.1 Introduction

Image-based brachytherapy is becoming the standard-of-care for cervical cancer. The GEC-ESTRO group first published guidelines on tumour volume delineation in 2005 [130, 131] and have held regular workshops to educate clinicians on these newly defined concepts. As with any new technique, interpretation of guidelines may vary with resulting differences in practice.

The GEC-ESTRO group have documented contouring variability amongst both GEC-ESTRO radiation oncologists [65] and worldwide including from the United Kingdom, India and USA [64]. Petrič et al evaluated delineation of GTV, CTV_{HR} and IR-CTV amongst 10 radiation oncologists and compared this with a Simultaneous Truth and Performance Level Estimation (STAPLE) contour [42] and expert consensus volume [64]. Delineation variability was noted for all volumes with the CTV_{HR} being the most consistent volume. This resulted in dosimetric differences with a mean relative standard deviation (rSD) of 8-10% for both GTV and CTV_{HR} [101].

Australian and New Zealand radiation oncologists have attended GEC-ESTRO workshops and subsequently implemented image-based brachytherapy in their own departments. The use of image-based brachytherapy has increased from 27% in 2005 [132] to 65% in 2009 [133]. The main imaging modality has been CT with only one department using MRI in 2005 increasing to three in 2009. The number of patients with cervical cancer treated in many departments can be low with 29% of centres treating 10 or less cervical brachytherapy patients per year in 2014 [134].

The practice of cervical brachytherapy is usually limited to one or two radiation oncologists in a department with limited opportunities for quality assurance on delineated volumes. The aims of this study were to evaluate translation of GEC-ESTRO guidelines into the Australian and New Zealand setting by measuring variability in CTV_{HR} delineation and assessing the dosimetric consequences of this.

3.2 Methods

Radiation oncologists currently using image-based brachytherapy were identified from a patterns-of-care survey conducted in 2014 [134] and invited to participate in a target volume delineation study. Of the 38 centres performing brachytherapy throughout Australia, New Zealand and Singapore, 21 used MRI at some point in brachytherapy planning. Clinicians from these 21 centres were invited to participate with those from 10 centres agreeing to do so. This included nine radiation oncologists and one radiation therapist with extensive experience in gynaecological brachytherapy. In addition, two radiologists specialising in pelvic MRI with experience in radiation therapy volume delineation also took part in the study.

The study was approved by the institutional ethics board. The datasets of 10 consecutive cervical cancer patients undergoing brachytherapy at a single institution were de-identified for use. The MRI images were acquired on fraction 1 or 2 of brachytherapy on a 3-Tesla wide bore MRI system (MAGNETOM Skyra, Siemens Healthcare, Erlangen, Germany) with 45 mT/m gradient strength and 200 T/m/s slew rate. Oblique sagittal and coronal T2 weighted half-fourier acquisition single-shot turbo spin echo (HASTE) and T2 weighted turbo spin echo (TSE) oblique axial imaging of the brachytherapy applicator set in vivo were acquired in all patients. The applicators used were tandem and ovoids in eight cases and tandem and cylinder in two cases.

Participants were given clinical details, a diagnostic MRI report of initial tumour extent, 'Examination Under Anaesthesia' findings at brachytherapy and a schematic diagram of findings at brachytherapy. Participants were asked to contour CTV_{HR} based on this information. This volume was chosen as CTV_{HR} is the most relevant volume to delineate in image-based brachytherapy. It contains macroscopic tumour at time of brachytherapy [130] and has a dose-response relationship with local control [65]. Two reference volumes were created for comparisons: a STAPLE [42] and a manually generated consensus volume (consensus) between two radiation oncologists and a radiologist. The

consensus contour was created after all individual contouring had been completed. The STAPLE contour is a probabilistic estimate of the 'true contour' computed from the manually drawn contours by measuring the performance of each individual segmentation.

Contour comparisons were performed using volume (mean and SD), the DSC and MASD. The DSC is a measure of spatial overlap between two volumes as defined by 2(A∩B)/A+B where 0 is no agreement and 1 is perfect agreement. The MASD is the average distance between surface voxels [32, 39]. A DSC close to 1 and MASD close to 0 mm is associated with minimal variability in delineation. For the purposes of this study, we defined DSC ≥ 0.7 and MASD ≤ 5 mm as showing good agreement in volume delineation [64]. Dosimetry was reported as mean values (μ) for the 12 observers with SD (σ). To compare the variability in dosimetry, rSD = σ/μ.

Two 8 Gy single fraction plans optimised to the consensus and STAPLE contours, respectively, were generated for each patient dataset by a radiation therapist and checked by a radiation oncologist. Standard applicator loadings were used to generate a conventional plan dosed to Point A. This was optimised to the consensus contours with the aim of ensuring a minimum of 8 Gy to CTV_{HR} dose to 90% of the volume (D90). The OAR tolerances were based on combined external beam (45 Gy/25 fractions) and brachytherapy doses, limiting bladder equi-effective dose (EQD2) to 2 cm^3 of the volume ($D2cm^3$) to $\leq$ 90 Gy and rectum and sigmoid $EQD2_3$ $D2cm^3 \leq 75$ Gy. This equated to bladder $D2cm^3 \leq 5.91$ Gy and rectum and sigmoid $D2cm^3 \leq 5.34$ Gy for this single fraction plan. If the OAR (bladder, rectum, sigmoid) DVHs exceeded 110% of acceptable dose, then a lower dose to the CTV_{HR} was accepted. This plan was subsequently applied to the STAPLE contour and optimised using the same criteria. For each case, D90 and dose to 100% of the volume (D100) coverage of individual participant's CTV_{HR} by the consensus and STAPLE plans was assessed.

Table 3.1: Mean $\pm$ SD of CTV_{HR} volumes, MASD and DSC.

Case	Stage	CTV_{HR}	STAPLE reference (mean)		Consensus reference (mean)		STAPLE vs Consensus	
		Volume (mm)	MASD (mm)	DSC	MASD (mm)	DSC	MASD (mm)	DSC
1	T3bN0M0	60.3 ± 19.2	3.56 ± 1.92	0.71 ± 0.15	2.80 ± 1.45	0.72 ± 0.09	3.41	0.72
2	T2bN1M0	37.5 ± 15.6	2.29 ± 1.39	0.79 ± 0.08	2.45 ± 1.94	0.76 ± 0.11	2.18	0.77
3	T2bN1M0	55.7 ± 14.7	4.54 ± 2.72	0.70 ± 0.18	5.22 ± 3.43	0.65 ± 0.12	7.39	0.58
4	T3bN0M0	63.9 ± 35.3	11.01 ± 6.38	0.53 ± 0.23	9.82 ± 7.48	0.47 ± 0.15	21.1	0.25
5	T3bN1M0	39.5 ± 21.4	3.29 ± 1.96	0.71 ± 0.11	4.05 ± 3.72	0.69 ± 0.14	4.90	0.64
6	T1b1N1M0	30.1 ± 9.5	3.75 ± 1.61	0.74 ± 0.11	3.87 ± 3.03	0.74 ± 0.14	5.29	0.62
7	T4N0M1	142.3 ± 71.6	13.30 ± 13.14	0.58 ± 0.30	13.72 ± 13.04	0.51 ± 0.30	16.32	0.57
8	T3bN1M0	100.6 ± 22.1	2.47 ± 0.88	0.85 ± 0.05	2.61 ± 1.17	0.81 ± 0.05	3.01	0.78
9	T2aN0M0	14.9 ± 4.2	1.95 ± 0.65	0.78 ± 0.08	1.84 ± 0.95	0.77 ± 0.07	3.27	0.67
10	T1b2N1M0	30.1 ± 9.5	3.12 ± 1.46	0.79 ± 0.07	3.02 ± 2.28	0.79 ± 0.09	4.43	0.74
Mean			4.93 ± 3.21	0.72 ± 0.14	4.94 ± 3.85	0.69 ± 0.12	7.13 ± 6.37	0.63 ± 0.15

3.3 Results

The stage distribution of patients included two patients with International Federation of Gynaecology and Obstetrics (FIGO) stage IB cancers, one FIGO stage IIA, two FIGO stage IIB, four FIGO Stage IIIB and one FIGO stage IVA (Table 3.1). Good responses to external beam radiotherapy noted at the time of examination for brachytherapy were documented for all patients except for patients 7 and 8. Two patients (1 and 9) were treated with a tandem and cylinder due to lower vaginal extension of cancer or inability to fit ovoids. The remainder were treated with tandem and ovoid applicators. The range of volumes contoured by the observers is shown in Figure 3.1. The largest variation volumetrically was seen for patient 7 who had a FIGO IVA cancer extending into adjacent bowel and the smallest variation for patient 9 who had a FIGO IIA cancer. There did not appear to be any relationship between the range of volumetric variation and initial tumour stage or response to external beam radiation therapy. One observer (observer A) contoured the largest or second largest volume in 6/10 cases and another (observer E) the smallest or second smallest volume in 7/10 cases (Figure 3.1). Visually the greatest variation in contouring occurred in the cranial and caudal directions, even in patients where delineation was more consistent. Representative examples for patients with low and high concordance indices are shown in Figures 3.2 and 3.3.

A mean DSC ≥ 0.7 was achieved in 8/10 cases when compared to the STAPLE reference and 6/10 cases when compared to the consensus reference (Table 3.1). A mean MASD ≤ 5 mm was recorded in 8/10 cases when compared to the STAPLE reference and 7/10 cases when compared to the consensus reference (Table 3.1). No systematic differences were noted between radiation oncologists/therapist and radiologists in terms of DSC and MASD. The concordance between the STAPLE and consensus reference volumes is also shown. Only four patients had DSC ≥ 0.7 and six patients MASD ≤ 5 mm.

For individual patients, the percentage of participants achieving DSC ≥ 0.7 was 25% and 8% for patient 4, 58% and 33% for patient 3, and 42% and 42% for patient 8, compared to

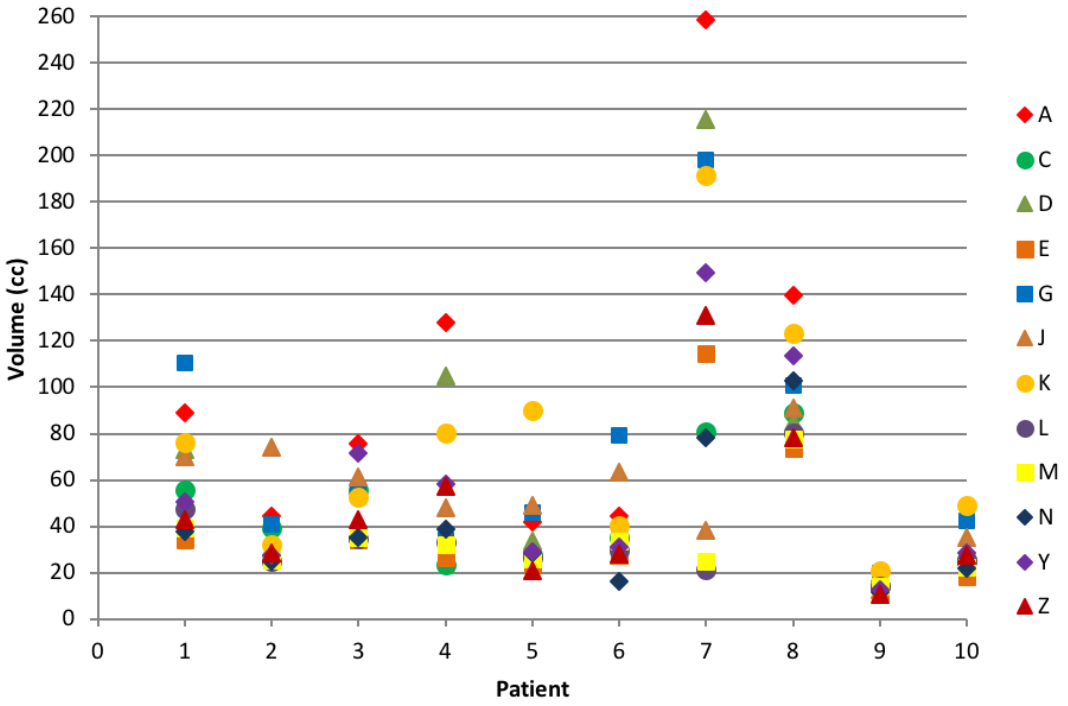


Figure 3.1: CTV_{HR} volume in cubic centimeters (cc) delineated by observers.

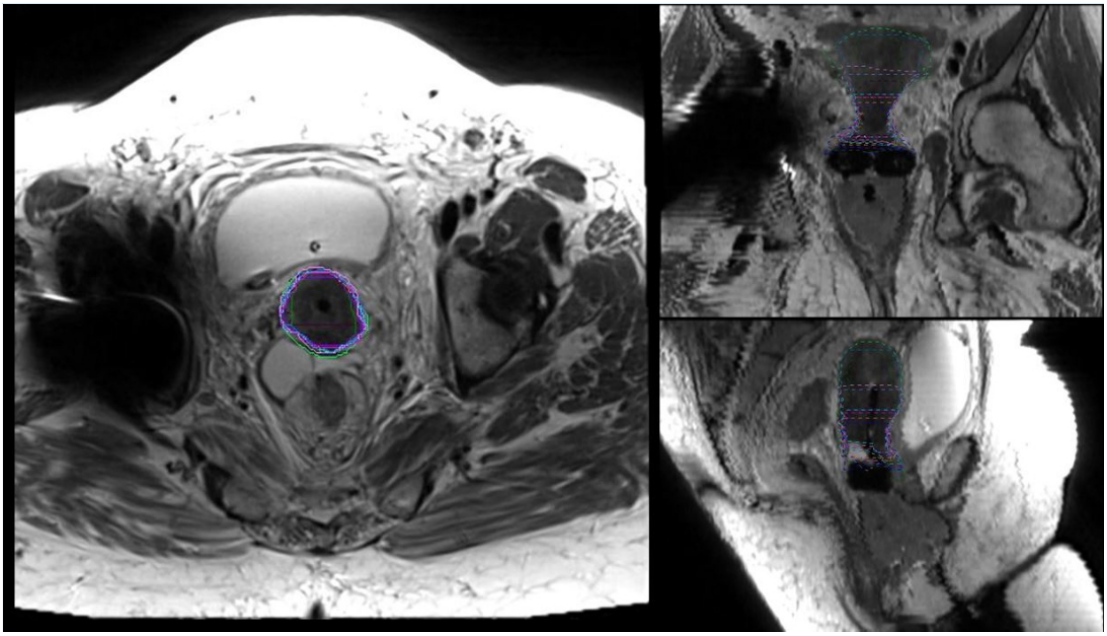


Figure 3.2: CTV_{HR} contours on Patient 4 where there was low concordance in delineation. The DSC was 0.47 and 0.53 and the MASD was 9.8 mm and 11.0 mm with reference to the consensus and STAPLE contours respectively.

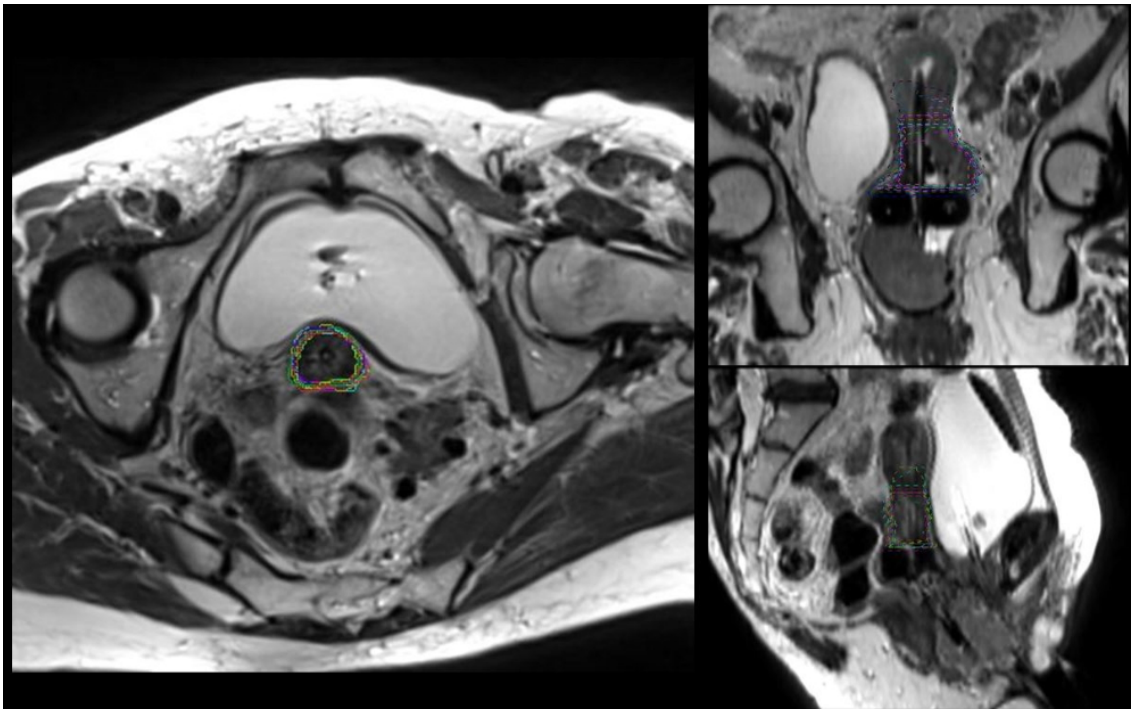


Figure 3.3: CTV_{HR} contours on Patient 9 where there was high concordance in delineation. The DSC was 0.77 and 0.78 and the MASD was 1.84 mm and 1.95 mm with reference to the consensus and STAPLE contours respectively.

the STAPLE and consensus references, respectively. For the remainder of the patients, the percentage achieving DSC $\geq$ 0.7 was 50% or greater. For patient 9, 100% of participants achieved DSC $\geq$ 0.7 compared to both reference volumes. A similar pattern was seen with MASD. Considering each individual contour (n = 120), a DSC $\geq$ 0.7 was achieved in 67% and 63% and MASD $\leq$ 5 mm achieved in 75% and 73% compared to the STAPLE and consensus references, respectively.

A consensus and STAPLE plan were optimised for each patient. The planning aim of CTV_{HR} D90 = 8 Gy was met in five patients for the consensus plan and two patients for the STAPLE plan (Table 3.2).

It was not achieved in the other patients due to unacceptably high doses to organs-at-risk and large CTV_{HR}s which could not be covered by non-interstitial brachytherapy. When the consensus and STAPLE plans were applied to individual participant contours, dose variation in D90 and D100 was seen. The smallest range of D90 ($\leq$ 3 Gy) was seen for patients 1, 9 and 10, whilst the largest range ($>$ 7 Gy) was seen in patients 4 and 7 (Figure

Table 3.2: Mean ± SD of CTV_{HR} D90. The dose achieved refers to the CTV_{HR} D90 dose achieved in the STAPLE and consensus plans, which were then applied to individual observer contours.

Case	STAPLE reference (Gy)		CONSENSUS reference (Gy)	
	Dose Achieved	CTV_{HR} D90	Dose Achieved	CTV_{HR} D90
1	4.05	4.15 ± 0.54	3.91	4.18 ± 0.59
2	9.52	9.83 ± 1.30	10.92	10.37 ± 1.36
3	4.68	6.37 ± 1.36	6.28	6.27 ± 1.56
4	2.66	7.38 ± 3.26	8.45	6.16 ± 2.81
5	6.64	7.91 ± 1.90	8.92	7.04 ± 1.60
6	6.28	7.74 ± 1.86	8.42	7.74 ± 1.86
7	1.17	3.07 ± 2.85	1.38	3.02 ± 3.11
8	5.59	6.11 ± 1.09	6.15	6.01 ± 1.03
9	8.03	8.64 ± 0.57	8.63	8.37 ± 0.63
10	7.18	7.34 ± 0.83	7.68	7.12 ± 0.85
Mean	5.58	6.84 ± 1.56	7.07	6.62 ± 1.54

3.4).

There was less variation in range of dose seen for D100. The smallest range of D100 ($<$ 2 Gy) was seen in patient 1 and the largest range ($>$ 4 Gy) in patients 2, 4, 5 and 6 (Figure 3.5). If we consider just the dose range for the middle 50% of participants, the greatest range for both D90 and D100 was seen in patient 4 and the least range in D90 for patients 1, 9 and 10 and D100 patient 1 and 7 for both STAPLE and consensus plans. Significant outlier doses were seen in D90 for patients 1, 7, 8, 9 and 10 and in D100 for patients 2, 7 and 10. The outlying observers were different for each patient.

The mean rSD of doses when STAPLE and consensus plans were applied to individual observer contours is shown in Figure 3.6. In most cases the mean rSD for CTV_{HR} D90 was less than 25% and D100 less than 35%. Values were greater for patients 4 and 7. The average mean rSD across all patients was 27% and 34% for the STAPLE CTV_{HR} D90 and D100, respectively, and 28% and 35% for the consensus CTV_{HR} D90 and D100. Delineation uncertainty resulted in an average dosimetric uncertainty of ±1.6 Gy per fraction for the STAPLE plan and ±1.5 Gy per fraction for the consensus plan based on an 8 Gy prescribed fraction (Table 3.2).

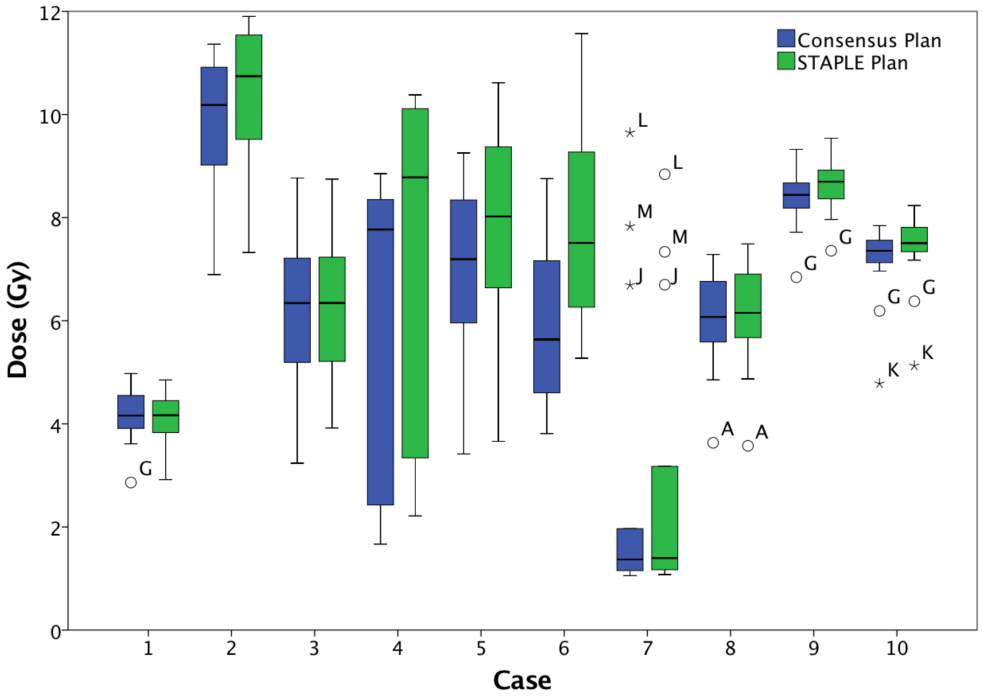


Figure 3.4: CTV_{HR} D90 doses for individual contours with application of STAPLE and consensus plans.

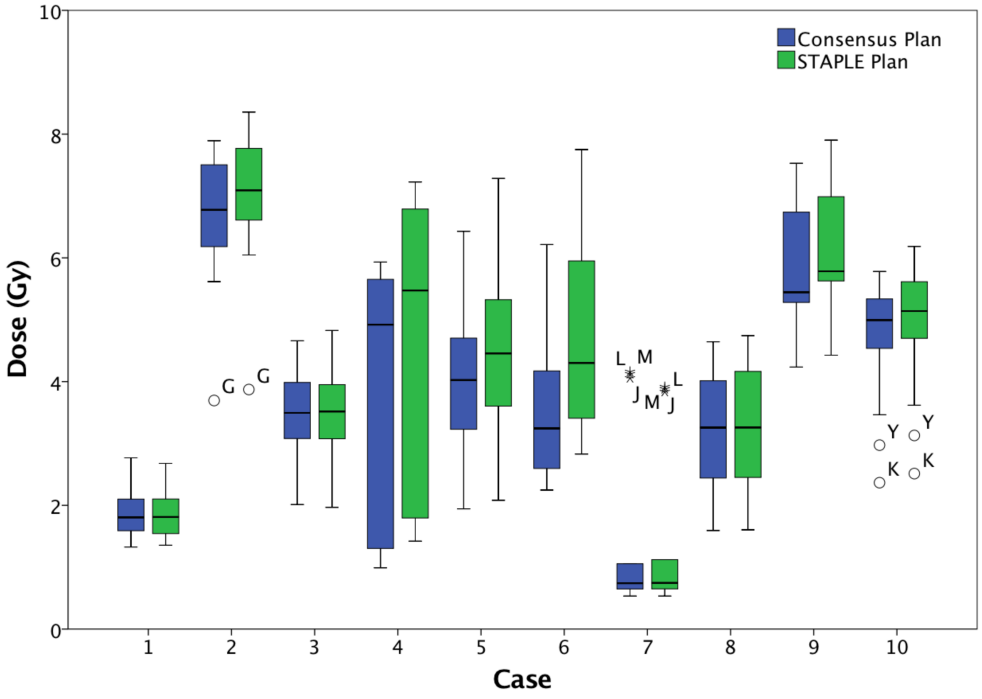


Figure 3.5: CTV_{HR} D100 doses for individual contours with application of STAPLE and consensus plans.

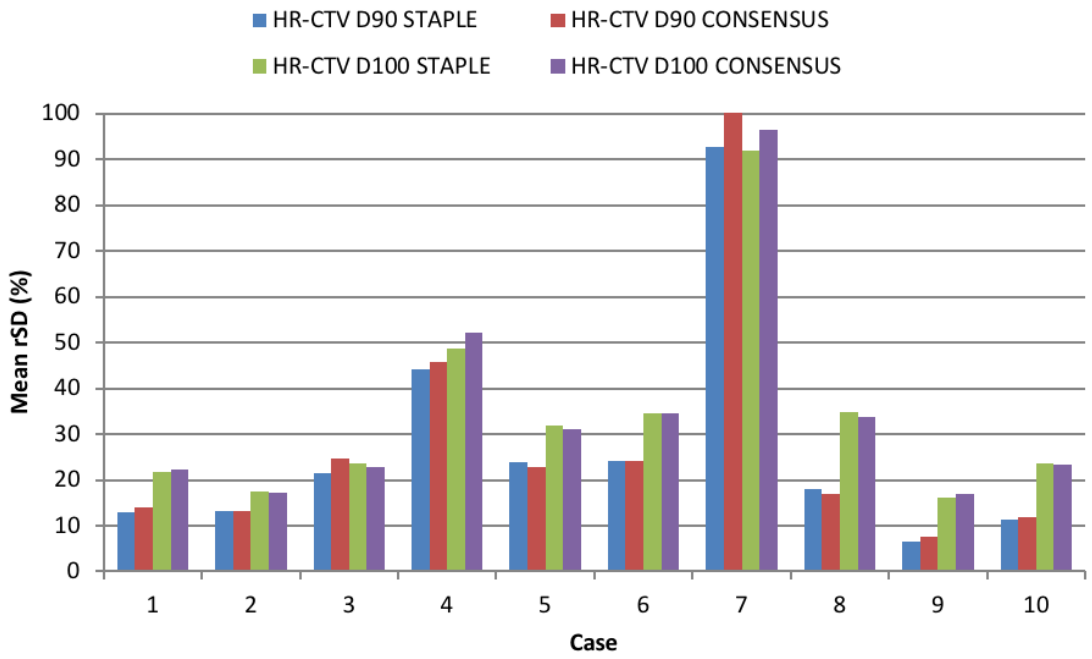


Figure 3.6: Mean rSD of doses when STAPLE and consensus plans where applied to individual observer contours.

3.4 Discussion

This study measured inter-observer variability in target volume delineation and resultant dosimetric effects for cervical brachytherapy in the Australian and New Zealand setting. We used unselected sequentially acquired patient datasets at a single centre. This is only the second study using 3-Tesla MRI to evaluate contouring variability for cervix brachytherapy, the previous one being from USA [66].

We found that in the majority of cases, delineation of CTV_{HR} was consistent, as defined by DSC ≥ 0.7 and MASD ≤ 5 mm. This is in contrast to Viswanathan et al [66] who evaluated delineation of CTV_{HR} on 3-Tesla MRI and CT for three cases using 23 observers. CIs were poor ranging 0.44 - 0.48 for CT and 0.38 - 0.42 for MRI, and were significantly higher for CT than MRI. This may be due to the larger number of observers in that study. In the current study the comparison metrics were acceptable in a slightly greater proportion of patients in reference to the STAPLE reference, as this volume is a probabilistic estimate of true volume generated from all observers. The consensus

reference was drawn by collaboration between two experienced radiation oncologists and a radiologist specialising in pelvic MRI.

Other studies published by the GEC-ESTRO group have used lower strength MRIs for imaging, either 0.2-Tesla or 1.5-Tesla [64, 65]. Dimopoulos et al. evaluated inter-observer variation in delineation of GTV, CTV_{HR} and CTV_{IR} in 19 patients with two observers [65]. The CI for GTV, CTV_{HR} and CTV_{IR} were 0.5, 0.6 and 0.6, respectively, showing moderate variation even amongst two experienced radiation oncologists working at centres which developed the GEC-ESTRO guidelines.

Petric et al [64] conducted a study involving observers from 10 institutions worldwide using MRI datasets of six patients. Two reference contours were used for comparisons, a STAPLE contour and an expert consensus contour. The mean CI for CTV_{HR} delineation was 0.76 for the STAPLE reference and 0.72 for the consensus reference which is slightly higher than the respective indices of 0.72 and 0.69 measured in the current study.

The choice of a reference volume can influence the results of both volume comparisons and dosimetric evaluation. The gold standard of a pathologically defined volume is impossible in this context hence surrogate volumes are chosen for reference. The choice of both an expert consensus volume and STAPLE volume is common. Our study shows that concordance of these volumes with each other is in fact moderate at best and similar to comparisons between two experienced GEC-ESTRO clinicians [65].

Visually, we observed that the greatest variation in delineation was in the cranio-caudal direction. This is similar to findings reported by Dimpoulos et al [65] and Petric et al [64]. We also noted some systematic differences in contouring with one observer consistently delineating larger volumes and one smaller volumes compared to the average.

The clinical impact of delineation variability can be evaluated by assessing the effect on dosimetry. Ideally a plan would be generated for each observers contour and applied to the consensus and STAPLE volumes to evaluate dose to target volumes and organs-at-risk. However, this is a considerably time consuming process.

Hence, the method chosen for dosimetric evaluation was identical to that published by Hellebust et al [101] whereby plans optimised to two reference contours, STAPLE and consensus were applied to individual observer contours. We found that the mean rSD varied from 27 to 28% for CTV_{HR} D90 and from 34 to 35% for CTV_{HR} D100. This is greater than that reported by Hellebust et al who found a mean rSD of 9-10% for CTV_{HR} D90 and 17-19% for CTV_{HR} D100 [101]. Dimopoulos et al [65] also evaluated dosimetry for plans based on contours by two observers and found no significant differences in the D90 or D100 for any of the volumes.

The average measured dosimetric variation was 1.5-1.6 Gy for a single brachytherapy insertion. The range of dosimetric variability was in most cases similar for the STAPLE and consensus plans. This may be a consequence of the applicators used which provided limited opportunities for dose optimisation for larger tumours due to lack of interstitial needles. Inter-patient variability was noted for whether planning aims were met. Dose variation was independent of initial disease stage and tumour response seen at the time of brachytherapy.

Despite acceptable volume concordance metrics for the majority of patients in this study, relative dose variation could be clinically significant. The effect of this uncertainty will increase with multiple fractions of brachytherapy. If plans are optimised per observer to improve this, one would expect the range of OAR doses to vary widely. This was not measured in this study as the organs-at-risk were delineated by a sole radiation therapist. Other studies have shown that rSD for $D2cm^3$ doses varies from 5 to 16% for bladder, 7 to 11% for rectum and 11 to 34% for sigmoid with larger variation being reported for delineation on CT [100–103]. Organ-at-risk volumes are often delineated by radiation therapists and should be checked by radiation oncologists prior to plan approval.

