Using an Ultrasound-Derived Model to Assist in Dosimetry for Prostate Cancer Treatment Through Brachytherapy

by

Cindy R. Tam

B.Sc., University of Victoria, 2001
M.Sc., McGill University, 2005

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

The Faculty of Graduate Studies

(Physics)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

August 2013

© Cindy R. Tam 2013
Abstract

Prostate cancer is the most common form of cancer affecting men in Canada. Patients with localized, early stage disease are often treated using prostate brachytherapy, a technique that involves surgically implanting small radioactive capsules or “seeds” in the prostate. Implantation is performed under the guidance of transrectal ultrasound (TRUS) imaging, and treatment is assessed postoperatively for quality assurance purposes. Pelvic CT imaging is used to evaluate the dose delivered to the target; however, it is a challenge to consistently and confidently identify prostate boundaries due to the poor soft tissue contrast on CT. This leads to large variability in CT-defined anatomical contours and calculated dosimetric quality assurance parameters, and has led to increased reliance on other imaging technologies such as MR. Meanwhile, TRUS typically provides high-quality anatomical visualization, but provides insufficient information for dose calculation purposes.

We have developed a new method to transfer ultrasound-based contours to CT images using mathematical modeling and a novel registration technique. The prostate model, derived from TRUS contours, is generated via two streams: one assumes a modified ellipsoid shape (model X), and the other performs a straightforward linear interpolation (model Y). Both are manipulated to account for expected deformations such as TRUS-probe compression and edema. Registration from TRUS to CT spatial coordinates is based on matched seed locations. We evaluate the quality of model-generated contours primarily by comparing the measured volume and dosimetric parameters to the observed variability range determined from manual CT contours. In 19 of model X, and 18 of model Y, cases out of 20, volumes produced were within the variability observed from 5 experienced physicians. However, dose parameters agreed in only a moderate number of cases (9–13), partly motivating a region-specific analysis. We found the least agreement in the posterior apex, with model contours tending to be larger. We discuss the possible reasons for this, as well as implications on the role of modeling in an applied clinical setting. Ultimately, the ultrasound-informed model shows promise, and has many benefits relative to other methods, such those based on CT or MR.
Preface

Aside from the material described below, the work presented in this thesis is original and unpublished.

A portion of the modeling algorithm described in Chapter 2 was previously published in the MSc thesis “Prostate segmentation in ultrasound images using image warping and ellipsoid fitting”, UBC, 2007 by S. Badiei, and the article “Semi-automatic segmentation for prostate interventions”, Med Image Anal, 15(2):226–237, 2011, principal investigator Dr. S. Mahdavi. The algorithm behind registration of matched seed locations was adapted from material presented in the article “Prostate brachytherapy postimplant dosimetry: automatic plan reconstruction of stranded implants”, Med Phys, 38(1):327–342, 2011, principal investigator Dr. N. Chng. Other parts of the computer code used in the analysis were adapted from existing but unpublished software written by N. Chng, R. Kosztyla, and J. Lobo.

This research was approved by the University of British Columbia–British Columbia Cancer Agency Research Ethics Board certificate number H10-03169, titled “Using ultrasound-based prostate modeling towards a new method of prostate brachytherapy quality assessment”. 


# Table of Contents

Abstract .......................................................... ii
Preface .............................................................. iii
Table of Contents ................................................... iv
List of Tables ........................................................ vi
List of Figures ......................................................... viii
List of Abbreviations and Terminology ............................... xii
Acknowledgments ...................................................... xiv

1 Introduction ......................................................... 1
  1.1 Prostate Cancer ............................................... 1
  1.2 Management of Prostate Cancer .......................... 3
    1.2.1 Prostatectomy ........................................ 3
    1.2.2 Radiation therapy ................................... 3
    1.2.3 Hormone therapy .................................... 5
  1.3 Prostate Brachytherapy at BCCA ......................... 5
    1.3.1 Treatment planning .................................. 6
    1.3.2 Implantation ......................................... 8
    1.3.3 Postimplant dosimetry .............................. 10
  1.4 Thesis Motivation .......................................... 12

2 Methods and Materials ........................................... 16
  2.1 Data ......................................................... 16
    2.1.1 Collection ........................................... 16
    2.1.2 Procedure ............................................ 17
    2.1.3 Software .............................................. 17
  2.2 Initialization ............................................... 18
    2.2.1 Prostate segmentation in ultrasound .............. 18
# Table of Contents

2.2.2 Seed segmentation in fluoroscopy ........................................ 19  
2.2.3 Seed segmentation in ultrasound ........................................ 19  
2.2.4 Seed segmentation in CT .................................................. 21  
2.3 Model Algorithm ............................................................... 21  
  2.3.1 Unwarping ................................................................. 23  
  2.3.2 Model X: Untapering and ellipse fitting ............................. 23  
  2.3.3 Model X: 3-D tapered ellipsoid fitting .............................. 24  
  2.3.4 Model X: Slicing and tapering ....................................... 24  
  2.3.5 Model Y: Linear interpolation ....................................... 25  
  2.3.6 Registration and edema ............................................... 25  
  2.3.7 CT contour slicing ..................................................... 25  
2.4 Evaluation ............................................................................ 26  
  2.4.1 STAPLE ...................................................................... 26  
  2.4.2 Volumetric and dosimetric analysis .................................. 28  
  2.4.3 Statistical testing ....................................................... 30  

3 Results .................................................................................... 33  
  3.1 STAPLE .............................................................................. 34  
  3.2 Volumetric and Dosimetric Analysis ...................................... 38  
    3.2.1 Total prostate dosimetry .............................................. 38  
    3.2.2 Sector dosimetry ...................................................... 38  
  3.3 Statistical Testing .............................................................. 40  

4 Discussion ............................................................................... 47  
  4.1 Modeling as an Alternative to MR-CT Fusion .......................... 49  
  4.2 Seed Localization in TRUS ................................................. 50  

5 Conclusions ............................................................................ 53  

Bibliography ............................................................................. 56  

Appendix  
A Data Tables and Figures ...................................................... 67
List of Tables

2.1 Summary of statistical terminology and the general relationship between parameters. ........................................... 28
2.2 Arrangement of raw data for the Friedman Test. .................. 31

3.1 Results of TRUS and fluoroscopy seed cloud reconstruction and comparison. Also indicated under column marked “Thin” are ThinStrand implant cases. \( N_{total} \) is the total number of seeds known to be present during intraoperative data collection, and \( N_{TRUS} \) and \( N_{fluoro} \) are the number of TRUS and fluoroscopy seeds successfully reconstructed, respectively. The mean offset between registered, matched seeds is calculated for \( N_{corr} \) number of seeds for which a correspondence was found. ................................................................. 35

3.2 Number of cases for which the model STAPLE parameter was within the 95% CI variability range of manual values, out of a maximum 20. ................................................................. 38

3.3 Number of cases for which the model total volume and dose parameter was within the 95% CI variability of manual values, out of a maximum 20. ................................................................. 38

3.4 Numbers of cases for which the model sector volume and dose parameter was within the 95% CI variability of manual values, out of a maximum 20. ................................................................. 40

3.5 Results of statistical tests. ................................................. 41

3.6 Absolute differences in the rank sums for the whole prostate volume between all pairs of observers. Note that values are symmetric across the main diagonal. Values in bold are greater than Fisher’s LSD critical value 19.3244 for volume; values marked with an asterisk (*) are greater than Nemenyi’s critical value 40.2860. ................................................................. 42
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.7</td>
<td>Absolute differences in the rank sums for total prostate ( V_{100} ) parameter between all pairs of observers. Refer to Table 3.6 caption for a description; here, Fischer’s LSD critical value is 25.5664.</td>
</tr>
<tr>
<td>3.8</td>
<td>Absolute differences in the rank sums for total prostate ( D_{90} ) parameter between all pairs of observers. Refer to Table 3.6 caption for a description; here, Fischer’s LSD critical value is 24.4543.</td>
</tr>
<tr>
<td>A.1</td>
<td>Prostate volume (cc) measured for all 20 patients, based on 5 manual (A–E) and 2 model (X and Y) sets of contours.</td>
</tr>
<tr>
<td>A.2</td>
<td>Prostate ( V_{100} ) (%) calculated for all 20 patients, based on 5 manual (A–E) and 2 model (X and Y) sets of contours.</td>
</tr>
<tr>
<td>A.3</td>
<td>Prostate ( D_{90} ) (Gy) calculated for all 20 patients, based on 5 manual (A–E) and 2 model (X and Y) sets of contours.</td>
</tr>
<tr>
<td>A.4</td>
<td>Ranks ( R(X_{ij}) ) and rank sums ( R_j ) of prostate volume given in Table A.1 for use in the Friedman statistical test and post-hoc analysis.</td>
</tr>
<tr>
<td>A.5</td>
<td>Ranks ( R(X_{ij}) ) and rank sums ( R_j ) of prostate ( V_{100} ) given in Table A.2 for use in the Friedman statistical test and post-hoc analysis.</td>
</tr>
<tr>
<td>A.6</td>
<td>Ranks ( R(X_{ij}) ) and rank sums ( R_j ) of prostate ( D_{90} ) given in Table A.3 for use in the Friedman statistical test and post-hoc analysis.</td>
</tr>
</tbody>
</table>
# List of Figures

1.1 The prostate and nearby organs, as viewed on a sagittal cross section. Source: National Cancer Institute [34]. Creator: Alan Hoofring.  

1.2 Example set of TRUS images from a preimplant volume study. Nine axial slices separated by 0.5 cm are shown, beginning above the prostate base (upper left), and moving inferiorly (left to right, then top to bottom) to below the apex (lower right).  

1.3 The same example set of TRUS images as in Figure 1.2 with contours overlaid indicating the CTV (red) and PTV (cyan).  

1.4 Example treatment plan, overlaid on the same set of TRUS images as in Figure 1.3. Also indicated are the implant template grid (yellow crosses), needle locations (open yellow circles), seed locations (filled cyan circles), and the 100% (green), 150% (orange), and 200% (magenta) isodose contours.  

1.5 Prostate brachytherapy being performed under ultrasound guidance. Source: Mayo Foundation for Medical Education and Research [25]. Used with permission from MayoClinic.com. All rights reserved.  

1.6 Example set of RO-defined contours overlaid on axial CT images (left) and rendered as 3-D surfaces (right). Anatomical structures included are the prostate (red), seminal vesicles (orange), bladder (yellow), urethra (yellow-green), and rectum (blue).  

1.7 An example demonstrating the large interobserver variability that can exist in CT contouring. Shown here are an axial CT scan slice located towards the prostate base without (top left) and with (top right) contours overlaid. Contours were manually defined by 5 experienced physicians. We also show the same contours as seen on a sagittal (bottom left) and coronal (bottom right) representation.
List of Figures

1.8 Midgland prostate slices as seen on intraoperative TRUS (left) and postoperative CT (right). Images are approximately scaled to the same dimensions. The intraoperative TRUS image was obtained after 3 strands had been implanted, while the postoperative CT was obtained after 22. Note how soft tissue boundaries are distinct on TRUS, but seeds locations are obscured, while the opposite is true on CT. 15

2.1 Sagittal ultrasound images, without (left) and with (right) an RO-defined contour overlaid. 19

2.2 A fluoroscopy coronal projection, containing 22 seeds. The TRUS probe and Foley catheter are also visible. 20

2.3 Flowchart illustrating the two modeling streams: the tapered ellipsoid model X, and linear interpolation model Y. We have emphasized in bold the modeling steps that are unique to this project. 22

2.4 Schematic diagram illustrating the 9 sector divisions. 29

2.5 Schematic showing the 9 sectors labeled as visualized on a sagittal view of the prostate. 30

3.1 Model X (left) and Y (right) final products are shown as colour surfaces, with the CT contours overlaid in coloured lines. Matched CT (solid squares) and ultrasound seeds (open squares) are indicated by a solid line connecting them. Note that extraprostatic seeds are intentionally placed in the posterior to achieve the treatment planning goals of delivering a high dose in the posterior peripheral zone. 33

3.2 An example set of evaluated STAPLE weight parameter $W_i$ maps. Note that voxel values tend to be nearly 0 or 1, so the consensus volume is largely independent of our choice of $W_i$ threshold. 36

3.3 Results of STAPLE analysis. Mean sensitivity $p$ and positive predictive values $PV(1)$ and 95% CIs derived from manual contours are shown (blue crosses and error bars), along with model X (red circles) and Y (green squares) values. 37
3.4 Example illustrating high sensitivity with low positive predictive value. Shown are contours generated by three manual “decisions” $D_A, D_B,$ and $D_C,$ the consensus region representing the estimated true segmentation $T,$ and one model decision $D_X.$ In this scenario, all decision contours result in roughly the same sensitivity, but by selecting a proportionally larger area outside of $T$, $D_X$ will have a relatively smaller positive predictive value. 

3.5 Volume and dosimetric parameter results for total prostate. Mean manual contouring values and 95% CIs (blue crosses and error bars) are shown along with values produced by model X (red circles) and model Y (green squares). Overplotted are the dosimetric “sub-optimal” QA thresholds (dotted lines): $V_{100} = 85\%$ and $D_{90} = 130$ Gy and 180 Gy. 

3.6 Colour sector diagrams representing the level of agreement in volume, $V_{100}$ and $D_{90}$ for models X and Y. The colour scale on the right indicates the number of cases, out of a maximum 20, for which the model parameter was within the 95% CI variability of manual values, in each sector as seen from the sagittal view (refer to Figure 2.5). 

3.7 Sagittal, axial and coronal view of a patient’s CT scan, with STAPLE (cyan) and model X (magenta) contours overlaid. Seed locations are also marked by cyan bars. The apical posterior sector (APS) tends to be larger on model-generated contours than it is typically defined manually. 

A.1 Volumetric and dosimetric results for the BAS. Mean manual contouring values and 95% CIs (blue crosses and error bars) are shown along with values derived from model X (red circles) and Y (green squares) contours. Overplotted are the dosimetric “sub-optimal” QA thresholds (dotted lines): $V_{100} = 85\%$ and $D_{90} = 130$ Gy and 180 Gy. 

A.2 Volumetric and dosimetric results for the MAS. Refer to Figure A.1 caption for a description. 

A.3 Volumetric and dosimetric results for the AAS. Refer to Figure A.1 caption for a description. 

A.4 Volumetric and dosimetric results for the BLS. Refer to Figure A.1 caption for a description. 