This study has some limitations. Of note there were no interstitial applicators used, and in two cases a tandem and cylinder were inserted and so dose coverage was compromised in some instances. The patients were unselected and so represent a typical clinical cohort

in whom the planning aims were not always met. There were only 10 patient datasets used and only one volume delineated (CTV_{HR}). However, both the number of datasets and observers is similar to other published studies. The choice of a single volume was to prevent contourer fatigue. The assessment of variation in dosimetry remains valid even if planning aims cannot be met.

Brachytherapy for cervical cancer remains a small practice in most Australian and New Zealand radiation oncology departments. Given the variation documented, it may be useful to involve a peer radiation oncologist or radiologist to aid in target volume delineation. Accurate volume delineation is crucial to the practice of brachytherapy and this may potentially suffer if performed infrequently. Treatment at low volume centres ($<$ 10 patients per year) can result in poor quality brachytherapy [135]. The Gynaecology Oncology Radiation Oncology Collaboration specialty interest group of the Royal Australian and New Zealand College of Radiologists will have a role in ensuring quality brachytherapy treatment for cervical cancer in the future.

Chapter 4

Dose Planning Variations Related to Delineation Uncertainties in MRI-Guided Brachytherapy for Locally Advanced Cervical Cancer

This chapter assesses the dosimetric variation that arises from inter-observer variation in HDR cervical brachytherapy. This chapter expands on Chapter 3 by investigating the dosimetric differences arising when individual plans are optimised to individual observer contours. The findings in this chapter contribute to answering the first aim of this thesis: to understand the impact of delineation uncertainty on radiotherapy efficacy (Chapter 1.1).

The content in this chapter is in preparation for publication. The candidate's contribution to this publication is in the form of data generation, collection and analysis as well as authorship of manuscript.

4.1 Introduction

Image-guided adaptive brachytherapy (IGABT) is the standard-of-care in the treatment of locally advanced cervical cancer, with MRI the imaging modality of choice. The excellent patient outcomes of IGABT for cervical cancer has been demonstrated in the EMBRACE studies (Retro-EMBRACE, EMBRACE and EMBRACE II) [136–140]. These studies are based on GEC-ESTRO guidelines on contouring and dose reporting [130, 131]. However, uncertainty in target volume delineation remains, regardless of adherence to protocols. The location and magnitude of delineation uncertainty has been extensively investigated for a number of anatomical sites in both external beam radiotherapy and brachytherapy contexts [21, 97, 107, 141]. The sharp dose gradients that characterise brachytherapy may result in an increased dosimetric impact of delineation uncertainty including the risk of geographical miss.

Hellebust et al investigated the dose reporting uncertainties due to contouring variation, when a single 'gold standard' cervical cancer brachytherapy plan was assessed on multiple contour sets delineated by 10 observers [101]. The observers were from centres participating in the EMBRACE study [64, 137]. A reported dosimetric difference of $\pm$ 5 Gy $EQD2_{10}$ was seen for the high risk clinical target volume and between $\pm$ 2-3 Gy $EQD2_3$ for the organs at risk (OARs).

The present study utilises the same gynaecological MRI brachytherapy dataset from Hellebust et al [101], but investigates the variation in dose planning that arises from target delineation uncertainties when individual plans are optimised to each individual contour set. As an additional aim, this study investigates whether there are correlations between contour variation metrics and dosimetric outcomes when plans are optimised to each individual contour [35].

4.2 Methods

4.2.1 Dataset

The material used for this analysis is from a study on contouring uncertainties performed in the frame of EMBRACE, completed and published in 2013 [64, 137]. The dataset consisted of MR images from six patients treated for FIGO stage IIb-IIIb cervical cancer. The patients were selected to represent common clinical scenarios and to reflect typical challenges relevant for contouring.

4.2.2 Target Volume Definition

Images were distributed to institutions that participated in the EMBRACE study and hence with considerable experience in MRI-based cervical brachytherapy. One observer from each participating institution delineated the target volumes and relevant OARs according to the GEC-ESTRO recommendations [130, 131]. Specific volumes contoured included the residual GTV, CTV_{IR}, CTV_{HR}, bladder, rectum and sigmoid.

A group of experts including four experienced radiation oncologists generated a gold standard expert consensus (EC) set of contours (targets and OARs). Additionally, the STAPLE was generated using in-house software [142] to provide a probabilistic estimate of the 'true' volume from the observer contours.

4.2.3 Treatment Planning

This study was performed retrospectively and is concerned with the brachytherapy component of a virtual treatment course consisting of combined external beam radiotherapy (45 Gy in 25 fractions) and four brachytherapy fractions of 7.2 Gy each (85.5 Gy $EQD2_{10}$ in total). The brachytherapy component comprised treatment via a ring and tandem applicator with interstitial needles inserted at appropriate

Table 4.1: Planning constraints as per EMBRACE II protocol [64].

	D90 CTV_{HR} $EQD2_{10}$	D98 CTV_{HR} $EQD2_{10}$	D98 GTV $EQD2_{10}$	D98 CTV_{IR} $EQD2_{10}$	D2cm^3 Bladder $EQD2_3$	D2cm^3 Sigmoid $EQD2_3$	D2cm^3 Rectum $EQD2_3$	Recto-vaginal point $EQD2_3$
Soft constraints	$>$ 90 Gy $<$ 95 Gy	$>$ 75 Gy	$>$ 95 Gy	$>$ 60 Gy	$<$ 80 Gy	$<$ 70 Gy	$<$ 65 Gy	$<$ 65 Gy
Hard constraints	$>$ 85.5 Gy	-	$>$ 90 Gy	-	$<$ 90 Gy	$<$ 75 Gy	$<$ 75 Gy	$<$ 75 Gy

locations for the specific disease. Applicator specifications for each patient are presented in Table A.1.

For each observers set of contours, a treatment plan was individually optimised resulting in 12 plans per patient (10 observer contours, EC contour and STAPLE contour). Each patient had a common applicator reconstruction for each plan. Plan optimisation was achieved by the same physicist with review by an oncologist specialising in cervical cancer treatment. The constraints used during treatment plan optimisation (soft and hard) were in line with the EMBRACE II protocol [137, 143] as outlined in Table 4.1. As a soft constraint, needle loading was minimised to less than 20% of the total dwell-weight. An additional focus was placed on de-escalating dose to the upper vagina as per EMBRACE II protocols and emerging evidence [137, 144–146]. The ratio of the vaginal total reference air kerma (TRAK) to the total TRAK was aimed, where possible, to be kept below 30-40%.

4.2.4 Inter-observer Variation Analysis

The inter-observer variation of the contours used in this study and the impact on reported dose arising from a gold standard plan has been assessed previously [64, 101]. This study aims to investigate the planning variation due to inter-observer delineation variation by optimising plans to specific delineations. The fundamental methodology difference of this study and the previous [101] is outlined in Figure 4.1. In the previous study, contouring variation resulted in reporting uncertainties when a single 'gold standard' plan is applied to all contour sets. In the current study, firstly, plans were optimised to each individual contour set, representing the variation in dose that would have been reported. All together 72 plans were made (10 contours sets plus EC and STAPLE for 6 patients). Secondly, these plans were applied to the EC and STAPLE contour sets and the variation in the dose that would have been delivered were found.

Treatment plans and DVH data were analysed using an in-house MATLAB code [147].

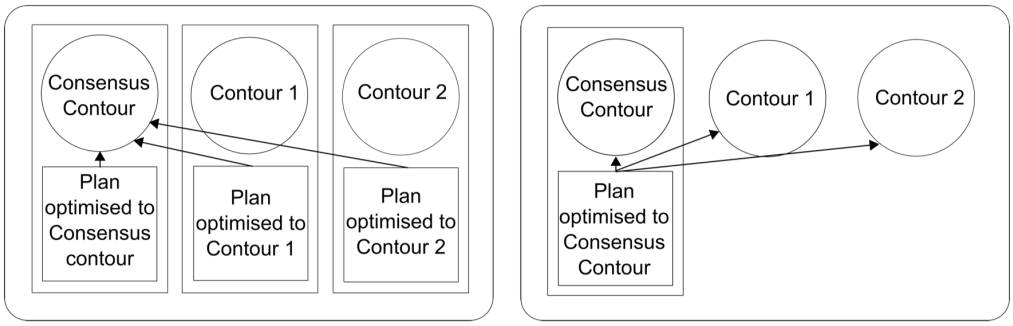


Figure 4.1: Methodology difference between this study (left) examining uncertainties in planned dose due to delineation uncertainties and Hellebust et al [101] (right) examining dose reporting variations due to delineation uncertainties. The arrows point to the contours on which the dosimetric parameters were calculated.

For a single fraction, D90 and the dose to 98% of the CTV_{HR} (D98) were determined for each optimised plan [148]. CTV_{IR} and GTV D98 were also determined for each plan. $D2cm^3$ of the volume was determined for the EC and STAPLE OARs (bladder, rectum and sigmoid) in the same fashion.

Incorporating the brachytherapy and external beam components of treatment, the total EQD2 was calculated for each of the DVH parameters with α/β = 10 Gy for target volumes and α/β = 3 Gy for OARs. For each DVH parameter, the mean and SD were determined. To assess the impact of inter-observer uncertainty over a typical brachytherapy course the rSD relative to the total brachytherapy dose was determined.

4.2.5 Correlation Analysis

The correlation between contouring metrics and dosimetric parameters was assessed to investigate whether contouring decisions have an impact on the dosimetric outcome. Contouring metrics as calculated in reference to both the STAPLE and consensus volumes include volume, shift in COM, CI [149] and maximum directions in the cardinal directions (MD-X, MD-Y, MD-Z where X is lateral, Y is anterior/posterior and Z is superior/inferior). Dosimetric parameters include the DVH parameters outlined

above. The two-sided Spearman's nonparametric rank-correlation coefficient (ρ) was determined for all metrics with significance indicated by $p < 0.05$ using IBM SPSS Statistics v19 [150].

4.3 Results

For details on the delineations, including an assessment of inter-observer variation, the reader is referred to Petric et al [64]. One of the plans generated was not assessed as it was incorrectly exported and subsequently inaccessible. The following presents the results of the remaining 71 plans.

4.3.1 Treatment Planning

Technical details of the generated treatment plans are presented in Table 4.2. Prescribed CTV_{HR} D90 for each plan based on the individual contours are presented in Table 4.3. The other resulting DVH parameters are listed in Table A.2. As described above, these figures represent the dose parameters that would have been *reported* due to contouring variation. 80.0% (57 out of 71) plans fulfilled the hard constraint of CTV_{HR} D90 $>$ 85.5 Gy $EQD2_{10}$). Furthermore, 45.0% (32 out of 71) plans fulfilled the soft constraint of $CTV_{HR} > 90$ Gy $EQD2_{10}$. Of the plans that did not achieve the hard constraints, 7 of these occurred for a single patient in which the proximity of the OARs in relation to the high dose region was especially challenging. Common planning problems arose when trying to achieve the soft constraint to the CTV_{IR} D98 without exceeding the bladder tolerance. Although no plan exceeded the hard constraint for the bladder, the soft constraint was exceeded in 57.7% (41 out of 71) of plans.

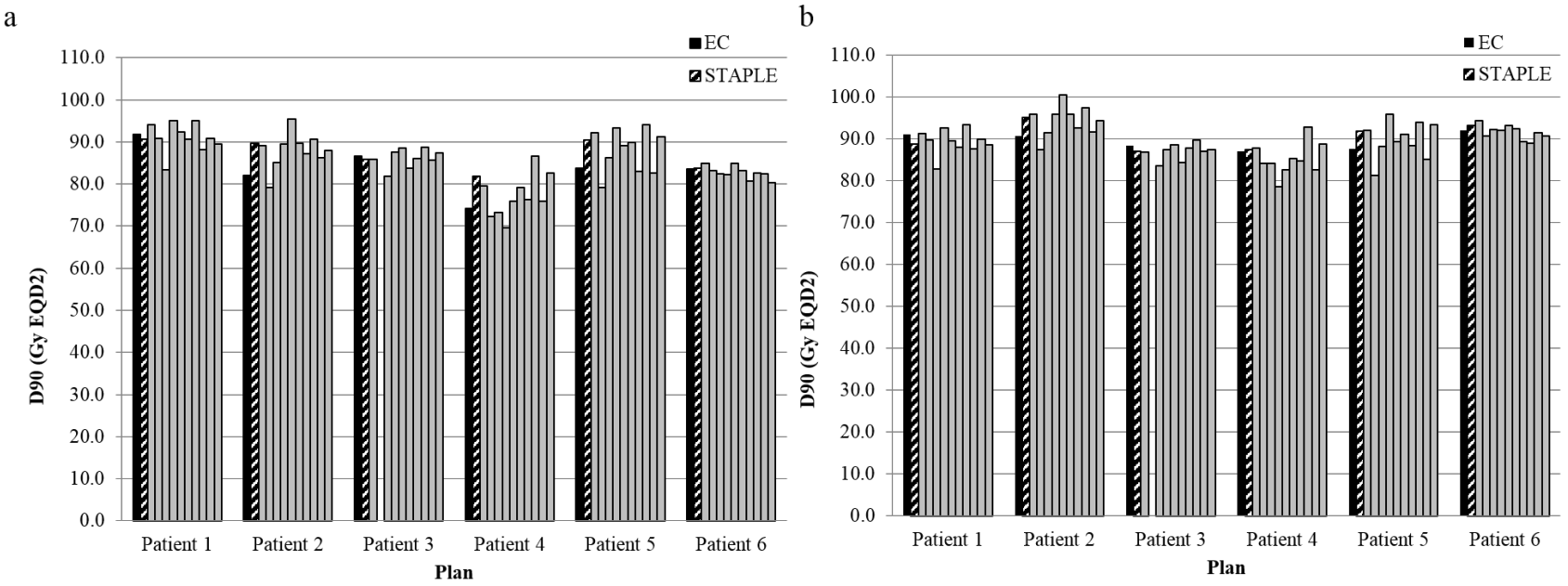
Table 4.2: Technical details of generated treatment plans. These values are for a decommissioned source therefore do not reflect a clinical treatment. This table is provided to enable relative difference comparisons between plans

Patient		Total Reference Air Kerma (cGy)	Total Dwell Time (s)	Vaginal Loading (%)
1	Mean	1.93	424.75	21.28
	SD	0.09	20.60	5.31
2	Mean	1.24	1459.48	33.37
	SD	0.09	107.08	5.36
3	Mean	1.54	1804.73	31.26
	SD	0.14	162.30	10.36
4	Mean	1.68	1720.98	22.87
	SD	0.09	608.40	2.99
5	Mean	1.81	2120.05	34.82
	SD	0.11	134.53	3.65
6	Mean	2.02	2374.94	23.27
	SD	0.08	90.21	3.67

Table 4.3: Prescribed CTV_{HR} D90 (Gy) for plans optimised to each observer contour. Italic font for the plans not fulfilling the hard constraint.

Patient	1	2	3	4	5	6
Master	91.0	90.6	88.3	87.1	87.6	92.1
STAPLE	90.6	89.8	85.8	*81.9*	90.5	*83.8*
Observer 1	90.6	89.3	86.3	*81.9*	91.0	85.9
Observer 2	90.1	96.6	*80.0*	85.6	*84.8*	*79.9*
Observer 3	91.5	92.3	-	88.9	92.0	85.6
Observer 4	94.2	93.6	89.8	86.8	93.6	85.9
Observer 5	89.9	89.4	*78.9*	85.9	90.5	*67.3*
Observer 6	89.7	*85.0*	85.7	*76.2*	91.0	87.4
Observer 7	90.0	90.2	86.0	*80.3*	90.5	87.4
Observer 8	94.7	92.4	92.3	*73.0*	95.0	85.5
Observer 9	94.5	92.5	91.6	*80.5*	91.0	86.8
Observer 10	93.6	90.6	90.0	*83.3*	90.3	86.0
Mean (Gy)	91.7	91.0	86.8	82.6	90.7	84.5
SD (Gy)	2.0	2.8	4.3	4.7	2.6	6.1
rSD (Gy)	4.1	6.0	10.2	12.2	5.6	15.1

Figure 4.2: CTV_{HR} D90 $EQD2_{10}$ for plans optimised to different contours applied on (a) STAPLE contour and (b) expert consensus contour. The order of observers is identical across both panels. The rSD is for the brachytherapy fractions only.



Mean (Gy)	91.0	87.6	86.2	77.3	87.9	82.9	Mean (Gy)	89.3	94.0	87.0	85.4	89.8	91.7
SD (Gy)	3.2	4.2	2.0	4.9	4.8	1.4	SD (Gy)	2.7	3.5	1.8	3.6	4.1	1.6
Rel. SD (%)	6.9	9.7	4.8	14.7	11.0	3.7	Rel. SD (%)	6.0	7.0	4.1	8.7	8.9	3.3

Table 4.4: Mean, SD and rSD of EQD2, averaged across all observers for each patient. The rSD is reported as percentages of the brachytherapy treatment only.

		STAPLE							EC						
	Case	1	2	3	4	5	6	Mean	1	2	3	4	5	6	Mean
CTVHR	Mean (Gy)	91.0	87.6	86.2	77.3	87.9	82.9	85.5	89.3	94.0	87.0	85.4	89.8	91.7	89.5
D90	SD (Gy)	3.2	4.2	2.0	4.9	4.8	1.4	3.4	2.7	3.5	1.8	3.6	4.1	1.6	2.9
	rSD (%)	6.9	9.7	4.8	14.7	11.0	3.7	8.5	6.0	7.0	4.1	8.8	8.9	3.3	6.4
CTVHR	Mean (Gy)	82.7	78.5	76.8	65.6	78.0	71.8	75.6	81.6	82.9	78.4	73.3	80.2	77.8	79.0
D98	SD (Gy)	2.9	4.1	1.8	3.5	5.9	1.7	3.3	2.4	2.8	1.6	2.6	4.7	2	2.7
	rSD (Gy)	7.5	12.1	5.5	16.5	17.5	6.2	10.9	6.5	7.2	4.6	9.1	13.2	5.9	7.8
CTVIR	Mean (Gy)	63.1	60.0	58.2	55.0	59.5	57.8	58.9	62.7	61.5	60.0	56.7	61.1	61.8	60.6
D98	SD (Gy)	1.7	1.8	0.9	1.5	2.3	0.7	1.5	1.5	1.5	1.1	1.1	2.1	0.8	1.4
	rSD (%)	8.9	11.3	6.2	13.5	15.4	4.9	10.0	8.3	9.0	7.1	9.1	12.2	4.5	8.4
GTV	Mean (Gy)	92.3	91.5	98.8	86.8	90.7	83.4	90.6	97.8	94.8	98.4	115.7	107.0	84.3	99.7
D98	SD (Gy)	2.8	6.6	2.7	3.3	3.5	2.7	3.6	2	3.1	2.6	5	4.3	3	3.3
	rSD (%)	5.7	13.9	4.9	7.7	7.4	6.8	7.7	3.7	6.0	4.8	7.0	6.8	7.5	6.0
Bladder	Mean (Gy)	78.0	82.9	79.0	88.7	82.0	87.0	83.0	77.8	80.2	75.0	85.3	82.7	86.5	81.2
$D2cm^3$	SD (Gy)	1.7	2.7	5.0	4.7	4.9	2.7	3.6	1.8	2.7	4.1	4.4	4.9	2.7	3.4
	rSD (%)	4.9	6.9	13.9	10.4	12.7	6.3	9.2	5.2	7.4	13.0	10.6	12.5	6.2	9.1
Rectum	Mean (Gy)	55.0	62.9	55.9	68.8	60.4	69.6	62.1	54.1	62.4	54.5	66.5	60.3	66.7	60.7
$D2cm^3$	SD (Gy)	1.5	2.9	1.9	1.7	2.2	2.4	2.1	1.4	2.7	1.7	1.5	2.2	2.0	1.9
	rSD (%)	13.0	14.7	14.8	6.6	12.8	9.2	11.9	12.7	14.2	15.2	6.4	12.9	8.7	11.7
Sigmoid	Mean (Gy)	70.9	50.6	72.2	55.9	70.0	67.6	64.5	68.2	50.4	68.6	55.1	66.6	65.9	62.5
$D2cm^3$	SD (Gy)	1.8	0.6	1.2	1.1	3.0	1.4	1.5	1.6	0.5	1.1	1.0	2.2	1.1	1.2
	rSD (%)	6.4	7.6	4.3	8.7	11.2	5.6	7.3	6.2	7.0	4.3	8.7	9.4	4.8	6.7

4.3.2 Inter-observer Variation Analysis

When individual contour based plans were applied to the gold standard contours (EC or STAPLE), the dosimetric variability can be seen in Figure 4.2. Tables 4.4 and A.3 provide further details for all contours and dosimetric parameters. The rSD in STAPLE CTV_{HR} D90 across patients ranged between 3.7 - 15.1% and for the EC CTV_{HR} D90, ranged between 3.4 - 9.0%. The rSD represents the dosimetric variability of the brachytherapy component of the treatment course only. This gives a representation of the variable of the *delivered* dose that may arise clinically depending on the clinician who delineates. The dosimetric variability depends on the gold standard used for analysis, with a greater range in variability observed for the STAPLE contours.

4.3.3 Correlation Analysis

Table 4.5 shows the Spearmans non parametric rank correlation coefficient and significance, relating the dosimetric and contouring parameters of the CTV_{HR}. The statistics for the other target contours and the OARs are shown in Tables A.4 and A.5. In summary, maximum dimensions in the cardinal directions correlate most frequently with dosimetric changes.

4.4 Discussion

For a single fraction of HDR intra-cavity cervix IGABT, the total dosimetric uncertainty in CTV_{HR} D90 can reach 12% [151, 152]. Factors that contribute to this dosimetric uncertainty include source strength specification (2%), treatment planning (3%), medium dosimetric corrections (1%), dose delivery (4%) and inter-fraction/intra-fraction changes (11%). Source strength specification impacts all levels of treatment planning requiring the methodology to be well documented and uniformly agreed upon in the brachytherapy community. Treatment planning uncertainties arise when adjusting input dosimetry

Table 4.5: Spearmans non-parametric correlation coefficient results for the EC and STAPLE CTV_{HR} contour. Statistically significant results shown in bold italics.

	x COM		y COM		z COM		COM		MD-X		MD-Y		MD-Z		Vol		CI	
	ρ	sig	ρ	sig	ρ	sig	ρ	sig	ρ	sig	ρ	sig	ρ	sig	ρ	sig	ρ	sig
EC D90	0.21	0.08	-0.01	0.94	0.09	0.46	0.03	0.82	-0.19	0.11	-0.07	0.59	0.05	0.68	-0.03	0.83	-0.17	0.17
EC D98	-0.16	0.19	0.02	0.87	-0.14	0.25	-0.06	0.64	***-0.38***	***< 0.01***	***-0.39***	***< 0.01***	***-0.25***	***0.04***	***-0.32***	***0.01***	0.04	0.72
STAPLE D90	0.21	0.09	-0.2	0.87	0.09	0.47	0.14	0.26	-0.19	0.12	***-0.27***	***0.02***	-0.18	0.13	-0.16	0.17	0.21	0.09
STAPLE D98	0.2	0.1	-0.04	0.72	0.08	0.51	0.1	0.39	***-0.27***	***0.02***	***-0.34***	***< 0.01***	-0.23	0.06	***-0.25***	***0.04***	0.21	0.08

parameters derived under reference conditions, for the patient specific circumstances. Treatment planning uncertainties will vary with source type and position relative to the source and can be minimised with fractionation. Medium dosimetric corrections account for the deviation of the patient specific conditions from uniform water. This includes the presence of high density materials and tissue/air interfaces near the treating volume. Uncertainties are dependent on the position of heterogeneities relative to the source. Dose delivery uncertainties include deviations in positional accuracy of the source as defined by the applicator positioning and afterloader performance. The temporal accuracy of the source is also included in this category. Inter- and intra-fraction changes are the largest contributors to the total dosimetric uncertainty, mainly due to contouring uncertainty [151, 152]. It is therefore especially important to investigate the dosimetric impact of inter-observer contouring variation in cervical brachytherapy treatments. With sharp dose gradients abutting normal tissue, a small change in the dose distribution due to inaccurate contouring may have substantial implications for patient treatment.

This study investigated the differences in prescribed dose arising from planning variation when plans are optimised to individual contours. Hellebust et al investigated the variation in reported dose by assessing a gold standard plan on individual contours [101] and other studies have investigated other dosimetric aspects of contouring uncertainty in the gynaecological context [101, 151–154].

4.4.1 Inter-observer Variation Analysis

As can be seen from Figure 4.2, there is variability between plans, patients and what gold standard contour is used to assess the dosimetric variation. This is in contrast to Hellebust et al who determined using a different methodology that the difference in variability in rSD between gold standards is negligible [101]. Using the STAPLE contour for dosimetric assessment of the CTV_{HR} D90 generally results in slightly lower mean doses and higher variation (rSD) within a single patient than using the EC contour.

The same is true for most target contour parameters as can be seen in Table A.3. This can be explained by the geometry and position of the contours. The STAPLE contour is inherently similar to the observers contours on which each plan is optimised to, because it is derived from these contours. Therefore similar doses may be expected. The EC in this cohort is independent of each of the observer contours, and generally smaller in the lateral and anterior/posterior directions than the STAPLE, hence the observed increase in dose. Furthermore, since the EC lies within the STAPLE volume in these dimensions, dose distributions optimised to a STAPLE-like contour will result in less variability in D90 for the EC.

Depending on the dosimetric parameter chosen, substantial variability over the course of the multi-modality treatment is observed. For CTV_{HR} D90, STAPLE contours resulted in 3.4 Gy $EQD2_{10}$ (8.1% rSD) variability. EC contours had smaller variability in CTV_{HR} D90, at 2.9 Gy $EQD2_{10}$ (6.1% rSD). The OAR variability was 1.2-3.6 Gy $EQD2_3$ for $D2cm^3$. These values are reported for the entire brachytherapy treatment. The single fraction variability as presented in Table A.3, is 6.0% and 4.5% for the STAPLE and EC CTV_{HR} D90 respectively. Single fraction variability for the OARs ranges between 4.4-7.7%.

The dosimetric differences of these results and that of Hellebust et al [101], despite using the same contours and applicator reconstructions, can be explained by different plan optimisations methods as seen in Figure 4.1. In the current study, the dosimetry for each observer is constrained by the ability to conform dose to each contour. In the previous study, a single 'gold standard' plan was prescribed to all contour sets. Hence, the dosimetry varies amongst observers due to deviation in each observer contour relative to the gold standard contour. The two approaches can therefore have substantially different results despite utilising the same dataset, specifically resulting in a lower SD in CTV_{HR} for this study than Hellebust et al [101]. In addition the present analysis includes intra-observer variations for the planning process itself however this has been minimised

by using the same planner for all 71 plans. These methodology differences also explain why in the present study, the rSD in Rectum D_{2cm3} is higher than the rSD in Sigmoid D_{2cm3}; but the opposite is true for Hellebust et al [101].

Tanderup et al identified that target contouring uncertainties (9%) account for almost the entire intra- and inter-fraction uncertainty budget, however this is referring to the variability in reported dose due to contouring uncertainties as determined in Hellebust et al [101, 151]. The variability in prescribed dose due to planning variations as determined in the current study (4.5-6.0%) was not included in this budget. Despite the inability to linearly add these uncertainties, the prominence of contouring uncertainties in the HDR cervix intra-cavity IGABT uncertainty budget is further reiterated. In the case of OARs, reported contouring uncertainties (5-11%) make up a smaller fraction of the overall 20-25% inter-intra-fraction changes uncertainty budget [151]. The current study reports the variability in prescribed dose due to planning variations as 4.4-7.7%, increasing the proportion of the inter- and intra-fraction changes budget that is due to contouring uncertainties. Organ motion remains a substantial component of the intra- inter-fraction changes uncertainty budget, more so than for target contours [155].

4.4.2 Correlation Analysis

Overall, Tables 4.5, A.4 and A.5 imply that the smaller the CTV_{HR} and CTV_{IR}, the larger the dose, with a decrease in the lateral, anterior/posterior and superior/inferior directions correlating most with an increase in the dosimetric parameters investigated. Adding a margin to account for these dosimetric variations due to differences in contour dimensions may be considered, however since the application of margin in any direction but the superior/inferior direction is not appropriate [156], the expansion should be limited to these directions.

It is important to note that although significant correlations were found in this study, no strong correlations ($\rho > |0.7|$) were observed. Most results indicated weak to moderate

correlations ($\rho > |0.3|$, ($\rho < |0.69|$) as indicated by the ρ value in Tables 4.5, A.4 and A.5.

4.5 Conclusion

There are differences in dosimetric outcomes arising from inter-observer variation in target delineation for cervical brachytherapy, resulting in approximately 2.9-3.4 Gy $EQD2_{10}$ variability in the CTV_{HR} D90 over the course of the treatment, depending on which gold standard contour is assessed. The maximum geometric dimensions of these delineations were most commonly correlated with dosimetry changes. Although the parameters that are significantly correlated are similar across gold standards, the direction of these correlations differ, indicating that the dosimetric outcomes are dependent on the contour that the plan is optimised to. This study reiterates the importance of carefully choosing a gold standard from which to benchmark, and highlights the potentially substantial dosimetric differences inter-observer uncertainty can have.

Chapter 5

Validation of Dose Planning Variations Related to Delineation Uncertainties in MRI-Guided Brachytherapy for Locally Advanced Cervical Cancer for the Australian Context

This chapter validates the dosimetric variation arising from inter-observer variation found in Chapter 4 on an Australian dataset, given the differences in brachytherapy practice between Australian and European centres. Spatially descriptive contouring metrics are quantified and compared to dosimetric metrics to determine whether correlations exist. The findings in this chapter contribute to answering the first aim of this thesis: to understand the impact of delineation uncertainty on radiotherapy efficacy (Chapter 1.1).

The content in this chapter is in preparation for publication. The candidate's contribution to this publication is in the form of data generation, collection and analysis as well as authorship of manuscript.

5.1 Introduction

Sources of uncertainty in brachytherapy are numerous, including source specification, treatment planning, dose delivery and inter/intra-fraction changes [151]. By far the biggest contributor to the total dosimetric uncertainty is the uncertainty in target delineation (contouring uncertainties) [101, 151]. Despite adherence to protocols, there is dosimetric variability due to contouring uncertainty, both in reported doses and prescribed doses arising due to resulting planning variations. Our previous work, presented in Chapter 4 investigated the variability in prescribed doses due to planning variations from contouring uncertainties. The study was conducted on a GEC-ESTRO dataset comprising European patient images and contours. The study found that inter-observer variation in CTV_{HR} definition produces substantial dosimetric variability, up to 3 Gy $EQD2_{10}$ for an HDR image guided adaptive brachytherapy course for the cervix. Additionally, correlations between commonly used delineation uncertainty metrics and dosimetric outcomes were found to exist, linking increased dose with smaller cardinal dimensions.

The current study aims to validate these findings in the Australian context. The validation of the GEC-ESTRO findings with Australian data will negate the impact of small practice differences and geographic-specific patient characteristics on dosimetric variations due to contouring. This means that solutions to mitigate these effects can be applied universally, despite local practice and patient demographic differences.

5.2 Methods

T2-weighted MRI from eight cervical cancer patients collected for the study (presented in Chapter 3) were utilised [157]. The CTV_{HR} was contoured by 12 independent observers from Australian institutions practicing MRI-based cervical cancer brachytherapy. Two gold standard contours were generated so that comparisons between observers can be

made on an expert consensus CTV_{HR} generated by a group of local experts in cervical cancer contouring and the STAPLE [42]. OARs were generated by a single observer for planning purposes.

Intra-cavitary (tandem and ovoid) brachytherapy plans were optimised according to local clinical practice. Virtual treatments for these patients consisted of 45 Gy in 25 fractions of external beam radiotherapy and 3 fractions of HDR brachytherapy [158]. Planning aims and limits for prescribed doses for each fraction are presented in Table 5.1.

Table 5.1: Planning aims and limits for prescribed dose for a single fraction of IGABT for cervix cancer.

	D90 CTV_{HR}	D2cm^3 Bladder	D2cm^3 Sigmoid	D2cm^3 Rectum
Planning Aims	$>$ 8 Gy	$<$ 5.8 Gy	$<$ 5.3 Gy	
Limits for Prescribed Dose		$<$ 6.5 Gy	$<$ 5.9 Gy	$<$ 5.9 Gy

The contouring variability in these contours was assessed via a number of metrics including the location of the COM relative to the gold standards, the CI with respect to the gold standards, the dimensions in the cardinal directions (MD-X, MD-Y, MD-Z) and the volume. The variability in the dose achieved for the CTV_{HR} for these plans was assessed by comparing the mean, SD and rSD. The correlation between the dosimetric metrics and contouring metrics was determined using a non-parametric Spearman's rank correlation.

5.3 Results

5.3.1 Planning

Table A.6 presents the reported doses for each optimised plan. 43/112 (38.4%) plans failed to achieve the single fraction planning aim of $>$ 8 Gy to CTV_{HR}. For one patient (patient 6), very little dose could be delivered to the CTV_{HR} without exceeding OAR limits due to unfavourable anatomy, with the STAPLE and EC CTV_{HR} receiving 0.9 Gy and 1.1 Gy respectively. It is acknowledged that this would not be used clinically. For

the OARs, bladder planning aims were exceeded in 94/112 (83.9%) of plans. Rectum and sigmoid colon planning aims were exceeded in 18/112 (16.1%) and 19/112 (17.0%) plans respectively.

5.3.2 Contouring Variability

Table 5.2 shows the contouring variability metrics for the observer contours and gold standards. The location of the x, y and z position of the COM relative to the gold standards, averaging across all observers and all patients, varied 0.16 cm, 0.27 cm and 0.87 cm respectively. The maximum dimensions demonstrated a variation ranging between 0.75 - 1.82 cm and a mean volume variation of 28.5 cm^3 was observed. The CI variation relative to the gold standards varied by 0.2 for both EC and STAPLE contours.

5.3.3 Dosimetric Variability

The dosimetric variability for the gold standard contours (EC or STAPLE) arising from plan optimisations based on different observers contours can be seen in Table 5.4. The rSD in STAPLE CTV_{HR} D90 across patients ranged between 3.20 - 11.77% and for the EC CTV_{HR} D90, ranged between 1.66 - 12.34%. The rSD represents the dosimetric variability of the brachytherapy component of the treatment course only. The magnitude in dosimetric variability is similar between gold standards. The EC exhibited greater mean doses than the STAPLE, however a similar variability in prescribed dose is observed.

Table 5.2: Contouring variability metrics.

	Patient	1	2	3	4	5	6	7	8	Mean
xCOM	Mean (cm)	-0.03	0.02	-0.14	0.02	0.09	0.15	0.00	0.09	0.02
STAPLE	SD (cm)	0.16	0.08	0.17	0.10	0.21	0.40	0.05	0.11	0.16
xCOM	Mean (cm)	0.16	0.00	0.17	0.06	-0.09	-0.58	0.02	-0.07	-0.04
EC	SD (cm)	0.16	0.08	0.17	0.10	0.21	0.40	0.05	0.11	0.16
yCOM	Mean (cm)	-0.01	-0.12	-0.09	0.03	0.03	-0.50	0.02	0.00	-0.08
STAPLE	SD (cm)	0.08	0.16	0.12	0.15	0.18	1.40	0.04	0.04	0.27
yCOM	Mean (cm)	0.00	0.17	0.19	-0.20	-0.22	-1.46	-0.03	0.00	-0.19
EC	SD (cm)	0.08	0.16	0.12	0.15	0.18	1.40	0.04	0.04	0.27
zCOM	Mean (cm)	0.16	0.40	1.35	0.12	0.06	-1.18	0.09	0.17	0.14
STAPLE	SD (cm)	0.39	0.79	1.19	0.63	0.41	2.94	0.26	0.37	0.87
zCOM	Mean (cm)	-0.03	0.18	-0.99	-0.45	-0.21	-1.95	0.16	-0.20	-0.44
EC	SD (cm)	0.39	0.79	1.19	0.63	0.41	2.94	0.26	0.37	0.87
Δ COM	Mean	0.34	0.61	1.55	0.48	0.38	2.8	0.22	0.37	0.84
STAPLE	SD	0.30	0.66	0.93	0.44	0.33	2.02	0.17	0.21	0.63
Δ COM	Mean (cm)	0.33	0.58	1.20	0.59	0.47	3.03	0.26	0.33	0.85
EC	SD (cm)	0.30	0.60	1.02	0.56	0.34	2.76	0.18	0.29	0.76
MD-X	Mean (cm)	5.30	6.16	5.91	5.12	4.95	7.41	4.18	4.95	5.50
	SD (cm)	0.63	2.22	1.29	0.35	1.05	2.38	0.51	0.68	1.14
MD-Y	Mean (cm)	4.25	4.64	4.41	4.89	4.65	7.16	2.84	3.48	4.54
	SD (cm)	0.51	1.40	0.29	0.70	0.52	1.73	0.47	0.38	0.75
MD-Z	Mean (cm)	3.77	5.51	5.44	5.44	4.36	7.14	3.16	4.19	4.88
	SD (cm)	0.88	1.82	2.86	2.86	1.21	2.87	0.80	1.25	1.82
Volume	Mean (cc)	35.95	51.36	56.74	56.74	38.14	138.83	15.84	30.33	52.99
	SD (cc)	13.44	18.83	39.31	39.31	17.63	86.03	4.42	9.04	28.50
CI	Mean	0.66	0.57	0.42	0.42	0.6	0.38	0.62	0.69	0.55
STAPLE	SD	0.13	0.22	0.28	0.28	0.18	0.27	0.16	0.12	0.20
CI EC	Mean	0.65	0.53	0.35	0.35	0.58	0.37	0.65	0.69	0.52
	SD	0.16	0.18	0.23	0.23	0.19	0.32	0.14	0.14	0.20

Table 5.3: Spearmans non-parametric rank correlation coefficient and significance. Statistically significant results shown in bold italics.

	x COM		y COM		z COM		C COM		MD-X		MD-Y		MD-Z		Vol		Vol	
	ρ	sig.	ρ	sig.	ρ	sig.	ρ	sig.	ρ	sig.	ρ	sig.	ρ	sig.	ρ	sig.	ρ	sig.
EC D90	***0.48***	***< 0.01***	-0.04	0.65	0.06	0.52	***-0.28***	***< 0.01***	***-0.22***	***0.02***	-0.15	0.12	***-0.38***	***< 0.01***	***-0.31***	***< 0.01***	0.18	0.05
EC D98	***0.46***	***< 0.01***	-0.06	0.55	0.09	0.33	***-0.30***	***< 0.01***	***-0.25***	***0.01***	***-0.20***	***0.04***	***-0.39***	***< 0.01***	***-0.35***	***< 0.01***	***0.20***	***0.03***
STAPLE D90	0.01	0.9	0.17	0.08	-0.17	0.07	***-0.58***	***< 0.01***	***-0.34***	***< 0.01***	***-0.55***	***< 0.01***	***-0.45***	***< 0.01***	***-0.48***	***< 0.01***	***0.44***	***< 0.01***
STAPLE D98	0.01	0.94	0.17	0.08	-0.17	0.08	***-0.59***	***< 0.01***	***-0.34***	***< 0.01***	***-0.55***	***< 0.01***	***-0.45***	***< 0.01***	***-0.48***	***< 0.01***	***0.44***	***< 0.01***

5.3.4 Correlations

Table 5.3 shows the Spearmans non parametric rank correlation coefficient and significance between the dosimetric and contouring parameters of the CTV_{HR}.

A linear regression analysis testing for interaction was used to determine whether correlations observed within the GEC-ESTRO dataset and the Australian dataset were statistically the same. Table 5.5 presents the results where significance greater than 0.05 for the linear regression indicates that the Spearman rank correlations are statistically the same between datasets.

5.4 Discussion

As contouring and treatment protocols for IGABT converge worldwide, there remains local practice differences that distinguish European and Australian clinical practice. The analysis of the GEC-ESTRO dataset found that inter-observer variation in CTV_{HR} definition produced up to 3 Gy $EQD2_{10}$ substantial dosimetric variability and that increased dose to the CTV_{HR} correlates with smaller cardinal dimensions. This study investigated the planning variation due to inter-observer variation in the Australian cohort and validated the findings from the GEC-ESTRO dataset on this cohort.

5.4.1 Similarities and Differences between Cohorts

The similarities and differences between the Australian and GEC-ESTRO cohorts are summarised in Table 5.6.

The Australian dataset does not utilise interstitial needles, therefore the ability to conform dose to individual tumour volumes is limited. This is most likely the main reason for the substantially lower prescribed doses achieved in this study compared to the study utilising the GEC-ESTRO interstitial brachytherapy dataset A.6. Other dataset differences include

Table 5.4: Dosimetric variability in EQD2 averaged across all observers for each patient. rSD is for the brachytherapy component only.

	Patient	1	2	3	4	5	6	7	8	Mean
CTV_{HR} D90 STAPLE	Mean (Gy)	84.20	59.92	51.40	68.88	62.73	46.61	77.52	72.39	65.46
	SD (Gy)	4.70	0.51	0.56	0.95	0.88	0.21	2.39	0.90	1.39
	rSD (%)	11.77	3.24	7.77	3.85	4.75	8.83	7.17	3.20	6.32
CTV_{HR} D90 EC	Mean (Gy)	97.28	66.63	81.28	86.57	84.38	47.13	84.67	75.37	77.92
	SD (Gy)	6.54	0.37	2.85	1.53	1.60	0.23	2.98	0.88	2.12
	rSD (%)	12.34	1.66	7.70	3.61	4.00	7.87	7.36	2.83	5.92
CTV_{HR} D98 STAPLE	Mean (Gy)	75.18	54.75	49.58	60.61	56.12	46.15	71.20	66.07	59.96
	SD (Gy)	3.31	0.36	0.39	0.71	0.51	0.18	2.01	0.73	1.02
	rSD (%)	10.69	3.47	7.29	4.34	4.27	9.28	7.47	3.36	6.27
CTV_{HR} D90 EC	Mean (Gy)	83.96	60.05	72.15	76.9	75.23	46.44	77.18	67.93	69.98
	SD (Gy)	4.34	0.31	2.07	1.19	1.23	0.16	2.50	0.78	1.57
	rSD (%)	10.92	1.96	7.41	3.63	3.98	7.53	7.59	3.28	5.79
Bladder $D2cm^3$	Mean (Gy)	67.95	80.10	75.35	79.65	79.68	79.68	76.92	80.13	77.43
	SD (Gy)	2.63	0.21	2.41	0.68	1.63	1.65	3.11	0.10	1.55
	rSD (%)	10.63	0.57	7.48	1.86	4.47	4.54	9.22	0.26	4.88
Sigmoid $D2cm^3$	Mean (Gy)	63.51	45.07	51.20	69.20	58.4	-	58.03	73.65	59.86
	SD (Gy)	2.46	0.07	0.64	0.49	1.36	-	1.21	1.12	1.05
	rSD (%)	12.11	3.64	7.98	1.87	8.96	-	8.16	3.67	6.63
Rectum $D2cm^3$	Mean (Gy)	59.54	65.78	64.77	61.17	55.71	72.12	69.46	60.37	63.62
	SD (Gy)	1.69	2.41	2.93	0.73	2.36	2.46	3.20	2.48	2.28
	rSD (%)	10.35	10.68	13.59	4.05	18.85	8.49	12.19	14.43	11.58

Table 5.5: Linear regression analysis results for significant correlations in the GEC-ESTRO study [64] and this study. Statistically significant results shown in bold italics.

	GEC-ESTRO Spearman Rank Correlation	Australian Spearman Rank Correlation	Regression Analysis
STAPLE	ρ (sig)	ρ (sig)	ρ (sig)
MD-X and D98	-0.27 (0.02)	-0.34 (< 0.01)	***1.29 (0.37)***
MD-Y and D98	-0.34 (< 0.01)	-0.55 (< 0.01)	***2.35 (0.13)***
MD-Y and D90	-0.27 (0.02)	-0.55 (< 0.01)	4.37 (0.02)
Vol and D98	-0.25 (0.04)	-0.48 (< 0.01)	***0.06 (0.32)***
EC	(sig)	(sig)	B (sig)
MD-X and D98	-0.38 (< 0.01)	-0.25 (0.01)	***2.33 (0.13)***
MD-Y and D98	-0.39 (< 0.01)	-0.20 (0.04)	3.84 (0.03)
MD-Z and D98	-0.25 (0.04)	-0.39 (< 0.01)	2.64 (0.04)
Vol and D98	-0.32 (0.01)	-0.35 (< 0.01)	***0.11 (0.05)***

Table 5.6: Similarities and differences between Australian and GEC-ESTRO IGABT cohorts.

GEC-ESTRO Dataset [64]	Australian Dataset [157]
Interstitial cervix IGABT	Intra-cavity cervix IGABT
Tandem and ring	Tandem and Ovoids
6 patients, 10 observers	8 patients, 12 observers
European contourers who have participated in EMBRACE II	Australian contourers adhering to local protocols.
CTV_{HR}, CTV_{IR}, GTV and OARs contoured by each observer	CTV_{HR} contoured by each observer. CTV_{IR}, GTV and OARs contoured by local radiation therapist
EC determined by multiple experts (> 3) through Europe, of different disciplines	EC determined by 3 radiation oncologists at one institution
7.2Gy in 4 fractions	8 Gy in 3 fractions

the use of ovoids in this current study as opposed to a ring. There is a similar presence of inter-observer variation across datasets however the types of contours generated differed. Specifically, the Australian dataset lacks multiple observer contours for OARs, as well as contours for the other target volumes recommended by the EMBRACE study (CTV_{IR} and GTV). The EC contour was determined from three Australian radiation oncologists from the same institution as opposed to the multi-discipline and multi-institutional approach in defining the EC contour for the GEC-ESTRO cohort. Although the planning aims are similar between studies, they are not identical and the fractionation schemes are different, with this dataset planned based on a virtual treatment of 8 Gy in 3 fractions as opposed to 7.2 Gy in 4 fractions. It is important to note that the same physicist generated the plans in both studies, negating the inter-planner variations.

5.4.2 Inter-observer Variation Analysis

Patients in which CTV_{HR} D90 planning aims were frequently met (patients 1, 3 and 7) exhibited greater dosimetric variability. This is due to favourable OAR positioning relative to the target contours, enabling greater flexibility and feasibility of dose escalation. For cases where the OARs were positioned close to the target volume, a smaller dosimetric variability was observed amongst inter-observer plans since further

dose optimisation is limited by OAR tolerances.

It is clear that the gold standard on which the plans are evaluated has an impact on the mean dose reported, specifically, plans assessed on the EC result in larger doses than those on the STAPLE. However, the amount of variability in prescribed dose is similar between gold standards. Over a full treatment course, the inter-observer variation in CTV_{HR} D90 results in 1.4 Gy $EQD2_{10}$ (6.3% rSD) variability in prescribed dose for the STAPLE contour, and 2.1 Gy $EQD2_{10}$ (5.9%) for the EC. This is similar, though slightly smaller variability than observed for the GEC-ESTRO cohort (3.4 Gy for the STAPLE and 2.9 Gy for the EC). The differences are likely due to the limited flexibility in dose optimisation for the Australian cohort, due to the lack of interstitial needles.

5.4.3 Correlations

Common types and magnitudes of correlations were found for the investigated dosimetric parameters (D90 and D98) between the Australian and GEC-ESTRO cohorts. Significant correlations of increased dose were observed in both cohorts for decreasing maximum dimensions and volume. In both cohorts, as a contour decreases in the cardinal directions and decreases in volume, greater dose was achieved. This was observed for both gold standards. The Australian cohort was also shown to correlate dose with decreased COM displacement and increased CI, however this was not observed in the GEC-ESTRO cohort. This implies that the Australian dataset is more sensitive to the dosimetric impact of inter-observer variation than the GEC-ESTRO dataset, in that any deviation from the 'true' volume showed a correlated decrease in achievable dose.

The observed correlations between maximum lateral dimensions and dose to 98% of the target volume, as well as volume and dose to 98% of the target were statistically the same between datasets for both gold standards. When using the STAPLE as a gold standard, the correlation between maximum anterior/posterior dimensions and dose to 98% of the target volume were verified to be the same in both datasets. This shows that these findings

are robust across different treatment modalities, contourers, treatment practices and other differences between datasets outlined in Table 5.6.

5.5 Conclusion

Despite numerous differences in treatment technique, fractionation schedule and geographical location of the datasets (which inherently include differences in training, clinical practice and patient demographics), the findings of the GEC-ESTRO dataset have been validated in the Australian cohort. Similar dosimetric variability is observed across an equi-effective treatment course, despite differences in prescribed dose. Correlations between increased dose and target contractions in the cardinal dimensions were observed. One distinction observed was the correlation between dose and COM location and CI observed in the Australian dataset, but not the GEC-ESTRO dataset. This indicates a greater dosimetric sensitivity to inter-observer variation. The validation of the GEC-ESTRO findings means that methods to mitigate the effects of inter-observer variation in the cervix IGABT context, such as a delineation margin, can be universally applied.

Chapter 6

Atlas-Based Segmentation Technique Incorporating Inter-Observer Delineation Uncertainty for Whole Breast

This chapter investigates an atlas-based auto-segmentation method for whole breast radiotherapy. This method improves on the accuracy of existing techniques by incorporating inter-observer variation into the segmentation process. The findings of this chapter contribute to answering the second aim of this thesis: to minimise delineation uncertainty by automating the contouring process (Chapter 1.1).

The content in this chapter has been published. The candidate's contribution to this publication is in the form of data generation, collection and analysis as well as manuscript authorship.

Bell, L. R., Dowling, J. A., Pogson, E. M., Metcalfe, P., Holloway, L. (2017), Atlas-based segmentation technique incorporating inter-observer delineation uncertainty for whole breast. J Phys: Conf Ser, 777: 012002. doi:10.1088/1742-6596/777/1/012002

6.1 Introduction

Internal anatomical changes during the course of radiotherapy limit treatment accuracy, as current treatments are planned on static images obtained prior to delivery. Adaptive radiotherapy is an increasingly investigated approach that aims to combat these anatomical changes by re-imaging and re-planning at multiple time points throughout the treatment course. However, this process has large workflow implications due to the time required to achieve this whilst maintaining plan quality.

Target delineation is the bottleneck of the planning process taking on average 18.6 min per patient for breast radiotherapy [69], and housing one of the largest uncertainties affecting modern radiotherapy accuracy [3, 159]. Segmentation techniques, boasting significant decreases in delineation times as compared to manual contouring, have been proposed as the solution to this limitation with several techniques implemented with impressive results [160, 161]. Atlas-based segmentation techniques in particular are suited for adaptive radiotherapy purposes as they are derived from real patient anatomy and verified clinician contours. However, current techniques generally fail to incorporate inter-observer variation, with single clinician contours frequently taken to be the gold standard delineation. With the largest uncertainty in modern radiotherapy treatment occurring during the target delineation process, it is essential to account for delineation uncertainty when devising an adaptive radiotherapy process.

This atlas-based segmentation technique improves on the accuracy of existing techniques by incorporating inter-observer variation into the segmentation process, whilst maintaining efficacy.

6.2 Methods

This segmentation technique, although developed and validated for a whole breast, is applicable to any target volume definition in which the contours are defined based on

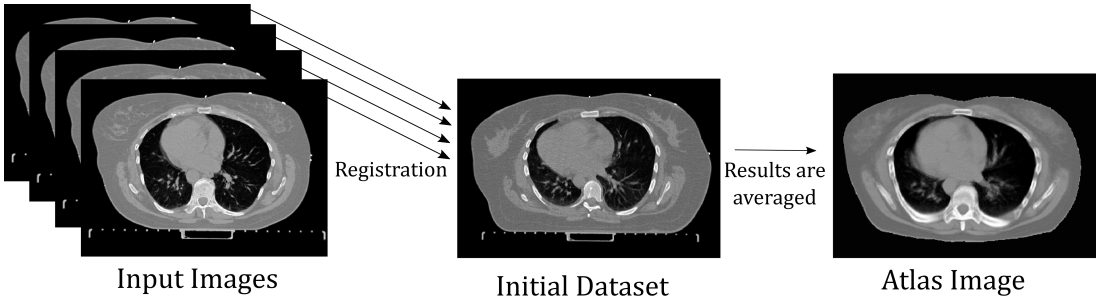


Figure 6.1: Method for atlas generation.

anatomy. The general process involves generating atlas, registering the atlas to a new patient to obtain good correspondence between equivalent regions, and mapping the atlas segmentation to the new image space.

6.2.1 Technique Development

The atlas generation method, outlined in Figure 6.1, has been utilised previously in medical imaging applications [162, 163]. An initial dataset representative of the patient cohort was chosen with remaining datasets in the cohort registered to the initial dataset using rigid and affine transformation [164]. The results were averaged and used as the subject for the next iteration in which rigid, affine registration and diffeomorphic demons non-rigid registration [165] was used. Three iterations were performed to produce the final atlas.

All CTVs except those of the left-out dataset were mapped to atlas space using the existing deformation fields and merged. The resulting contouring probability model was thresholded to 50% to obtain an auto-segmented CTV contour.

6.2.2 Validation

The technique was validated on a whole breast radiotherapy cohort containing datasets from 28 patients (14 right, 14 left), collected for a previous study [22]. Non-contrast CT scans were used with whole breast CTV delineated by eight independent observers.

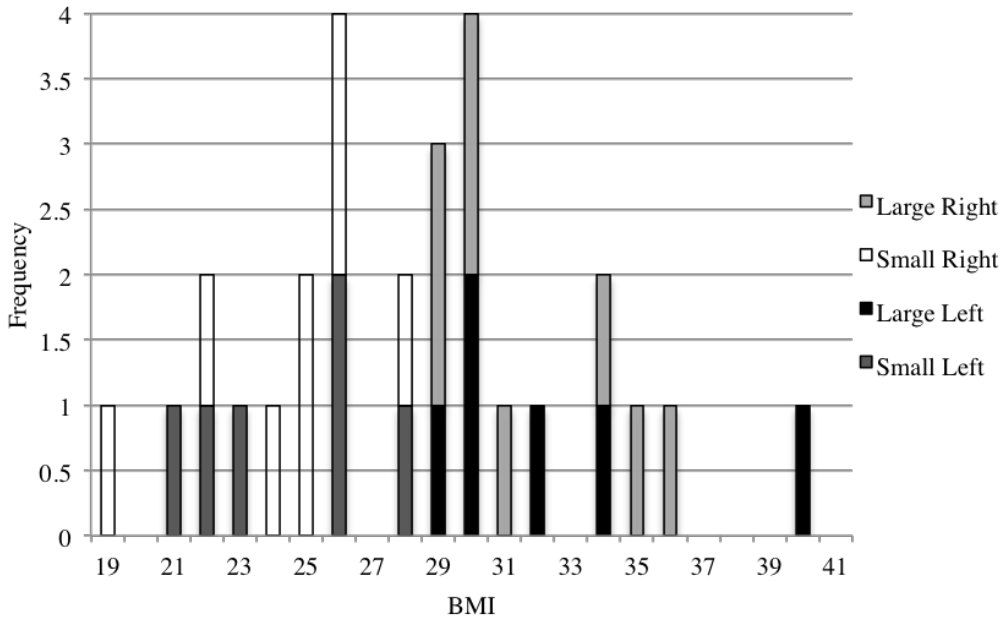


Figure 6.2: Distribution of patient cohort within volume categories.

Observers followed a delineation protocol and were blind to other observer contours [22, 37].

Since atlas accuracy correlates with similar anatomy of input patients [166], the cohort was divided into categories, grouping similar patients. Body mass index (BMI) was used as a surrogate for similar patient anatomy, with small/large patients classed as </> 29 BMI and datasets were further divided according to laterality. Figure 6.2 shows the distribution of patient datasets.

An atlas was generated from the 6-9 datasets in each category using the MILXView platform [162] using a leave-one-out approach. The atlas was registered using rigid, affine transformation followed by non-rigid registration to the 'left out' dataset as shown in Figure 6.3. The auto segmentation was mapped to this dataset from atlas space, and clipped such that it did not extend past the patient surface.

A gold standard, consensus contour was generated from the observer CTVs on the left-out dataset in each category using the STAPLE algorithm [42] in MILXView. To assess similarity, the gold standard contour was compared to the auto-segmentation using the

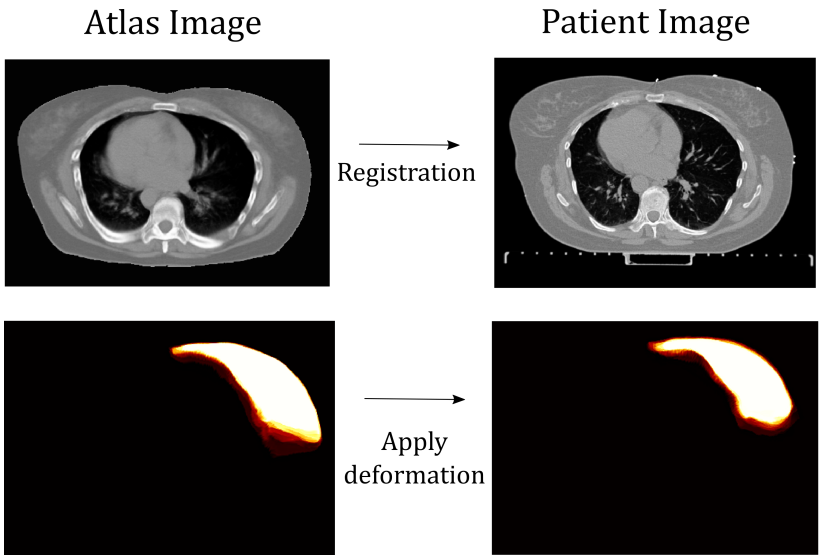


Figure 6.3: Applying the auto-segmentation method to a new patient.

DSC [167] and MASD [168]. A comparison between the smallest and largest CTV with the auto-segmentation was also made to determine the range in accuracy.

6.3 Results

Average atlas' with corresponding whole breast CTV segmentations were successfully generated for each category.

The average atlas for each category with corresponding probability maps is shown in Figure 6.4. The whole breast CTV auto-segmentation is visually compared to the consensus, smallest and largest contour in Figure 6.5. The time required to auto-segment, including registration of the atlas to the required patient and propagating contours, was on average 3 min and 43 seconds. The DSC and MASD values indicating the accuracy of the auto-segmentation in delineating the whole breast CTV and accounting for inter-observer variation is shown in Table 6.1.

The regions of largest deviation between the auto-segmentation and consensus volume is visually seen in Figure 6.6 medially in the small BMI category, and at superior/inferior slices for the large BMI category.

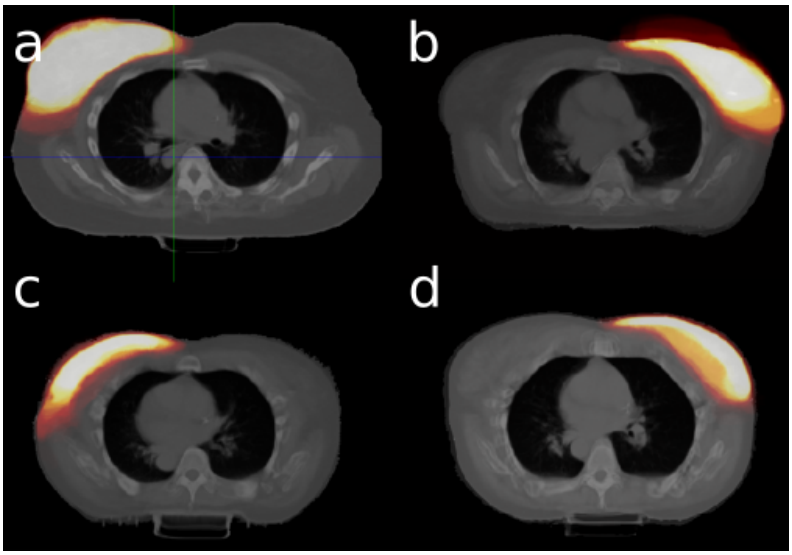


Figure 6.4: Probability maps of observer contours for a) large right, b) large left, c) small left and d) small right atlas'. Heat map indicates regions of highest probability.

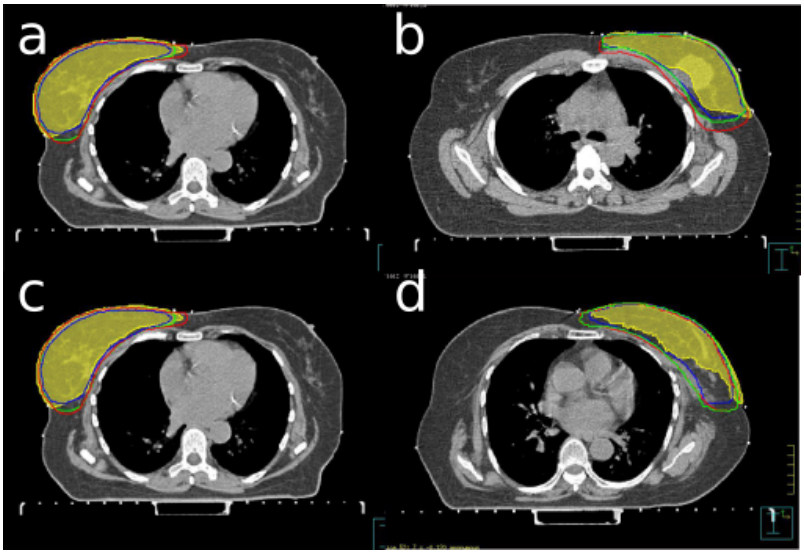


Figure 6.5: Auto-segmentation (yellow colour wash) for left-out patients. A) large right, b) large left, c) small left and d) small right categories. STAPLE (green), smallest (blue) and largest (red) CTVs are shown for comparison.

Table 6.1: Similarity metrics comparing whole breast segmentation with manual target volumes.

	DSC				MASD (mm)			
	Large Left	Large Right	Small Left	Small Right	Large Left	Large Right	Small Left	Small Right
Consensus	0.81	0.85	0.71	0.79	7.99	3.47	6.17	5.16
Smallest CTV	0.77	0.86	0.71	0.79	9.28	3.64	4.81	5.16
Largest CTV	0.81	0.80	0.71	0.79	7.50	4.71	6.13	5.07

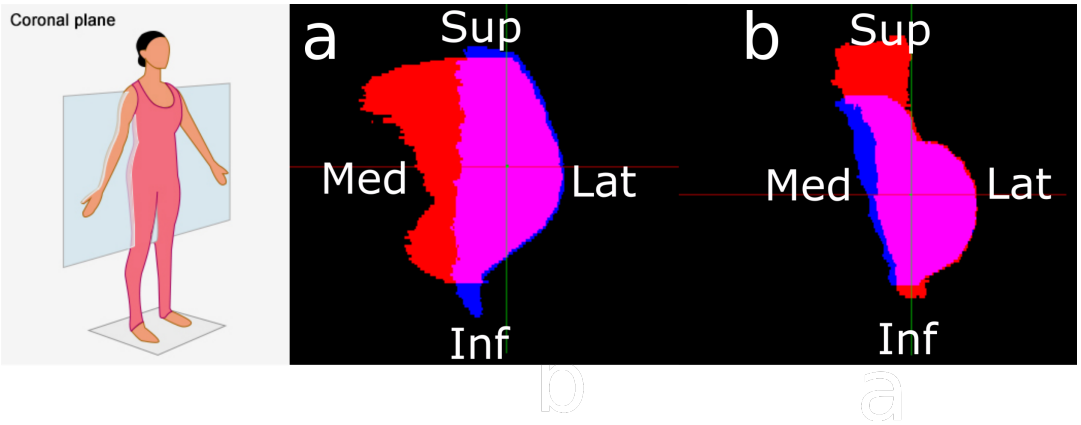


Figure 6.6: Differences in whole breast auto-segmentation (blue) and gold standard contour (red) for large BMI patients (right) and small BMI (left) in the coronal plane. Common regions are shown in pink.

6.4 Discussion

Coverage probability maps overlaid on average atlas' showed high observer agreement for the majority of the whole breast segmentation. Observer agreement was lowest in the anterior-medial and posterior-lateral directions as is supported in literature [69, 169].

Auto-segmented contours generally compared well with the gold standard for each category. Insufficiencies in auto-segmentation were observed close to the chest wall for large and small BMI left breast patients. The authors hypothesise that this is due to larger segmentation variations amongst observers and patients due to the close proximity of the heart. Manual review by of this key area by a clinician is recommended. Improvements in segmentation in these instances could be achieved through optimisation of the probability map thresholding value.

Nevertheless, auto-segmentation accuracy was high, producing high concordance (DSC $>$ 0.7) with target volumes in all volume categories. Although MASDs were often greater than 5mm, which is the tolerance quoted in Chapter 3, they were limited to less than 1 cm with location of largest surface differences specific to BMI category. For the purposes of this study, this is deemed acceptable due the the unique PTV shape of whole

breast volumes, where distances between contours can be large in the anterior/medial and posterior/lateral regions, despite being qualitatively similar. The largest differences for large BMI patients occurred in the medial direction and for small BMI patients in the superior direction.