A.5 Volumetric and dosimetric results for the MLS. Refer to Figure A.1 caption for a description.
List of Figures

A.6  Volumetric and dosimetric results for the ALS. Refer to Figure A.1 caption for a description. ............................... 78
A.7  Volumetric and dosimetric results for the BPS. Refer to Figure A.1 caption for a description. ............................... 79
A.8  Volumetric and dosimetric results for the MPS. Refer to Figure A.1 caption for a description. ............................... 80
A.9  Volumetric and dosimetric results for the APS. Refer to Figure A.1 caption for a description. ............................... 81
List of Abbreviations and Terminology

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-D/3-D</td>
<td>2-/3-Dimensional</td>
</tr>
<tr>
<td>AAPM</td>
<td>American Association of Physicists in Medicine</td>
</tr>
<tr>
<td>AAS</td>
<td>Apex Anterior Sector</td>
</tr>
<tr>
<td>ADT</td>
<td>Androgen Deprivation Therapy</td>
</tr>
<tr>
<td>ALS</td>
<td>Apex Lateral Sector</td>
</tr>
<tr>
<td>APS</td>
<td>Apex Posterior Sector</td>
</tr>
<tr>
<td>ASQ</td>
<td>Anterior Superior Quadrant</td>
</tr>
<tr>
<td>BAS</td>
<td>Base Anterior Sector</td>
</tr>
<tr>
<td>BCCA</td>
<td>British Columbia Cancer Agency</td>
</tr>
<tr>
<td>BLS</td>
<td>Base Lateral Sector</td>
</tr>
<tr>
<td>BPS</td>
<td>Base Posterior Sector</td>
</tr>
<tr>
<td>CT</td>
<td>Computed Tomography</td>
</tr>
<tr>
<td>CTV</td>
<td>Clinical Target Volume</td>
</tr>
<tr>
<td>$D_{90}$</td>
<td>The minimum dose delivered to the hottest 90% of the volume</td>
</tr>
<tr>
<td>DRE</td>
<td>Digital Rectal Exam</td>
</tr>
<tr>
<td>DVH</td>
<td>Dose Volume Histogram</td>
</tr>
<tr>
<td>Gy</td>
<td>Gray</td>
</tr>
<tr>
<td>HDR</td>
<td>High Dose Rate</td>
</tr>
</tbody>
</table>
List of Abbreviations and Terminology

LDR  Low Dose Rate
MAS  Midgland Anterior Sector
MLS  Midgland Lateral Sector
MPS  Midgland Posterior Sector
MR   Magnetic Resonance
PSA  Prostate Specific Antigen
PTV  Planning Target Volume
QA   Quality Assurance
RO   Radiation Oncologist
STAPLE Simultaneous Truth and Performance Level Estimation
TG   Task Group
TRUS Transrectal Ultrasound

$V_{100}$  The volume receiving at least $100\%$ of the prescription dose
Acknowledgments

First and foremost, I would like to thank my supervisor Ingrid Spadinger, whose wealth of knowledge, guidance, and support were invaluable to me and the success of my research. Many thanks are also due to my co-supervisor Steven Thomas and departmental supervisor Alex MacKay, for their unceasing encouragement. I owe special thanks to Nick Chng, whose expertise in programming was indispensable to my project, and who seemed to know the answer to just about everything I had to ask about anything. Special thanks also to Jim Morris, who donated much of his valuable time and expertise to this work. I am extremely grateful to the following physicians at the BCCA who volunteered to participate in my study and kindly donated their skills and time: Muneeza Akbar, Arthur Cheung, Juanita Crook, Mira Keyes, Mitchell Liu, Michael McKenzie, Jim Morris, and Tom Pickles. Many others have generously contributed valuable advice and assistance, including Sara Mahdavi, Rob Kosztyla, Julio Lobo, Mehdi Moradi and Tim Salcudean. I am indebted to BCCA staff who were instrumental during the collection of my data, especially radiation therapists, who took on responsibilities I never asked of them, and made my life easier in the process.

Finally, I thank my family for their constant support, especially my husband Aaron, and our son Elijah, whose immeasurable contribution was simply his arrival in the middle of all this.
Chapter 1

Introduction

1.1 Prostate Cancer

Prostate cancer is the most commonly diagnosed form of cancer affecting men in Canada, accounting for approximately 25% of new cancer diagnoses so far in 2013 [12]. The prostate, a male reproductive gland, secretes part of the fluid in semen. It is located inferior to the bladder and anterior to the rectum, with the urethra passing through it, as seen in Figure 1.1. Also superior and posterior to the gland are the seminal vesicles. The androgen testosterone stimulates the reproduction and growth of prostate tissue cells, including both normal and cancerous cells. Symptoms of prostate cancer may include urinary or erectile dysfunction; however, since it is common for the prostate to naturally become enlarged as men age, these symptoms do not necessarily indicate the presence of cancer. Conversely, many men with prostate cancer are asymptomatic. The typical prostate gland in healthy young adult men is roughly the size and shape of a walnut: approximately 3 by 4 by 2 cm in dimension (~20–30 cc), and weighing ~20 g [45]. The superior-most aspect of the gland is commonly referred to as the base, and the inferior-most as the apex; this terminology will be used throughout this work. The prostate can generally be described by 4 main regions: the central zone, peripheral zone, transitional zone, and anterior fibromuscular stroma [47]. Prostate cancers are usually adenocarcinomas, occurring in epithelial cells of glandular tissue, and most commonly, although not exclusively, involve the peripheral zone. The primary risk factor for prostate cancer is age, increasing for men over 50; other risk factors include family history and race, and possibly diet and lifestyle.

Testing for prostate cancer typically involves a physical exam by a physician (a digital rectal exam, DRE) and a blood test for elevated levels of prostate specific antigen (PSA). Normal PSA levels are considered less than ~4.0 ng/mL, although this value is highly dependent on age [76], and a high PSA level alone is not necessarily indicative of disease. Whether regular PSA screening for men ought to be recommended is a subject of continuing controversy [28]. Diagnosis is usually made through pathological testing of
1.1. Prostate Cancer

biopsied material, obtained from ~6–12 core samples using needle extraction under the guidance of transrectal ultrasound (TRUS) imaging. If present, a clinical stage is determined according to the TNM classification system, which describes the site and size of the primary tumour (T1–T4), nodal involvement (N0 or N1) and metastatic spread (M0 or M1). Also, a Gleason score of 2–10 is calculated from the grade assigned to the biopsy material, which describes the level of cell differentiation and indicates how aggressive a tumour may be. Patients are stratified into low, intermediate or high risk groups based on tumour stage, PSA level, and Gleason score [2, 20].

Figure 1.1: The prostate and nearby organs, as viewed on a sagittal cross section. Source: National Cancer Institute [34]. Creator: Alan Hoofring.
1.2 Management of Prostate Cancer

Once prostate cancer has been diagnosed, the most common forms of management are active surveillance, or radical treatment with the intent to cure through prostatectomy and/or radiation therapy. Prostate cancer is a relatively slowly progressing disease with high overall survival rates. For this reason, active surveillance is often recommended for low risk patients because it avoids, or at least delays, the side effects of intervention and prevents unnecessary over-treatment of patients with favourable risk disease \[36\]. Under active surveillance, patients are routinely monitored for disease progression through recurrent PSA tests, DREs, and biopsy, and regularly assessed for continuation of this form of management.

1.2.1 Prostatectomy

Radical prostatectomy is the surgical removal of the prostate gland, seminal vesicles, and other surrounding tissue. Low to intermediate risk patients with localized disease (ie. tumour is confined to the prostate) who have \(\geq 10\) years life expectancy are generally eligible for this treatment method, provided they are deemed fit for surgery under anaesthetic. High risk patients may also be eligible for surgery in combination with other forms of treatment. Open surgery, with the incision in the lower abdomen, is the most common surgical procedure, although surgery may also be performed laparoscopically. At the beginning of the operation, pelvic lymph nodes are dissected and assessed for microscopic invasion; the outcome of pathology testing determines whether the procedure should be carried out, and if so, how much tissue should be removed. Potential side effects of radical prostatectomy, beyond those related to the trauma of surgery such as pain or bleeding, include temporary urinary incontinence typically lasting 6–12 months (patients are typically catheterized for \(\sim 1–2\) weeks following surgery), erectile dysfunction, although in some cases nerve sparing surgery to retain sexual function may be an option, and permanent infertility.

1.2.2 Radiation therapy

Radiation therapy, or radiotherapy, uses high energy ionizing radiation (X-rays or \(\gamma\)-rays) to kill cancer cells by causing lethal DNA strand breaks. However, radiation can damage normal cells as well: thus, the primary goal of radiotherapy is to deliver a sufficiently high dose to the tumour while inflicting the least amount of damage possible to healthy tissue. The main
1.2. Management of Prostate Cancer

Approaches to achieving this balance involve limiting the amount of normal tissue exposed in the targeted volume, and utilizing the differing radiobiological characteristics of cancerous and normal cells [75]. Fractionation, the technique of delivering small doses over a protracted period of time, amplifies both the survival abilities of normal tissue and the radiosensitivity of tumour tissue [17, 33]. For prostate cancer, radiation therapy is delivered through one of two ways: external beam radiation, or brachytherapy.

External beam radiation is the most common form of radiotherapy for cancer treatment in general, utilizing radiation sourced from outside the body, typically generated by a linear particle accelerator known as a linac, that is directed to a localized area inside the patient. The complexity of a treatment delivered through external beam can range from 2 or 4 rectangular “beams” directed to the general area of the patient’s tumour, to treatment plans that involve moving the linac unit and beam-shaping collimators while the beam is on, with the goal of irradiating a precise, complex volume that tightly conforms to the shape of the target. Typically, patients receive treatment in 20 to 40 fractions over the course of 4 to 8 weeks. Prostate cancer patients with stage T1–T3 disease are generally eligible for external beam therapy, although usually only in combination with other forms of treatment for higher stage patients. For advanced stage patients (T4), radiation may be used to relieve symptoms of disease (i.e. palliative, rather than radical, therapy). External beam radiation is also used in combination with (adjuvant) or after (salvage therapy) prostatectomy or brachytherapy. General side effects of radiation therapy may include fatigue, weight loss, or loss of appetite, and more specific to the region are short-term urinary and bowel irritation or dysfunction, normally lasting weeks to months, and erectile dysfunction that may be temporary or permanent.

Brachytherapy, from the Greek prefix meaning “short distance”, delivers treatment through radioactive sources that are implanted directly into the patient at or near the tumour site. Historically, the first uses of radiation therapy were brachytherapy, following the discovery of radium in the late 19th century [17, 74]. The primary advantage of this technique is that a high dose can be deposited close to the target location, with the exposure rapidly falling off with increasing distance from the source, so that nearby structures may be spared. Because of its highly localized nature, patients must have disease that is confined to the prostate (i.e. low or intermediate risk, stages T1–T2c), to be eligible for brachytherapy. In addition, prostate anatomy must be within set size constraints so as to avoid pubic arch interference during source implantation, and patients must be suitable candidates for surgery under anesthesia. Brachytherapy treatments are typically
1.3. Prostate Brachytherapy at BCCA

categorized as being either low dose rate (LDR) or high dose rate (HDR), which describes the radioactive source being used. For HDR treatments, the method of delivery is temporary, in which catheters are used to place the radioactive source at specific locations for limited amount of time. LDR is most frequently delivered through permanent implantation of radioactive sources. For prostate brachytherapy, permanent implant LDR treatment is more common, although HDR is becoming more widely used. Permanent implant brachytherapy is an out-patient procedure requiring an operation under anesthesia to implant the sources. Side effects of brachytherapy are short term discomfort during the recovery from surgery, and radiation toxicity symptoms similar to external beam but reportedly less severe, except for urinary symptoms which tend to be more severe and longer lasting (3–12 months or more) [1, 61].

1.2.3 Hormone therapy

In some cases, hormone therapy known as androgen deprivation therapy (ADT), which blocks either the production or effect of androgens on prostate cells, may be considered. It is used in conjunction with other treatment types, or alone as a management option for advanced or recurrent cancer. The goal is to relieve symptoms such as prostate swelling, and to slow or reverse tumour growth. For some patients, ADT may also improve the effectiveness of radiation therapy, and in these cases ADT is administered for 3–6 months before, during, and after treatment for up to 3 years (or more) [61].

1.3 Prostate Brachytherapy at BCCA

The Provincial Prostate Brachytherapy Program at the British Columbia Cancer Agency (BCCA) was founded in 1997, and since then, over 4000 LDR implants have been performed [51]. Initially, only low-risk or “low-tier” intermediate risk patients were treated; since 2009, prostate brachytherapy has become standard treatment for all low and immediate risk patients, with some receiving hormone therapy before and after their implant [35]. Current eligibility criteria is clinical stage $\leq T2c$, PSA $\leq 20$ ng/mL, and Gleason score $\leq 7$. Recently, the disease-free survival rates from this program based on the first 1006 consecutive patients enrolled were reported to be 96.7% for 5-years and 94.1% 10-years [51]. Certain high risk patients are eligible to receive LDR brachytherapy in combination with external beam and hormone
therapy. Furthermore, at the BCCA Centre for the Southern Interior, HDR prostate brachytherapy is offered as part of a clinical trial.

General recommendations on treatment and quality assurance procedures for permanent implant prostate brachytherapy are described in the reports of The American Association of Physicists in Medicine (AAPM) Task Groups #64 [94] and #137 [58]. Various other professional organizations have published guidelines on performing and analysing permanent prostate brachytherapy: The American Brachytherapy Society (ABS) [21, 52, 53], The American Society of Radiation Oncology (ASTRO) and American College of Radiology (ACR) [69]. Here, we outline the standard clinical procedures followed in the BCCA program.

1.3.1 Treatment planning

Prostate brachytherapy at the BCCA is carried out following the “Seattle method” [7], based on a technique first reported by Holm et al. in 1993 [32], for which treatment is a 2-step procedure. First is the treatment planning step, wherein patients undergo a transrectal ultrasound (TRUS) imaging volume study: a series of parallel axial image “slices” captured at 0.5 cm intervals that encompass the entire prostate anatomy from above the base to below the apex, see Figure 1.2. On this volume study, a preliminary set of organ delineations or “contours” is defined. Contours are generated at BCCA using in-house semi-automatic contouring software, which creates contours for the clinical target volume (CTV). The CTV includes the prostate and a small portion of the seminal vesicles. These are reviewed by the radiation oncologist (RO), and modified if necessary. Once approved, the planning target volume (PTV) is generated, which includes the CTV plus margins, that are added mainly to account for targeting uncertainties; see Figure 1.3.