6.5 Conclusion

This atlas based auto-segmentation technique incorporating inter-observer variation has been shown to be accurate (DSC $<$ 0.7) and efficient (time $<$ 5 min) as is necessary for the clinical efficacy of adaptive radiotherapy. There is scope for further optimisation of correlation with gold standard segmentations as well as computation time. This method is a feasible solution to auto-segmentation adaptive radiotherapy with the potential for viable application to other anatomical sites.

Chapter 7

A Comparison of Coordinate Systems for Use in Determining a Radiotherapy Delineation Margin for Whole Breast

This chapter investigates the most appropriate coordinate system to define a whole breast radiotherapy delineation uncertainty margin, given that a conventional three-plane cartesian margin may not adequately account for delineation uncertainty. The findings of this chapter contribute to answering the third aim of this thesis: to manage delineation uncertainty using an uncertainty margin (Chapter 1.1).

7.1 Introduction

Radiotherapy treatment techniques have improved significantly with the implementation of IMRT, VMAT and image-guided radiotherapy. Further improvements in radiotherapy

The content in this chapter has been published. The candidate's contribution to this publication is in the form of data generation, collection and analysis as well as manuscript contribution.

Pogson, E. M. and Bell, L. and Batumalai, V., Koh, E. S., Delaney, G., Metcalfe, P. and Holloway, L. (2014), A comparison of coordinate systems for use in determining a radiotherapy delineation margin for whole breast. J Phys: Conf Ser, 489: 012057. doi:10.1088/1742-6596/489/1/012057

are achieved by setting clear delineation protocols, minimising set-up errors and imaging errors, and imaging for internal motion management [6]. Delineation uncertainty arising from inter-observer variability has traditionally been ignored as larger systematic errors dominate. Inter-observer variability is still a significant uncertainty, even with clear delineation protocols and uncertainties from delineation, may now play a significant role in influencing the success of these improved techniques.

Although the inclusion of delineation margins in PTV margin recipes is well acknowledged in the literature as necessary, particularly with the introduction of more conformal delivery techniques calculation of these margins is rarely attempted. Studies have been performed for cancer sites such as the prostate [14, 170] and the brain [171] to optimise set-up and motion margins ignoring the inter-observer delineation uncertainties. These have shown that margins are necessary to reduce the likelihood of geographical misses, reduce effects from systematic errors (under-dosage) and reduce effects from random errors (blurring) [6]. Margin formulae using the combined preparation/systematic standard deviation (Σ) and execution/random standard deviation (σ) such as $2.5\Sigma+0.7\sigma-3$ mm [10], incorporate inter-observer variability within the preparation errors (Σ). However this is performed with one value over the whole structure equally or with 3 values in 3 planes. As the inter-observer variation may be accurately determined over the whole structure (multiple angles and points), perhaps this adaptation should be performed before the other margin formula is applied. The appropriate amount or weight that this should have compared to the whole margin formula needs to be investigated further.

Introduction of an additional margin, or ensuring PTV margins include the inter-observer variability margin, to account for breast delineation uncertainties would reduce the clinical impact of inter- and intra- observer variation. This has previously been performed using cartesian coordinates for sites such as the prostate and seminal vesicles [172], partial breast [173], whole breast [174] and performed using spherical coordinates

for sites such prostate [38] and bladder [125]. Shape of the target will play a significant role in what coordinate system should be used.

7.2 Methods

Datasets from 9 patients from a previous study were utilised [22]. Volumes of the CTV were contoured by four breast cancer radiation oncologists (with experience ranging from 6 to over 20 years in oncology) and four radiation therapists (with experience ranging from 3 to 13 years). One-sided whole breast CTV were outlined on transverse slices with a standard window level (0) and width (500). Radio-opaque wire was used to aid delineation. Each observer was allocated a unique letter to de-identify their contours. Observers were blind to other observer contours. The SD in delineation uncertainty using cylindrical coordinates for 360 degrees in 1 degree increments and in cartesian coordinates (the maximum in each X and Y direction) for whole breast was compared. As the CTV structure may be easily elongated many papers calculate the deviation in the cardinal directions (from the geometric center coordinates in 6 directions) and perform this adaption enabling some change over the entire volume. This has been generally undertaken to account for tumour motion, however for inter-observer error a more detailed expansion at many directions may be necessary. For this study 360 interpolated points were calculated for each slice in cylindrical coordinates and compared to the cartesian cardinal points and extensions. The interpolation at 1 degree increments were compared with the original structures to ensure the shape of the structures remained constant. The SD of this was calculated for slices as in Equation 7.1 for cylindrical coordinates and Equation 7.4 for cartesian coordinates.

7.2.1 Cylindrical Coordinates

The contours were input into the Computational Environment for Radiotherapy Research (CERR) software platform (v5.1) within the MATLAB environment as DICOM-RT

structure sets and associated cartesian coordinates (X,Y,Z) were determined for each patient and observer at the selected slices [147, 175]. The cartesian coordinates were then converted into cylindrical coordinates for analysis. The SD of contour positions were determined for cylindrical coordinates at θ = 1° increments for 5 slices (Z = -2 cm, -1 cm, 0 cm, 1 cm and 2 cm from the origin) for each patient using equation 7.1, where N is the total number of observers per patient and $\bar{r}$ is the radius averaged over all observations, Z is the slice or height of the structure in cm, j is the observer, i is the patient number and θ is the cylindrical angle ranging from 0°-360°.

$$SD = \sum^{j}_{inter-observer,radial}(\theta,Z) = \sqrt{\frac{\sum_{i=1}^{N}(r_i^j(\theta,Z) - \overline{r^j_{obs}(\theta,Z)^2})}{N-1}} \tag{7.1}$$

The mean COM was used as the origin for the right breast, for the left the COM was shifted 1 cm superiorly to avoid clipping of the structure. If an observer did not have a contour on a distal slice (e.g. Z = 2 cm) the number of observers was reduced to those present. A margin to account for margin delineation uncertainty (MDU) was then calculated as in equation 7.2 where SD is the standard deviation, i the patient number and θ, Z indicates the direction. This is not recommended and is simply a preliminary study into the validity of including inter-observer margins in multiple directions in PTV margin recipes.

$$MDU^i_{\theta,Z} = 4 \times SD^i_{\theta,Z} \tag{7.2}$$

This margin was selected such that 95% of the contoured union volume would be included for any of the initial contours assuming a Gaussian distribution of contour variation. The margin to be used on a new patient would be calculated using all 9 previous patient datasets as outlined in equation 7.3 where P is the total number of patients.

$$MDU_{\theta,Z}^{general} = \frac{\sum_{j=0}^{P} MDU_{\theta,Z}^{i}}{P} \tag{7.3}$$

7.2.2 Cartesian Coordinates

The SD and MDU were determined for cartesian coordinates for medial-lateral and anterior-posterior positions, following previously published methodology [174]. If inter-observer variability is considered in studies the SD is calculated usually in only 6 directions as treatment planning systems typically allow extension only in the anterior (A), posterior (P), superior (S), inferior (I), lateral (L) and medial (M) directions by one value for all points (from the average COM). The SD is outlined in equation 7.4 where x_d is the pixel position, in cm, at the edge of the contour in the appropriate direction (d = A, P, S, I, M, or L); and x_{COM} is the average centre of mass coordinate also in that direction (d).

$$SD = \sqrt{\frac{\sum_{i=1}^{N} [(x_{i,d} - x_{i,d,COM}) - \overline{(x_{i,d} - x_{i,COM})}]^2}{N-1}} \tag{7.4}$$

The MDU is found as in equations 7.2 and 7.3, where d is the directions replacing and Z (A, P, S, I, L or M). The MDU was added onto the observer contour in all 6 directions to generate a CTV with a margin to account for inter-observer errors. For all methodologies the MDU would be clipped to the skin, lung, heart, muscle interface and the medial border if necessary.

7.3 Results

Figure 7.1 displays the interpolation points and original points of eight contours on one patient at slice Z = 0 cm for cylindrical interpolation at 1 degree increments, indicating consistency through the interpolation with minimal clipping or divergence.

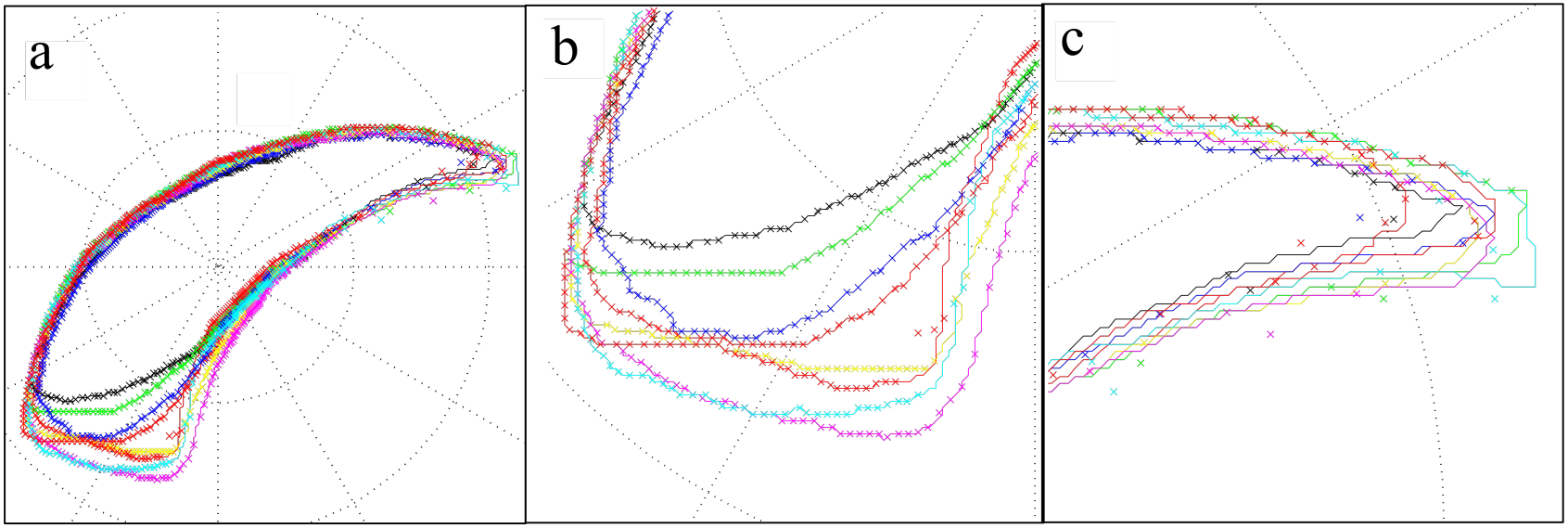


Figure 7.1: (a) Interpolated data points in cylindrical coordinates for patient 1 at slice Z = 0 cm. The interpolated points are shown as crosses and the original data is shown as solid lines. The same colour has been used for each observer. Parts (b) and (c) are expanded views of the structure shown in (a).

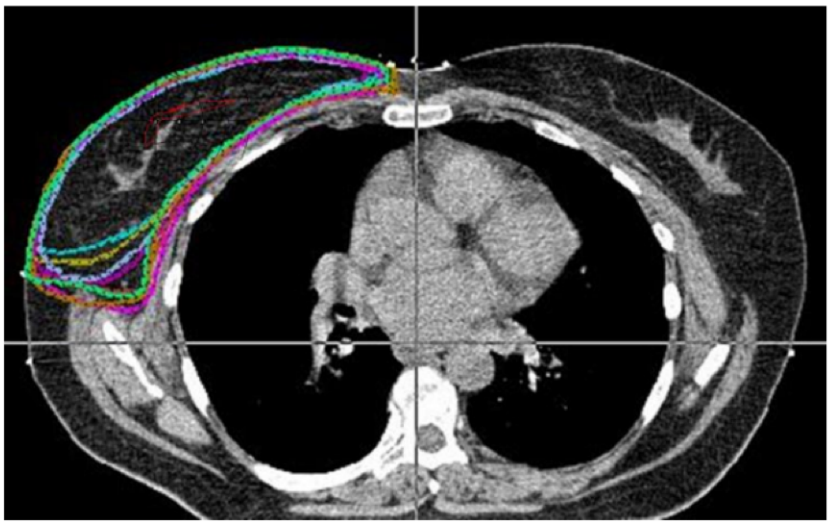


Figure 7.2: The contours at Z = 0 cm for patient 1, with the overall uncertainty for all patients at Z = 0 cm overlaid in red.

The interpolation used linear interpolation at 1 degree increments. The overall uncertainty for Z = 0 cm is plotted for patient 1 in Figure 7.2. Figure 7.2 highlights the anatomical borders that are most affected by the uncertainty.

At slice Z = 0 cm considering cylindrical coordinates (lateral = 0°), the SD peaked medially reaching 3.55 cm at 15° for the right breast, and 1.44 cm at 171° for the left (see Figure 7.3). The SD of the remaining slices maintained a similar distribution, with variation in the peak occurring within 10° of the Z = 0 cm positions. In comparison, for cartesian coordinates at slice Z = 0 cm, the largest SD in the medial-lateral and anterior-posterior directions was 0.54/0.57 cm and 1.03/0.67cm respectively for right/left breasts. The average MDU for the cardinal directions in cartesian coordinates is displayed in Table 7.1 [174].

It can be seen in Figure 7.4 that for every contour, the additional margin of 4 SD is in line

Table 7.1: SD and MDU to account for delineation uncertainty in cartesian coordinates.

Direction	Anterior	Posterior	Medial	Lateral	Superior	Inferior
Right SD (cm)	0.40	0.22	1.15	1.48	0.52	0.79
Left SD (cm)	0.32	0.46	1.70	1.13	0.46	0.64
Left MDU (cm)	1.28	1.86	1.13	1.70	1.83	2.57
Right MDU (cm)	1.61	0.87	1.15	1.48	2.07	3.17

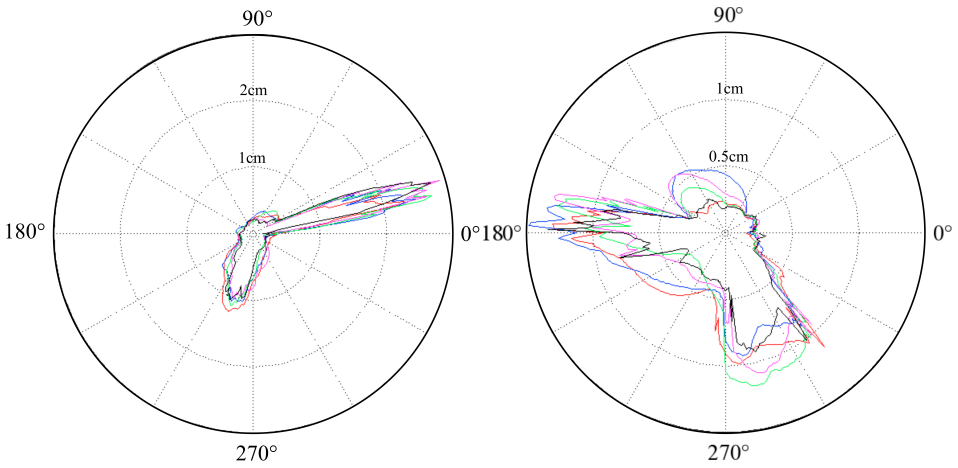


Figure 7.3: The uncertainty (1SD) derived from 5 patients with right breast cancer (a) and 4 with left breast cancer (b) utilising 8 observers calculated in cylindrical coordinates for Z = -2 cm, -1 cm, 0 cm, 1 cm, 2 cm.

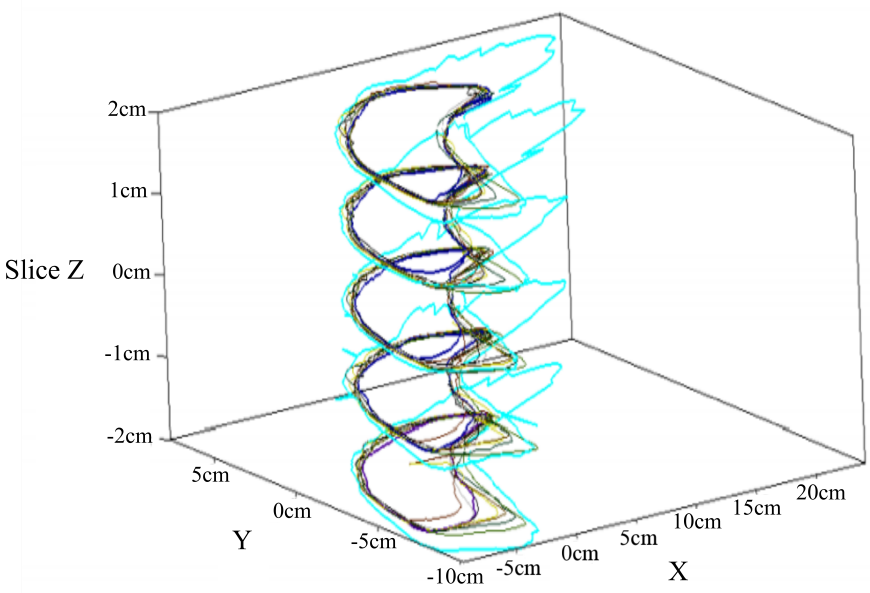


Figure 7.4: Patient 1 (right breast) contours are all within the minimum MDU calculated from cylindrical coordinates. All 8 observers lie within the minimum observers contour with the addition of the MDU. The thin lines are observers 1-7, observer 8 contour (minimum contour) is the thick dark blue line. The light blue thick line is the MDU of observer 8. Each observers contour is of a different colour consistent through all slices.

with the statistical account of 95% for any observers contour. This was also the case for Z = -2 cm, -1 cm, 1 cm and 2 cm as shown in Figure 7.4 for patient 1. This is the case for all the patients, only one patient is shown as an example for clarity. The mean SD was calculated separately for left and right breast patients.

7.4 Conclusion

It has been shown that there is more accuracy in calculating a standard deviation at multiple points (1 degree increments) to extend the CTV to account for inter-observer error, rather than extending X, Y and Z by their cardinal variances. Margins are larger in some directions using cylindrical coordinates than if the contour was extended by these cardinal variances. This highlights the regions where inter-observer variation is high such as both edges of the contours for both left and right breast patients. At slice Z = 0 cm considering cylindrical coordinates (lateral=0°), the SD peaked medially reaching 3.55cm at 15° for the right breast, and 1.44cm at 171° for the left. The SD of the remaining slices maintained a similar distribution, with variation in the peak occurring within 10° of the Z = 0 cm positions. By comparison, for cartesian coordinates at Z = 0 cm, the largest SD in the medial-lateral and anterior-posterior directions was 0.54/0.57cm and 1.03/0.67cm respectively for right/left breasts. Future work will calculate and compare cartesian coordinates at every point to cylindrical coordinate calculations. Sensitivity to COM variations and a spherical coordinate analysis method is also being investigated. Incorporation into a margin recipe that includes organ motion, set-up errors and random errors should be investigated for clinical implementation. A method of extending contours outside six directions (e.g. 360 directions) used in treatment planning systems also requires investigation.

Chapter 8

Defining and Assessing an Anisotropic Delineation Margin

This chapter defines and assesses an anisotropic delineation margin in both polar and spherical coordinate systems in order to account for the spatially varying nature of this uncertainty. A whole breast radiotherapy cohort was used as a proof of concept. The findings of this chapter contribute to answering the third aim of this thesis: to manage delineation uncertainty using an uncertainty margin (Chapter 1.1).

8.1 Introduction

Studies investigating margins to account for delineation uncertainty are few. One study has investigated this concept, utilising contour coverage as a function of the applied margin to determine an isotropic GTV to PTV delineation margin for non-small cell lung cancer [159]. Another introduces the delineation margin concept on which this work is

The content in this chapter has been published. The candidate's contribution to this publication is in the form of data generation, collection and analysis as well as manuscript authorship.

Bell, L. R., Pogson, E. M., Metcalfe, P. E. and Holloway, L. (2016), Defining and assessing an anisotropic delineation margin for modern radiotherapy. Med. Phys., 43: 6644-6653. doi:10.1118/1.4967942

based, for whole breast radiotherapy in cartesian coordinates [174]. Work presented in Chapter 7 compared coordinate systems for defining a whole breast radiotherapy delineation margin at selected slices in polar and cartesian coordinates [169]. However, a margin to account for delineation uncertainty may need to be anisotropically defined outside the cardinal directions to account for the spatially varying nature of this uncertainty [38, 125, 127, 128, 169]. Existing uncertainty margins vary only in the cardinal directions, therefore the uncertainty in all directions may not be appropriately accounted for.

This study investigates an anisotropic delineation margin applied to whole breast radiotherapy as a proof of concept, utilising an available inter-observer variation cohort [22]. This is the first study to determine and assess an anisotropic delineation margin that accounts for the largest contouring uncertainties where they occur. For comparison with current margin practices, this approach was compared to a conventional approach that extends the whole breast CTV in the cardinal directions. Spherical and polar coordinates were both used for the anisotropic approach, as either may be appropriate in describing the whole breast CTV shape.

8.2 Methods

8.2.1 Margin Definition and Methodology

The MDU utilises the radial SD (Σ) in observer contours averaged across datasets. A 'leave-one-out' approach is used in which the margin is determined for all datasets within the cohort but the subject. Margins were defined to enable each dataset to be the subject. The SD was then multiplied by a weighting factor (MDU = $\alpha\Sigma_{av}$). This approach is an adaptation of the first systematic term of existing margin recipes [9, 14], as delineation uncertainty is systematically carried through all treatment fractions in current techniques. The weighting factor was chosen to be α=2, as it is a commonly referenced and accepted

coefficient for systematic uncertainties in published margin recipes. Furthermore, it is the smallest published coefficient hence producing a conservative delineation margin and it is representative of 95% confidence interval in which observer contours lie [9, 14]. The expansion of a CTV at relevant points by the MDU results in a target volume that accounts for delineation uncertainty (PTV_{del}).

The MDU aims to encompass all malignant cells (the true target volume), regardless of which observer CTV it is applied to. This work has assumed that the true target volume is somewhere between the smallest and largest observer contour. In assessing the tissue missed by the MDU the conservative approach of assessing if the union volume of all CTVs is encompassed when the MDU is added was taken. To account for the fact that this will also include healthy tissue, the amount of included extra healthy tissue (determined to be healthy, not CTV, by all observers) has also been assessed.

It is not apparent whether similar sized CTVs across datasets are required for derivation of a satisfactory MDU. Hence, the utilisation of datasets with similar target volumes only (a volume specific approach), as opposed to the whole cohort when defining a MDU was compared to determine whether the margin method is dependent on similar anatomy. The volume specific approach involved the patient cohort being divided according to small, medium and large average CTV volume as well as right and left laterality, with MDUs defined for each.

8.2.2 Implementation

As a proof of concept, datasets from 21 breast cancer patients (9 right, 12 left) undergoing breast conservation surgery and adjuvant breast radiotherapy were utilised. Approval was granted by the Sydney South West Human Research Ethics Committee. This cohort is a subset of a tangential breast radiotherapy cohort collected for a previous study, comparing radiation therapist and radiation oncologists volumes, utilising CT datasets of identical slice thicknesses [22]. Each patient underwent a non-contrast CT

scan with a 2.5 mm slice thickness in the supine position for the purposes of breast radiotherapy. Eight independent observers from the same institution, including four radiation oncologists and four radiation therapists, delineated the whole breast CTV according to a delineation protocol adapted from Struikmans et al [22, 37] and were blinded to other observer contours.

Image visualisation, data extraction and data manipulation was achieved using CERR [147, 175]. To effectively use polar and spherical coordinate systems for an anisotropic approach, the CT origin was shifted to the average COM of all observer CTVs on each transverse slice, and in three dimensions respectively [176]. Data points describing each CTV were similarly shifted and converted into both polar and spherical coordinates. The nearest neighbour interpolation method was used to obtain radii at every whole angle for comparison of corresponding CTV points across different observers and patients.

The SD for both anisotropic approaches in defining the MDU was determined in polar and spherical coordinates. For polar coordinates, the SD in observer contours on each transverse slice was determined radially according to Equation 7.1, where N is the total number of observers per patient (N=8), j is the observer, i is the patient number, $r_i(\theta, Z)$ is the radius of the contour at a particular theta (θ) projection and z slice, $\overline{r_{obs(\theta,Z)}}$ is the average radius of all observations at the same θ projection and Z slice [127].

The SD in observer contours for spherical coordinates was taken radially in three-dimensions according to Equation 8.1 [125, 127, 176]. $r_i(\theta, \phi)$ is the radius of the contour at a particular 3D projection defined by θ and ϕ, $\overline{r_{obs(\theta,\phi)}}$ is the average radius of all observations at the same 3D projection.

$$SD = \sum^{j}_{inter-observer,spherical}(\theta, \phi) = \sqrt{\frac{\sum_{i=1}^{N}(r_i^j(\theta, \phi) - \overline{r_{obs}^j(\theta, \phi)^2})}{N-1}} \tag{8.1}$$

For a conventional margin definition approach, the SD in observer contours on each transverse slice was determined using Equation 8.1 in the right, left, anterior, posterior,

superior and inferior directions. To minimise the impact of outliers, SD were determined at the three angles surrounding these directions (theta angles of 0°, 90°, 180° and 270° and phi angles of 0° and 180°) and averaged to define the margin.

The gold standard (GS), consensus CTV for each patient was determined from all observers using the STAPLE algorithm [42]. The STAPLE algorithm computes a probabilistic estimation of the true segmentation based on the estimated performance level of each input contour. This STAPLE volume is considered the gold standard computed using probabilistic estimations of all 8 observers contours. Datasets were assigned a volume category based on average CTV volume, and the calculated MDU for that category was added to the GS volume in each dataset to generate the $PTV_{del,\ GS}$. The MDU was additionally applied to the largest and smallest observer CTVs for a best case $PTV_{del,\ smallest}$ and worst case $PTV_{del,\ largest}$ respectively.

The anisotropic margin was applied by adding the MDU to the radial magnitude of the target volume at each angle. The conventional margin was applied by inputting the six MDU magnitudes into the Pinnacle^3 treatment planning system (v9.8) to expand the GS, smallest and largest CTV volumes in the cardinal directions, as is current clinical practice for CTV expansion. All PTV_{del} were clipped to the patient surface to provide a more clinically relevant volume.

Polar and spherical coordinate systems require an origin to lie within the structure they define. Since whole breast CTVs are approximately crescentric in shape, these coordinate systems are not always suited for describing breast CTVs. A particular angular projection may describe more than one point on the surface of a whole breast CTV as seen in Figure 8.1. If the standard deviation is calculated between two points describing different areas of the CTV surface, the resulting margin will not represent the true inter-observer variation.

This effect occurs through-plane as well since whole breast CTVs are also crescentric in the superior-inferior direction. To account for this, the SD was limited to half the

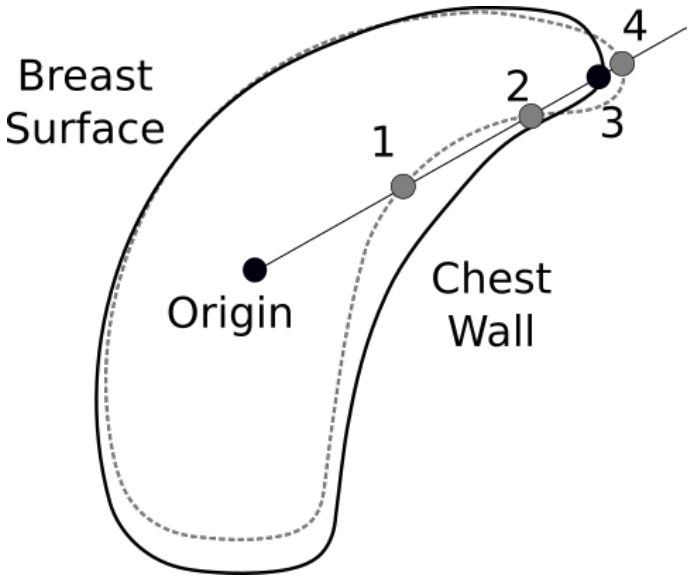


Figure 8.1: Potential coordinate system failure on a transverse slice. A single projection may intersect more than one point for each whole breast CTV. In the case that point 1 is used, the SD between the solid and dotted contours along that projection will be large, reflecting the separation of points 1 and 3. The true SD should be taken between points 3 and 4 since they lie at similar locations on the CTVs. This will affect margin calculations with the margin size overestimated in regions where this occurs.

maximum Hausdorff distance [177] between CTVs, calculated pairwise using the MILXView platform [162], as this is approximately similar to the true SD in observer contours. The frequency of coordinate system failure was determined by finding the number of times the MDU was limited. To determine the effect this coordinate system failure has on the overall volume, the DSC [167] comparing the original structure and the interpolated structure was determined.

8.2.3 Assessment

The MDU derived from all datasets in the cohort and the volume specific MDU derived from the datasets in each volume category were statistically compared to justify the use of volume categories. The approach that produced the most favourable results in terms of accounting for contouring variation, whilst including minimal extra healthy tissue (extra-included tissue, EIT) and including minimal malignant tissue (missed tissue, MT), was used for the rest of the study. Statistical analysis was performed using a Wilcoxon Signed

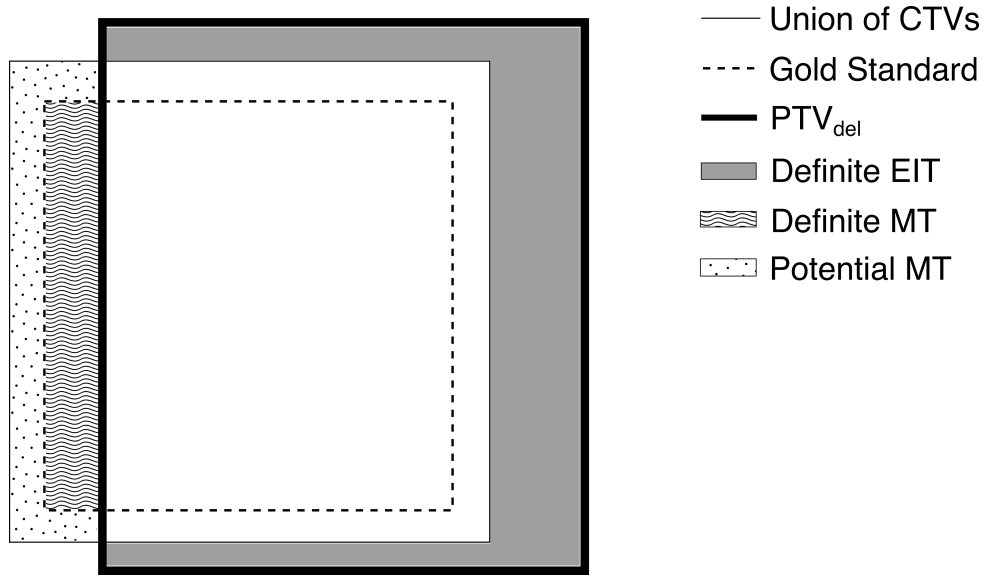


Figure 8.2: Example schematic outlining the components of MT and EIT resulting from the delineation margin applied to a target volume. Potential EIT not depicted as this only occurs when the coordinate system fails.

Rank test in IBM SPSS Statistics v21.0 with $p < 0.05$ significance [150].

The coverage of PTV_{del} with each original CTV (Equation 8.2) was determined to gauge the robustness of the inter-observer uncertainty encompassment for different margin approaches. Original CTVs were used instead of interpolated volumes to minimise potential errors arising from coordinate system failures. Successful coverage exceeded 95% [11].

$$Coverage(\%) = \frac{PTV_{del} \cap CTV_i}{CTV_i} \times 100 \tag{8.2}$$

MT and EIT were defined for all datasets using Equations 8.3, 8.4, 8.5, see Figure 8.2.

Definite and potential components to EIT and MT were defined based on whether the tissue lies within the GS volume.

Definite MT is tissue not included in PTV_{del} but lies within both the union of CTVs and the GS volume.

$$DefiniteMT(\%) = \frac{GS - GS \cap PTV_{del}}{GS} \times 100 \tag{8.3}$$

Ideally, there would be no MT within the GS volume.

Potential MT is tissue not included in PTV_{del} or the GS volume, but is within the union of CTVs.

$$PotentialMT(\%) = \frac{CTV_{Union} - (CTV_{Union} \cap PTV_{del}) - GS + (GS \cap PTV_{del})}{GS} \times 100 \tag{8.4}$$

The number of instances exceeding 5% potential MT averaged across all patients would be a logical tolerance based on the remaining 5% of tissue left in a successful CTV coverage of $> 95\%$. However, the number of instances exceeding 10% potential MT was considered for a more lenient approach incorporating the impact of interpolation errors.

Definite EIT refers to tissue that is included within PTV_{del} but not in the GS or the union of CTVs.

$$DefiniteEIT(\%) = \frac{PTV_{del} - (CTV_{Union} \cap PTV_{del})}{GS} \times 100 \tag{8.5}$$

For each patient, the GS volume was expanded by 5 mm in all cardinal directions to obtain a PTV volume in line with published clinical whole breast radiotherapy practices [75, 178, 179]. The percentage volume difference between the resulting volume and the GS volume was determined to be on average 37%. The tolerated amount definite EIT was therefore accepted to be less than 37% of the union of CTVs.

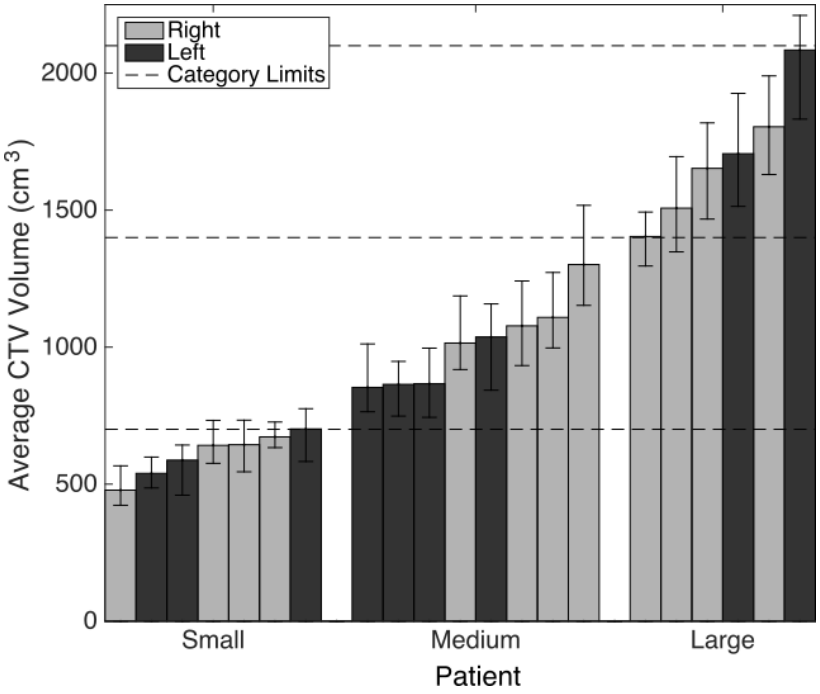


Figure 8.3: Whole breast CTV laterality, average volume and range (bars) for each patient in the cohort. Each patient's category based on average whole breast volume.

8.3 Results

8.3.1 Implementation

The laterality, range of volumes and average volume for each patient within the patient cohort are outlined in Figure 8.3. Small, medium and large volume categories were defined as 0-700 cm^3, 701-1400 cm^3 and 1401-2100 cm^3 respectively based on the distribution of CTV volumes across the cohort.

The margin magnitude in the cardinal directions and other selected directions for each margin approach and volume category are presented in Table 8.1. Although no polar margin expansion was explicitly applied in the z direction, the variation in start and finish slices of contours used in margin determination resulted in a variable Z component to the margin application.

Application of the different margin approaches is visually depicted in Figure 8.4, showing the three-dimensional encompassment of the GS contour by $PTV_{del,GS}$.

Table 8.1: MDU magnitudes in cardinal directions (mm). Right and left categories are subdivided into small, medium and large categories.

Category	System	Right (mm)	Left (mm)	Ant (mm)	Post (mm)	Sup (mm)	Inf (mm)
Right Small	Spherical	5.4	5.4	4.5	7.2	5.2	11.2
	Polar	5.0	6.1	3.0	9.1	-	-
	Conventional	4.1	1.6	4.2	9	6.7	8.3
Right Medium	Spherical	7.2	7.2	6.3	11.5	5.2	8.3
	Polar	6.5	9.7	7.2	11.2	-	-
	Conventional	5.7	2.4	6.3	11.4	6.2	7.2
Right Large	Spherical	12.8	12.8	10.1	13.5	3.5	3.9
	Polar	14.2	16.1	10.7	17.6	-	-
	Conventional	9.8	3.8	10.1	13.4	3.7	3.7
Left Small	Spherical	4.6	4.6	3.2	5.9	5.8	7.3
	Polar	4.5	4.5	3.6	6.1	-	-
	Conventional	5.0	5.9	3.2	5.8	6.3	6.8
Left Medium	Spherical	14.5	14.5	12.0	13.9	9.7	15.4
	Polar	13.1	10.0	12.9	12.7	-	-
	Conventional	15.5	15.8	12	14.1	11.6	13.6
Left Large	Spherical	6.1	6.1	5.1	12.6	8.8	6.9
	Polar	8.9	5.0	4.0	12.5	-	-
	Conventional	6.4	6.9	5.1	12.7	8.1	7.6

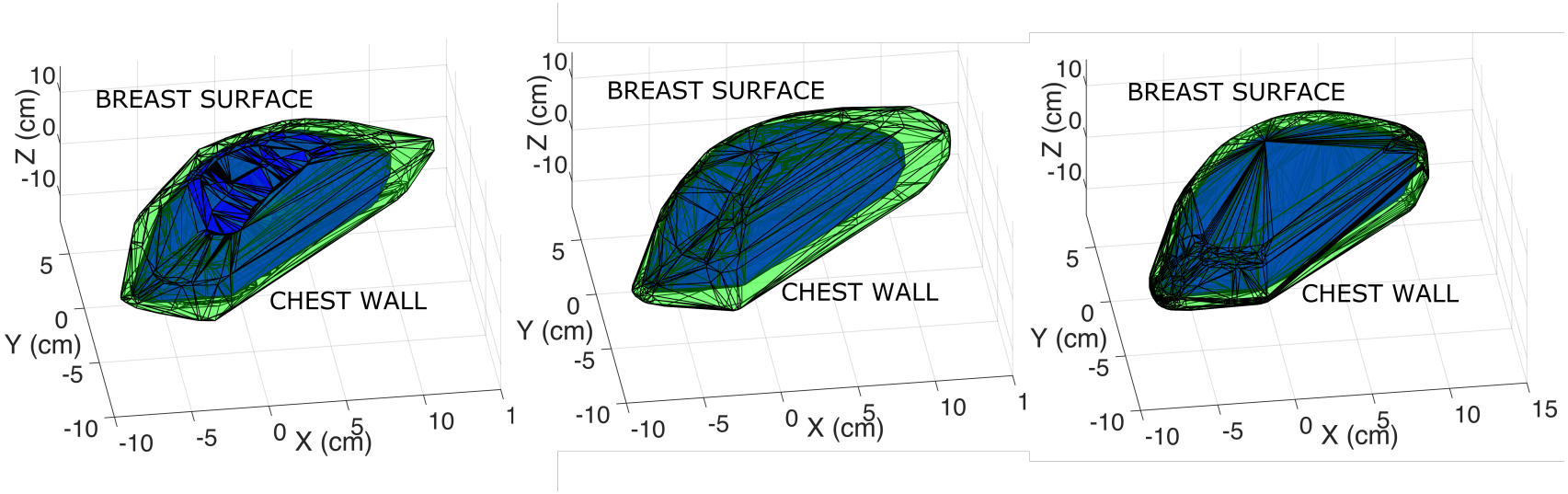


Figure 8.4: $PTV_{del,GS}$ (green) defined using (a) polar, (b) spherical and (c) cartesian coordinates compared to the GS volume (blue) in three dimensions for a medium sized left breast patient. X corresponds to lateral (negative)/medial (positive), y corresponds to anterior (positive)/posterior (negative) and z corresponds to superior (positive)/inferior (negative).

The main visual difference in margin approaches occurred in the anterior-medial and posterior-lateral directions in which inter-observer variation for whole breast radiotherapy is known to be largest [23].

Consistent coordinate system failure, corresponding to the maximum MDU values averaged over all datasets in each volume category (Table A.7), occurred most often between theta angles 0° and 90° for polar defined margins and between 270° and 360° for spherically defined margins. Excluding one outlier patient, this occurred on average in (1.35 $\pm$ 2.29)% of data points for polar margins and (0.56 $\pm$ 1.34)% of data points for spherical margins. The impact of this coordinate system failure on the overall volume was minimal as evidenced by high DSC values for the original structures and the interpolated structures (Table A.8). Average DSC across all structures when using a polar interpolation method was 0.95 $\pm$ 0.05 and for spherical interpolation was 0.98 $\pm$ 0.02. A DSC of 1 indicates identical volumes.

8.3.2 Assessment

Comparing the MDU derived using the whole cohort to the MDU derived using the volume specific method, a volume specific MDU resulted in significantly greater scores for all margin approaches ($p < 0.001$ each for polar, spherical and cartesian). Although there was significantly more definite EIT included for volume specific, spherically defined ($p < 0.001$), there was significantly less definite MT for polar margins ($p = 0.04$) and less potential MT for spherical MDUs ($p = 0.01$).

These results show that the use of volume categories in MDU derivation produced the most favourable results overall compared to utilising the whole cohort. This is likely because a margin defined on target volumes that varies largely in shape and size, is expected to account for delineation uncertainty in an average target volume well, but outlier volumes poorly. Volume categories were therefore utilised for the delineation margin assessment. The results of the CTV coverage, EIT and MT are presented for a

Table 8.2: Successful coverage of CTVs due to application of delineation margin to different target volumes. 'Av' is shortened from 'average'.

	Small	GS	Large
Polar	110/168, Av = 95.4%	150/168, Av = 98.1%	152/168, Av = 98.5%
Spherical	85/168, Av = 94.0%	151/168, Av = 98.3%	161/168, Av = 99.2%
Conventional	138/168, Av = 96.9%	164/168, Av = 98.8%	164/168, Av = 98.9%

volume specific MDU.

Applying the MDU to the smallest, largest and GS volumes, the coverage of PTV_{del} with each CTV was determined. Successful coverage was achieved for most CTVs regardless of target volume size, with greater than 95% coverage occurring in 412/504 (81.7%) cases (21 patients, 8 observers, 3 target volumes) for polar defined margins and 397/504 (78.8%) cases for spherically defined margins. For the conventional method of margin expansion, greater than 95% overlap occurred in 466/480 (92.4%) cases. The results were averaged over all contours and all patients in each volume category and are presented in Figure 8.5.

Table 8.2 outlines the differences in the number of CTVs successfully covered by the MDU specific to the target volume the margin was applied to, as well as the average coverage in each category. Each category contains 168 data points corresponding to 21 datasets and 8 observers.

Average coverage of CTVs by a MDU is high, with $> 95\%$ for all margin approaches and all target volumes, except for spherical MDUs applied to small target volumes. Coverage is observed to increase and the SD decreases as target volumes increase. This can be explained by considering the relative size of the CTVs being encompassed. PTV_{del},largest can be expected to encompass more CTVs than PTV_{del},smallest because the target volume it is applied to is larger than all other CTVs. This is not expected to occur in all cases however, since target volume may be shifted, and the shape may vary. Perfect overlap of all CTVs in all cases may not be achieved if CTVs are shifted relative to PTV_{del}.

The statistical differences in CTV overlap between pairs of margin approaches were

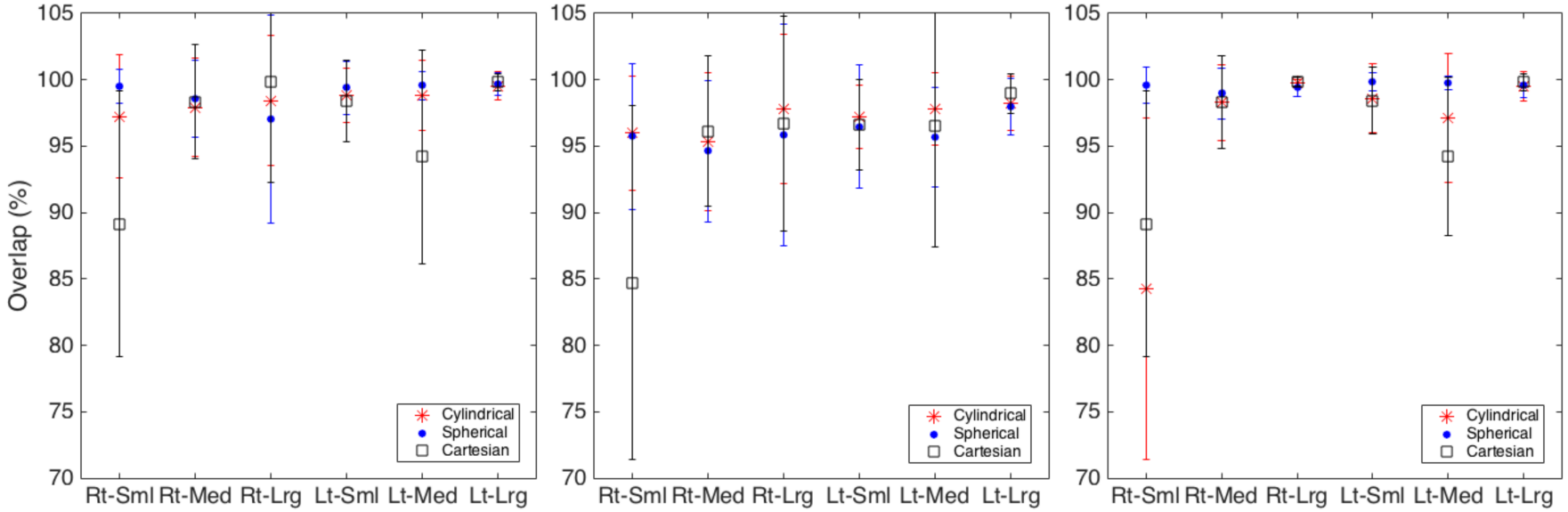


Figure 8.5: Overlap of PTV_{del} with observer CTVs, averaged (with SD) within each volume category. Delineation margin applied to (a) small, (b) GS and (c) large target volumes.

assessed (Table A.9). Summarising, conventional margin scores are significantly greater than both anisotropic approaches for small target volumes. For GS target volumes, conventional margin scores are significantly greater than cylindrical margin scores. For large target volumes, spherical margin scores are significantly greater than conventional margin scores, which in turn are greater than polar margin scores. These significance results may be due to the fact that as whole breast CTVs increase in volume, they become more spherical and hence the SD in delineation uncertainty for these shapes is better described in spherical coordinates.

The amount of MT averaged across all structures and patients in each volume category is presented in Figure 8.6.

The amount of potential MT within the 10% limit in 39/63 (61.9%) cases (21 datasets and 3 target volumes) for polar margins, 36/63 (57.1%) cases for spherical margins and 42/60 (71.4%) cases for conventionally defined margins. This indicates that in the majority of cases, all delineation margin approaches missed a tolerated amount of potential tissue. Generally, margins applied to small target volumes and GS volumes failed to meet tolerance more frequently than large target volumes because $PTV_{del,largest}$ often includes more tissue, reducing the chance of missing tissue. For small target volumes, right breast patients had larger amounts of potential MT when using a conventionally defined margin than left breasted patients perhaps due to differences in contouring of especially crescent shaped target volumes close to the chest wall, where particular attention may be paid to left breast volumes because of the proximity of the heart.

For definite MT, 0/63 (0%) cases met the zero tolerance criteria for polar margins, however 41/63 (65.1%) cases missed less than 1% of the GS volume. For spherically defined margins, 31/63 (49.2%) cases met the zero tolerance criteria and 44/63 (69.8%) missed less than 1% of the GS volume. For conventionally defined margins, 1/63 (1.6%) cases met the zero tolerance criteria and 57/63 (90.5%) missed less than 1% of the GS

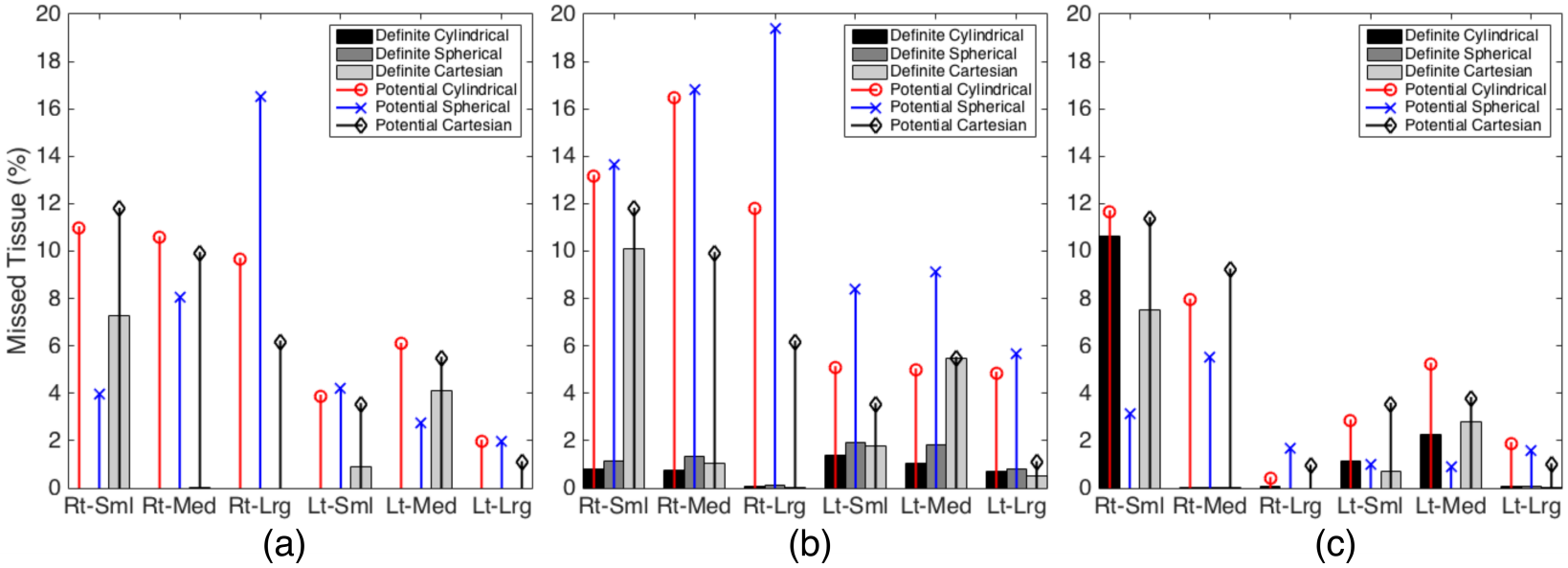


Figure 8.6: Definite and potential components of the tissue missed by the delineation margin averaged within each volume category for polar, spherically and cartesian defined margins. Delineation margin applied to (a) small, (b) GS and (c) large target volumes.

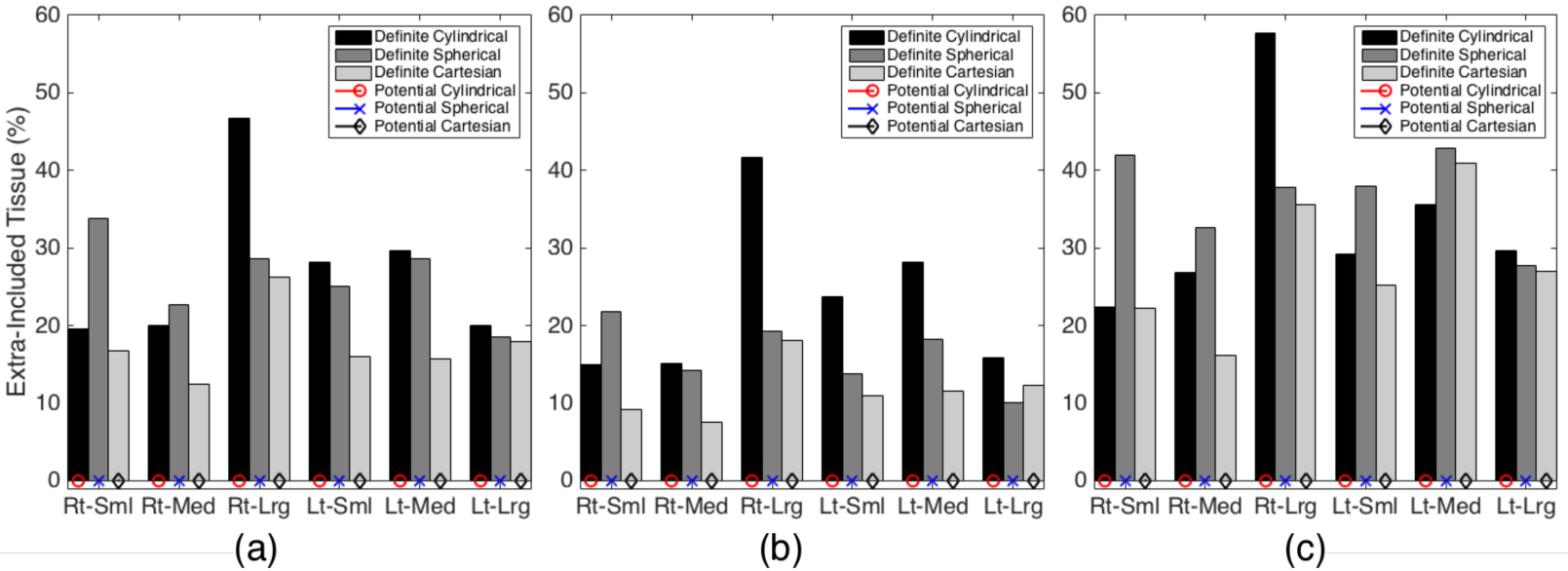


Figure 8.7: Definite and potential components of the extra tissue encompassed by the delineation margin averaged within each volume category for polar, spherically and cartesian defined margins. Delineation margin applied to (a) small, (b) GS and (c) large target volumes.

volume. These results show that although some tissue was missed, it is a minimal fraction of the GS volume in the majority of cases.

The amount of EIT included in PTV_{del} averaged across all structures and patients in each volume category is presented in Figure 8.7. Definite and potential components are shown.

The amount of definite EIT fell within the $< 37\%$ tolerance for the majority of patients and target volumes, with 60/63 (95.2%) cases including less than 37% of the GS for polar margins, 56/63 (88.9%) cases for spherical margins and 61/63 (96.8%) cases for conventionally defined margins. This indicates that no delineation margin approach includes excessive normal tissue. Generally, MDUs applied to the largest volume failed to meet tolerance more frequently than other target volumes probably because remaining CTVs smaller and more likely to be easily encompassed, resulting in the inclusion of greater excess normal tissue.

The results of the pairwise statistical analysis comparing margin approaches for definite EIT, potential MT and definite MT is presented in Tables A.10, A.11 and A.12. The Wilcoxon Signed Rank Test was used with $p< 0.05$ significance and scores are presented as the data is negatively skewed.

The results from Tables A.10, A.11 and A.12 indicate that a conventional margin misses less definite MT for small volumes and includes less EIT for GS volumes. Although missing less potential MT, a polar approach misses less definite EIT than a conventional approach for large target volumes. As mentioned earlier, this trend towards the benefit of an anisotropic with increasing target volume may be due to use of a coordinate system that better describes the changing CTV shape for large target volumes.

The location of MT and EIT relative to organs at risk is important. Figure 8.8 visually depicts the encompassment of $PTV_{del,GS}$ with the union of CTVs for a dataset representative of the patient cohort. There is little MT observed overall which reflects the low potential MT and definite MT values determined. EIT occurs mostly in the

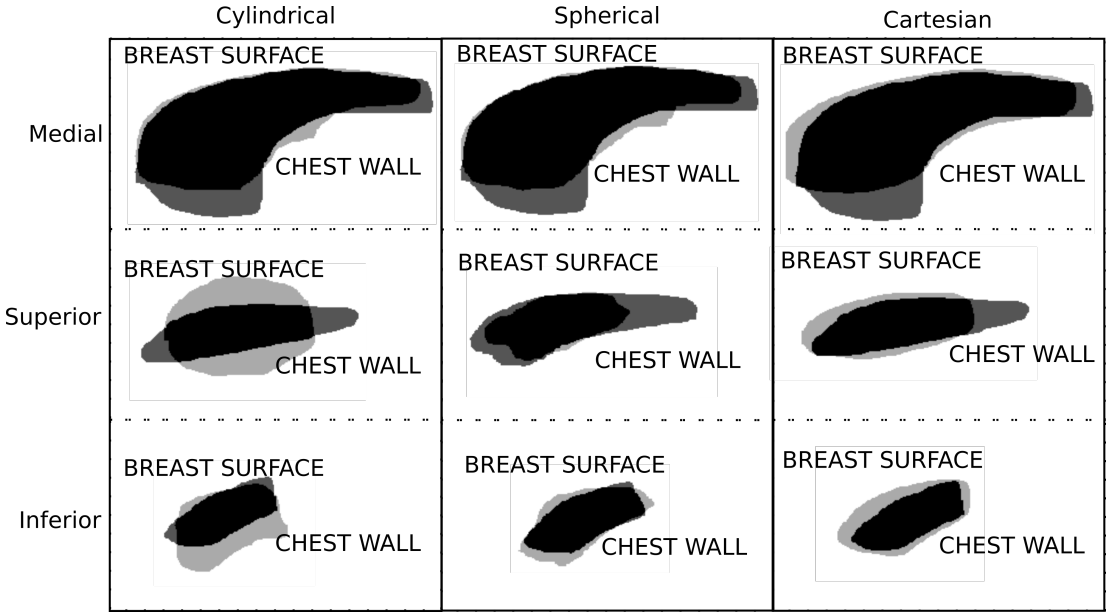


Figure 8.8: Comparison of total missed tissue (dark grey), total extra tissue (light grey) due to polar, spherically and cartesian defined $PTV_{del,GS}$ (black) for a medium volume left breast patient. Medial, superior and inferior slices are shown.

anterior-medial and posterior-lateral directions for polar and spherical approaches and occurs in all directions for the conventional approach. The added tissue at the chest wall for a conventional margin may impact lung doses. EIT at the breast surface is minimal due to clipping of the margin here.

8.4 Discussion

It is acknowledged that there is justification for a delineation margin in radiotherapy [14, 159, 169]. Although derived for CTV shifts and displacements as opposed to shape deformations, a 2.5Σ margin for systematic uncertainties has been shown to be valid in accounting for delineation uncertainty [14, 180]. Furthermore, Peulen et al validated that the required MDU in non-small cell lung tumours is 2.8Σ and Pogson et al proposed a whole breast radiotherapy margin of 2Σ [159, 169]. However these studies defined margins that did not explicitly investigate the spatial dependence of delineation uncertainty across the whole target surface and therefore may not have adequately accounted for delineation uncertainty in regions of largest variation. Other studies have

utilised polar coordinates to determine inter-observer variability [38, 127] and have proposed isotropic MDU approaches [174], however this is the first study to combine these and assess the outcome. This study uses whole breast as an example case and provides the first investigation on an anisotropic delineation margin accounting for contouring uncertainty where it is largest. The anisotropic approach has been compared to conventional margin application approaches where cardinal cartesian margins are used enabling an assessment on the need to determine delineation uncertainty margins differently to other uncertainties in clinical margin definitions. This study also presents a novel approach to assessing the impact of treatment margins considering both cancerous tissue that may not be included within a margin and normal tissue that is included within a margin.

8.4.1 Implementation

The margin magnitudes determined in this study, range from 3.2-14.5 mm for spherical margins (average 8.3 mm), 3.0-17.6 mm for polar margins (average 8.9 mm) and 1.6-15.8 mm for cartesian defined margins (average 7.8 mm) across all volume categories and cardinal directions. Peulen et al, obtained an isotropic GTV to PTV margin for non-small cell lung cancer ranging from 3.4-5.9 mm [159]. This is smaller than the values obtained in this study however the SD determined in their study was relatively small (1.2-1.8 mm) and the anatomical site tested was different. A typical whole breast radiotherapy CTV to PTV margin expansion used in clinical practice is 5 mm expansion in the cardinal directions. The average magnitudes of this delineation margin are larger than that of clinical margins however this may be attributed to the fact that clinical margins often do not incorporate delineation uncertainty and do not spatially vary across the target volume surface.

8.4.2 Assessment

A delineation margin for whole breast radiotherapy has been effectively implemented, showing acceptable observer CTV coverage ($> 95\%$ in the majority of cases) and minimal EIT and MT for a weighting factor of α=2. Both anisotropic and conventional approaches produced favourable results, however defining a margin in the cardinal directions as is done clinically, is most appropriate for small and GS volumes. For small target volumes, a conventional approach produces significantly greater overlap of CTVs, less definite EIT and less potential MT. For GS target volumes, significantly greater overlap and less definite MT is achieved for a conventional approach. This is despite the largest delineation uncertainty occurring in regions outside these six main directions. This suggests that delineation uncertainty margins for target volumes < 1400 cm^3 are not sensitive to inter-observer variation outside the cardinal directions.

For large target volumes however, there is a need for a delineation margin to be anisotropic to account for spatially varying inter-observer variation. Significantly greater overlap occurs with this approach as opposed to a conventional approach, and significantly less definite EIT and definite MT occur. Specifically, a spherically defined margin is most appropriate, with more favourable significant results as compared to a polar approach. This difference for large target volumes in the most appropriate margin approach may be due to the fact that large target volumes are more spherical in nature and are therefore better described using spherical coordinates.

The implementation of the proposed delineation margin may impact patient outcomes. A study classifying local disease recurrences after breast conserving surgery and radiotherapy found that most recurrences occur in the upper two quadrants of the breast [181]. Radiotherapy in the anterior region of breast CTV must therefore be highly precise to enhance effectiveness of the treatment. The assessed anisotropic delineation margin will help achieve this goal for large target volumes, as it explicitly accounts for the higher inter-observer variation in the anterior-medial direction of whole breast target

volume. For all other target volumes, a conventional delineation margin approach is adequate to account for this variation.

This study has proposed and implemented a MDU that adequately accounts for inter-observer variation in whole breast radiotherapy. Future work may involve optimising the weighting factor improve coverage, ET and MT across the entire cohort.

8.4.3 Limitations

There are limitations in using a radial-based method for defining the spatial variation in target volume delineation, with radial projections potentially inadequate in describing the target shape. However, a radial-based method is necessary in order to easily and consistently re-produce points at which to define and apply the margin across different structures and different patients. Acknowledging these limitations, this study utilises basic coordinate systems in this first instance of defining a delineation margin, so as not to overcomplicate the method. More advanced mathematics could have been used, such as multiple coordinate systems, however this would likely result in the same problems due to inherent inconsistencies in whole breast radiotherapy contours across observers and patients.

This inherent limitation was found to have minimal impact on the approach, with on average, $< 1.5\%$ of data points for a polar margin and $< 0.6\%$ for a spherical margin affected. High DSC values (> 0.94 for polar and > 0.98 for spherical) between interpolated structures and original structures support this. Using this method for defining a delineation margin for more spherically shaped volumes like prostate, would likely yield an even smaller error [176].

A cohort consisting of N = 8 observers was assumed to be adequate in producing a normal distribution for SD calculation since there was noticeable variability in delineation. In the event that a smaller number of observers were to be used or there was insufficient variation amongst observers, for example, 8 observers that consistently delineate smaller

target volumes compared to their peers, the delineation margin would be highly dependent on the input delineations and hence limited in its widespread accuracy. As the number of observers increases, the more applicable the margin is in effectively accounting for a range in delineation uncertainty.

This study was designed as a preliminary investigation into methods for accounting for delineation uncertainty in a spatially varying sense. The whole breast CT dataset was used primarily as it was an available dataset in which to develop this methodology. This methodology assumes an idealistic scenario where setup and motion uncertainties are minimal, reflecting the direction modern radiotherapy is approaching. A more realistic scenario for current radiotherapy practice would be to include delineation uncertainty with methods for accounting for existing treatment uncertainties. The optimum way to combine an anisotropic delineation margin with existing uncertainty margins and in treatment planning systems, warrants further investigation.

Chapter 9

Discussion

This thesis has presented a body of work investigating the magnitude of inter-observer variation, its impact on the radiotherapy process and how best to manage its presence. Although two specific treatment sites were investigated, the concepts and methods presented can be applied to any radiotherapy treatment site.