The medical physicist manually generates a preoperative implant plan (or “preplan”), in which the placement of brachytherapy sources to obtain the optimum treatment is determined. For all patients, 144 Gy as a minimum peripheral dose is prescribed to cover at least 98% of the PTV [50], where 1 Gray (Gy) is the SI unit of absorbed dose (equivalent to 1 J/kg). The same radioactive isotope—Iodine-125 ($^{125}$I), which has a half-life of 59.4 days and generates photons of energy $\sim$30 keV [59]—is used for all prostate brachytherapy treatments. The treatment planning algorithm aims to achieve an inverted horseshoe-shaped continuous region of high dose (150%, or 216 Gy) by increasing the source density in the posterior peripheral zone, while maintaining a lower maximum dose around the urethra; see Figure 1.4. To make the planning and implantation procedures sim-
1.3. Prostate Brachytherapy at BCCA

Figure 1.2: Example set of TRUS images from a preimplant volume study. Nine axial slices separated by 0.5 cm are shown, beginning above the prostate base (upper left), and moving inferiorly (left to right, then top to bottom) to below the apex (lower right).

pler, faster, and less prone to errors, source positions are planned according to a regular template grid pattern, and are symmetric with respect to the midsagittal plane. Other techniques to improve efficiency and reduce errors involve minimizing the number of needle insertions required by maximizing the number of sources per needle, and using uniformly spaced, stranded sources as much as possible (needles and strands are described in §1.3.2).

Dose calculation is performed following the protocols outlined in the AAPM Task Group #43 (TG-43) reports [56, 65]. These documents provide recommendations on the calculation of the 3-dimensional (3-D) dose-rate distribution to water as a function of distance and orientation from low energy photon-emitting brachytherapy sources, taking into account the source strength, geometry, and the effects of attenuation and scatter due to the source material itself. Once a treatment plan has been approved by the medical physicist and RO, the sources are ordered, and arrive pre-loaded in needles for the individualized treatment.
1.3. Prostate Brachytherapy at BCCA

Figure 1.3: The same example set of TRUS images as in Figure 1.2, with contours overlaid indicating the CTV (red) and PTV (cyan).

1.3.2 Implantation

The second step in the treatment procedure is the brachytherapy source implant itself. Typically, this takes place 4–6 weeks following the date of the TRUS volume study. Implantation is performed by the RO, and assisted by radiation therapists and nurses, in a surgical procedure with the patient under anesthesia and placed in lithotomy position. Patients are usually catheterized with a Foley catheter and contrast medium placed in the urethra, for imaging purposes.

Individual brachytherapy sources are in the form of small “seeds” (terms will be used interchangeably throughout this work), approximately the size of a rice grain, spaced together at fixed, predetermined center-to-center distances (usually 1 cm) using tissue absorbable spacers. Seeds and spacers are encased in a braided, tissue absorbable sleeve (the “strand”) that helps maintain the correct seed spacing after implantation. Strands are sterilized prior to the procedure, and implanted through the patient’s perineum using hollow needles that are inserted using a template grid for guidance, following
1.3. Prostate Brachytherapy at BCCA

Figure 1.4: Example treatment plan, overlaid on the same set of TRUS images as in Figure 1.3. Also indicated are the implant template grid (yellow crosses), needle locations (open yellow circles), seed locations (filled cyan circles), and the 100% (green), 150% (orange), and 200% (magenta) isodose contours.

the same pattern as used in planning. Strand configurations (ie. number of seeds and seed spacings) vary from needle to needle and are based on individualized treatment plans. Typically, 80–150 seeds on 18–30 strands are implanted per treatment.

Real-time TRUS imaging is employed during patient set up to ensure that anatomical information matches between planning and implant data, and also to provide for guidance during needle insertions (see Figure 1.5). X-ray fluoroscopic imaging, from which coronal projections are obtained, is also utilized for guidance [94]. The two imaging modalities are complimentary: sources can be easily localized on X-ray fluoroscopy, while soft tissue anatomy is better visualized on TRUS. The procedure typically requires ~1 hour depending on the number of needles required, and patients are released from the clinic on the same day.
1.3. Prostate Brachytherapy at BCCA

Figure 1.5: Prostate brachytherapy being performed under ultrasound guidance. Source: Mayo Foundation for Medical Education and Research [25]. Used with permission from MayoClinic.com. All rights reserved.

1.3.3 Postimplant dosimetry

Each treatment undergoes quality assurance (QA) as part of routine procedures in the prostate brachytherapy program. QA is performed not only to determine the quality of individual treatments, but to provide feedback for improvement of future treatments and evaluate the quality and outcomes of the program as a whole [53]. Postimplant dosimetry, which is the measurement of dose received by regions of interest (again following the AAPM TG-43 protocol), is one component of QA, and is based on CT (computed tomography) imaging. Patients receive a pelvic CT scan in supine position following their implantation. At BCCA, this is currently done on the same day as the implant, and is referred to as Day-0 CT. As a side note, general recommendations state that dosimetric evaluation should ideally be performed ∼4 weeks postimplant, when the effects of prostatic edema are reduced [94]; however, the choice to obtain implant-day CT is based on
1.3. Prostate Brachytherapy at BCCA

Figure 1.6: Example set of RO-defined contours overlaid on axial CT images (left) and rendered as 3-D surfaces (right). Anatomical structures included are the prostate (red), seminal vesicles (orange), bladder (yellow), urethra (yellow-green), and rectum (blue).

practical considerations for patient accessibility, as well as the benefits of receiving immediate feedback on treatment quality [57]. Manual contouring of patient anatomy, including the prostate as well as nearby structures such as the bladder, urethra, seminal vesicles, and rectum, is performed on Day-0 CT by the RO. Accurate 3-D source positions are localized on CT using commercial software; more recently, in-house software that also performs a “plan reconstruction”, wherein individual sources are identified and assigned a strand and seed number according to the treatment preplan [15], has been adopted into routine procedures.

Treatment quality is primarily assessed through dosimetric parameters, which are determined from the quantitative dose distribution graph known as a dose volume histogram (DVH). On a DVH, the differential or cumulative dose to a target or organ at risk can be plotted for any volumetric structure. As related to the target volume (ie. the prostate), the most important metrics utilized in postimplant dosimetry that can be retrieved from the DVH are

• \( V_{100} \): the volume receiving at least 100\% of the prescription dose, reported in cubic centimeters (cc) or as a percentage of total volume, and

• \( D_{90} \): the minimum dose delivered to the hottest 90\% of the volume, reported in Gy or as a percentage of prescription dose.
1.4. Thesis Motivation

Cases for which $V_{100} < 85\%$ and $D_{90} < 130$ Gy or $> 180$ Gy are classified as being dosimetrically “suboptimal”, and may be subject to review by a QA committee [35], along with cases of high toxicity to organs at risk, such as the rectum and urethra. Suboptimal cases are usually re-evaluated on Day-30 with additional CT, and often magnetic resonance (MR), imaging.

At some centers in the BCCA Prostate Brachytherapy Program, MR imaging is regularly obtained on Day-30 and incorporated into dosimetry procedures, for its superior soft tissue edge detecting abilities [9]. This is achieved through the use of coregistration or “fusion” with CT. However MR is not routinely available at all centers, due to limited resources. If obtained, a dose calculation is performed on anatomical structures, manually contoured on MR by the RO, based on seeds as located on CT. The two image sets are fused by matching CT seeds to the negative seed “voids” on MR.

1.4 Thesis Motivation

One of the significant challenges to accurate dose reporting for postimplant QA stems from the limitations of using CT for visualization of patient anatomy. The poor soft-tissue contrast observed on CT presents difficulties in identification of the boundaries of structures such as the prostate, and can lead to large interobserver and intraobserver variability; see Figure 1.7. This is a well documented problem. The AAPM’s TG-137 report §III [57] contains a summary and literature review on the choice of imaging modality (ultrasound, CT and MR) and its impact on contoured volume variability leading to inconsistencies in reported dose. From early on, concerns over CT-based target delineation and its effect on dosimetric parameters have been raised, particularly as compared to the relative consistency found on MR images [6, 24]. In general, prostate volumes defined on CT are larger than they appear on ultrasound and MR [54, 66, 73], both of which allow superior anatomical visualization. Difficulty is most pronounced at the superior and inferior aspects: superiorly, confusion stems from the overlap between the prostate base and the bladder neck, as well as indistinguishability from the adjacent seminal vesicles, while at the apex, the levator ani muscles, genitourinary diaphragm and neurovascular bundles are often mistakenly included [46, 64, 66]. Furthermore, imaging artifacts from the seeds themselves, and user bias and subjectivity on the part of the person contouring can have a non-negligible effect on organ definition. This has a direct impact on the reliability of dose metrics assessed from CT contours alone.
1.4. Thesis Motivation

Figure 1.7: An example demonstrating the large interobserver variability that can exist in CT contouring. Shown here are an axial CT scan slice located towards the prostate base without (top left) and with (top right) contours overlaid. Contours were manually defined by 5 experienced physicians. We also show the same contours as seen on a sagittal (bottom left) and coronal (bottom right) representation.

[29, 37]. Despite this, CT-based postimplant dosimetry is still currently the standard of care in prostate brachytherapy.

More recently, techniques that include prostate delineation on MR have been encouraged by professional organizations such as the AAPM [57] and ABS [21], and put into practice at a number of treatment facilities, includ-
1.4. Thesis Motivation

ing some clinics in the BCCA as mentioned above. Typically, this is done through fusion of Day-30 MR to CT, with registration performed by either seed-to-seed void matching, or comparison of bony and soft tissue anatomy, all leading to greater reproducibility [3, 19, 42, 62, 85]. However, as already alluded to, the cost and availability of MR prohibits its regular use in all clinics. This motivates our search for a technique that could reduce Day-0 CT-based dosimetric variability, which relies only on resources already available in the clinic.

TRUS imaging is routinely used during preimplant and intraoperative procedures, and provides excellent soft tissue contrast compared to CT (see Figure 1.8); however, its use in postoperative dosimetric evaluation is limited by poor seed visualization, making it extremely difficult to locate all of the seeds following a full implant [30]. Seed segmentation in ultrasound is hindered by specularity (high acoustic reflectivity from the smooth, small-diameter surface), clutter and confusion from other nearby highly reflecting objects such as calcifications, and shadowing. It is a well established problem, having been studied and reported on by many groups attempting to devise automated techniques, particularly because of the potential benefits for intraoperative dosimetry if coregistered with fluoroscopy [eg. 23, 48, 83, 86, 88].

Attempts have also been made to fuse ultrasound to CT. Several authors have reported on postimplant dosimetry based on CT seed locations and preimplant TRUS contours, that had been registered using soft tissue landmarks such as the urethra (with Foley catheterization) or rectal surface [8, 44, 55]. However, since the data are obtained separated by weeks or longer, during which the patient has undergone a brachytherapy implant, which is known to cause edema [71, 90], significant differences in the patient anatomy are expected to exist between the two. Furthermore, the presence of the TRUS imaging probe, causing deformation of the patient’s anatomy by compressing the posterior aspect of the prostate, and the fact that TRUS is obtained with the patient in lithotomy position rather than supine, are problems that persist even if the differences due to edema are reduced by obtaining CT and TRUS closer in time [26]. Dosimetry performed on simultaneous TRUS and CT images obtained in the same pose, and fused based on the transducer itself and refined using seed positions, has been shown to be feasible [80]; however, it is unclear whether dose metrics in this set up truly represent the actual dose received by the patient under normal (ie. uncompressed) conditions, not to mention the added inconvenience and discomfort experienced by the patient due to additional TRUS imaging.

Thus, our objective is to develop a method for improving CT-based
1.4. Thesis Motivation

Figure 1.8: Midgland prostate slices as seen on intraoperative TRUS (left) and postoperative CT (right). Images are approximately scaled to the same dimensions. The intraoperative TRUS image was obtained after 3 strands had been implanted, while the postoperative CT was obtained after 22. Note how soft tissue boundaries are distinct on TRUS, but seeds locations are obscured, while the opposite is true on CT.

postimplant dosimetry that uses resources already in place, namely intraoperative TRUS imaging. Anatomical information based on TRUS will inform a mathematical model, which can be registered to CT image coordinates using matched seed positions, and deformed to generate prostate contours that reflect patient anatomy as seen in CT space. Contours based on a mathematical model, unlike those derived from raw TRUS images, take into account effects such as probe compression and edema. Modeling and registration is made possible by software already part of clinical practice at BCCA [15, 40]. TRUS contouring will be aided by sagittal ultrasound, while seed localization is assisted by intraoperative fluoroscopic imaging, and made feasible by capturing only a small number of seeds (~20) early on during the implant procedure. Except for the additional time required to collect data in the operating room (typically ~10–15 minutes), patients undergo standard brachytherapy treatment as they would under normal conditions. We aim to produce consistent contours that can be generated semi-automatically, from which dose metrics can be more objectively derived.
Chapter 2

Methods and Materials

2.1 Data

2.1.1 Collection

Study subjects

The data for this research was collected from a cohort of 21 patients undergoing prostate brachytherapy at the BCCA Vancouver Center between February and September, 2011. Ethics board approval for our study was received, and informed patient consent was obtained from all study subjects. Treatment was performed following routine clinical procedures, with standard materials and equipment. One subject’s data was omitted from the analysis for reasons described in §3, resulting in only 20 case subjects.

Radiation Sources

The main brachytherapy implant sources used were RAPIDStrand (Oncura, Arlington Heights, IL), containing OncoSeed Model 6711 sources. The source (or seed) design is a cylindrical titanium capsule 4.5 mm long and 0.8 mm in diameter, containing radioactive $^{125}\text{I}$ adsorbed onto a silver rod [59, 65]. The strand and spacing material typically takes 60 to 90 days to be absorbed. Strands are loaded into 18-gauge hollow needles (1.270 mm outer diameter) for implantation.

Recently, a thinner seed and strand model, also by Oncura (ThinStrand Model 9011) [68], was introduced into clinical practice at BCCA. Thin-Strands are planned and used similarly to RAPIDStrand, but are loaded into 20-gauge needles (0.9081 mm outer diameter).

Imaging Units

Ultrasound data were obtained with the BK (BK Medical, Herlev, Denmark) Flex Focus 1202 ultrasound scanner, and transducer Type 8848 containing a convex array for imaging in the transverse plane, and a linear array for
2.1. Data

the sagittal plane. The data were acquired in B-mode, which produces 2-D brightness pixel maps.

Fluoroscopy images were obtained with the GE OEC Series 9800 Plus (GE Healthcare, Little Chalfont, UK) mobile C-arm unit, which produced projection images of the patient in the coronal plane. CT images were obtained with the GE Lightspeed RT 16 CT Simulator.

2.1.2 Procedure

The following intraoperative data collection steps were taken during each subject’s implant operation. After 3–4 strands had been implanted, corresponding to the 14–28 seeds that were to be used as registration seeds, we obtained:

- a TRUS volume study,
- a sagittal ultrasound scan, on or close to the prostate midline,
- five fluoroscopic coronal projections, with the C-arm rotated laterally in the transverse plane at the following angles: $-10^\circ$, $-5^\circ$, $0^\circ$, $5^\circ$ and $10^\circ$ relative to the anterior-posterior axis. The precise angle was recorded using a digital protractor.

Following the implant, each patient received a pelvic CT scan, as part of routine QA procedures, consisting of $\sim 40$–$60$ transverse images separated by 0.25 cm. All imaging data was either collected as or converted to the DICOM (Digital Imaging and Communications in Medicine) file format.

2.1.3 Software

Throughout this research, the commercial software VariSeed (Varian Medical Systems, Palo Alto, CA), was used extensively. It is the standard software used clinically at BCCA for permanent implant prostate brachytherapy treatment planning and dose calculation. VariSeed has many functions, a few of which were particularly important to this research:

- manual segmentation of anatomical structures, or “contouring”, and contour manipulation
- seed localization in a 3-D coordinate system
- calculation of isodoses and dose volume histograms for segmented structures
2.2. Initialization

Other software tools developed in-house and used in clinical practice at BCCA were used in this research as well, and will be referred to by the names AutoContour and PlanReconstruction. AutoContour is a semi-automatic contouring program that generates prostate contours on the preimplant TRUS volume study based on user-selected initialization points, edge detection algorithms, and a 3-D ellipsoidal mathematical model [4, 40]. PlanReconstruction performs an automatic 3-D reconstruction of stranded seed trains segmented on CT images, matching each seed to the corresponding preimplant planned seed and needle number [15].

Software developed by collaborators in the Department of Electrical and Computer Engineering, UBC, was used to reconstruct seed locations in 3-D coordinates from coronal fluoroscopy images using a back projection algorithm (described in §2.4 of [39]).

All other software for this analysis was written in Matlab (Mathworks, Natick, MA).

2.2 Initialiation

Here we describe the initialization steps undertaken to prepare the raw data for input to the modeling routine.

2.2.1 Prostate segmentation in ultrasound

An initial set of prostate contours on the intraoperative TRUS volume study was generated following standard clinical guidelines using the program AutoContour. Initialization points, demarcating TRUS probe center and six defined points along the prostate boundary (see [40] for a description), were also exported. Automatic contours were reviewed and manually modified in VariSeed by an RO according to BCCA treatment planning protocols.

A sagittal image of the prostate, obtained to aid in the localization of prostate boundaries in the base and apex region, was manually contoured by the same RO who reviewed the TRUS contours. We then estimated (by eye, because an exact alignment was not given by the ultrasound imaging unit) a coordinate shift between the sagittal and axial images to align their anterior-posterior and superior-inferior boundaries as best as possible. The RO-modified axial contours were then deformed in the anterior-posterior direction to match the sagittal image, on which the anterior boundary was more distinct, especially at the base and apex. No lateral modifications were made. The adjusted prostate contours formed the basis of the prostate model shape.
2.2. Initialization

Figure 2.1: Sagittal ultrasound images, without (left) and with (right) an RO-defined contour overlaid.

2.2.2 Seed segmentation in fluoroscopy

In order to assist with and validate seed segmentation in ultrasound, intraoperative fluoroscopy images were obtained at the same time as the ultrasound volume study. Seeds on fluoroscopic images were localized in 3-D space from five coronal projections obtained at known angles of rotation. Note that fluoroscopy images were obtained with the ultrasound TRUS probe in place, to better match the patient’s anatomical conditions during ultrasound.

A manual reconstruction was performed by hand: the seed cloud was examined and each seed was assigned the appropriate strand number based on the preplan. This process was possible because only a small number of seeds had been implanted at this stage, and their relative positioning was known from the treatment plan. The absolute distance scale of the seed coordinates was not obtained at the time of the implant; however, since we were primarily interested in their position relative to ultrasound, this was not necessary. To assign an absolute distance scale to seed coordinates, we used the fact that neighbouring seeds on a strand had known separations.

2.2.3 Seed segmentation in ultrasound

Intraoperative seeds were manually segmented on TRUS, guided by the fluoroscopic seed reconstruction, the expected preplan location of seeds relative to the implantation template, and the known seed separation. VariSeed was used to facilitate this process: each slice was examined, and seeds were
assigned a strand number and coordinate located on the slice that they appeared the brightest on. Occasionally, the end of a strand extended beyond the first or last slice of the volume study. Most often, this occurred towards the inferior, due to dragging by the needle as it was retracted from the patient, although there are other obstacles that prevent implants from perfectly reproducing the treatment plan in practice (eg. see §1.2.5 of [16] for an overview). In the ~50% of cases for which this occurred, only a subset of seeds could be segmented. Omitted seeds (typically 2–3) were identified, using the fluoroscopy reconstruction as a reference when available, and taken into account in subsequent seed matching procedures. Intraoperative ultrasound seed locations were exported, and used in the modeling for registration purposes.
Validation

Due to the large expected uncertainties in seed segmentation on TRUS images, we validated seed reconstruction by comparison to fluoroscopy. Seeds that could be identified on both modalities were matched, and a registration between their 3-D coordinates was computed, using singular value decomposition \[13\]. The measured rotation and translation was applied to the fluoroscopy seed coordinates to produce the best possible alignment with corresponding ultrasound seeds, and a mean 3-D offset value was calculated from all known pairs. This represented the estimated error on TRUS seed positions.

2.2.4 Seed segmentation in CT

Postoperative CT images were obtained as part of routine QA procedures. A 3-D seed reconstruction of the completed implant was generated based on the preimplant treatment plan using the in-house clinical software Plan-Reconstruction \[15\]. In the clinic, physicists ensure that a complete plan reconstruction is achieved, i.e. that all seeds are accounted for. Having this allowed us to identify precisely which, out of the 80–150 seeds in the full implant, belonged to the 3–4 strands containing our 14–28 registration seeds. Reconstructed CT seed locations were exported and used in the modeling for registration.

2.3 Model Algorithm

In this section, the steps used to create an ultrasound-based model of the prostate are described. A flowchart summarizing the main components is provided in Figure 2.3. We generated model contours via two different streams, which from here on will be referred to as model X and Y. Many of the steps are the same for both versions, and where they diverge is illustrated in Figure 2.3. We indicate in the titles of the following subsections if the procedure was performed for one model set only; otherwise, they are common to both.

The procedures to generate a 3-D prostate model were devised by S. Badiei \[4\] and S. Mahdavi \[40\], and the procedure to match and register seed coordinates was written by N. Chng \[15\]. Details of these algorithms have been described extensively in their publications. What follows is a brief summary of the main components of our procedure, and we have highlighted in Figure 2.3 the components unique to this study.
2.3. Model Algorithm

The input data, having been prepared as described in §2.2, include DICOM images, contours, seed locations and identities, initialization parameters, and the ultrasound sagittal-axial coordinate shift.

In model coordinates, the $x$-axis refers to the lateral direction, increasing from patient right to left, the $y$-axis refers to the anterior-posterior direction, increasing towards the posterior, and the $z$-axis refers to the superior-inferior direction, increasing towards the superior.
2.3. Model Algorithm

2.3.1 Unwarping

The presence of the TRUS probe causes a posterior deformation in patient anatomy, that we refer to as “warping”. Contour and seed coordinates on ultrasound images must be unwarped, since the probe effect is not present in the CT image, and because the modeling algorithm requires that a 2-D elliptical fit be performed to the contour points. The probe is assumed to cause a radial compression that is maximal at the posterior medial aspect of the prostate, and decreases in amplitude with distance from the probe center, and with angle away from the midline. In polar coordinates with the origin at the probe center, the sinusoidal Gaussian function

\[ r_{\text{new}} = r - r \sin(\theta) \exp \left( \frac{-r^2}{2\sigma^2} \right) \]  

(2.1)

describes the radius \( r_{\text{new}} \) of an unwarped point originally with radius \( r \), where \( \theta \) is the angle from the horizontal axis (\( \theta = 90^\circ \) along the midline). \( \sigma \) is the stretch variable in the radial direction, and is measured from the input initialization parameters [4]. Assuming a uniform deformation along the probe axis, equation (2.1) and measured \( \sigma \) can be used to unwarp all input contour points and seed locations.

2.3.2 Model X: Untapering and ellipse fitting

Once the warp compression has been removed, prostate anatomy as seen on transverse ultrasound resembles an ellipse that often tapers towards the anterior. This 2-D contour can be described as an ellipse containing a tapering parameter \( t_1 \), and represented as

\[
\left( \frac{x'}{a_x} \right)^2 + \left( \frac{y'}{a_y} \right)^2 = 1
\]

(2.2)

\[
x' = \frac{x - x_0}{a_x t_1} (y - y_0) + 1
\]

\[
y' = (y - y_0)
\]

where \( x_0, y_0 \) are the ellipse center coordinates, and \( a_x, a_y \) are the radii on the \( x \) and \( y \) axes, respectively [40]. If \( t_1 = 1 \), the anterior tip comes to a point at a \( 90^\circ \) angle; if \( t_1 = 0 \), the equation describes a straightforward ellipse (see Figure 3 in [40] for an illustration). Equation (2.2) was fit to each unwarped axial contour to find \( t_1 \); this is similar to the original procedure.
2.3. Model Algorithm

in AutoContour, except here the fit is performed on intraoperative contours rather than on edge detection boundaries. Contours were then untapered, and a 2-D ellipse was fit to the unwarped, untapered contour points on each transverse ultrasound slice.

2.3.3 Model X: 3-D tapered ellipsoid fitting

In order to create a smooth, continuous 3-D surface, an ellipsoid that tapers in the $-z$ direction (ie. towards the apex) is fit to the 2-D ellipses. The cross-section of the tapered ellipsoid is itself an ellipse. The tapered ellipsoid is defined as

$$\left(\frac{x'}{a_x}\right)^2 + \left(\frac{y'}{a_y}\right)^2 + \left(\frac{z'}{a_z}\right)^2 = 1 \tag{2.3}$$

where $x_0, y_0, z_0$ are the coordinates of the ellipsoid center, $a_x, a_y, a_z$ are the radii on the $x, y$ and $z$ axes, respectively, and $t_2, t_3$ are the respective tapering parameters in the $x$ and $y$ directions along the axis of the probe \[40\].

2.3.4 Model X: Slicing and tapering

Having found the tapered ellipsoid model parameters, we can construct a finely-spaced grid of points representing the prostate surface. First, the centers of the 2-D ellipses from \[2.3.2\] are used to define a line in the 3-D coordinate system. This line will define the centers of the final contours drawn from the tapered ellipsoid model, accounting for any possible pitch and yaw that may exist in the main axis of the ultrasound prostate anatomy relative to CT anatomy.

Next, the tapered ellipsoid model is finely sliced in the $z$-direction, generating a regular grid of closely spaced ellipse boundary points at small slice separations. In order to construct a model surface that closely resembles the ultrasound anatomy as much as possible, we re-apply the anterior tapering that was removed in \[2.3.2\]. The parameter $t_1$ at the $z$-slice locations is estimated by interpolating between the values measured on TRUS slices, and used to apply tapering to model ellipses in the anterior direction. We define our model surface as this fine grid of points.
2.3. Model Algorithm

2.3.5 Model Y: Linear interpolation

A second ultrasound-based model was generated without fitting the surface to a tapered ellipsoid, but by using simple linear interpolation. This model, referred to as “model Y”, was primarily motivated by oncologists’ interest in seeing how spatially transformed TRUS contours would appear on CT images with minimal manipulation applied. Note that the posterior unwarping deformation is treated in the same way for this model.

Without the use of 3-D ellipsoid model parameters, model Y can not be described on a fine spatial scale, being limited by sparsely sampled TRUS slice contours. The midline sagittal contour, however, can provide important information about the prostate base and apex regions missing from the transverse volume study. We define the superior-most and inferior-most points of the sagittal contour as the base and apex, respectively, and a coarse surface model is constructed by linearly interpolating between the base, unwarped axial contours, and apex.

2.3.6 Registration and edema

To transform the prostate model from ultrasound to CT coordinates, a seed-based rigid registration is calculated between the unwarped TRUS and CT seed positions. Since the seed correspondence is already known, computing the rotation and translation factors is relatively straightforward using singular value decomposition [13]. The computed transformation is then applied to the ultrasound model and seeds: thus, the model surface and seed positions are defined in CT image coordinates.

Expansion of the prostate due to edema is expected to occur between intraoperative ultrasound and postoperative CT. We estimate the edema factor as the ratio of the mean radial components of matched seeds around a point defined as the center of expansion (COE): this assumes a uniform expansion around COE. For model X, the COE is defined as the best-fit center of the 3-D tapered ellipsoid \( x_0, y_0, z_0 \) from §2.3.3 for model Y, the COE is defined as the geometric center of the coarse surface model from §2.3.5. If the edema factor is \( >1 \), radial expansion is applied to model coordinates.

2.3.7 CT contour slicing

Final CT contours are generated by extracting axial slices of the transformed model at the \( z \) coordinates corresponding to the axial CT scan slices. This
produces a set of $x,y$ contour coordinates that can be imported and overlaid on CT images in VariSeed or other software for further analysis.

2.4 Evaluation

Routine clinical procedures for prostate brachytherapy patients require the treating RO to contour patient anatomy on postoperative CT for dose calculation purposes. We collected these prostate contours, and also obtained 4 additional sets of CT prostate contours for each patient from volunteer physicians who were recruited to participate in this research. All volunteers were experienced physicians trained in prostate segmentation, and were blinded to each other's contributions, as well as to other imaging modalities (ie. ultrasound) and patient identity.

Thus, our data set for analysis consists of prostate contours for 20 patients (or subjects), from 7 “observers”, which includes both manual and model contour-generators.

2.4.1 STAPLE

The Simultaneous Truth and Performance Level Estimation (STAPLE) algorithm \cite{87} uses an iterative expectation-maximization algorithm to simultaneously compute a probabilistic estimate of the true segmentation, and a quantitative assessment of the performance level of each “rater”, based on the input from multiple raters. The true segmentation is an unknown binary variable $T_i \in \{0, 1\}$ for each voxel $i$ of $N$ total voxels in the CT scan volume, and is either present (1) or absent (0). For each subject, we input to the algorithm a collection of voxel sets, each representing the CT image prostate segmentation “decisions” by one of the 5 manual contour generators. Decisions are also binary $D_i \in \{0, 1\}$, meaning that voxels are either included in the manually contoured region or not, and are directly observable.