Inter-observer variation can be quantified, understood, minimised and managed by following the methods outlined in this thesis. Chapters 3 and 5 provide examples of the ways inter-observer variation can be quantified using a number of volumetric and dimensional metrics. It is important to quantify the magnitude of inter-observer variation such that its relative importance with respect to other uncertainties in the radiotherapy treatment process can be determined. Also, the SD of this uncertainty is required if it is to be incorporated into uncertainty margins. However, once the magnitude of inter-observer variation is known, it is clinically important to determine the impact it has on treatment outcomes. Chapters 3, 4 and 5 outline how to determine to what extent inter-observer variation impacts DVH parameters in a patients treatment plan. Chapters 4 and 5 further demonstrate how DVH parameters can be correlated with contouring metrics such that a delineation can be evaluated with respect to the resulting treatment plan. With the magnitude and impact of inter-observer variation known, attempts to minimise its impact can be implemented. Chapter 6 provides an example of one such

method to minimise inter-observer variation by automating the delineation process whilst maintaining the value provided by subjective interpretation of imaging datasets by clinicians. However, inter-observer variation remains despite these efforts as long as the 'ground truth' target volume remains unknown. Hence, Chapter 7 and 8 investigate methods to account for the remaining inter-observer variation by defining and applying a margin to account for it. This method aligns with the current procedure of CTV to PTV expansions for other uncertainties.

The cumulative outcome of this body of work is a workflow for managing inter-observer variation that can be useful for clinical trials and in individual treatment centres; first assessing current variation, considering the dosimetric impact of this and then investigating margins to account for this where necessary. Adhering to this workflow before commencing clinical trials would enable consistency in treatment planning and dose reporting such that trial interventions can be clearly assessed. Individual treatment centres can follow this workflow to characterise and customise their procedures for their specific patient cohort.

9.1 Research Aims

The specific aims of this body of work are addressed below.

Aim 1. Understand the impact of delineation uncertainty on radiotherapy efficacy

Research questions:

What is the magnitude of inter-observer variation for high dose rate cervical cancer brachytherapy in the Australian context?

Chapter 3 quantified the variability in CTV_{HR} delineation for HDR cervical cancer brachytherapy in the Australian context and assessed the dosimetric consequences of this variation on a gold standard target volume.

Inter-observer variation has been shown to be present in the case of HDR cervical cancer brachytherapy. Chapter 3 found that overall, CTV_{HR} contouring was relatively consistent with an average DSC of ≥ 0.7 occurring in 8/10 cases assessed on the STAPLE gold standard, and in 6/10 cases assessed on the consensus contour. However, there was a **wide range in variability** when considering each individual contour, with an average MASD of 7.13 ± 6.37 mm and an average DSC of 0.63 ± 0.15.

Chapter 5 quantified spatially descriptive contouring metrics for the CTV_{HR} inter-observer variation in the same dataset used in Chapter 3. The location of the lateral, anterior/posterior and superior/inferior positions of the COM relative the gold standards COM varied on average 0.16 cm, 0.27 cm and 0.87cm respectively. The maximum dimensions demonstrated 0.75 - 1.82cm variation and a mean volume variation of 28.5 cm^3 was observed.

How does inter-observer variation affect the dosimetric parameters obtained during the radiotherapy treatment planning process?

The inter-observer variability in HDR cervix brachytherapy observed in the Chapter 3 translated to a dosimetric variability. A gold standard plan optimised to the gold standard contours was assessed on each observer contour. The average relative standard deviation in D90 across all patients was 27% and 28% respectively for the STAPLE and consensus plans. The average standard deviation in D100 was 34% and 35% respectively for the STAPLE and consensus plans. **The average dosimetric uncertainty in CTV_{HR} for an individual fraction (with a prescription of 8 Gy) due to inter-observer variation was approximately 1.5 Gy.**

The dose planning variation that arises from inter-observer variation when individual plans are optimised to individual contours was assessed in

Chapter 4. **Over the course of a four-fraction treatment, 2.9 - 3.4 Gy $EQD2_{10}$ uncertainty was observed for the CTV_{HR}** depending on which gold standard contour was used for reporting. The single fraction variability was 4.5 - 6.0%. This is substantially smaller than the results found in Chapter 3, however it is pertinent to highlight to the reader that the methodologies of these two studies were fundamentally different as outlined in Figure 4.1.

Chapter 5 utilised the dataset of Chapter 3, but applied the same methodology as in Chapter 4. **For an entire four fraction brachytherapy course, CTV_{HR} D90 varied by 1.9 Gy for the STAPLE contour and 2.8 Gy for the STAPLE contour.** This is slightly smaller variability than observed in Chapter 4, likely due to the differences in the datasets as outlined in Table 5.6.

Do common metrics used to assess contouring consistency correlate with dosimetric outcomes?

Volume and maximum dimension metrics used to quantify the magnitude of inter-observer variation were shown in Chapter 4 to correlate with CTV_{HR} D98. Specifically, a decrease in the lateral, anterior/posterior and superior/inferior dimensions correlate most with an increase in D98. These correlations were weak to moderate in strength. This was validated in Chapter 5 using an independent dataset. The correlations between lateral dimensions and D98, as well as volume and D98 were found to be statistically the same between both datasets.

Aim 2. Minimise delineation uncertainty by automating the contouring process

Research questions:

Can an accurate automatic segmentation method be defined to reduce the magnitude of delineation uncertainty?

Chapter 6 proposed and investigated **an atlas-based auto-segmentation method that was shown to accurately contour whole breast external beam radiotherapy target contours** in less than 4 minutes on average. The auto-segmented contours had good similarity when compared to the gold standard STAPLE contour with a DSC > 0.7 and a MASD < 1 cm. The utilisation of this approach in the radiotherapy treatment planning process would eliminate inter-observer variation arising from manual contouring.

Aim 3. Manage delineation uncertainty using an uncertainty margin

Research questions:

What is the best coordinate system to define a delineation uncertainty margin for whole breast external beam radiotherapy?

For a whole breast external beam radiotherapy dataset, a cylindrical coordinate system is most appropriate for defining a delineation margin than cartesian coordinates due to the hemi-spherical shape of the target volumes. Chapter 7 demonstrated that for the central slice, the standard deviation in contour dimension peaked medially at 15° (3.55 cm) for right breast datasets and 171° (1.44 cm) for left breast datasets, indicating the largest inter-observer variation occurs anterior-medially. Cartesian coordinates cannot accurately describe the hemispherical breast shape in these regions, therefore are inadequate at defining an anisotropic delineation uncertainty margin.

Chapter 8 extended the work in Chapter 7 to investigate spherical

coordinate systems. This study highlighted the difficulty in using any coordinate system based on radial projections from a central point for contours that are hemispherical. Nevertheless, **a spherically defined delineation uncertainty margin was deemed to be most appropriate for large breast target volumes** ($>$ 1400 cm^3) due to the increase coverage of the margins. This is likely due to the added superior-inferior information available that is otherwise lost with polar coordinate systems. Defining a margin in the cardinal dimensions (cartesian coordinate system) as is done clinically for other uncertainties, is most appropriate for small volumes ($<$1400 cm^3) as these volumes are less sensitive to inter-observer variation outside the cardinal directions.

Can inter-observer variation be accounted for appropriately with an uncertainty margin?

A whole breast external beam radiotherapy delineation margin to account for inter-observer variation has been defined and applied in Chapter 8. Delineation uncertainty margin magnitudes ranged from 3.2-14.5 mm for spherically defined margins 3.0-17.6 mm for polar margins and 1.6 -15.8 mm for Cartesian defined margins. Acceptable CTV observer coverage was determined with $>$ 95% of observer contours including in the delineation margin for all coordinate systems. 'Extra-included tissue', that is, tissue not contoured by any observer and 'missed tissue' which refers to tissue contoured by at least one observer, but not covered by the delineation uncertainty margin were within the accepted tolerances for the majority of cases.

9.2 Future Work

At the conclusion of this work, there is scope for further investigation in a number of areas.

The atlas-based method for auto-segmentation proposed in Chapter 6 utilised an approach that was the most advanced and common approach at the time of publication. In the following years, automatic segmentation techniques have expanded into the realm of machine-learning techniques. An investigation utilising a similar methodology but applying machine-learning algorithms to delineate target volumes is warranted. This approach would offer increased accuracy and increased speed in delineation.

The delineation margin proposed in Chapter 8 relies on an origin based coordinate system. This dependence introduces limitations for non-spherical target volumes. This work could be expanded to investigate deformedly registered datasets in which the margin is defined based on surface variations between observers. Additionally, with an idealistic move away from binary contours to probabilistic target regions, a conceptual delineation uncertainty probability map could be defined as opposed to delineation uncertainty margin. This would enable treatment planning isodoses to be conformed to isolines representing the likelihood of observer agreement. As such, the impact of inter-observer variation would be further reduced.

Finally, this thesis investigated specific aims on available retrospective datasets. It would be interesting to repeat the work for other treatment sites, especially the delineation margin study which could be assessed for robustness on different treatment sites.

Chapter 10

Conclusion

The resulting outcome of this thesis is a workflow that can be utilised to improve the accuracy of patient cancer treatments.

This thesis has demonstrated that inter-observer variation is present and for some specific patients, substantial, for high-dose rate cervical cancer brachytherapy. Although the presence of any uncertainty is reason for pause in radiotherapy, the more pertinent question is whether the presence of inter-observer variation has an impact on the treatment plan delivered to the patient. For both European GEC-ESTRO and Australian datasets, the dosimetric impact was demonstrated to be substantial, indicating that methods to manage this uncertainty should be investigated.

Acknowledging that inter-observer variation should be reduced where possible, this thesis investigated a method to automatically contour target volumes (with specific application to whole breast radiotherapy) such that inter-observer variation arising from human judgment might be eliminated. It was shown that an atlas-based auto-segmentation algorithm can quickly and accurately delineate target volumes. However, manual review of the auto-contour is recommended and manual adjustments would re-introduce inter-observer variation.

After demonstrating that inter-observer variation has a large dosimetric impact and that its presence remains, even after minimising the human element in contouring, this thesis

has investigated a margin to manage this uncertainty. Margins used in radiotherapy are the current convention for accounting for uncertainties in the treatment process. A proof of concept delineation margin was proposed and investigated for whole breast radiotherapy. Despite limitations in this method, it is a successful proof of concept in effectively managing delineation uncertainty in radiotherapy.

Overall, this thesis acts as a guide in how to quantify, understand and manage inter-observer variation when implementing a new technique in a clinical department or clinical trial such that radiotherapy can be delivered accurately and safely for every patient.

Bibliography

[1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*,

[2] M. Barton, S. Jacob, J. Shafiq, K. H. W. Wong, S. Thompson, G. Delaney, and T. Hanna, "Review of optimal radiotherapy utilisation rates," Department of Health and Ageing Australian Government, Report, 2013.

[3] C. F. Njeh, "Tumor delineation: The weakest link in the search for accuracy in radiotherapy," *Journal of Medical Physics/Association of Medical Physicists of India*, vol. 33, no. 4, pp. 136–140, 2008.

[4] "ICRU Report 50: Prescribing, recording and reporting photon beam therapy," Report, 1993.

[5] "ICRU Report 62: Prescribing, recording and reporting photon beam therapy (supplement to ICRU Report 50)," Tech. Rep.

[6] M. van Herk, "Errors and margins in radiotherapy," *Seminars in Radiation Oncology*, vol. 14, no. 1, pp. 52–64, 2004.

[7] Royal College of Radiologists, *On target: Ensuring geometric accuracy in radiotherapy*, 2008.

[8] J. Stroom, K. Gilhuijs, S. Vieira, W. Chen, J. Salguero, E. Moser, and J. Sonke, "Combined recipe for clinical target volume and planning target volume margins," *International Journal of Radiation Oncology Biology Physics*, vol. 88, no. 3, pp. 708–714, 2014.

[9] J. C. Stroom, H. C. J. de Boer, H. Huizenga, and A. G. Visser, "Inclusion of geometrical uncertainties in radiotherapy treatment planning by means of coverage probability," *International Journal of Radiation Oncology Biology Physics*, vol. 43, no. 4, pp. 905–919, 1999.

[10] M. van Herk, P. Remeijer, and J. V. Lebesque, "Inclusion of geometric uncertainties in treatment plan evaluation," *International Journal of Radiation Oncology Biology Physics*, vol. 52, no. 5, pp. 1407–1422, 2002.

[11] T. Craig, J. Battista, V. Moiseenko, and J. Van Dyk, "Considerations for the implementation of target volume protocols in radiation therapy," *International Journal of Radiation Oncology Biology Physics*, vol. 49, no. 1, pp. 241–250, 2001.

[12] A. Bel, M. van Herk, and J. V. Lebesque, "Target margins for random geometrical treatment uncertainties in conformal radiotherapy," *Medical Physics*, vol. 23, no. 9, pp. 1537–1545, 1996.

[13] J. A. Antolak and I. I. Rosen, "Planning target volumes for radiotherapy: How much margin is needed?" *International Journal of Radiation Oncology Biology Physics*, vol. 44, no. 5, pp. 1165–1170, 1999.

[14] M. van Herk, P. Remeijer, C. Rasch, and J. V. Lebesque, "The probability of correct target dosage: Dose-population histograms for deriving treatment margins in radiotherapy," *International Journal of Radiation Oncology Biology Physics*, vol. 47, no. 4, pp. 1121–1135, 2000.

[15] A. L. McKenzie, M. v. Herk, and B. Mijnheer, "The width of margins in radiotherapy treatment plans," *Physics in Medicine and Biology*, vol. 45, no. 11, p. 3331, 2000.

[16] B. C. Parker, A. S. Shiu, M. H. Maor, F. F. Lang, H. H. Liu, R. A. White, and J. A. Antolak, *PTV margin determination in conformal SRT of intracranial lesions*. 2002.

[17] R. Topolnjak, C. van Vliet-Vroegindeweij, J.-J. Sonke, D. Minkema, P. Remeijer, J. Nijkamp, P. Elkhuizen, and C. Rasch, "Breast-conserving therapy: Radiotherapy margins for breast tumor bed boost," *International Journal of Radiation Oncology Biology Physics*, vol. 72, no. 3, pp. 941–948, 2008.

[18] A. McKenzie, M. van Herk, and B. Mijnheer, "Margins for geometric uncertainty around organs at risk in radiotherapy," *Radiotherapy and Oncology*, vol. 62, no. 3, pp. 299–307, 2002.

[19] J. Bernier, E. J. Hall, and A. Giaccia, "Radiation oncology: A century of achievements," *Nature Reviews Cancer*, vol. 4, p. 737, 2004.

[20] E. Weiss and C. F. Hess, "The impact of gross tumor volume (GTV) and clinical target volume (CTV) definition on the total accuracy in radiotherapy," *Strahlentherapie und Onkologie*, vol. 179, no. 1, pp. 21–30, 2003.

[21] L. J. Peters, B. O'Sullivan, J. Giralt, T. J. Fitzgerald, A. Trotti, J. Bernier, J. Bourhis, K. Yuen, R. Fisher, and D. Rischin, "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*, vol. 28, no. 18, pp. 2996–3001, 2010.

[22] V. Batumalai, E. S. Koh, G. P. Delaney, L. C. Holloway, M. G. Jameson, G. Papadatos, and D. M. Lonergan, "Interobserver variability in clinical target volume delineation in tangential breast irradiation: A comparison between radiation oncologists and radiation therapists," *Clinical Oncology*, vol. 23, no. 2, pp. 108–113, 2011.

[23] R. Moeckli, W. Jeanneret Sozzi, R.-O. Mirimanoff, M. Ozsahin, A. Zouhair, J.-F. Valley, and F. Bochud, "Physical considerations on discrepancies in target volume delineation," *Zeitschrift für Medizinische Physik*, vol. 19, no. 4, pp. 224–235, 2009.

[24] P. Giraud, S. Elles, S. Helfre, Y. De Rycke, V. Servois, M.-F. Carette, C. Alzieu, P.-Y. Bondiau, B. Dubray, E. Touboul, M. Housset, J.-C. Rosenwald, and J.-M. Cosset, "Conformal radiotherapy for lung cancer: Different delineation of the gross tumor volume (GTV) by radiologists and radiation oncologists," *Radiotherapy and Oncology*, vol. 62, no. 1, pp. 27–36, 2002.

[25] R. Steenbakkers, J. C. Duppen, I. Fitton, K. Deurloo, L. J. Zijp, E. Comans, A. Uitterhoeve, P. Rodrigus, G. Kramer, J. Bussink, K. De Jaeger, J. Belderbos, P. Nowak, M. van Herk, and C. Rasch, "Reduction of observer variation using matched ct-pet for lung cancer delineation: A three-dimensional analysis," *International Journal of Radiation Oncology*Biology*Physics*, vol. 64, no. 2, pp. 435–448, 2006.

[26] C. Rasch, R. Steenbakkers, I. Fitton, J. C. Duppen, P. Nowak, F. A. Pameijer, A. Eisbruch, J. Kaanders, F. Paulsen, and M. van Herk, "Decreased 3D observer variation with matched CT-MRI, for target delineation in nasopharynx cancer," *Radiation Oncology*, vol. 5, no. 1, p. 21, 2010.

[27] N. Usmani, R. Sloboda, W. Kamal, S. Ghosh, N. Pervez, J. Pedersen, D. Yee, B. Danielson, A. Murtha, J. Amanie, and T. Monajemi, "Can images obtained with high field strength magnetic resonance imaging reduce contouring variability of the prostate?" *International Journal of Radiation Oncology Biology Physics*, vol. 80, no. 3, pp. 728–734, 2011.

[28] E. Weiss, J. Wu, W. Sleeman, J. Bryant, P. Mitra, M. Myers, T. Ivanova, N. Mukhopadhyay, V. Ramakrishnan, M. Murphy, and J. Williamson, "Clinical evaluation of soft tissue organ boundary visualization on cone-beam computed tomographic imaging," *International Journal of Radiation Oncology Biology Physics*, vol. 78, no. 3, pp. 929–936, 2010.

[29] C. W. Hurkmans, J. H. Borger, B. R. Pieters, N. S. Russell, E. P. M. Jansen, and B. J. Mijnheer, "Variability in target volume delineation on CT scans of the

breast," *International Journal of Radiation Oncology Biology Physics*, vol. 50, no. 5, pp. 1366–1372, 2001.

[30] D. Genovesi, G. Ausili Céfaro, M. Trignani, A. Vinciguerra, A. Augurio, M. Di Tommaso, F. Perrotti, A. De Paoli, P. Olmi, V. Valentini, and M. Di Nicola, "Interobserver variability of clinical target volume delineation in soft-tissue sarcomas," *Cancer/Radiothérapie*, vol. 18, no. 2, pp. 89–96, 2014.

[31] G. Persson, D. Nygaard, C. Hollensen, F. Munck-af-Rpsemschold, L. Mouritsen, A. Due, A. Berthelsen, J. Nyman, E. Markova, A. Roed, H. Roed, S. Korreman, and L. Specht, "Interobserver delineation variation in lung tumour stereotactic body radiotherapy," *The British Journal of Radiology*, vol. 85, no. 1017, e654–e660, 2012.

[32] I. Fotina, C. Lütgendorf-Caucig, M. Stock, R. Pötter, and D. Georg, "Critical discussion of evaluation parameters for inter-observer variability in target definition for radiation therapy," *Strahlentherapie und Onkologie*, vol. 188, no. 2, pp. 160–167, 2012.

[33] M. G. Jameson, L. C. Holloway, P. J. Vial, S. K. Vinod, and P. E. Metcalfe, "A review of methods of analysis in contouring studies for radiation oncology," *Journal of Medical Imaging and Radiation Oncology*, vol. 54, no. 5, pp. 401–410, 2010.

[34] S. K. Vinod, M. G. Jameson, M. Min, and L. C. Holloway, "Uncertainties in volume delineation in radiation oncology: A systematic review and recommendations for future studies," *Radiotherapy and Oncology*, vol. 121, no. 2, pp. 169–179, 2016.

[35] M. G. Jameson, S. Kumar, S. K. Vinod, P. E. Metcalfe, and L. C. Holloway, "Correlation of contouring variation with modeled outcome for conformal non-small cell lung cancer radiotherapy," *Radiotherapy and Oncology*, vol. 112, no. 3, pp. 332–336, 2014.

[36] D. Roach, M. G. Jameson, J. A. Dowling, M. A. Ebert, P. B. Greer, A. M. Kennedy, S. Watt, and L. C. Holloway, "Correlations between contouring similarity metrics and simulated treatment outcome for prostate radiotherapy," *Physics in Medicine and Biology*, vol. 63, no. 3, p. 035 001, 2018.

[37] H. Struikmans, C. Wárlám-Rodenhuis, T. Stam, G. Stapper, R. Tersteeg, G. H. Bol, and C. Raaijmakers, "Interobserver variability of clinical target volume delineation of glandular breast tissue and of boost volume in tangential breast irradiation," *Radiotherapy and Oncology*, vol. 76, no. 3, pp. 293–299, 2005.

[38] C. Rasch, I. Barillot, P. Remeijer, A. Touw, M. van Herk, and J. V. Lebesque, "Definition of the prostate in CT and MRI: A multi-observer study," *International Journal of Radiation Oncology Biology Physics*, vol. 43, no. 1, pp. 57–66, 1999.

[39] G. G. Hanna, A. R. Hounsell, and J. M. O'Sullivan, "Geometrical analysis of radiotherapy target volume delineation: A systematic review of reported comparison methods," *Clinical Oncology*, vol. 22, no. 7, pp. 515–525, 2010.

[40] P. Jaccard, "Etude comparative de la distribution florale dans une portion des alpes et des jura," *Bull Soc Vaudoise Sci Nat*, vol. 37, pp. 547–579, 1901.

[41] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.

[42] S. K. Warfield, K. H. Zou, and W. M. Wells, "Simultaneous truth and performance level estimation (STAPLE): An algorithm for the validation of image segmentation," *Medical Imaging, IEEE Transactions on*, vol. 23, no. 7, pp. 903–921, 2004.

[43] Z. Gao, D. Wilkins, L. Eapen, C. Morash, Y. Wassef, and L. Gerig, "A study of prostate delineation referenced against a gold standard created from the visible human data," *Radiotherapy and Oncology*, vol. 85, no. 2, pp. 239–246, 2007.

[44] E. M. Pogson, J. Begg, M. G. Jameson, C. Dempsey, D. Latty, V. Batumalai, A. Lim, K. Kandasamy, P. E. Metcalfe, and L. C. Holloway, "A phantom assessment of achievable contouring concordance across multiple treatment planning systems," *Radiotherapy and Oncology*, vol. 117, no. 3, pp. 438–441, 2015.

[45] M. Giezen, E. Kouwenhoven, A. N. Scholten, E. G. Coerkamp, M. Heijenbrok, W. P. A. Jansen, M. E. Mast, A. L. Petoukhova, and H. Struikmans, "Magnetic resonance imaging-versus computed tomography-based target volume delineation of the glandular breast tissue (clinical target volume breast) in breast-conserving therapy: An exploratory study," *International Journal of Radiation Oncology Biology Physics*, vol. 81, no. 3, pp. 804–811, 2011.

[46] C. Weltens, J. Menten, M. Feron, E. Bellon, P. Demaerel, F. Maes, W. Van den Bogaert, and E. van der Schueren, "Interobserver variations in gross tumor volume delineation of brain tumors on computed tomography and impact of magnetic resonance imaging," *Radiotherapy and Oncology*, vol. 60, no. 1, pp. 49–59, 2001.

[47] A. S. Gerlich, J. M. van der Velden, A. Kotte, C. L. Tseng, G. Fanetti, W. S. C. Eppinga, N. Kasperts, M. P. W. Intven, F. A. Pameijer, M. E. P. Philippens, H. M. Verkooijen, and E. Seravalli, "Inter-observer agreement in GTV delineation of bone metastases on CT and impact of MR

imaging: A multicenter study," *Radiotherapy and Oncology*, vol. 126, no. 3, pp. 534–540, 2018.

[48] C. Rasch, R. Steenbakkers, and M. van Herk, "Target definition in prostate, head, and neck," *Seminars in Radiation Oncology*, vol. 15, no. 3, pp. 136–145, 2005.

[49] C. Rasch, R. Keus, F. A. Pameijer, W. Koops, V. de Ru, S. Muller, A. Touw, H. Bartelink, M. van Herk, and J. V. Lebesque, "The potential impact of CT-MRI matching on tumor volume delineation in advanced head and neck cancer," *International Journal of Radiation Oncology Biology Physics*, vol. 39, no. 4, pp. 841–848, 1997.

[50] H. Ashamalla, S. Rafla, K. Parikh, B. Mokhtar, G. Goswami, S. Kambam, H. Abdel-Dayem, A. Guirguis, P. Ross, and A. Evola, "The contribution of integrated PET/CT to the evolving definition of treatment volumes in radiation treatment planning in lung cancer," *International Journal of Radiation Oncology Biology Physics*, vol. 63, no. 4, pp. 1016–1023, 2005.

[51] C. B. Caldwell, K. Mah, Y. C. Ung, C. E. Danjoux, J. M. Balogh, S. N. Ganguli, and L. E. Ehrlich, "Observer variation in contouring gross tumor volume in patients with poorly defined non-small-cell lung tumors on CT: The impact of 18FDG-hybrid PET fusion," *International Journal of Radiation Oncology Biology Physics*, vol. 51, no. 4, pp. 923–931, 2001.

[52] A. van Baardwijk, G. Bosmans, L. Boersma, J. Buijsen, S. Wanders, M. Hochstenbag, R. van Suylen, A. Dekker, C. Dehing-Oberije, R. Houben, S. M. Bentzen, M. van Kroonenburgh, P. Lambin, and D. De Ruysscher, "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*, vol. 68, no. 3, pp. 771–778, 2007.

[53] K. Morarji, A. Fowler, S. K. Vinod, I. Ho Shon, and J. M. Laurence, "Impact of FDG-PET on lung cancer delineation for radiotherapy," *Journal of Medical Imaging and Radiation Oncology*, vol. 56, no. 2, pp. 195–203, 2012.

[54] S. L. Breen, J. Publicover, S. De Silva, G. Pond, K. Brock, B. O'Sullivan, B. Cummings, L. Dawson, A. Keller, J. Kim, J. Ringash, E. Yu, A. Hendler, and J. Waldron, "Intraobserver and interobserver variability in GTV delineation on FDG-PET-CT images of head and neck cancers," *International Journal of Radiation Oncology Biology Physics*, vol. 68, no. 3, pp. 763–770, 2007.

[55] K. Karki, S. Saraiya, G. D. Hugo, N. Mukhopadhyay, N. Jan, J. Schuster, M. Schutzer, L. Fahrner, R. Groves, K. M. Olsen, J. C. Ford, and E. Weiss, "Variabilities of magnetic resonance imaging-, computed tomography-, and

positron emission tomography-computed tomography-based tumor and lymph node delineations for lung cancer radiation therapy planning," *International Journal of Radiation Oncology Biology Physics*, vol. 99, no. 1, pp. 80–89, 2017.

[56] H. Yamazaki, H. Shiomi, T. Tsubokura, N. Kodani, T. Nishimura, N. Aibe, H. Udono, M. Nishikata, Y. Baba, M. Ogita, K. Yamashita, and T. Kotsuma, "Quantitative assessment of inter-observer variability in target volume delineation on stereotactic radiotherapy treatment for pituitary adenoma and meningioma near optic tract," *Radiation Oncology*, vol. 6, no. 1, p. 10, 2011.

[57] P. Bowden, R. Fisher, M. Mac Manus, A. Wirth, G. Duchesne, M. Millward, A. McKenzie, J. Andrews, and D. Ball, "Measurement of lung tumor volumes using three-dimensional computer planning software," *International Journal of Radiation Oncology Biology Physics*, vol. 53, no. 3, pp. 566–573, 2002.

[58] J. Van de Steene, N. Linthout, J. de Mey, V. Vinh-Hung, C. Claassens, M. Noppen, A. Bel, and G. Storme, "Definition of gross tumor volume in lung cancer: Inter-observer variability," *Radiotherapy and Oncology*, vol. 62, no. 1, pp. 37–49, 2002.

[59] A. V. Louie, G. Rodrigues, J. Olsthoorn, D. Palma, E. Yu, B. Yaremko, B. Ahmad, I. Aivas, and S. Gaede, "Inter-observer and intra-observer reliability for lung cancer target volume delineation in the 4D-CT era," *Radiotherapy and Oncology*, vol. 95, no. 2, pp. 166–171, 2010.

[60] H. Vorwerk, G. Beckmann, M. Bremer, M. Degen, B. Dietl, R. Fietkau, T. Gsänger, R. M. Hermann, M. K. Alfred Herrmann, U. Höller, M. van Kampen, W. Körber, B. Maier, T. Martin, M. Metz, R. Richter, B. Siekmeyer, M. Steder, D. Wagner, C. F. Hess, E. Weiss, and H. Christiansen, "The delineation of target volumes for radiotherapy of lung cancer patients," *Radiotherapy and Oncology*, vol. 91, no. 3, pp. 455–460, 2009.

[61] L. Kepka, K. Bujko, D. Garmol, J. Palucki, A. Zolciak-Siwinska, Z. Guzel-Szczepiorkowska, L. Pietrzak, K. Komosinska, A. Sprawka, and A. Garbaczewska, "Delineation variation of lymph node stations for treatment planning in lung cancer radiotherapy," *Radiotherapy and Oncology*, vol. 85, no. 3, pp. 450–455, 2007.

[62] M. C. Aznar, T. Girinsky, A. K. Berthelsen, B. Aleman, M. Beijert, M. Hutchings, Y. Lievens, P. Meijnders, P. Meidahl Petersen, D. Schut, M. V. Maraldo, R. van der Maazen, and L. Specht, "Interobserver delineation uncertainty in involved-node radiation therapy (INRT) for early-stage hodgkin lymphoma: On behalf of the radiotherapy committee of the EORTC lymphoma group," *Acta Oncologica*, vol. 56, no. 4, pp. 608–613, 2017.

[63] S. Mercieca, J. S. A. Belderbos, K. De Jaeger, D. A. X. Schinagl, N. van der Voort Van Zijp, J. Pomp, J. Theuws, J. Khalifa, P. van de Vaart, and M. van Herk, "Interobserver variability in the delineation of the primary lung cancer and lymph nodes on different four-dimensional computed tomography reconstructions," *Radiotherapy and Oncology*, vol. 126, no. 2, pp. 325–332, 2018.

[64] P. Petrič, R. Hudej, P. Rogelj, M. Blas, K. Tanderup, E. Fidarova, C. Kirisits, D. Berger, J. C. A. Dimopoulos, R. Pötter, and T. P. Hellebust, "Uncertainties of target volume delineation in mri guided adaptive brachytherapy of cervix cancer: A multi-institutional study," *Radiotherapy and Oncology*, vol. 107, no. 1, pp. 6–12, 2013.

[65] J. C. A. Dimopoulos, V. D. Vos, D. Berger, P. Petric, I. Dumas, C. Kirisits, C. B. Shenfield, C. Haie-Meder, and R. Pötter, "Inter-observer comparison of target delineation for MRI-assisted cervical cancer brachytherapy: Application of the GYN GEC-ESTRO recommendations," *Radiotherapy and Oncology*, vol. 91, no. 2, pp. 166–172, 2009.

[66] A. N. Viswanathan, B. Erickson, D. K. Gaffney, S. Beriwal, S. K. Bhatia, O. Lee Burnett, D. P. D'Souza, N. Patil, M. G. Haddock, A. Jhingran, E. L. Jones, C. A. Kunos, L. Lee, L. Lin, N. Mayr, I. Petersen, P. Petric, L. Portelance, W. Small, J. B. Strauss, K. Townamchai, A. H. Wolfson, C. M. Yashar, and W. Bosch, "Comparison and consensus guidelines for delineation of clinical target volume for CT- and MR-based brachytherapy in locally advanced cervical cancer," *International Journal of Radiation Oncology Biology Physics*, vol. 90, no. 2, pp. 320–328, 2014.

[67] D. H. Wu, "Interobserver variation in cervical cancer tumor delineation for image-based radiotherapy planning among and within different specialties," *Journal of applied clinical medical physics*, vol. 6, no. 4, pp. 106–110, 2005.

[68] M. A. Pitkänen, K. A. Holli, A. T. Ojala, and P. Laippala, "Quality assurance in radiotherapy of breast cancer variability in planning target volume delineation," *Acta Oncologica*, vol. 40, no. 1, pp. 50–55, 2001.

[69] V. K. Reed, W. A. Woodward, L. Zhang, E. A. Strom, G. H. Perkins, W. Tereffe, J. L. Oh, T. K. Yu, I. Bedrosian, G. J. Whitman, T. A. Buchholz, and L. Dong, "Automatic segmentation of whole breast using atlas approach and deformable image registration," *International Journal of Radiation Oncology Biology Physics*, vol. 73, no. 5, pp. 1493–1500, 2009.