At each iterative step, STAPLE evaluates a weight variable $W_i$, representing the conditional probability of the true segmentation at each voxel $i$ being equal to 1, given the decisions $D_i$ and an estimate of the rater performance from a previous iteration. $W_i$ is a continuous parameter between 0 and 1, ie. $W_i \in [0, 1]$, and is defined as

$$W_i^{(k-1)} \equiv f(T_i = 1|D_i, p^{(k-1)}, q^{(k-1)})$$ (2.4)

where $k$ is the iteration number, and $(p^k, q^k)$ is the estimate of the performance level parameters at iteration $k$. Rater performance parameters are
2.4. Evaluation

also evaluated at each iteration, based on the previous iteration’s estimated true segmentation. The sensitivity is defined as \( p_j = \Pr(D_{ij} = 1 | T_i = 1) \) and the specificity as \( q_j = \Pr(D_{ij} = 0 | T_i = 0) \), where \( p_j, q_j \in [0,1] \); these characterize the quality of each rater \( j = 1, \ldots, 5 \). Iteration ends when \( W_i \) converges.

Following the final iterative step, we obtain a consensus segmentation by choosing a threshold value of \( W_i \). To assess the quality of rater performance, the sensitivity and specificity parameters \( p_j, q_j \) based on the final \( W_i \) are evaluated for all observers, of which there are now \( J = 1, \ldots, 7 \) in total (5 manual plus 2 model). Additional performance parameters evaluated are the statistical measures known as the predictive values, defined as \( \text{PV}_J(s) = \Pr(T_i = s | D_{iJ} = s) \), \( \forall s \in 0, 1 \). The positive predictive value corresponds to the case \( s = 1 \), the negative predictive value to \( s = 0 \). From Bayes theorem, we can express \( \text{PV}_J(s) \) in terms of the known quantities \( W, p \) and \( q \):

\[
\text{PV}_J(1) = \frac{\Pr(T_i = 1 | D_{iJ} = 1)}{\Pr(T_i = 1) \cdot \Pr(D_{iJ} = 1 | T_i = 1) + \Pr(T_i = 0) \cdot \Pr(D_{iJ} = 1 | T_i = 0)}
= \frac{f \cdot p}{f \cdot p + (1-f) \cdot (1-q)}
\]

and

\[
\text{PV}_J(0) = \frac{\Pr(T_i = 0 | D_{iJ} = 0)}{\Pr(T_i = 0) \cdot \Pr(D_{iJ} = 0 | T_i = 0) + \Pr(T_i = 1) \cdot \Pr(D_{iJ} = 0 | T_i = 1)}
= \frac{(1-f) \cdot q}{(1-f) \cdot q + f \cdot (1-p)}
\]

where

\[
f = \Pr(T_i = 1) = \frac{1}{N} \sum_{i}^{N} W_i
\]

is the global probability of the true segmentation, and \( W, p \) and \( q \) are understood to be binary random variables, ie. the conditional probability \( f(T_i = 0 | D_i, p, q) \) is \( 1 - f(T_i = 1 | D_i, p, q) \).

The relationship between true segmentation voxels, rater decisions, and performance parameters is summarized in Table 2.1. Later in this work, the terminology True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) will be used to discuss those fractions of voxels that relate to the STAPLE parameters.
2.4. Evaluation

| $D = 1$ | $T = 1$ | $T = 0$ | $PV(1) = \Pr(T = 1|D = 1)$ |
| $D = 0$ | FN | TN | $PV(0) = \Pr(T = 0|D = 0)$ |

Sensitivity

$S = \Pr(D = 1|T = 1)$

Specificity

$S = \Pr(D = 0|T = 0)$

Table 2.1: Summary of statistical terminology and the general relationship between parameters.

**Statistical comparison**

Our goal is to determine whether model-generated CT contours of the prostate can be considered equivalent in quality to those manually-generated, knowing that there is often considerable variability in manual contouring. The benefit of possessing multiple manual contours for each subject is that the interobserver variability can therefore be quantitatively measured. For each subject and STAPLE performance parameter, we estimate the 95% confidence interval (CI) around the sample mean value of the manual segmentations, from the estimated standard error of the mean $\tilde{s}_x$ and the critical value from the Student’s t-distribution $t_\alpha$ corresponding to the $\alpha = 0.05$ level of significance for a non-directional (two-tailed) test:

$$95\%\ CI = \bar{x} \pm t_\alpha \tilde{s}_x.$$ (2.5)

We use the t-distribution in determining the range, which makes the simple assumption that within-subject interobserver measurements are normally distributed. The performance parameters evaluated for model X and Y are then compared against this confidence interval; if within the 95% CI, the model is classified as being “in agreement”.

**2.4.2 Volumetric and dosimetric analysis**

Dosimetric parameters are ultimately what determines the success or failure of a particular treatment in terms of QA. Given each observer’s set of contours, the total prostate volume and corresponding $V_{100}$ and $D_{90}$ can be simply computed using the treatment planning software VariSeed. Within-subject interobserver 95% CIs for manual segmentations are evaluated using the same method as (2.4.1) for the volumetric and dosimetric parameters (ie. volume, $V_{100}$ and $D_{90}$) as a comparison standard against which to compare the model-generated parameters.
2.4. Evaluation

Analysis of the dose delivered to prostate volume as a whole is of limited usefulness because no information about the spatial distribution of dose is given. Therefore, we performed a region-specific analysis by dividing the prostate gland into 9 sectors, similar to the analysis of Mahdavi et al [40, 41]. First the base, midgland and apex regions are defined, and then each is further subdivided into the anterior, lateral (containing both right and left quadrants together) and posterior regions; see Figure 2.4. The reference volume chosen to define the sector axes was the STAPLE consensus volume from §2.4.1 since it represents the best estimate of the “true segmentation”. Thus, the same set of axes in absolute CT coordinates are used to create sectors for all observer segmentations, allowing us to compare “apples to apples”. Similar to the total prostate analysis, we compare model volume, $V_{100}$ and $D_{90}$ parameters to the 95% CI of the manual segmentation means for each sector.

The 9 sectors are named the BAS, MAS, AAS, BLS, MLS, ALS, BPS, MPS and APS, where the first letter in this convention stands for base, midgland, or apex, and the second letter stands for anterior, lateral, or posterior. When viewed sagittally, the sector divisions appear as they do in Figure 2.5 which we include for ease of reference.

It should be emphasized that while differing degrees of variability are expected in each sector, we are not interested in absolutely quantifying or comparing them. Rather, we wish to determine whether model-generated contours of the whole prostate and its sectors are generally within the observed variability.

Figure 2.4: Schematic diagram illustrating the 9 sector divisions.
2.4. Evaluation

Sagittal view

Figure 2.5: Schematic showing the 9 sectors labeled as visualized on a sagittal view of the prostate.

2.4.3 Statistical testing

We also perform traditional statistical tests on the total prostate volume and dosimetric parameters, which is the typical route taken for studies of inter-observer variability. However, simple tests, such as Student’s t-test, are not optimal in data such as ours because in order to test for differences between observers, results from multiple subjects (which naturally have wide ranges of variability) must be combined. Such inter-patient variability could dominate over the interobserver variability we are interested in. Furthermore, it is probable that the values of some parameters of interest, such as $V_{100}$, are not Normally distributed across subjects, so use of parametric tests that assume a Gaussian data set is inappropriate.

We therefore explore the Friedman test, which is a repeated-measures multivariate test for non-parametric data that utilizes ranks; Bewick et al. [5] demonstrate a worked example. In this test, one can detect differences between $k$ different experiments, sometimes called “treatments”, where each experiment is performed once on $b$ different sample groups or “blocks” of data. In our study, patients are blocks, observer contours are treatments, and the volume, $V_{100}$ or $D_{90}$ are the values obtained from each treatment.
2.4. Evaluation

<table>
<thead>
<tr>
<th>Blocks (Patients)</th>
<th>Treatments (Contours)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 2 3 ... k</td>
</tr>
<tr>
<td>1</td>
<td>X_{11} X_{12} ... X_{1k}</td>
</tr>
<tr>
<td>2</td>
<td>X_{21} X_{22} ... X_{2k}</td>
</tr>
<tr>
<td>...</td>
<td>... ... ... ... ...</td>
</tr>
<tr>
<td>b</td>
<td>X_{b1} X_{b2} ... X_{bk}</td>
</tr>
</tbody>
</table>

Table 2.2: Arrangement of raw data for the Friedman Test.

performed on each block. The actual numerical values measured are unimportant, because data are ranked between 1 and k within each block. A notable feature of the Friedman test is that it assumes differences between blocks are large, so no comparisons between blocks are made.

In this analysis, the data take the form of \( b \times k \) mutually independent random variables, with \( X_{ij} \) being the observation in block \( i = 1,2,\ldots,b \) associated with treatment \( j = 1,2,\ldots,k \). Table 2.2 demonstrates how data are arranged.

Values within each row \( i \) (block) are assigned a rank, denoted by \( R(X_{ij}) \), from 1 to \( k \), and the sum of ranks from the \( j \)th column (treatment) is

\[
R_j = \sum_{i=1}^{b} R(X_{ij}). \tag{2.6}
\]

We wish to test the null hypothesis \( H_0 \): that the order of rankings within a block are equally likely, i.e., each treatment has equivalent effect. The alternative hypothesis \( H_1 \) is that at least one treatment yields values that are consistently higher or lower than one other. The Friedman test statistic is defined as

\[
T = \frac{12}{bk(k+1)} \sum_{j=1}^{k} R_j^2 - 3b(k+1). \tag{2.7}
\]

For large sample sizes, as in the case here, we reject \( H_0 \) at the \( \alpha \) level of significance if \( T \) is greater than the critical value, approximated by the \( \chi^2 \) distribution for \( k - 1 \) degrees of freedom.

Post-hoc analysis

If the null hypothesis is rejected, we then perform a post-hoc multiple comparisons test, which identifies which pairs of treatments differ significantly. The procedure requires calculating pairwise absolute differences of the rank
sums given by equation 2.6. All pairs of treatments \( u, v \) simultaneously differ at the experiment-wise error rate \( \alpha \) for which the following inequality is true:

\[
|R_u - R_v| \geq r(\alpha, k, b),
\]

(2.8)

where \( r(\alpha, k, b) \) is the critical value defined by the particular test being performed; we examine results from two tests, described below. The total number of possible pairwise comparisons equals \( k(k - 1)/2 \), where \( u, v \) combinations are given by \( u < v, u = 1, \ldots, k-1, v = u+1, \ldots, k \).

The Nemenyi test, described in detail by Hollander et al. [31], is considered conservative, i.e. less able to detect small differences. Assuming large sample statistics, the critical value is given as

\[
r(\alpha, k, b) \approx q(\alpha, k, \infty) \sqrt{\frac{bk(k + 1)}{12}} \quad (2.9)
\]

where \( q(\alpha, k, \infty) \) is the studentized range statistic (Table A.10, [31]) associated with \( \infty \) degrees of freedom within groups (i.e. large \( b \)).

The Fisher least significant difference (LSD) test modified for nonparametric data, described by Conover [18] and used by Bewick et al. [5], is less conservative, and defines the critical value as

\[
r(\alpha, k, b) = t_\alpha \sqrt{\frac{2 \left( b \sum_{j=1}^{k} \sum_{i=1}^{b} [R(X_{ij})]^2 - \sum_{j=1}^{k} R_j^2 \right)}{(b - 1)(k - 1)}} \quad (2.10)
\]

where \( t_\alpha \) is the \( 1 - \alpha/2 \) quantile of the Student’s t-distribution with \( (b - 1)(k - 1) \) degrees of freedom (Table A.21, [18]).
Chapter 3

Results

Prostate models based on two ultrasound-informed methods were successfully generated for 20 case subjects. Sagittal delineation was not possible in one of the original 21 cases due to poor image quality, so this data set was omitted from the rest of the analysis, as mentioned in §2.1.1. Registration from TRUS to CT coordinates was performed based on matched seed locations, and the model was sliced at 0.25 cm intervals corresponding to the CT scan spacing. Shown in Figure 3.1 are example models generated using the 3-D tapered ellipsoid (model X) and linear interpolation (model Y) methods.

Figure 3.1: Model X (left) and Y (right) final products are shown as colour surfaces, with the CT contours overlaid in coloured lines. Matched CT (solid squares) and ultrasound seeds (open squares) are indicated by a solid line connecting them. Note that extraprostatic seeds are intentionally placed in the posterior to achieve the treatment planning goals of delivering a high dose in the posterior peripheral zone.

In the following evaluation, patients will be referred to by a number from 1 to 20, manual CT contours are referred to by a letter from A to E, and
3.1 STAPLE

the two model contours by X and Y.

Ultrasound seed validation

Seed localization in ultrasound was aided by fluoroscopy whenever possible. Table 3.1 lists the total number of seeds \( N_{\text{total}} \) implanted for each patient at the time of intraoperative data collection, and the number that could be found and identified on the TRUS volume study \( N_{\text{TRUS}} \) and coronal fluoroscopy projections \( N_{\text{fluoro}} \). An average of 93% (range 80%–100%) of seeds were successfully reconstructed on TRUS. In 4 out of the 20 cases, the fluoroscopy seed cloud could not be reconstructed at all; this may have been due to the use of ThinStrands which were noticeably more difficult to detect, or a misalignment of the C-arm out of the patient’s transverse axis causing problems with the simple back projection algorithm. In the remaining 16 cases, an average of 85% (range 50%–100%) of seeds were reconstructed, with failures possibly resulting from the obscuring effect of the ultrasound probe’s presence in the coronal projection.

Matched seed coordinates in TRUS and fluoroscopy were registered and the 3-D relative mean offset was calculated. On average, the number of seeds for which a TRUS-fluoroscopy correspondence was made \( N_{\text{corr}} \) is 80% (range 50%–100%). Using the fluoroscopy positions as the reference frame, we estimate the error in ultrasound seed locations from the average mean offset, which was found to be 0.24 cm (range 0.07–0.49 cm).

3.1 STAPLE

From the STAPLE weight parameter \( W_i \), we estimate the true segmentation from the consensus volume defined at the cutoff threshold of \( W_i \geq 0.5 \), the value used in an example by Warfield et al. [87]. Although seemingly arbitrary, it was observed that the evaluated voxel weights were generally binary in nature, tending towards values either >0.9 or <0.1; therefore, the consensus volume is highly insensitive to the chosen threshold of \( W_i \) (see Figure 3.2).

Having assigned a value of \( T_i = 0 \) or 1 to each CT voxel, the consensus volume representing the STAPLE “truth” is established. From this, performance parameters are calculated for all 7 observers from the definitions given in §2.4.1. It was found that because of the inherent large true negative (TN) fraction by nature of the CT voxel space being so large, the specificity, and therefore the negative predictive value, were always near unity; that is, \( q_J = \Pr(D_{i,J} = 0|T_i = 0) \sim 1 \) and \( \text{PV}_J(0) \sim 1 \) for all \( J \). Thus, we focus on
3.1. STAPLE

<table>
<thead>
<tr>
<th>Patient</th>
<th>Thin</th>
<th>(N_{\text{total}})</th>
<th>(N_{\text{TRUS}})</th>
<th>(N_{\text{fluoro}})</th>
<th>Mean offset (cm)</th>
<th>(N_{\text{corr}})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>0.10</td>
<td>14</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>15</td>
<td>12</td>
<td>12</td>
<td>0.07</td>
<td>12</td>
</tr>
<tr>
<td>3</td>
<td>Y</td>
<td>20</td>
<td>17</td>
<td>18</td>
<td>0.28</td>
<td>15</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>22</td>
<td>22</td>
<td>20</td>
<td>0.14</td>
<td>20</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>20</td>
<td>16</td>
<td>13</td>
<td>0.12</td>
<td>11</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>20</td>
<td>18</td>
<td>20</td>
<td>0.13</td>
<td>18</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>20</td>
<td>19</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>22</td>
<td>18</td>
<td>19</td>
<td>0.20</td>
<td>18</td>
</tr>
<tr>
<td>9</td>
<td>Y</td>
<td>20</td>
<td>20</td>
<td>10</td>
<td>0.12</td>
<td>10</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>22</td>
<td>22</td>
<td>19</td>
<td>0.25</td>
<td>19</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>28</td>
<td>28</td>
<td>28</td>
<td>0.39</td>
<td>28</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td>22</td>
<td>20</td>
<td>22</td>
<td>0.28</td>
<td>20</td>
</tr>
<tr>
<td>13</td>
<td></td>
<td>24</td>
<td>24</td>
<td>14</td>
<td>0.44</td>
<td>14</td>
</tr>
<tr>
<td>14</td>
<td>Y</td>
<td>22</td>
<td>22</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>15</td>
<td>Y</td>
<td>20</td>
<td>20</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>16</td>
<td></td>
<td>22</td>
<td>20</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>17</td>
<td>Y</td>
<td>20</td>
<td>19</td>
<td>15</td>
<td>0.31</td>
<td>14</td>
</tr>
<tr>
<td>18</td>
<td>Y</td>
<td>28</td>
<td>25</td>
<td>17</td>
<td>0.49</td>
<td>14</td>
</tr>
<tr>
<td>19</td>
<td></td>
<td>20</td>
<td>20</td>
<td>19</td>
<td>0.2</td>
<td>19</td>
</tr>
<tr>
<td>20</td>
<td></td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>0.24</td>
<td>18</td>
</tr>
</tbody>
</table>

Table 3.1: Results of TRUS and fluoroscopy seed cloud reconstruction and comparison. Also indicated under column marked “Thin” are ThinStrand implant cases. \(N_{\text{total}}\) is the total number of seeds known to be present during intraoperative data collection, and \(N_{\text{TRUS}}\) and \(N_{\text{fluoro}}\) are the number of TRUS and fluoroscopy seeds successfully reconstructed, respectively. The mean offset between registered, matched seeds is calculated for \(N_{\text{corr}}\) number of seeds for which a correspondence was found.

The results of STAPLE parameter analysis for all 20 subjects are plotted in Figure 3.3, and summarized in Table 3.2. The sensitivity parameters measured for both model-generated segmentations are in agreement with

the sensitivity \(p_J\) and positive predictive value \(PV_J(1)\). To better understand the distinction between them, consider the following definitions: the sensitivity is the probability that the observer chose all the true structure present, while the positive predictive value is the probability that structure is truly present when the observer chose it. Sensitivity is synonymous with “recall” and \(PV(1)\) with “precision” in information retrieval terminology.

The results of STAPLE parameter analysis for all 20 subjects are plotted in Figure 3.3, and summarized in Table 3.2. The sensitivity parameters measured for both model-generated segmentations are in agreement with
3.1. STAPLE

Figure 3.2: An example set of evaluated STAPLE weight parameter $W_i$ maps. Note that voxel values tend to be nearly 0 or 1, so the consensus volume is largely independent of our choice of $W_i$ threshold.

the manual in nearly all cases, while the positive predictive value agrees in only 6 (model X) and 8 (model Y) cases, tending towards values lower than the manually determined 95% CI. A similar sensitivity with relatively low PV(1) can be interpreted as follows. Manual and model observers are generally contouring (and omitting) the same number of voxels inside the consensus volume, meaning the TF and FN numbers are similar. However, the model is selecting a larger fraction of voxels outside the consensus region (FP) than the typical manual contourer, i.e. the model is “overcontouring” relative to the consensus segmentation. This is illustrated in Figure 3.4.
3.1. STAPLE

Figure 3.3: Results of STAPLE analysis. Mean sensitivity $p$ and positive predictive values $PV(1)$ and 95% CIs derived from manual contours are shown (blue crosses and error bars), along with model X (red circles) and Y (green squares) values.
3.2 Volumetric and Dosimetric Analysis

3.2.1 Total prostate dosimetry

The measured volume and dosimetric parameters for the total prostate volume are summarized in Table 3.3, and plotted in Figure 3.5. We include all values in the Appendix for completeness; see Tables A.1, A.2, and A.3.

In 19 cases for model X, and 18 for model Y, out of 20, the prostate volume generated by modeling methods are within the expected variability range determined through manual segmentation, while there is only moderate agreement between model and manual dosimetric parameters. Model contours tended to produce slightly larger volumes compared to manual segmentation.

<table>
<thead>
<tr>
<th>Model</th>
<th>Sensitivity</th>
<th>PV(1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>20</td>
<td>6</td>
</tr>
<tr>
<td>Y</td>
<td>19</td>
<td>8</td>
</tr>
</tbody>
</table>

Table 3.2: Number of cases for which the model STAPLE parameter was within the 95% CI variability range of manual values, out of a maximum 20.

It is notable that at least one case (Patient 19) designated “suboptimal” by routine QA classification schemes, i.e. $V_{100} < 85\%$ and $D_{90} < 130$ Gy based on manual contouring, would have received an acceptable assessment, and subsequently not been subject to review, had the contours been model generated. Conversely, $D_{90}$ for Patients 2 and 12 yields a borderline dosimetry assessment for at least one of the model-generated contours, whereas the average manual contour dose assessment was acceptable.

3.2.2 Sector dosimetry

We break down the analysis into 9 sectors. Plots of the mean volumetric and dosimetric parameter value and 95% CIs for each patient and sector

<table>
<thead>
<tr>
<th>Model</th>
<th>Volume</th>
<th>$V_{100}$</th>
<th>$D_{90}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>19</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>Y</td>
<td>18</td>
<td>12</td>
<td>13</td>
</tr>
</tbody>
</table>

Table 3.3: Number of cases for which the model total volume and dose parameter was within the 95% CI variability of manual values, out of a maximum 20.
3.2. Volumetric and Dosimetric Analysis

Figure 3.4: Example illustrating high sensitivity with low positive predictive value. Shown are contours generated by three manual “decisions” $D_A, D_B,$ and $D_C$, the consensus region representing the estimated true segmentation $T$, and one model decision $D_X$. In this scenario, all decision contours result in roughly the same sensitivity, but by selecting a proportionally larger area outside of $T$, $D_X$ will have a relatively smaller positive predictive value.

(similar to Figure 3.5) are presented in the Appendix, see Figures A.1–A.9. On a handful of occasions for both manual and model generated contours, the measured volume for a particular sector was zero: this is a consequence of the sector divisions for all prostate segmentations being defined in absolute coordinates based on the STAPLE prostate axes. Null results were excluded from the analysis. All sector results are summarized in Table 3.4. To help visualize the results, we present all 9 sectors in a single diagram for each parameter (volume, $V_{100}$ and $D_{90}$) and modeling method (X and Y), using a colour scale to represent level of agreement, in Figure 3.6. For ease of reference, refer to Figure 2.5 which illustrates the sector locations as they appear in the sagittal view.

Upon examination, we find that the apical posterior sector (APS) exhibits the least agreement between model and manual contours. Model-generated volume and dose parameters for fewer than half of the cases were
Table 3.4: Numbers of cases for which the model sector volume and dose parameter was within the 95% CI variability of manual values, out of a maximum 20.

<table>
<thead>
<tr>
<th>Sector</th>
<th>Model</th>
<th>Volume</th>
<th>(V_{100})</th>
<th>(D_{90})</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAS</td>
<td>X</td>
<td>11</td>
<td>13</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Y</td>
<td>10</td>
<td>12</td>
<td>13</td>
</tr>
<tr>
<td>MAS</td>
<td>X</td>
<td>13</td>
<td>17</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Y</td>
<td>10</td>
<td>15</td>
<td>13</td>
</tr>
<tr>
<td>AAS</td>
<td>X</td>
<td>16</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>Y</td>
<td>16</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>BLS</td>
<td>X</td>
<td>15</td>
<td>11</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Y</td>
<td>16</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>MLS</td>
<td>X</td>
<td>14</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Y</td>
<td>15</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>ALS</td>
<td>X</td>
<td>15</td>
<td>17</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>Y</td>
<td>16</td>
<td>17</td>
<td>14</td>
</tr>
<tr>
<td>BPS</td>
<td>X</td>
<td>18</td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>Y</td>
<td>18</td>
<td>16</td>
<td>15</td>
</tr>
<tr>
<td>MPS</td>
<td>X</td>
<td>14</td>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Y</td>
<td>15</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>APS</td>
<td>X</td>
<td>9</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Y</td>
<td>8</td>
<td>9</td>
<td>9</td>
</tr>
</tbody>
</table>

within the manual 95% CI: 9 cases for volume, 10 for \(V_{100}\) and 9 for \(D_{90}\) for model X, and 8, 9 and 9 cases for model Y, respectively. Figure A.9 reveals that in cases of poor agreement, modeling often tends to overestimate the volume relative to manual contours by including regions further to the posterior and inferior of manually-defined boundaries, see Figure 3.7. This has the effect of decreasing the associated \(V_{100}\) and \(D_{90}\) to this sector. To a lesser degree, the same model overestimation of the volume is observed for the midgland posterior sector (MPS) as well (see Figure A.8), although the number of cases for which values exceeded the 95% CI variability range is less extreme.

### 3.3 Statistical Testing

We test for statistical differences in volumetric and dosimetric parameters between observers. Evaluation up to this point has been focused on mea-
3.3. Statistical Testing

Suring how frequently model-determined parameters are in agreement with manual. Statistical testing on the other hand quantifies the significance level of their agreement, and importantly, can be used to determine if either model’s segmentation is more significantly different from the group than any one manual segmentation.

We perform the Friedman test and post-hoc analysis outlined in §2.4.3 on the volume, \( V_{100} \) and \( D_{90} \) values. Values and ranks for each subject (blocks, \( b = 20 \)) produced by each observer’s contours (treatment, \( k = 7 \)) are presented in the Appendix, see Tables A.1–A.6. Results are reported in Table 3.5. The Friedman test statistic \( T \) is calculated with 6 degrees of freedom and the corresponding two-tailed \( p \)-values are shown. For all parameters, we reject the null hypothesis at a significance level \( p < 0.01 \), meaning that contouring between all observers and models are not equivalent.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Friedman</th>
<th>Post-hoc</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( T )</td>
<td>( p )-value</td>
</tr>
<tr>
<td>Volume</td>
<td>61.8857</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>( V_{100} )</td>
<td>18.2786</td>
<td>0.0056</td>
</tr>
<tr>
<td>( D_{90} )</td>
<td>26.9357</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

Table 3.5: Results of statistical tests.

Post-hoc tests reveal which observer’s segmentations differ at the chosen level of significance \( \alpha = 0.05 \). For Nemenyi’s test, given by equation 2.9, we obtain from tables the appropriate Studentized range statistic \( q(0.05, 7, \infty) = 4.170 \) for infinite degrees of freedom, and calculate a critical value of \( r(0.05, 7, 20) = 40.2860 \) for all three parameters. Using the modified Fisher’s LSD test, we obtain from tables \( t_{0.05} = 1.9810 \) for 114 degrees of freedom, and calculate the critical values from equation 2.10, reported in Table 3.5.

We find the absolute differences in the sum of ranks (equation 2.6) between each pair of observers, for comparison against the minimum critical differences calculated from post-hoc tests. All pairwise differences for volume, \( V_{100} \) and \( D_{90} \) are presented in Tables 3.6, 3.7 and 3.8, respectively. Highlighted are those whose difference in rank sum are greater than Fisher’s LSD (in bold) and Nemenyi’s (with asterisk) critical values, and thus considered significantly different.