[70] J. Yang, W. A. Woodward, V. K. Reed, E. A. Strom, G. H. Perkins, W. Tereffe, T. A. Buchholz, L. Zhang, P. Balter, L. E. Court, X. A. Li, and L. Dong,

"Statistical modeling approach to quantitative analysis of interobserver variability in breast contouring," *International Journal of Radiation Oncology Biology Physics*, vol. 89, no. 1, pp. 214–221, 2014.

[71] M. Mast, E. Coerkamp, M. Heijenbrok, A. Scholten, W. Jansen, E. Kouwenhoven, J. Nijkamp, S. de Waard, A. Petoukhova, and H. Struikmans, "Target volume delineation in breast conserving radiotherapy: Are co-registered CT and MR images of added value?" *Radiation Oncology*, vol. 9, no. 1, p. 65, 2014.

[72] M. Jolicoeur, M.-L. Racine, I. Trop, L. Hathout, D. Nguyen, T. Derashodian, and S. David, "Localization of the surgical bed using supine magnetic resonance and computed tomography scan fusion for planification of breast interstitial brachytherapy," *Radiotherapy and Oncology*, vol. 100, no. 3, pp. 480–484, 2011.

[73] T. Shaikh, T. Chen, A. Khan, N. J. Yue, T. Kearney, A. Cohler, B. G. Haffty, and S. Goyal, "Improvement in interobserver accuracy in delineation of the lumpectomy cavity using fiducial markers," *International Journal of Radiation Oncology Biology Physics*, vol. 78, no. 4, pp. 1127–1134, 2010.

[74] F. van der Leij, P. H. M. Elkhuizen, T. M. Janssen, P. Poortmans, M. van der Sangen, A. N. Scholten, C. van Vliet-Vroegindeweij, and L. J. Boersma, "Target volume delineation in external beam partial breast irradiation: Less inter-observer variation with preoperative- compared to postoperative delineation," *Radiotherapy and Oncology*, vol. 110, no. 3, pp. 467–470, 2014.

[75] A. M. van Mourik, P. H. M. Elkhuizen, D. Minkema, J. C. Duppen, and C. van Vliet-Vroegindeweij, "Multiinstitutional study on target volume delineation variation in breast radiotherapy in the presence of guidelines," *Radiotherapy and Oncology*, vol. 94, no. 3, pp. 286–291, 2010.

[76] M. D. den Hartogh, M. Philippens, I. E. van Dam, C. E. Kleynen, R. J. Tersteeg, R. M. Pijnappel, A. N. Kotte, H. M. Verkooijen, M. A. van den Bosch, M. A. van Vulpen, B. van Asselen, and H. D. van den Bongard, "MRI and CT imaging for preoperative target volume delineation in breast-conserving therapy," *Radiation Oncology*, vol. 9, no. 1, p. 63, 2014.

[77] W. Jeanneret-Sozzi, R. Moeckli, J. Valley, A. Zouhair, E. M. Ozsahin, and R.-O. Mirimanoff, "The reasons for discrepancies in target volume delineation: A SASRO study on head-and-neck and prostate cancers," *Strahlentherapie Und Onkologie: Organ Der Deutschen Rontgengesellschaft*, vol. 182, no. 8, pp. 450–457, 2006.

[78] E. Khoo, K. Schick, A. W. Plank, M. Poulsen, W. W. G. Wong, M. Middleton, and J. M. Martin, "Prostate contouring variation: Can it be fixed?" *International Journal of Radiation Oncology Biology Physics*, vol. 82, no. 5, pp. 1923–1929, 2012.

[79] H. J. Choi, Y. S. Kim, S. H. Lee, Y. S. Lee, G. Park, J. H. Jung, B. C. Cho, S. H. Park, H. Ahn, C.-S. Kim, S. Y. Yi, and S. D. Ahn, "Inter- and intra-observer variability in contouring of the prostate gland on planning computed tomography and cone beam computed tomography," *Acta Oncologica (Stockholm, Sweden)*, vol. 50, no. 4, pp. 539–546, 2011.

[80] G. M. Villeirs, K. Van Vaerenbergh, L. Vakaet, S. Bral, F. Claus, W. J. De Neve, K. L. Verstraete, and G. O. De Meerleer, *Interobserver delineation variation using CT versus combined CT plus MRI in intensity-modulated radiotherapy for prostate cancer*, Generic, 2005.

[81] T. Nyholm, J. Jonsson, K. Söderström, P. Bergström, B. Zackrisson, A. Carlberg, G. Frykholm, C. F. Behrens, P. F. Geertsen, R. Trepiakas, S. Hanvey, A. Sadozye, J. Ansari, H. McCallum, J. Frew, and R. McMenemin, "Variability in prostate and seminal vesicle delineations defined on magnetic resonance images, a multi-observer, -center and -sequence study," *Radiation Oncology*, vol. 8, no. 1, 2013.

[82] K. Nakamura, Y. Shioyama, S. Tokumaru, N. Hayashi, N. Oya, Y. Hiraki, K. Kusuhara, T. Toita, H. Suefuji, N. Hayabuchi, H. Terashima, M. Makino, and K. Jingu, *Variation of clinical target volume definition among Japanese radiation oncologists in external beam radiotherapy for prostate cancer*, Generic, 2008.

[83] B. Seddon, M. Bidmead, J. Wilson, V. Khoo, and D. Dearnaley, "Target volume definition in conformal radiotherapy for prostate cancer: Quality assurance in the MRC RT-01 trial," *Radiotherapy and Oncology*, vol. 56, no. 1, pp. 73–83, 2000.

[84] J. E. Livsey, J. P. Wylie, R. Swindell, V. S. Khoo, R. A. Cowan, and J. P. Logue, "Do differences in target volume definition in prostate cancer lead to clinically relevant differences in normal tissue toxicity?" *International Journal of Radiation Oncology Biology Physics*, vol. 60, no. 4, pp. 1076–1081, 2004.

[85] X. A. Li, A. Tai, D. W. Arthur, T. A. Buchholz, S. Macdonald, L. B. Marks, J. M. Moran, L. J. Pierce, R. Rabinovitch, A. Taghian, F. Vicini, W. Woodward, and J. R. White, "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*, vol. 73, no. 3, pp. 944–951, 2009.

[86] F. Spoelstra, S. Senan, C. Le Péchoux, S. Ishikura, F. Casas, D. Ball, A. Price, D. De Ruysscher, and J. R. van Sörnsen de Koste, "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*, vol. 76, no. 4, pp. 1106–1113, 2010.

[87] S. Dewas, J.-E. Bibault, P. Blanchard, C. Vautravers-Dewas, Y. Pointreau, F. Denis, M. Brauner, and P. Giraud, "Delineation in thoracic oncology: A prospective study of the effect of training on contour variability and dosimetric consequences," *Radiation Oncology*, vol. 6, no. 1, p. 118, 2011.

[88] S. W. Loo, W. M. C. Martin, P. Smith, S. Cherian, and T. W. Roques, "Interobserver variation in parotid gland delineation: A study of its impact on intensity-modulated radiotherapy solutions with a systematic review of the literature," *The British Journal of Radiology*, vol. 85, no. 1016, pp. 1070–1077, 2012.

[89] M. Min, D. Roos, E. Keating, M. Penniment, S. Carruthers, L. Zanchetta, K. Wong, J. Shakeshaft, and S. Baxi, "External evaluation of the radiation therapy oncology group brachial plexus contouring protocol: Several issues identified," *Journal of Medical Imaging and Radiation Oncology*, vol. 58, no. 3, pp. 360–368, 2014.

[90] C. Rasch, A. Eisbruch, P. Remeijer, L. Bos, M. Hoogeman, M. van Herk, and J. V. Lebesque, "Irradiation of paranasal sinus tumors, a delineation and dose comparison study," *International Journal of Radiation Oncology Biology Physics*, vol. 52, no. 1, pp. 120–127, 2002.

[91] J. Stanley, P. Dunscombe, H. Lau, P. Burns, G. Lim, H.-W. Liu, R. Nordal, Y. Starreveld, B. Valev, J. Voroney, and D. P. Spencer, "The effect of contouring variability on dosimetric parameters for brain metastases treated with stereotactic radiosurgery," *International Journal of Radiation Oncology Biology Physics*, vol. 87, no. 5, pp. 924–931, 2013.

[92] L. Perna, C. Cozzarini, E. Maggiulli, G. Fellin, T. Rancati, R. Valdagni, V. Vavassori, S. Villa, and C. Fiorino, "Inter-observer variability in contouring the penile bulb on CT images for prostate cancer treatment planning," *Radiation Oncology*, vol. 6, no. 1, p. 123, 2011.

[93] E. L. Lorenzen, C. W. Taylor, M. Maraldo, M. H. Nielsen, B. V. Offersen, M. R. Andersen, D. O'Dwyer, L. Larsen, S. Duxbury, B. Jhitta, S. C. Darby, M. Ewertz, and C. Brink, "Inter-observer variation in delineation of the heart and left anterior descending coronary artery in radiotherapy for breast cancer: A multi-centre study from Denmark and the UK," *Radiotherapy and Oncology*, vol. 108, no. 2, pp. 254–258, 2013.

[94] M. Feng, J. M. Moran, T. Koelling, A. Chughtai, J. L. Chan, L. Freedman, J. A. Hayman, R. Jagsi, S. Jolly, J. Larouere, J. Soriano, R. Marsh, and L. J. Pierce, "Development and validation of a heart atlas to study cardiac exposure to radiation following treatment for breast cancer," *International Journal of Radiation Oncology Biology Physics*, vol. 79, no. 1, pp. 10–18, 2011.

[95] F. Lobefalo, M. Bignardi, G. Reggiori, A. Tozzi, S. Tomatis, F. Alongi, A. Fogliata, A. Gaudino, P. Navarria, L. Cozzi, M. Scorsetti, and P. Mancosu, "Dosimetric impact of inter-observer variability for 3D conformal radiotherapy and volumetric modulated arc therapy: The rectal tumor target definition case," *Radiation Oncology*, vol. 8, no. 1, pp. 1–10, 2013.

[96] M. Feng, C. Demiroz, K. A. Vineberg, A. Eisbruch, and J. M. Balter, "Normal tissue anatomy for oropharyngeal cancer: Contouring variability and its impact on optimization," *International Journal of Radiation Oncology Biology Physics*, vol. 84, no. 2, e245–e249, 2012.

[97] B. E. Nelms, W. A. Tomé, G. Robinson, and J. Wheeler, "Variations in the contouring of organs at risk: Test case from a patient with oropharyngeal cancer," *International Journal of Radiation Oncology Biology Physics*, vol. 82, no. 1, pp. 368–378, 2012.

[98] R. Kosztyla, R. Olson, H. Carolan, S. Balkwill, V. Moiseenko, and W. Kwan, "Evaluation of dosimetric consequences of seroma contour variability in accelerated partial breast irradiation using a constructed representative seroma contour," *International Journal of Radiation Oncology Biology Physics*, vol. 84, no. 2, pp. 527–532, 2012.

[99] E. P. M. Jansen, J. Nijkamp, M. Gubanski, P. A. R. M. Lind, and M. Verheij, "Interobserver variation of clinical target volume delineation in gastric cancer," *International Journal of Radiation Oncology Biology Physics*, vol. 77, no. 4, pp. 1166–1170, 2010.

[100] A. L. Damato, K. Townamchai, M. Albert, R. J. Bair, R. A. Cormack, J. Jang, A. Kovacs, L. J. Lee, K. S. Mak, K. L. Mirabeau-Beale, K. W. Mouw, J. G. Phillips, J. L. Pretz, A. L. Russo, J. H. Lewis, and A. N. Viswanathan, "Dosimetric consequences of interobserver variability in delineating the organs at risk in gynecologic interstitial brachytherapy," *International Journal of Radiation Oncology Biology Physics*, vol. 89, no. 3, pp. 674–681, 2014.

[101] T. P. Hellebust, K. Tanderup, C. Lervåg, E. Fidarova, D. Berger, E. Malinen, R. Pötter, and P. Petrič, "Dosimetric impact of interobserver variability in MRI-based delineation for cervical cancer brachytherapy," *Radiotherapy and Oncology*, vol. 107, no. 1, pp. 13–19, 2013.

[102] A. E. Saarnak, M. Boersma, B. N. F. M. van Bunningen, R. Wolterink, and M. J. Steggerda, "Inter-observer variation in delineation of bladder and rectum contours for brachytherapy of cervical cancer," *Radiotherapy and Oncology*, vol. 56, no. 1, pp. 37–42, 2000.

[103] F. K. Duane, B. Langan, C. Gillham, L. Walsh, G. Rangaswamy, C. Lyons, M. Dunne, C. Walker, and O. McArdle, "Impact of delineation uncertainties on dose to organs at risk in CT-guided intracavitary brachytherapy," *Brachytherapy*, vol. 13, no. 2, pp. 210–218, 2014.

[104] J. M. Freilich, P. E. Spiess, M. C. Biagioli, D. C. Fernandez, E. J. Shi, D. C. Hunt, S. Gupta, and R. B. Wilder, "Lipiodol as a fiducial marker for image-guided radiation therapy for bladder cancer," *International Brazilian Journal of Urology*, vol. 40, pp. 190–197, 2014.

[105] S. Mercieca, J. S. A. Belderbos, A. van Baardwijk, S. Delorme, and M. van Herk, "The impact of training and professional collaboration on the interobserver variation of lung cancer delineations: A multi-institutional study," *Acta Oncologica*, pp. 1–9, 2018.

[106] E. K. Wong, P. T. Truong, H. A. Kader, A. M. Nichol, L. Salter, R. Petersen, E. S. Wai, L. Weir, and I. A. Olivotto, "Consistency in seroma contouring for partial breast radiotherapy: Impact of guidelines," *International Journal of Radiation Oncology Biology Physics*, vol. 66, no. 2, pp. 372–376, 2006.

[107] D. M. Mitchell, L. Perry, S. Smith, T. Elliott, J. P. Wylie, R. A. Cowan, J. E. Livsey, and J. P. Logue, "Assessing the effect of a contouring protocol on postprostatectomy radiotherapy clinical target volumes and interphysician variation," *International Journal of Radiation Oncology Biology Physics*, vol. 75, no. 4, pp. 990–993, 2009.

[108] M. Mukesh, R. Benson, R. Jena, A. Hoole, T. Roques, C. Scrase, C. Martin, G. A. Whitfield, J. Gemmill, and S. Jefferies, "Interobserver variation in clinical target volume and organs at risk segmentation in post-parotidectomy radiotherapy: Can segmentation protocols help?" *The British Journal of Radiology*, vol. 85, no. 1016, e530–e536, 2012.

[109] B. De Bari, M. Dahele, M. Palmu, S. Kaylor, L. Schiappacasse, and M. Guckenberger, "Short interactive workshops reduce variability in contouring treatment volumes for spine stereotactic body radiation therapy: Experience with the ESTRO FALCON programme and EduCase training tool," *Radiotherapy and Oncology*, vol. 127, no. 1, pp. 150–153, 2018.

[110] U. Nestle, D. De Ruysscher, U. Ricardi, X. Geets, J. Belderbos, C. Pöttgen, R. Dziadiuszko, S. Peeters, Y. Lievens, C. Hurkmans, B. Slotman, S. Ramella, C.

Faivre-Finn, F. McDonald, F. Manapov, P. M. Putora, C. LePéchoux, and P. Van Houtte, "ESTRO ACROP guidelines for target volume definition in the treatment of locally advanced non-small cell lung cancer," *Radiotherapy and Oncology*, vol. 127, no. 1, pp. 1–5, 2018.

[111] P. Tai, J. Van Dyk, J. Battista, E. Yu, L. Stitt, J. Tonita, O. Agboola, J. Brierley, R. Dar, C. Leighton, S. Malone, B. Strang, P. Truong, G. Videtic, C. S. Wong, R. Wong, and Y. Youssef, "Improving the consistency in cervical esophageal target volume definition by special training," *International Journal of Radiation Oncology Biology Physics*, vol. 53, no. 3, pp. 766–774, 2002.

[112] E. Szumacher, N. Harnett, S. Warner, V. Kelly, C. Danjoux, R. Barker, M. Woo, K. Mah, I. Ackerman, A. Dubrowski, S. Rose, and J. Crook, "Effectiveness of educational intervention on the congruence of prostate and rectal contouring as compared with a gold standard in three-dimensional radiotherapy for prostate," *International Journal of Radiation Oncology Biology Physics*, vol. 76, no. 2, pp. 379–385, 2010.

[113] K. Schick, T. Sisson, J. Frantzis, E. Khoo, and M. Middleton, "An assessment of OAR delineation by the radiation therapist," *Radiography*, vol. 17, no. 3, pp. 183–187, 2011.

[114] J. Breunig, S. Hernandez, J. Lin, S. Alsager, C. Dumstorf, J. Price, J. Steber, R. Garza, S. Nagda, E. Melian, B. Emami, and J. C. Roeske, "A system for continual quality improvement of normal tissue delineation for radiation therapy treatment planning," *International Journal of Radiation Oncology Biology Physics*, vol. 83, no. 5, e703–e708, 2012.

[115] J. Nijkamp, D. F. M. de Haas-Kock, J. C. Beukema, K. J. Neelis, D. Woutersen, H. Ceha, T. Rozema, A. Slot, H. Vos-Westerman, M. Intven, P. H. Spruit, Y. van der Linden, D. Geijsen, K. Verschueren, M. B. van Herk, and C. A. M. Marijnen, "Target volume delineation variation in radiotherapy for early stage rectal cancer in the netherlands," *Radiotherapy and Oncology*, vol. 102, no. 1, pp. 14–21, 2012.

[116] J. E. Bekelman, S. Wolden, and N. Lee, "Head-and-neck target delineation among radiation oncology residents after a teaching intervention: A prospective, blinded pilot study," *International Journal of Radiation Oncology Biology Physics*, vol. 73, no. 2, pp. 416–423, 2009.

[117] L. J. Stapleford, J. D. Lawson, C. Perkins, S. Edelman, L. Davis, M. W. McDonald, A. Waller, E. Schreibmann, and T. Fox, "Evaluation of automatic atlas-based lymph node segmentation for head-and-neck cancer," *International Journal of Radiation Oncology Biology Physics*, vol. 77, no. 3, pp. 959–966, 2010.

[118] A. V. Young, A. Wortham, I. Wernick, A. Evans, and R. D. Ennis, "Atlas-based segmentation improves consistency and decreases time required for contouring postoperative endometrial cancer nodal volumes," *International Journal of Radiation Oncology Biology Physics*, vol. 79, no. 3, pp. 943–947, 2011.

[119] N. Sharma and L. M. Aggarwal, "Automated medical image segmentation techniques," *Journal of medical physics*, vol. 35, no. 1, pp. 3–14, 2010.

[120] C. Chu, J. De Fauw, N. Tomasev, B. Paredes, C. Hughes, J. Ledsam, T. Back, H. Montgomery, G. Rees, R. Raine, K. Sullivan, S. Moinuddin, D. D'Souza, O. Ronneberger, R. Mendes, and J. Cornebise, "Applying machine learning to automated segmentation of head and neck tumour volumes and organs at risk on radiotherapy planning CT and MRI scans," *F1000Research*, vol. 5, no. 2104, 2016.

[121] Y. Jinzhong, W. A. Woodward, V. K. Reed, E. A. Strom, G. H. Perkins, W. Tereffe, T. A. Buchholz, Z. Lifei, P. Balter, L. E. Court, X. A. Li, and D. Lei, "Statistical modeling approach to quantitative analysis of interobserver variability in breast contouring," *International Journal of Radiation Oncology Biology Physics*, vol. 89, no. 1, pp. 214–221, 2014.

[122] R. J. Myerson, M. C. Garofalo, I. El Naqa, R. A. Abrams, A. Apte, W. R. Bosch, P. Das, L. L. Gunderson, T. S. Hong, J. J. J. Kim, C. G. Willett, and L. A. Kachnic, "Elective clinical target volumes for conformal therapy in anorectal cancer: A radiation therapy oncology group consensus panel contouring atlas," *International Journal of Radiation Oncology Biology Physics*, vol. 74, no. 3, pp. 824–830, 2009.

[123] G. Sharp, K. D. Fritscher, V. Pekar, M. Peroni, N. Shusharina, H. Veeraraghavan, and J. Yang, "Vision 20/20: Perspectives on automated image segmentation for radiotherapy," *Medical Physics*, vol. 41, no. 5, p. 050 902, 2014.

[124] J. Jaymin, C. Mudit, Z. Xinyan, H. P. Robert, M. S. Jeffrey, J. F. Matthew, M. M. Tiffany, R. Justin, D. Anees, E. Eric, R. E. Bree, J. O. Jeffrey, J. C. Walter, G. S. Hui-Kuo, R. C. Ian, and R. P. Kirtesh, "Does size matter? investigating the optimal planning target volume margin for postoperative stereotactic radiosurgery to resected brain metastases," *Journal of Neurosurgery JNS*, pp. 1–7, 2018.

[125] G. Meijer, C. Rasch, P. Remeijer, and J. Lebesque, "Three-dimensional analysis of delineation errors, setup errors, and organ motion during radiotherapy of bladder cancer," *International Journal of Radiation Oncology Biology Physics*, vol. 55, no. 5, p. 1277, 2003.

[126] D. Genovesi, G. A. Céfaro, A. Vinciguerra, A. Augurio, M. Di Tommaso, R. Marchese, U. Ricardi, A. R. Filippi, T. Girinsky, K. Di Biagio, M. Belfiglio,

E. Barbieri, and V. Valentini, "Interobserver variability of clinical target volume delineation in supra-diaphragmatic hodgkin's disease," *Strahlentherapie und Onkologie*, vol. 187, no. 6, pp. 357–366, 2011.

[127] P. Remeijer, C. Rasch, J. V. Lebesque, and M. van Herk, "A general methodology for three-dimensional analysis of variation in target volume delineation," *Medical Physics*, vol. 26, no. 6, pp. 931–940, 1999.

[128] J. C. Stroom and B. J. M. Heijmen, "Geometrical uncertainties, radiotherapy planning margins, and the ICRU-62 report," *Radiotherapy and Oncology*, vol. 64, no. 1, pp. 75–83, 2002.

[129] M. Austin-Seymour, I. Kalet, J. McDonald, S. Kromhout-Schiro, J. Jacky, S. Hummel, and J. Unger, "Three dimensional planning target volumes: A model and a software tool," *International Journal of Radiation Oncology Biology Physics*, vol. 33, no. 5, pp. 1073–1080, 1995.

[130] C. Haie-Meder, R. Pötter, E. Van Limbergen, E. Briot, M. De Brabandere, J. Dimopoulos, I. Dumas, T. P. Hellebust, C. Kirisits, S. Lang, S. Muschitz, J. Nevinson, A. Nulens, P. Petrow, and N. Wachter-Gerstner, "Recommendations from gynaecological (GYN) GEC-ESTRO working group (I): Concepts and terms in 3D image based 3D treatment planning in cervix cancer brachytherapy with emphasis on MRI assessment of GTV and CTV," *Radiotherapy and Oncology*, vol. 74, no. 3, pp. 235–245, 2005.

[131] R. Pötter, C. Haie-Meder, E. V. Limbergen, I. Barillot, M. D. Brabandere, J. Dimopoulos, I. Dumas, B. Erickson, S. Lang, A. Nulens, P. Petrow, J. Rownd, and C. Kirisits, "Recommendations from gynaecological (GYN) GEC ESTRO working group (II): Concepts and terms in 3D image-based treatment planning in cervix cancer brachytherapy-3D dose volume parameters and aspects of 3D image-based anatomy, radiation physics, radiobiology," *Radiotherapy and Oncology*, vol. 78, no. 1, pp. 67–77, 2006.

[132] S. Van Dyk, D. Byram, and D. Bernshaw, "Brachytherapy for cancer of the cervix: An Australian and New Zealand survey of current treatment techniques," *Journal of Medical Imaging and Radiation Oncology*, vol. 52, no. 6, pp. 588–597, 2008.

[133] S. Van Dyk, D. Byram, and D. Bernshaw, "Use of 3d imaging and awareness of GEC-ESTRO recommendations for cervix cancer brachytherapy throughout Australia and New Zealand," *Journal of Medical Imaging and Radiation Oncology*, vol. 54, no. 4, pp. 383–387, 2010.

[134] K. Lim, S. Van Dyk, S. Vinod, J. Veera, P. Khaw, and L. Ohanessian, "Patterns of practice for cervix cancer brachytherapy in Australia and New Zealand," *International Journal of Gynaecological Cancer*, vol. 24, p. 629, 2014.

[135] S. R. Thompson, G. P. Delaney, G. S. Gabriel, and M. B. Barton, "Patterns of care study of brachytherapy in New South Wales: Cervical cancer treatment quality depends on caseload," *Journal of contemporary brachytherapy*, vol. 6, no. 1, pp. 28–32, 2014.

[136] *Embrace*, `www.embracestudy.dk`.

[137] R. Pötter, K. Tanderup, C. Kirisits, A. de Leeuw, K. Kirchheiner, R. Nout, L. T. Tan, C. Haie-Meder, U. Mahantshetty, B. Segedin, P. Hoskin, K. Bruheim, B. Rai, F. Huang, E. Van Limbergen, M. Schmid, N. Nesvacil, A. Sturdza, L. Fokdal, N. B. K. Jensen, D. Georg, M. Assenholt, Y. Seppenwoolde, C. Nomden, I. Fortin, S. Chopra, U. van der Heide, T. Rumpold, J. C. Lindegaard, and I. Jürgenliemk-Schulz, "The EMBRACE II study: The outcome and prospect of two decades of evolution within the GEC-ESTRO GYN working group and the EMBRACE studies," *Clinical and Translational Radiation Oncology*, vol. 9, pp. 48–60, 2018.

[138] A. Sturdza, R. Pötter, L. U. Fokdal, C. Haie-Meder, L. T. Tan, R. Mazeron, P. Petric, B. Šegedin, I. M. Jurgenliemk-Schulz, C. Nomden, C. Gillham, O. McArdle, E. Van Limbergen, H. Janssen, P. Hoskin, G. Lowe, E. Tharavichitkul, E. Villafranca, U. Mahantshetty, P. Georg, K. Kirchheiner, C. Kirisits, K. Tanderup, and J. C. Lindegaard, "Image guided brachytherapy in locally advanced cervical cancer: Improved pelvic control and survival in retroembrace, a multicenter cohort study," *Radiotherapy and Oncology*, vol. 120, no. 3, pp. 428–433, 2016.

[139] N. B. K. Jensen, R. Pötter, K. Kirchheiner, L. Fokdal, J. C. Lindegaard, C. Kirisits, R. Mazeron, U. Mahantshetty, I. M. Jürgenliemk-Schulz, B. Segedin, P. Hoskin, and K. Tanderup, "Bowel morbidity following radiochemotherapy and image-guided adaptive brachytherapy for cervical cancer: Physician- and patient reported outcome from the EMBRACE study," *Radiotherapy and Oncology*, vol. 127, no. 3, pp. 431–439, 2018.

[140] L. Fokdal, R. Pötter, K. Kirchheiner, J. C. Lindegaard, N. B. K. Jensen, C. Kirisits, C. Chargari, U. Mahantshetty, I. M. Jürgenliemk-Schulz, B. Segedin, P. Hoskin, and K. Tanderup, "Physician assessed and patient reported urinary morbidity after radio-chemotherapy and image guided adaptive brachytherapy for locally advanced cervical cancer," *Radiotherapy and Oncology*, vol. 127, no. 3, pp. 423–430, 2018.

[141] B. Segedin and P. Petrič, "Uncertainties in target volume delineation in radiotherapy - are they relevant and what can we do about them?" *Radiology and oncology*, vol. 50, no. 3, pp. 254–262, 2016.

[142] C. Lervåg, *RTKIT GitHub Repository*, `https://github.com/dicom/rtkit.`, 2012.

[143] K. Tanderup, L. U. Fokdal, A. Sturdza, C. Haie-Meder, R. Mazeron, E. van Limbergen, I. Jürgenliemk-Schulz, P. Petrič, P. Hoskin, W. Dörr, S. M. Bentzen, C. Kirisits, J. C. Lindegaard, and R. Pötter, "Effect of tumor dose, volume and overall treatment time on local control after radiochemotherapy including MRI guided brachytherapy of locally advanced cervical cancer," *Radiotherapy and Oncology*, vol. 120, no. 3, pp. 441–446, 2016.

[144] K. S. Nkiwane, R. Pötter, K. Tanderup, M. Federico, J. C. Lindegaard, and C. Kirisits, "Single line source with and without vaginal loading and the impact on target coverage and organ at risk doses for cervix cancer stages IB, II, and IIIB: Treatment planning simulation in patients treated with MRI-guided adaptive brachytherapy in a multicentre study (EMBRACE)," *Brachytherapy*, vol. 12, no. 4, pp. 317–323, 2013.

[145] S. Mohamed, J. C. Lindegaard, A. A. C. de Leeuw, I. Jürgenliemk-Schulz, K. Kirchheiner, C. Kirisits, R. Pötter, and K. Tanderup, "Vaginal dose de-escalation in image guided adaptive brachytherapy for locally advanced cervical cancer," *Radiotherapy and Oncology*, vol. 120, no. 3, pp. 480–485, 2016.

[146] K. Kirchheiner, R. A. Nout, J. C. Lindegaard, C. Haie-Meder, U. Mahantshetty, B. Segedin, I. M. Jürgenliemk-Schulz, P. J. Hoskin, B. Rai, W. Dörr, C. Kirisits, S. M. Bentzen, R. Pötter, and K. Tanderup, "Dose-effect relationship and risk factors for vaginal stenosis after definitive radio(chemo)therapy with image-guided brachytherapy for locally advanced cervical cancer in the EMBRACE study," *Radiotherapy and Oncology*, vol. 118, no. 1, pp. 160–166, 2016.

[147] The MathWorks Inc. R2014b, *Matlab*, Natik, Massachusetts, 2014.

[148] L. C. Holloway, J.-A. Miller, S. Kumar, B. M. Whelan, and S. K. Vinod, "Comp Plan: A computer program to generate dose and radiobiological metrics from dose-volume histogram files," *Medical Dosimetry*, vol. 37, no. 3, pp. 305–309, 2012.

[149] E. Kouwenhoven, M. Giezen, and H. Struikmans, "Measuring the similarity of target volume delineations independent of the number of observers," *Physics in Medicine and Biology*, vol. 54, no. 9, pp. 2863–2873, 2009.

[150] IBMCorp, "IBM SPSS Statistics for Windows Version 19," Armonk, NY, Tech. Rep., 2010.

[151] K. Tanderup, N. Nesvacil, R. Pötter, and C. Kirisits, "Uncertainties in image guided adaptive cervix cancer brachytherapy: Impact on planning and prescription," *Radiotherapy and Oncology*, vol. 107, no. 1, pp. 1–5, 2013.

[152] C. Kirisits, M. J. Rivard, D. Baltas, F. Ballester, M. De Brabandere, R. van der Laarse, Y. Niatsetski, P. Papagiannis, T. P. Hellebust, J. Perez-Calatayud, K. Tanderup, J. L. M. Venselaar, and F.-A. Siebert, "Review of clinical brachytherapy uncertainties: Analysis guidelines of GEC-ESTRO and the AAPM," *Radiotherapy and Oncology*, vol. 110, no. 1, pp. 199–212, 2014.

[153] M. R. Arnesen, K. Bruheim, E. Malinen, and T. P. Hellebust, "Spatial dosimetric sensitivity of contouring uncertainties in gynecological 3D-based brachytherapy," *Radiotherapy and Oncology*, vol. 113, no. 3, pp. 414–419, 2014.

[154] G. Eminowicz, V. Rompokos, C. Stacey, and M. McCormack, "The dosimetric impact of target volume delineation variation for cervical cancer radiotherapy," *Radiotherapy and Oncology*, vol. 120, no. 3, pp. 493–499, 2016.

[155] N. Nesvacil, K. Tanderup, T. P. Hellebust, A. De Leeuw, S. Lang, S. Mohamed, S. V. Jamema, C. Anderson, R. Pötter, and C. Kirisits, "A multicentre comparison of the dosimetric impact of inter- and intra-fractional anatomical variations in fractionated cervix cancer brachytherapy," *Radiotherapy and Oncology*, vol. 107, no. 1, pp. 20–25, 2013.

[156] K. Tanderup, R. Pötter, J. C. Lindegaard, D. Berger, A. Wambersie, and C. Kirisits, "PTV margins should not be used to compensate for uncertainties in 3D image guided intracavitary brachytherapy," *Radiotherapy and Oncology*, vol. 97, no. 3, pp. 495–500, 2010.