With Fisher’s LSD test being highly sensitive to small effects, we find that most of the pairwise differences across our three parameters are considered significant, which makes it difficult to identify any standout observers. The more conservative Nemenyi test identifies a smaller subset of those re-
3.3. Statistical Testing

results that are the most significant. Focusing on Nemenyi’s test results, we examine the volume parameter and find that model X is identified as significantly different in 1 pair, and model Y in 2. This is less than the manual segmentation C, which significantly disagrees in all 6 possible pairwise combinations. The $V_{100}$ parameters for model X and Y are not found to be significantly different relative to any other observers. For $D_{90}$ parameter, model Y is significantly different from observer C, but so too are observer A and E. In fact, the overall trend indicated by post-hoc analysis is that observer C is the outlier. Removing C from the comparison, it is still apparent that X and Y are not more different than any other observer by either test on any of the parameters: observers D (in volume), A and E (in $V_{100}$ and $D_{90}$) are equivalent to or exceed X and Y in numbers of significant pairwise differences.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>X</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>-</td>
<td>1</td>
<td>63*</td>
<td>36</td>
<td>2</td>
<td>20</td>
<td>13</td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>-</td>
<td>64*</td>
<td>35</td>
<td>3</td>
<td>19</td>
<td>14</td>
</tr>
<tr>
<td>C</td>
<td>63*</td>
<td>64*</td>
<td>-</td>
<td>99*</td>
<td>61*</td>
<td>83*</td>
<td>50*</td>
</tr>
<tr>
<td>D</td>
<td>36</td>
<td>35</td>
<td>99*</td>
<td>-</td>
<td>38</td>
<td>16</td>
<td>49*</td>
</tr>
<tr>
<td>E</td>
<td>2</td>
<td>3</td>
<td>61*</td>
<td>38</td>
<td>-</td>
<td>22</td>
<td>11</td>
</tr>
<tr>
<td>X</td>
<td>20</td>
<td>19</td>
<td>83*</td>
<td>16</td>
<td>22</td>
<td>-</td>
<td>33</td>
</tr>
<tr>
<td>Y</td>
<td>13</td>
<td>14</td>
<td>50*</td>
<td>49*</td>
<td>11</td>
<td>33</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 3.6: Absolute differences in the rank sums for the whole prostate volume between all pairs of observers. Note that values are symmetric across the main diagonal. Values in bold are greater than Fisher’s LSD critical value 19.3244 for volume; values marked with an asterisk (*) are greater than Nemenyi’s critical value 40.2860.
### 3.3. Statistical Testing

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>X</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td>35</td>
<td>44*</td>
<td>38</td>
<td>8</td>
<td>34</td>
<td>16</td>
</tr>
<tr>
<td>B</td>
<td>35</td>
<td></td>
<td>9</td>
<td>3</td>
<td>27</td>
<td>1</td>
<td>19</td>
</tr>
<tr>
<td>C</td>
<td>44*</td>
<td>9</td>
<td></td>
<td>6</td>
<td>36</td>
<td>10</td>
<td>30</td>
</tr>
<tr>
<td>D</td>
<td>38</td>
<td>3</td>
<td>6</td>
<td></td>
<td>30</td>
<td>4</td>
<td>22</td>
</tr>
<tr>
<td>E</td>
<td>8</td>
<td>27</td>
<td>36</td>
<td>30</td>
<td>26</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>34</td>
<td>1</td>
<td>10</td>
<td>4</td>
<td>26</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td>16</td>
<td>19</td>
<td>30</td>
<td>22</td>
<td>8</td>
<td>18</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.7: Absolute differences in the rank sums for total prostate $V_{100}$ parameter between all pairs of observers. Refer to Table 3.6 caption for a description; here, Fischer’s LSD critical value is 25.5664.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>X</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td>40</td>
<td>54*</td>
<td>40</td>
<td>5</td>
<td>31</td>
<td>12</td>
</tr>
<tr>
<td>B</td>
<td>40</td>
<td></td>
<td>14</td>
<td>0</td>
<td>35</td>
<td>9</td>
<td>28</td>
</tr>
<tr>
<td>C</td>
<td>54*</td>
<td>14</td>
<td></td>
<td>14</td>
<td>49*</td>
<td>23</td>
<td>42*</td>
</tr>
<tr>
<td>D</td>
<td>40</td>
<td>0</td>
<td>14</td>
<td></td>
<td>35</td>
<td>9</td>
<td>28</td>
</tr>
<tr>
<td>E</td>
<td>5</td>
<td>35</td>
<td>49*</td>
<td>35</td>
<td></td>
<td>26</td>
<td>7</td>
</tr>
<tr>
<td>X</td>
<td>31</td>
<td>9</td>
<td>23</td>
<td>9</td>
<td>26</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td>12</td>
<td>28</td>
<td>42*</td>
<td>28</td>
<td>7</td>
<td>19</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.8: Absolute differences in the rank sums for total prostate $D_{90}$ parameter between all pairs of observers. Refer to Table 3.6 caption for a description; here, Fischer’s LSD critical value is 24.4543.
3.3. Statistical Testing

Figure 3.5: Volume and dosimetric parameter results for total prostate. Mean manual contouring values and 95% CIs (blue crosses and error bars) are shown along with values produced by model X (red circles) and model Y (green squares). Overplotted are the dosimetric “sub-optimal” QA thresholds (dotted lines): $V_{100} = 85\%$ and $D_{90} = 130$ Gy and 180 Gy.
3.3. Statistical Testing

Figure 3.6: Colour sector diagrams representing the level of agreement in volume, $V_{100}$ and $D_{90}$ for models X and Y. The colour scale on the right indicates the number of cases, out of a maximum 20, for which the model parameter was within the 95% CI variability of manual values, in each sector as seen from the sagittal view (refer to Figure 2.5).
3.3. Statistical Testing

Figure 3.7: Sagittal, axial and coronal view of a patient’s CT scan, with STAPLE (cyan) and model X (magenta) contours overlaid. Seed locations are also marked by cyan bars. The apical posterior sector (APS) tends to be larger on model-generated contours than it is typically defined manually.
Chapter 4

Discussion

The goal of this study is to determine if an ultrasound-based model of the prostate can provide assistance in performing post-implant dosimetry on CT images. We have generated model-based contours on CT images and evaluated their utility through comparison with manual CT contours, the current standard of care at BCCA and the most direct and cost-effective method for dosimetric evaluation as stated in TG-64 [94], despite the known segmentation difficulties (see also TG-137 [58]). Overall, model-based volumes agree well with manual contours, tending to be slightly larger than the manual mean, which partly motivated our region-specific analysis. Dosimetric parameters $V_{100}$ and $D_{90}$ agree in only a moderate number of cases, although even when model values are not within estimated manual variability, they do not fall far beyond 95% CIs. In cases where this is not true (for example, model contours for Patient 15 show significant deviation from manual values), a low STAPLE positive predictive value confirms that the model segmentation suffers from poor precision.

Comparing the two modeling streams, we observe that STAPLE, volume and dosimetric parameter results are very similar, although the tapered ellipsoid modeling method (X) produces slightly larger volumes than simple linear interpolation (Y). This is unsurprising, given that the ellipsoid model fits a smooth, convex surface shape to the known contour points. This result is also reflected in the higher sensitivity (ie. recall) parameters for model X from the STAPLE analysis.

With regard to our sector specific analysis, the sectors that exhibit the lowest agreement between model and manual contours, namely the midgland and apical posterior (MPS and APS), are regions where the probe warp deformation has its greatest effect. The fact that model-generated contours tend to “over-contour” this sector may indicate that the unwarping procedure requires modification, ie. the posterior may need to be “flattened” in order to resemble manual CT contours. The original intent of the unwarping algorithm was to facilitate 2-D ellipse-fitting to the axial contours [4], not necessarily to reproduce prostate anatomy, so it is perhaps unsurprising that the algorithm would require adjustment. However, it is also
important to keep in mind that our reference for comparison, ie. manual CT contouring, is itself subject to large uncertainties. Smith et al. [73] analysed manually drawn contours on 3-dimensional TRUS (3DTRUS), CT and MR images and found that CT contouring showed the greatest degree of inter- and intra-observer variability. They also report that MR contours at the apex were typically as large as, or larger than, the base, suggesting that perhaps prostate anatomy does not taper in the inferior direction as drastically as interpreted on other types of imaging. Additionally, Gao et al. [27] report that CT contours by radiation oncologists tend to systematically omit some of the prostate posterior, while including anterior normal tissue, when compared with photographic images of a human cadaver. In the future, a more in-depth analysis of model-based contouring might include MR imaging for validation purposes. It would also be interesting to generate a set of physician-reviewed model-based contours, similar to the analysis of Mahdavi et al. [41], and evaluate their level of acceptance by a group of expert observers, compared against other anonymized, non-model-based contours.

As a general observation, our data confirms previous findings in research carried out at this institution that the anterior superior quadrant (ASQ), corresponding to our BAS and MAS, is being routinely underdosed compared to other regions [70]. It is clear that $V_{100}$ and $D_{90}$ to the BAS, as seen in Figure A.1 is well below suboptimal cutoff values of 85% and 130 Gy, respectively, for the majority of patient cases. This is observed for the MAS in Figure A.2 as well, but to a lesser degree. Poor dose coverage in the ASQ may be attributed to a number of factors, including intentional avoidance of excessive dose in the urethra and bladder neck, needle drag or splay during implantation, or contouring uncertainties [70]. However, dosimetric parameters for the whole prostate [50] or the ASQ [79] are not predictive of biochemical relapse in a large cohort study from the same institution. This has been observed in other investigations as well [eg. 10, 11], in contrast to the majority of reports demonstrating that dose response [eg. 63, 82], as well as contouring and seed segmentation uncertainties [37], are tied to biological outcomes. Reasons for this discrepancy, and the role of traditional dose metrics in assessing treatment quality, are matters of ongoing debate [49, 81].

Another notable observation is that TRUS-based model contours generate larger volumes overall relative to manual CT contours. This is perhaps unexpected on first glance, since CT volumes are generally reported to be larger than TRUS (eg. postimplant CT/TRUS volume ratio 1.13 [77], 1.30 [73]). However, as previously discussed, the model unwarping procedure
4.1 Modeling as an Alternative to MR-CT Fusion

adds volume to the raw TRUS contours, which likely accounts for some of the effect.

Finally, Friedman testing further demonstrates the equivalence between model-based and manually-generated CT contours. From a statistical standpoint, both modeling methods generated contours that are no more different from the group as a whole than any individual observer. If prostate brachytherapy treatment quality continues to be assessed mainly on dosimetric parameters such as $V_{100}$ and $D_{90}$ that are derived from CT contours, an ultrasound-based model could provide some assistance in the contouring process, if not replace it outright. In a clinical setting, model-based contours could be used as a starting point to guide physicians, who could then further modify and approve the final segmentation.

4.1 Modeling as an Alternative to MR-CT Fusion

MR-CT fusion has long been considered a solution to the difficulties associated with delineation of soft tissue on CT alone [64, 66]. Typically, fusion-based dosimetry is performed on Day-30 MR and CT images, when the effects of edematic expansion of the prostate have lessened [57]. Between different imaging modalities, MR typically yields the least variability [19, 24, 73], although the medium itself is not without uncertainties. For example, De Brabandere et al. [22] report a “surprisingly large” degree of interobserver contouring variability on T2-weighted MR images, and suggest that the fusion step itself may also be a “weak link” in the procedure. Possible disadvantages aside, MR imaging is not routinely available for all prostate brachytherapy patients, and to obtain it requires extra time and effort on behalf of the patient, not to mention the additional cost. Ultrasound, on the other hand, is regularly utilized for intraoperative guidance during prostate brachytherapy implants, and yields comparable contouring results. Smith et al. [73] report that in volume, length and variability, prostate contouring on 3DTRUS is most consistent with MR. In a study considering MR images for preimplant treatment planning, Liu et al. [38] find strong similarities between US and MR (volume ratio 0.99 ± 0.08), with small differences in volume and dimension being significant intraobserver, but not interobserver. Greater TRUS contour variability observed postimplant (eg. 13% median standard deviation of the volume vs. 7% preimplant [72]) may be attributed to the presence of seeds or intraprostatic hemorrhage obscuring prostate boundaries [92]. Thus, the early intraoperative TRUS images obtained in this work (after ∼19% of seeds have been implanted, on average)
4.2 Seed Localization in TRUS

might be spared some of the seed-induced confusion.

The correspondence between MR and TRUS therefore suggests that contours derived from TRUS-based modeling could provide a suitable substitute for the often-cited “gold standard” of MR-CT fusion, when such resources are not readily available.

This is not the first study to explore the option of incorporating TRUS into postoperative dosimetry. Xue et al. [92] demonstrated the feasibility of dosimetry performed on TRUS directly by segmenting both the prostate and seeds directly on TRUS, but found that interobserver contour variability on postimplant TRUS had a prohibitive effect on the calculated $V_{100}$ and $D_{90}$, showing variations that are less than, but approach, those observed on CT. A follow-up statistical analysis, however, found TRUS and CT-based dosimetry to be indistinguishable [14]. Others [8, 44, 55] have fused preimplant TRUS to postimplant CT based on urethral or rectal anatomy and found comparable dosimetric parameters to MR-CT fusion; however, probe deformation and volume effects such as edema are described as some of the limitations of this technique. Many groups have reported on combining TRUS with fluoroscopy and shown promising results, for the purposes of intraoperative planning [23, 48, 86] or dosimetry [78, 83]. In the present study, our approach using mathematical modeling based on intraoperative TRUS is novel and comes with many benefits. Prostate edge detection is less susceptible to user bias and image degradation caused by $\sim$100 seeds postimplant. Our model-based procedure eliminates the need for manual image fusion, and uses intraoperative, rather than preoperative, TRUS. Registration is based on matched seed locations, rather than anatomical features that may potentially be poorly defined, such as the urethra without catheterization. Furthermore, utilization of a model allows deformations such as probe warping and edema to be taken into account, and possibly corrected for. Finally, we reiterate that truly only resources that are currently in place are needed, eliminating any inconvenience to the patient (ie. fiducial markers, special imaging equipment, urethrography, or postoperative TRUS volume study are not required).

4.2 Seed Localization in TRUS

One of the main limitations to using ultrasound for model-based CT contouring is in the ability to accurately calculate a coordinate registration through matched seeds. Seed localization and identification on CT is very accurate thanks to in-house software [15], the results of which are then manually
verified, but on TRUS this is not a straightforward task [30]. Regardless, ultrasound seed localization is essential to the methods of this algorithm, and we have attempted to quantify the uncertainty, estimating \( \sim 0.24 \) cm error in seed locations.

A variety of factors contribute to the uncertainty. Most notably, brachytherapy seeds on ultrasound are difficult to distinguish from other objects exhibiting similar echogenicity, such as the connecting brachytherapy strand material or soft tissue calcifications, leading to confusion and detection of false positives and false negatives. The high acoustic reflectivity (specularity) of seed casings also creates shadows more distal from the transducer probe, and rotation of cylindrical seeds out of the implant axis can render them “invisible” on a transverse scan. Despite these challenges, significant effort has been directed towards development of algorithms to automatically detect seeds in TRUS (including here at UBC), with preliminary results showing success in controlled test environments [eg. 43, 88, 89, 93]. Other authors have reported on the accuracy of performing manual seed segmentation, with varying levels of success ranging from full reconstruction [91] to only 20%–25% of seeds being identified [60]. For our purposes, without access to the tools utilized by some groups, we are limited to data and techniques that are clinically available, namely manual identification on axial B-mode images. Recall however that with only a partial implant in place, many of the problems encountered when attempting a full postimplant reconstruction, such as confusion or shadowing, are alleviated. In addition, the use of sagittal imaging allowed better visualization of patient anatomy, particularly at the base and apex. Future work may potentially include the use of more advanced sagittal ultrasound techniques, such as those described by Wen et al. [89].

The use of flexible ThinStrands may have compounded some of the negative effects in 6 cases: implanted sources appeared to undergo increased rotation and displacement, and this, combined with their smaller diameters, may have lead to greater difficulty in detection for both ultrasound and fluoroscopy. This is despite studies on phantoms examining multiple imaging platforms that have reported that thin seeds are as easily detected (and possibly even more so) as standard thickness seeds [67, 68]. However, perhaps some of the stated advantages were lost when applied to our clinical patient data, in contrast to other findings [84], possibly due to inexperience on the part of the implanting physician.