[157] S. K. Vinod, K. Lim, L. Bell, J. Veera, L. Ohanessian, E. Juresic, N. Borok, P. Chan, R. Chee, V. Do, G. Govindarajulu, S. Sridharan, C. Johnson, D. Moses, S. Van Dyk, and L. Holloway, "High-risk CTV delineation for cervix brachytherapy: Application of GEC-ESTRO guidelines in Australia and New Zealand," *Journal of Medical Imaging and Radiation Oncology*, vol. 61, no. 1, pp. 133–140, 2017.

[158] R. Pearcey, M. Brundage, P. Drouin, J. Jeffrey, D. Johnston, H. Lukka, G. MacLean, L. Souhami, G. Stuart, and D. Tu, "Phase III trial comparing radical radiotherapy with and without cisplatin chemotherapy in patients with advanced squamous cell cancer of the cervix," *Journal of Clinical Oncology*, vol. 20, no. 4, pp. 966–972, 2002.

[159] H. Peulen, J. Belderbos, M. Guckenberger, A. Hope, I. Grills, M. van Herk, and J. Sonke, "Target delineation variability and corresponding margins of peripheral early stage NSCLC treated with stereotactic body radiotherapy," *Radiotherapy and Oncology*, vol. 114, no. 3, pp. 361–366, 2015.

[160] S. Ghose, L. Holloway, K. Lim, P. Chan, J. Veera, S. K. Vinod, G. Liney, P. B. Greer, and J. Dowling, "A review of segmentation and deformable registration methods applied to adaptive cervical cancer radiation therapy treatment planning," *Artificial Intelligence in Medicine*, vol. 64, no. 2, pp. 75–87, 2015.

[161] J. A. Dowling, J. Fripp, S. Chandra, J. P. W. Pluim, J. Lambert, J. Parker, J. Denham, P. B. Greer, and O. Salvado, "Fast automatic multi-atlas segmentation of the prostate from 3D MR images," in *Prostate Cancer Imaging. Image Analysis and Image-Guided Interventions*, A. Madabhushi, J. Dowling, H. Huisman, and D. Barratt, Eds., Springer Berlin Heidelberg, 2011, pp. 10–21.

[162] J. A. Dowling, J. Lambert, J. Parker, O. Salvado, J. Fripp, A. Capp, C. Wratten, J. W. Denham, and P. B. Greer, "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*, vol. 83, no. 1, e5–e11, 2012.

[163] P. B. Greer, J. A. Dowling, J. A. Lambert, J. Fripp, J. Parker, J. W. Denham, C. Wratten, A. Capp, and O. Salvado, "A magnetic resonance imaging-based workflow for planning radiation therapy for prostate cancer," *Medical Journal of Australia*, vol. 194, no. S4, S24–S27, 2011.

[164] S. Ourselin, A. Roche, G. Subsol, X. Pennec, and N. Ayache, "Reconstructing a 3D structure from serial histological sections," *Image and Vision Computing*, vol. 19, no. 1, pp. 25–31, 2001.

[165] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache, "Diffeomorphic demons: Efficient non-parametric image registration," *NeuroImage*, vol. 45, no. 1, Supplement 1, S61–S72, 2009.

[166] P. Aljabar, R. A. Heckemann, A. Hammers, J. V. Hajnal, and D. Rueckert, "Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy," *NeuroImage*, vol. 46, no. 3, pp. 726–738, 2009.

[167] R. Abreu, P. Zoeteweij, and A. J. c. V. Gemund, "An evaluation of similarity coefficients for software fault localization," in *2006 12th Pacific Rim International Symposium on Dependable Computing (PRDC'06)*, 2006, pp. 39–46.

[168] K. O. Babalola, B. Patenaude, P. Aljabar, J. Schnabel, D. Kennedy, W. Crum, S. Smith, T. Cootes, M. Jenkinson, and D. Rueckert, "An evaluation of four automatic methods of segmenting the subcortical structures in the brain," *NeuroImage*, vol. 47, no. 4, pp. 1435–1447, 2009.

[169] E. M. Pogson, L. Bell, V. Batumalai, E. S. Koh, G. Delaney, P. Metcalfe, and L. Holloway, "A comparison of coordinate systems for use in determining a

radiotherapy delineation margin for whole breast," *Journal of Physics: Conference Series*, vol. 489, p. 012 057, 2014.

[170] T. F. Mutanga, H. C. de Boer, G. J. van der Wielen, M. S. Hoogeman, L. Incrocci, and B. J. Heijmen, "Margin evaluation in the presence of deformation, rotation, and translation in prostate and entire seminal vesicle irradiation with daily marker-based setup corrections," *International Journal of Radiation Oncology Biology Physics*, vol. 81, no. 4, pp. 1160–1167, 2011.

[171] M. Heesters, H. Wijrdeman, H. Struikmans, T. Witkamp, and M. Moerland, "Brain tumor delineation based on CT and MR imaging. implications for radiotherapy treatment planning," *Strahlentherapie und Onkologie Organ der Deutschen Rontgengesellschaft*, vol. 169, no. 12, pp. 729–733, Dec. 1993.

[172] C. Fiorino, M. Reni, A. Bolognesi, G. M. Cattaneo, and R. Calandrino, "Intra- and inter-observer variability in contouring prostate and seminal vesicles: Implications for conformal treatment planning," *Radiotherapy and Oncology*, vol. 47, no. 3, pp. 285–292, 1998.

[173] R. P. Petersen, P. T. Truong, H. A. Kader, E. Berthelet, J. C. Lee, M. L. Hilts, A. S. Kader, W. A. Beckham, and I. A. Olivotto, "Target volume delineation for partial breast radiotherapy planning: Clinical characteristics associated with low interobserver concordance," *International Journal of Radiation Oncology Biology Physics*, vol. 69, no. 1, pp. 41–48, 2007.

[174] L. C. Holloway, M. G. Jameson, V. Batumalai, E. Koh, G. Papadatos, D. Lonergan, and G. P. Delaney, "Estimating a delineation uncertainty margin to account for inter-observer variability in breast cancer radiotherapy," *International Journal of Radiation Oncology Biology Physics*, vol. 78, no. 3, S741, 2010.

[175] J. O. Deasy, A. I. Blanco, and V. H. Clark, "Cerr: A computational environment for radiotherapy research," *Medical Physics*, vol. 30, no. 5, pp. 979–985, 2003.

[176] W. Y. Song, B. Chiu, G. S. Bauman, M. Lock, G. Rodrigues, R. Ash, C. Lewis, A. Fenster, J. J. Battista, and J. V. Dyk, "Prostate contouring uncertainty in megavoltage computed tomography images acquired with a helical tomotherapy unit during image-guided radiation therapy," *International Journal of Radiation Oncology Biology Physics*, vol. 65, no. 2, pp. 595–607, 2006.

[177] D. P. Huttenlocher, W. J. Rucklidge, and G. A. Klanderman, "Comparing images using the hausdorff distance under translation," in *Proceedings 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1992, pp. 654–656.

[178] H. P. van der Laan, W. V. Dolsma, J. H. Maduro, E. W. Korevaar, M. Hollander, and J. A. Langendijk, "Three-dimensional conformal simultaneously integrated boost technique for breast-conserving radiotherapy," *International Journal of Radiation Oncology Biology Physics*, vol. 68, no. 4, pp. 1018–1023, 2007.

[179] R. 1005, *A phase III trial of accelerated whole breast irradiation with hypofractionation plus concurrent boost versus standard whole breast irradiation plus sequential boost for early-stage breast cancer.* 2012.

[180] A. McKenzie, "Defining the PTV and PRV- new ideas about old problems," *Radiotherapy and Oncology*, vol. 73, S203, 2004.

[181] E. Huang, T. A. Buchholz, F. Meric, S. Krishnamurthy, N. Q. Mirza, F. C. Ames, B. W. Feig, H. M. Kuerer, M. I. Ross, S. E. Singletary, M. D. McNeese, E. A. Strom, and K. K. Hunt, "Classifying local disease recurrences after breast conservation therapy based on location and histology," *Cancer*, vol. 95, no. 10, pp. 2059–2067, 2002.

Appendix A

Supplementary Material

A.1 Chapter 4

Table A.1: Cervix brachytherapy applicator specifications.

Case	Applicator Length (mm)	Ring Diameter (mm)	Number of Needles
1	30	40	9
2	34	60	5
3	30	60	5
4	30	60	9
5	30	40	5
6	30	40	5

Table A.2: $EQD2_{10}$ (Gy) prescribed doses for plans optimised to each observer contour.

		CTV_{HR} D90	CTV_{HR} D98	CTV_{IR} D98	GTV D98	Bladder D2 cm^3	Sigmoid D2 cm^3	Rectum D2 cm^3
Patient 1	Master	91.0	83.1	62.6	100.4	77.7	53.6	69.1
	STAPLE	90.6	83.2	64.1	91.1	77.0	56.9	69.8
	Observer 1	90.6	80.8	65.4	98.1	72.0	53.9	69.1
	Observer 2	90.1	81.5	59.4	98.5	77.2	53.6	68.9
	Observer 3	91.5	81.7	61.3	92.7	77.2	65.7	58.5
	Observer 4	94.2	84.2	62.1	96.4	73.0	53.9	67.0
	Observer 5	89.9	79.6	60.4	95.1	76.7	55.3	69.1
	Observer 6	89.7	79.6	60.5	97.9	83.1	54.7	69.2
	Observer 7	90.0	81.9	63.0	90.4	77.3	54.7	66.0
	Observer 8	94.7	85.4	61.7	89.6	78.4	64.8	62.6
	Observer 9	94.5	86.3	62.9	95.7	77.1	53.6	69.1
	Observer 10	93.6	85.4	63.6	104.7	75.3	54.0	68.5
Patient 2	Master	90.6	82.4	60.2	93.8	74.9	60.5	49.7
	STAPLE	89.8	80.5	60.3	94.6	83.7	60.1	50.4
	Observer 1	89.3	78.8	62.3	102.3	82.5	64.0	50.9
	Observer 2	96.6	89.8	57.3	96.7	79.9	60.5	50.5
	Observer 3	92.3	83.2	59.3	94.3	76.0	57.2	46.6
	Observer 4	93.6	83.4	60.1	94.5	82.2	60.0	50.0
	Observer 5	89.4	80.2	64.8	97.5	81.4	63.7	50.2
	Observer 6	85.0	71.8	57.0	96.4	85.8	70.0	52.0
	Observer 7	90.2	80.7	61.8	93.8	82.9	62.9	47.7
	Observer 8	92.4	82.7	60.0	83.0	84.4	61.3	49.1
	Observer 9	92.5	82.6	58.9	94.6	78.6	61.6	49.9
	Observer 10	90.6	81.2	62.7	94.7	81.2	62.9	49.6

Table A.2 – *Continued on next page*

Table A.2 – *Continued from previous page*

		CTV_{HR} D90	CTV_{HR} D98	CTV_{IR} D98	GTV D98	Bladder D2 cm^3	Sigmoid D2 cm^3	Rectum D2 cm^3
Patient 3	Master	88.3	80.2	59.9	102.2	75.5	53.5	68.6
	STAPLE	85.8	76.4	57.3	100.0	74.5	54.1	70.8
	Observer 1	86.3	75.4	63.2	108.3	82.5	53.7	70.0
	Observer 2	80.0	69.7	55.1	93.3	74.1	56.5	70.9
	Observer 3	-	-	-	-	-	-	-
	Observer 4	89.8	81.9	60.3	102.7	70.5	53.2	68.8
	Observer 5	78.9	69.1	58.9	88.3	78.4	56.2	70.7
	Observer 6	85.7	73.8	56.1	98.5	80.8	55.5	70.8
	Observer 7	86.0	73.5	58.1	92.3	74.6	59.3	70.3
	Observer 8	92.3	80.7	59.5	94.2	71.1	63.1	69.7
	Observer 9	91.6	79.2	58.6	103.0	74.5	53.7	69.4
	Observer 10	90.0	81.8	61.1	99.1	76.4	56.4	69.9
Patient 4	Master	87.1	75.0	57.3	117.9	85.7	67.1	55.0
	STAPLE	81.9	69.1	56.1	88.8	85.9	67.9	56.0
	Observer 1	81.9	68.3	58.2	110.3	85.6	70.4	57.5
	Observer 2	85.6	70.9	55.2	126.7	86.0	67.2	56.5
	Observer 3	88.9	76.0	56.2	88.0	83.0	68.7	52.5
	Observer 4	86.8	76.2	55.6	110.9	85.9	68.5	53.9
	Observer 5	85.9	74.0	58.7	103.1	83.1	66.2	51.9
	Observer 6	76.2	63.7	54.9	98.2	85.9	67.5	54.7
	Observer 7	80.3	68.3	57.5	102.1	85.9	70.0	52.6
	Observer 8	73.0	63.2	54.5	67.3	87.1	67.8	51.4
	Observer 9	80.5	65.5	54.2	123.1	85.8	69.5	55.0
	Observer 10	83.3	70.7	56.8	84.8	85.3	70.9	56.5

Table A.2 – *Continued on next page*

Table A.2 – *Continued from previous page*

		CTV_{HR} D90	CTV_{HR} D98	CTV_{IR} D98	GTV D98	Bladder D2 cm^3	Sigmoid D2 cm^3	Rectum D2 cm^3
Patient 5	Master	87.6	80.0	60.6	100.4	78.0	62.3	66.0
	STAPLE	90.5	83.0	62.0	94.3	82.0	60.4	69.9
	Observer 1	91.0	77.1	60.7	126.2	84.0	57.6	68.3
	Observer 2	84.8	74.7	54.8	113.5	82.2	57.8	70.5
	Observer 3	92.0	84.2	61.3	94.6	76.8	56.5	62.7
	Observer 4	93.6	85.7	58.6	94.5	82.9	60.5	65.0
	Observer 5	90.5	75.6	58.3	109.6	77.2	60.4	69.0
	Observer 6	91.0	80.5	59.3	105.8	77.3	58.3	70.1
	Observer 7	90.5	80.0	61.2	94.5	80.2	61.1	66.8
	Observer 8	95.0	83.8	56.8	91.3	84.0	60.8	66.7
	Observer 9	91.0	80.2	59.5	120.3	78.2	57.9	61.8
	Observer 10	90.3	80.8	61.2	103.3	78.0	61.2	68.1
Patient 6	Master	92.1	78.7	63.2	84.5	83.7	68.7	66.1
	STAPLE	83.8	72.0	58.1	83.4	85.6	70.4	67.6
	Observer 1	85.9	74.7	60.7	98.5	81.0	66.3	64.5
	Observer 2	79.9	67.0	55.2	75.9	85.7	70.1	65.3
	Observer 3	85.6	66.7	55.0	107.1	85.9	69.9	55.3
	Observer 4	85.9	69.0	58.1	91.9	85.1	67.4	65.4
	Observer 5	67.3	59.3	54.0	72.8	85.8	69.7	64.6
	Observer 6	87.4	76.5	58.2	89.7	85.0	70.1	68.6
	Observer 7	87.4	76.8	59.4	88.9	85.2	68.3	64.8
	Observer 8	85.5	75.6	51.7	80.5	85.8	67.0	63.8
	Observer 9	86.8	72.5	57.6	92.9	83.9	70.1	66.0
	Observer 10	86.0	72.0	58.8	91.2	85.7	66.0	62.9

Table A.3: Single fraction mean, SD and rSD for all dosimetric parameters and all target contours, averaged across all observers for each patient.

		STAPLE							EC						
	Case	1	2	3	4	5	6	Mean	1	2	3	4	5	6	Mean
CTV_{HR} D90	Mean (Gy)	7.9	7.4	7.3	6.1	7.5	6.9	7.2	7.7	8.2	7.4	7.2	7.7	7.9	7.7
	SD (Gy)	0.4	0.5	0.2	0.7	0.6	0.2	0.4	0.3	0.4	0.2	0.4	0.5	0.2	0.3
	rSD (%)	4.8	6.9	3.4	10.6	7.8	2.6	6.0	4.3	4.8	2.9	6.2	6.3	2.3	4.5
CTV_{HR} D98	Mean (Gy)	6.8	6.3	6.1	4.4	6.2	5.4	5.9	6.7	6.9	6.3	5.6	6.5	6.2	6.4
	SD (Gy)	0.4	0.6	0.2	0.6	0.8	0.3	0.5	0.3	0.3	0.2	0.4	0.6	0.3	0.4
	rSD (%)	5.4	8.8	4.0	12.6	13.3	4.6	8.1	4.7	5.1	3.4	6.7	10.0	4.3	5.7
CTV_{IR} D98	Mean (Gy)	4.0	3.5	3.2	2.6	3.4	3.1	3.3	4.0	3.8	3.5	2.9	3.7	3.8	3.6
	SD (Gy)	0.3	0.3	0.2	0.3	0.4	0.1	0.3	0.3	0.3	0.2	0.2	0.4	0.1	0.2
	rSD (%)	6.9	9.0	5.0	11.2	12.5	4.0	8.1	6.4	6.9	5.7	7.3	9.9	3.5	6.6
GTV D98	Mean (Gy)	8.0	7.9	8.7	7.3	7.8	6.9	8.1	8.6	8.3	8.7	10.5	9.5	7.1	8.8
	SD (Gy)	0.3	0.8	0.3	0.4	0.4	0.3	0.4	0.2	0.4	0.3	0.5	0.4	0.4	0.4
	rSD (%)	4.0	10.1	3.3	5.4	5.2	4.8	5.3	2.6	4.3	3.4	4.8	4.4	5.0	4.1
Bladder D2 cm^3	Mean (Gy)	5.3	5.7	5.3	6.2	5.6	6.1	5.7	5.2	5.5	5.0	5.9	5.7	6.0	5.5
	SD (Gy)	0.2	0.2	0.4	0.4	0.4	0.2	0.3	0.2	0.2	0.4	0.4	0.4	0.2	0.3
	rSD (%)	3.0	4.2	8.3	6.0	7.4	3.7	5.4	3.2	4.5	7.9	6.2	7.3	3.6	5.4
Rectum D2 cm^3	Mean (Gy)	2.6	3.7	2.7	4.4	3.4	4.4	3.5	2.5	3.6	2.5	4.1	3.4	4.1	3.4
	SD (Gy)	0.2	0.3	0.3	0.2	0.3	0.3	0.3	0.2	0.3	0.3	0.2	0.3	0.2	0.3
	rSD (%)	8.6	9.3	9.9	4.2	8.5	5.8	7.7	8.5	9.1	10.3	4.1	8.6	5.5	7.7
Sigmoid D2 cm^3	Mean (Gy)	4.6	1.9	4.7	2.8	4.5	4.2	3.8	4.3	1.8	4.3	2.6	4.1	4.0	3.5
	SD (Gy)	0.2	0.1	0.1	0.2	0.3	0.2	0.2	0.2	0.1	0.1	0.2	0.3	0.1	0.2
	rSD (%)	4.0	5.4	2.7	5.8	7.1	3.5	4.8	4.0	5.0	2.7	5.8	6.1	3.1	4.4

Table A.4: Spearmans non-parametric correlation coefficient results for the EC contour. Statistically significant results shown in bold italics.

	x COM		y COM		z COM		C COM		MD-X		MD-Y		MD-Z		Vol		CI	
	ρ	sig	ρ	sig	ρ	sig	ρ	sig	ρ	sig	ρ	sig	ρ	sig	ρ	sig	ρ	sig
CTV_{IR} D98	0.15	0.22	-0.07	0.59	0.19	0.11	-0.02	0.86	-0.19	0.11	0.03	0.79	0.08	0.5	0.04	0.73	0.11	0.38
GTV D98	0.21	0.08	0.06	0.63	0.1	0.42	0.01	0.96	0.23	0.06	-0.16	0.18	**-0.4**	**<0.01**	-0.18	0.14	**-0.33**	**0.01**
Bladder D2 cm^3	-0.13	0.27	***-0.31***	***0.01***	***0.36***	***<0.01***	***0.26***	***0.03***	***-0.37***	***<0.01***	-0.23	0.06	***-0.34***	***<0.01***	***-0.69***	***<0.01***	***-0.47***	***<0.01***
Rectum D2 cm^3	-0.22	0.07	0.09	0.46	***-0.3***	***0.01***	-0.21	0.07	***-0.52***	***<0.01***	***-0.48***	***<0.01***	***0.58***	***<0.01***	***0.37***	***<0.01***	0.03	0.8
Sigmoid D2 cm^3	***0.24***	***0.04***	-0.12	0.31	0.07	0.58	-0.12	0.32	***-0.28***	***0.02***	-0.2	0.1	***0.47***	***<0.01***	0.03	0.82	0.16	0.18

Table A.5: Spearmans non-parametric correlation coefficient results for the STAPLE contour. Statistically significant results shown in bold italics.

	x COM		y COM		z COM		C COM		MD-X		MD-Y		MD-Z		Vol		CI	
	ρ	sig	ρ	sig	ρ	sig	ρ	sig	ρ	sig	ρ	sig	ρ	sig	ρ	sig	ρ	sig
CTV_{IR} D98	0.18	0.13	0.14	0.25	-0.07	0.55	0.12	0.34	***-0.29***	***0.02***	***-0.31***	***0.01***	-0.22	0.06	-0.21	0.08	0.1	0.41
GTV D98	-0.01	0.95	-0.06	0.60	0.17	0.15	0.10	0.43	-0.19	0.11	-0.19	0.11	***-0.34***	***<0.01***	-0.20	0.09	***0.34***	***<0.01***
Bladder D2 cm^3	-0.03	0.79	0.02	0.85	-0.06	0.60	-0.16	0.19	***-0.32***	***<0.01***	-0.22	0.07	***-0.40***	***<0.01***	***-0.70***	***<0.01***	***-0.36***	***<0.01***
Rectum D2 cm^3	0.18	0.46	0.19	0.12	-0.11	0.36	-0.03	0.81	***-0.50***	***<0.01***	***-0.52***	***<0.01***	***0.60***	***<0.01***	***0.40***	***<0.01***	-0.02	0.86
Sigmoid D2 cm^3	-0.21	0.08	0.05	0.68	0.05	0.71	0.09	0.44	***-0.24***	***0.05***	-0.20	0.10	***0.43***	***<0.01***	0.05	0.67	0.04	0.73

A.2 Chapter 5

Table A.6: Single fraction prescribed doses (Gy) for plans optimised to each observer contour.

		CTV_{HR} D90	Bladder D2 cm^3	Rectum D2 cm^3	Sigmoid D2 cm^3
Patient 1	STAPLE	9.5	5.7	4.4	5.0
	EC	10.9	5.5	4.2	4.7
	Observer 1	8.3	4.9	3.7	4.3
	Observer 2	8.4	5.3	4.2	4.7
	Observer 3	9.4	4.8	3.8	4.3
	Observer 4	9.4	4.9	3.7	4.1
	Observer 5	8.0	5.7	4.3	5.3
	Observer 6	9.2	5.3	4.2	4.6
	Observer 7	9.9	4.9	3.9	4.5
	Observer 8	9.1	5.0	3.7	4.4
	Observer 9	9.1	4.9	3.9	4.4
	Observer 10	9.6	4.7	3.6	4.1
	Observer 11	9.8	4.7	3.7	4.0
	Observer 12	9.6	5.1	4.0	4.4
Patient 2	STAPLE	4.3	6.5	5.9	0.9
	EC	5.8	6.5	5.2	0.9
	Observer 1	4.3	6.5	5.2	0.8
	Observer 2	5.6	6.5	4.9	0.8
	Observer 3	4.8	6.5	4.7	0.8
	Observer 4	6.4	6.5	4.8	0.8
	Observer 5	5.9	6.5	4.7	0.8
	Observer 6	8.0	6.5	5.2	0.8
	Observer 7	6.5	6.5	4.7	0.8
	Observer 8	6.0	6.5	4.1	0.8
	Observer 9	6.4	6.5	4.6	0.8
	Observer 10	6.5	6.4	4.8	0.8
	Observer 11	3.8	6.5	4.7	0.8
	Observer 12	5.4	6.5	4.4	0.8

Table A.6 – *Continued on next page*

Table A.6 – *Continued from previous page*

		CTV_{HR} D90	Bladder D2 cm^3	Rectum D2 cm^3	Sigmoid D2 cm^3
Patient 3	STAPLE	2.5	6.5	5.3	2.6
	EC	8.4	5.8	4.8	2.2
	Observer 1	2.1	6.4	5.0	2.6
	Observer 2	8.1	5.7	3.9	2.4
	Observer 3	2.3	6.2	4.8	2.6
	Observer 4	8.0	6.0	4.8	2.5
	Observer 5	8.2	5.8	4.2	2.3
	Observer 6	8.0	6.0	5.3	2.5
	Observer 7	2.9	5.6	4.7	2.4
	Observer 8	8.5	5.8	4.7	2.4
	Observer 9	8.2	6.4	5.0	2.7
	Observer 10	8.6	5.9	4.2	2.3
	Observer 11	5.6	5.9	4.6	2.5
	Observer 12	7.0	5.9	4.4	2.5
Patient 4	STAPLE	6.2	6.5	4.4	5.4
	EC	8.9	6.2	4.2	5.2
	Observer 1	5.7	6.4	4.1	5.2
	Observer 2	7.6	6.5	4.0	5.1
	Observer 3	6.7	6.5	4.2	5.3
	Observer 4	9.0	6.4	4.0	5.2
	Observer 5	6.3	6.5	4.2	5.3
	Observer 6	8.6	6.4	4.4	5.3
	Observer 7	4.0	6.5	4.2	5.3
	Observer 8	8.2	6.4	4.1	5.2
	Observer 9	7.7	6.5	4.2	5.3
	Observer 10	7.1	6.4	4.2	5.3
	Observer 11	6.9	6.5	4.2	5.3
	Observer 12	8.6	5.9	3.8	4.8
Patient 5	STAPLE	4.9	6.5	3.7	3.9
	EC	8.7	6.5	3.7	3.9
	Observer 1	5.2	6.5	3.8	4.0
	Observer 2	6.9	6.5	3.2	3.8
	Observer 3	8.1	6.5	3.1	3.7
	Observer 4	7.6	6.5	2.7	3.4
	Observer 5	5.8	6.5	3.6	4.0
	Observer 6	5.0	6.5	3.7	3.9
	Observer 7	4.2	6.5	3.5	3.9
	Observer 8	5.5	6.5	2.8	3.5
	Observer 9	4.8	6.5	2.7	3.5
	Observer 10	8.2	5.9	3.4	3.5
	Observer 11	6.3	6.5	3.2	3.7
	Observer 12	6.7	6.5	2.8	3.4

Table A.6 – *Continued on next page*

Table A.6 – *Continued from previous page*

		CTV_{HR} D90	Bladder D2 cm^3	Rectum D2 cm^3	Sigmoid D2 cm^3
Patient 6	STAPLE	0.9	6.4	5.9	-
	EC	1.1	6.5	5.5	-
	Observer 1	0.9	6.5	4.9	-
	Observer 2	1.6	6.5	5.8	-
	Observer 3	0.9	6.5	5.6	-
	Observer 4	1.3	6.5	5.8	-
	Observer 5	1.0	6.5	5.8	-
	Observer 6	5.6	6.5	5.6	-
	Observer 7	1.1	6.5	5.9	-
	Observer 8	9.0	5.9	5.2	-
	Observer 9	7.1	6.5	5.7	-
	Observer 10	1.2	6.5	5.7	-
	Observer 11	1.0	6.5	5.6	-
	Observer 12	1.2	6.5	5.2	-
Patient 7	STAPLE	8.0	6.5	5.9	3.9
	EC	8.3	5.9	5.2	3.5
	Observer 1	8.0	5.6	5.3	3.4
	Observer 2	8.0	5.9	5.0	3.6
	Observer 3	8.4	5.9	5.2	3.6
	Observer 4	8.5	5.9	5.1	3.6
	Observer 5	7.9	6.5	4.4	3.8
	Observer 6	8.6	6.5	5.8	3.9
	Observer 7	8.7	6.5	5.9	3.9
	Observer 8	8.7	6.5	5.4	3.9
	Observer 9	8.1	6.5	5.5	3.9
	Observer 10	8.7	5.9	5.3	3.5
	Observer 11	8.3	6.5	5.1	3.8
	Observer 12	8.7	6.5	5.2	3.9
Patient 8	STAPLE	6.7	6.4	4.4	5.9
	EC	7.0	6.5	3.6	5.6
	Observer 1	7.0	6.5	3.7	5.9
	Observer 2	7.0	6.5	4.3	5.7
	Observer 3	7.2	6.5	4.4	5.9
	Observer 4	7.3	6.5	3.9	5.7
	Observer 5	6.2	6.5	4.4	5.9
	Observer 6	7.4	6.5	4.5	5.9
	Observer 7	6.0	6.5	4.4	5.8
	Observer 8	7.4	6.5	4.1	5.9
	Observer 9	7.1	6.5	3.4	5.5
	Observer 10	7.4	6.5	3.8	5.9
	Observer 11	6.8	6.2	4.3	5.9
	Observer 12	6.7	6.5	3.9	5.9

A.3 Chapter 8

The maximum magnitude of the MDU (corresponding to potential coordinate system failures) averaged over the patients in each volume category and the frequency of occurrences as a percentage of total margin data points is presented in Table A.7. The impact of these potential failures is assessed by comparing the interpolated structures with the original structures using DSC (Table A.8).

Table A.7: Frequency and magnitude of maximum MDU for each patient.

Patient	Polar (%)	Spherical (%)	Maximum MDU (mm)
1	3.10	0.81	17.91
2	0.00	0.00	19.13
3	1.76	0.05	21.80
4	52.45	44.33	17.07
5	0.04	0.00	13.88
6	0.02	0.00	21.43
7	9.38	0.54	27.78
8	2.41	0.11	20.89
9	0.10	0.06	13.73
11	0.60	0.10	15.14
12	3.31	1.88	13.65
13	0.00	0.00	11.74
14	0.03	0.12	21.13
15	0.05	0.02	14.04
16	0.00	0.00	15.04
17	3.42	5.84	10.43
18	1.64	1.33	14.03
19	0.48	0.19	10.59
20	0.21	0.07	12.51
21	0.39	0.09	19.71
22	0.15	0.00	18.69

Table A.8: Average DSC for each patient of interpolated structures and original structures.

Patient	Polar	Spherical
1	0.96	0.98
2	0.96	0.98
3	0.88	0.96
4	0.97	0.98
5	0.98	0.99
6	0.97	0.97
7	0.97	0.93
8	0.96	0.99
9	0.90	0.98
11	0.97	0.98
12	0.83	0.97
13	0.97	0.99
14	0.93	0.96
15	0.98	0.99
16	0.97	0.99
17	0.96	0.98
18	0.97	0.98
19	0.97	0.98
20	0.95	0.97
21	0.83	0.95
22	0.96	0.99

Table A.9: Significant differences in overlap assessed pairwise between polar (pol), spherical (sph) and conventional (conv) margin approaches.

Target Volume	Approaches Compared	p value	Z statistic
Small	Conv > Pol	<0.001	-8.53
Small	Conv > Pol	<0.001	-8.81
Small	Pol > Sph	<0.001	-4.02
GS	Conv > Pol	<0.001	-6.75
GS	Sph > Pol	<0.001	-5.19
Large	Conv > Pol	0.03	-3.02
Large	Sph > Conv	<0.001	-6.68
Large	Sph > Pol	<0.001	-9.45

Table A.10: Significant differences in potential MT assessed pairwise between polar (pol), spherical (sph) and conventional (conv) margin approaches.

Target Volume	Approaches Compared	p value	Z statistic
GS	Pol > Conv	<0.001	-3.58
GS	Sph > Conv	0.001	-3.47
Large	Pol > Conv	<0.001	-3.92
Large	Sph > Conv	<0.001	-3.88
Large	Pol > Sph	0.002	-3.06

Table A.11: Significant differences in definite MT assessed pairwise between polar (pol), spherical (sph) and conventional (conv) margin approaches.

Target Volume	Approaches Compared	p value	Z statistic
Small	Pol > Conv	<0.001	-3.81
Small	Sph > Conv	<0.001	-3.58
Large	Conv > Sph	<0.001	-3.66
Large	Pol > Sph	0.001	-3.47

Table A.12: Significant differences in definite EIT assessed pairwise between polar (pol), spherical (sph) and conventional (conv) margin approaches.

Target Volume	Approaches Compared	p value	Z statistic
GS	Sph > Conv	0.001	-3.32
GS	Sph > Pol	0.011	-2.54
Large	Conv > Pol	0.02	-2.28
Large	Sph > Conv	<0.001	-3.84
Large	Sph > Pol	<0.001	-3.88