Even when seeds are adequately visualized, sparsely sampled TRUS slices (0.5 cm separation), leading to a high probability of seeds falling between slices or beyond the base or apex slice, hinders accurate identification of
the correct strand and seed number. Perhaps future studies should include collection of a larger volume TRUS study, particularly extending inferior to the prostate apex. Despite this, it was possible to locate \( \sim 93\% \) of seeds on average, mainly because of our knowledge of the seeds present at the time of imaging, and our reliance on fluoroscopic visualization.
Chapter 5

Conclusions

Postimplant dosimetry is of utmost importance for the continuous assurance of treatment quality to patients undergoing brachytherapy for the treatment of prostate cancer. However, dosimetry based on manually-drawn CT contours, the current standard of care, is known to suffer from large uncertainties due to the considerable intra- and interobserver variability. Clearly, there is a need for methods that can improve the reliability of calculated dose metrics. To that aim, we have demonstrated the feasibility of using ultrasound-based contours to model patient anatomy and produce contours for CT-based dosimetry. This was possible due to our ability to perform a full source plan reconstruction on postoperative CT images, allowing a matched seed-to-seed correspondence between TRUS and CT to be obtained, from which a coordinate registration could be calculated. Mathematical modeling, based on a modified 3-dimensional ellipsoid shape, permits anatomy as seen on one imaging modality (TRUS) to be deformed to better represent its appearance on another (CT).

In general, model-based prostate contours are within the variability range observed in a set of five manual contours, all generated by experienced physicians well versed in prostate contouring on CT images. We evaluate the quality of model-generated contouring in several ways. In our analysis based on the STAPLE algorithm and definitions of sensitivity and positive predictive value, we find that model-generated contours select all the voxels that most likely represent true structure (ie. the sensitivity of the model, or recall) with the same frequency as manual contours, in nearly all cases (≥19 out of 20). However, modeling also selected more voxels that are likely non-prostate structure (ie. it has a lower positive predictive value, or precision) than manual contours in over half of the cases.

Volume and dosimetric analysis reveals that modeling tends to predict a total prostate volume that is slightly greater than, but still statistically consistent with, manual contouring. This increased volume, however, translates to less agreement of calculated dose parameters $V_{100}$ and $D_{90}$, and motivates our region-specific analysis. By dividing the prostate volume into 9 sectors, we are able to determine that the posterior sectors of the apex and midgland
are the most responsible for this dosimetric disagreement, with the model
tending to over-contour these regions relative to the mean manual value,
leading to an underestimate of the dose metrics in these sectors. We hy-
pothesize that these sectors, corresponding to the region most affected by
TRUS probe compression, may imply that our algorithm to “unwarp” the
anatomical deformation requires further refinement. However, our standard
for comparison is itself subject to uncertainty, which must be factored into
any interpretation of the results.

Statistical testing using the Friedman test for repeated-measures, non-
parametric data indicates that, based on volumetric and dosimetric param-
eters, the contours generated by our 7 total “observers” are not equivalent.
Subsequent post-hoc analysis reveals that model-based contours are no more
different from the group as a whole than any individual manual-contour-
generator, at a significance level of $\alpha = 0.05$. From a purely statistical
standpoint, this seems to imply that dose metrics derived directly from
model-based contours are no more “uncertain” than those from manually
drawn contours; however, our knowledge of the spatial geometry suggests
cautions should be exercised before drawing such a conclusion.

There are several important benefits to creating CT contours from a
model that is based on TRUS images. Contours are generated in a consistent
way using semi-automatic algorithms, so they are less subject to individual
bias or confusion from image artifacts, such as those originating from the
seeds themselves. Another significant strength of our model-based method
is that only resources that are already available in the clinic are required,
which has the obvious benefit of being convenient and noninvasive for the
patient, as well as cost-effective from a public health standpoint. Although
MR is frequently cited as being the “gold standard” for postimplant prostate
dosimetry, strong similarities between anatomical information retrieved from
MR and ultrasound images suggests that a TRUS-based-model may provide
a suitable alternative, in cases where MR is not an option. Also, compared
to methods that rely on raw TRUS images fused to CT, our technique
involving a deformable model and seed-based registration may produce a
better representation of true patient anatomy.

Several limitations of this method do exist, the most significant being the
ongoing challenge of seed localization on TRUS. Seeds are notoriously diffi-
cult to see on B-mode ultrasound images, even with careful manual inspec-
tion. Therefore, any attempt to automate this procedure would prove dif-
cult, if not impossible, without the introduction of specialized equipment,
a step that would compromise the ease-of-use and “readily available” assets
of this technique. Although there are currently other manually-performed
Chapter 5. Conclusions

steps, such as the adjustment of axial images to match sagittal boundaries, these could likely be automated with a moderate amount of technical manipulation.

If ultrasound seed localization could be performed with relative ease and on a reasonable timescale, we propose that model-based contours could be applied in a clinical setting. It is likely that manual modification, or at least review, by a physician would be recommended, however, the technique of using semi-automatic contouring as a starting point for physicians is not uncommon, and in fact, is the basis for routine treatment planning practices at some centers in the BCCA.

In future studies, it would be advantageous to adapt the sagittal ultrasound image collection procedure (eg. perhaps obtaining additional perisagittal images, in addition to the midsagittal) to provide better visualization of organ boundaries, and to devise a more accurate registration method between sagittal and transverse ultrasound. To capture more extraprostatic seeds, collecting TRUS image slices inferior to the apex slice of a standard volume study would be simple to execute and immensely useful. A refined unwarping algorithm to modify the prostate model posterior also warrants examination, guided perhaps by information on prostate anatomy provided by surgical research. Finally, access to MR images of the patient cohort for validation purposes would simplify, and perhaps improve, the evaluation of the model algorithm.
Bibliography


Bibliography


63


## Appendix A

### Data Tables and Figures

<table>
<thead>
<tr>
<th>Patient</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>X</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>29.33</td>
<td>35.53</td>
<td>19.59</td>
<td>27.18</td>
<td>29.87</td>
<td>26.17</td>
<td>25.57</td>
</tr>
<tr>
<td>2</td>
<td>48.47</td>
<td>43.76</td>
<td>32.02</td>
<td>45.50</td>
<td>49.22</td>
<td>40.59</td>
<td>39.60</td>
</tr>
<tr>
<td>3</td>
<td>39.35</td>
<td>36.86</td>
<td>28.54</td>
<td>41.14</td>
<td>40.91</td>
<td>31.40</td>
<td>29.54</td>
</tr>
<tr>
<td>4</td>
<td>46.88</td>
<td>50.93</td>
<td>21.74</td>
<td>44.78</td>
<td>47.03</td>
<td>47.42</td>
<td>46.13</td>
</tr>
<tr>
<td>5</td>
<td>48.44</td>
<td>51.86</td>
<td>15.97</td>
<td>50.23</td>
<td>38.61</td>
<td>53.97</td>
<td>51.80</td>
</tr>
<tr>
<td>6</td>
<td>43.08</td>
<td>37.80</td>
<td>16.79</td>
<td>39.14</td>
<td>33.23</td>
<td>23.69</td>
<td>22.68</td>
</tr>
<tr>
<td>7</td>
<td>57.69</td>
<td>59.24</td>
<td>28.14</td>
<td>64.00</td>
<td>49.65</td>
<td>53.03</td>
<td>49.80</td>
</tr>
<tr>
<td>8</td>
<td>30.29</td>
<td>31.75</td>
<td>19.95</td>
<td>35.51</td>
<td>32.79</td>
<td>29.37</td>
<td>28.19</td>
</tr>
<tr>
<td>9</td>
<td>38.92</td>
<td>36.28</td>
<td>20.40</td>
<td>45.28</td>
<td>39.70</td>
<td>45.11</td>
<td>43.28</td>
</tr>
<tr>
<td>10</td>
<td>39.23</td>
<td>33.55</td>
<td>17.15</td>
<td>43.27</td>
<td>37.38</td>
<td>35.15</td>
<td>33.25</td>
</tr>
<tr>
<td>11</td>
<td>55.67</td>
<td>57.31</td>
<td>29.47</td>
<td>68.86</td>
<td>63.76</td>
<td>66.45</td>
<td>63.55</td>
</tr>
<tr>
<td>12</td>
<td>33.20</td>
<td>21.31</td>
<td>9.70</td>
<td>27.52</td>
<td>21.15</td>
<td>23.33</td>
<td>21.50</td>
</tr>
<tr>
<td>13</td>
<td>43.52</td>
<td>43.87</td>
<td>27.76</td>
<td>49.40</td>
<td>41.19</td>
<td>46.88</td>
<td>44.37</td>
</tr>
<tr>
<td>14</td>
<td>45.03</td>
<td>48.19</td>
<td>28.00</td>
<td>59.98</td>
<td>53.87</td>
<td>55.12</td>
<td>51.65</td>
</tr>
<tr>
<td>15</td>
<td>33.66</td>
<td>44.10</td>
<td>26.84</td>
<td>52.00</td>
<td>50.01</td>
<td>64.61</td>
<td>60.81</td>
</tr>
<tr>
<td>16</td>
<td>38.31</td>
<td>45.86</td>
<td>24.21</td>
<td>46.94</td>
<td>38.19</td>
<td>48.05</td>
<td>43.97</td>
</tr>
<tr>
<td>17</td>
<td>35.43</td>
<td>45.39</td>
<td>20.41</td>
<td>47.07</td>
<td>44.68</td>
<td>48.46</td>
<td>45.35</td>
</tr>
<tr>
<td>18</td>
<td>48.38</td>
<td>47.63</td>
<td>30.54</td>
<td>53.51</td>
<td>52.61</td>
<td>58.05</td>
<td>53.95</td>
</tr>
<tr>
<td>19</td>
<td>45.78</td>
<td>44.59</td>
<td>29.57</td>
<td>53.49</td>
<td>38.69</td>
<td>46.99</td>
<td>44.20</td>
</tr>
<tr>
<td>20</td>
<td>43.86</td>
<td>32.43</td>
<td>22.40</td>
<td>47.58</td>
<td>37.77</td>
<td>33.11</td>
<td>31.32</td>
</tr>
</tbody>
</table>

Table A.1: Prostate volume (cc) measured for all 20 patients, based on 5 manual (A–E) and 2 model (X and Y) sets of contours.
### Table A.2: Prostate $V_{100}$ (%) calculated for all 20 patients, based on 5 manual (A–E) and 2 model (X and Y) sets of contours.

<table>
<thead>
<tr>
<th>Patient</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>X</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>96.19</td>
<td>88.50</td>
<td>83.07</td>
<td>93.66</td>
<td>90.54</td>
<td>94.49</td>
<td>94.11</td>
</tr>
<tr>
<td>2</td>
<td>93.66</td>
<td>91.63</td>
<td>83.77</td>
<td>94.36</td>
<td>90.87</td>
<td>86.35</td>
<td>86.17</td>
</tr>
<tr>
<td>3</td>
<td>93.38</td>
<td>93.08</td>
<td>84.09</td>
<td>92.19</td>
<td>93.47</td>
<td>92.76</td>
<td>93.10</td>
</tr>
<tr>
<td>4</td>
<td>93.31</td>
<td>87.08</td>
<td>71.52</td>
<td>92.71</td>
<td>89.55</td>
<td>88.11</td>
<td>89.18</td>
</tr>
<tr>
<td>5</td>
<td>94.95</td>
<td>89.83</td>
<td>98.39</td>
<td>93.72</td>
<td>94.86</td>
<td>88.88</td>
<td>90.96</td>
</tr>
<tr>
<td>6</td>
<td>93.18</td>
<td>90.70</td>
<td>96.59</td>
<td>92.60</td>
<td>97.20</td>
<td>98.77</td>
<td>98.95</td>
</tr>
<tr>
<td>7</td>
<td>90.71</td>
<td>84.17</td>
<td>87.05</td>
<td>88.23</td>
<td>89.59</td>
<td>92.30</td>
<td>91.77</td>
</tr>
<tr>
<td>8</td>
<td>91.88</td>
<td>85.68</td>
<td>74.98</td>
<td>89.17</td>
<td>93.52</td>
<td>92.75</td>
<td>92.39</td>
</tr>
<tr>
<td>9</td>
<td>93.55</td>
<td>91.92</td>
<td>91.38</td>
<td>88.35</td>
<td>91.84</td>
<td>91.85</td>
<td>93.11</td>
</tr>
<tr>
<td>10</td>
<td>98.54</td>
<td>95.38</td>
<td>97.54</td>
<td>96.04</td>
<td>97.65</td>
<td>95.26</td>
<td>95.52</td>
</tr>
<tr>
<td>11</td>
<td>99.16</td>
<td>99.60</td>
<td>99.48</td>
<td>98.62</td>
<td>99.12</td>
<td>99.11</td>
<td>99.15</td>
</tr>
<tr>
<td>12</td>
<td>97.54</td>
<td>99.39</td>
<td>99.77</td>
<td>98.11</td>
<td>99.95</td>
<td>96.05</td>
<td>97.45</td>
</tr>
<tr>
<td>13</td>
<td>92.65</td>
<td>91.54</td>
<td>87.87</td>
<td>90.51</td>
<td>96.07</td>
<td>94.93</td>
<td>95.38</td>
</tr>
<tr>
<td>14</td>
<td>99.05</td>
<td>98.18</td>
<td>99.77</td>
<td>98.74</td>
<td>97.99</td>
<td>96.91</td>
<td>97.97</td>
</tr>
<tr>
<td>15</td>
<td>99.97</td>
<td>98.11</td>
<td>100.00</td>
<td>98.65</td>
<td>98.75</td>
<td>88.46</td>
<td>91.08</td>
</tr>
<tr>
<td>16</td>
<td>98.50</td>
<td>95.65</td>
<td>88.34</td>
<td>93.37</td>
<td>94.67</td>
<td>92.15</td>
<td>93.56</td>
</tr>
<tr>
<td>17</td>
<td>98.47</td>
<td>97.12</td>
<td>96.46</td>
<td>96.61</td>
<td>96.87</td>
<td>91.33</td>
<td>91.72</td>
</tr>
<tr>
<td>18</td>
<td>90.74</td>
<td>90.11</td>
<td>90.49</td>
<td>93.52</td>
<td>92.90</td>
<td>94.77</td>
<td>96.51</td>
</tr>
<tr>
<td>19</td>
<td>71.70</td>
<td>74.74</td>
<td>46.21</td>
<td>66.00</td>
<td>77.19</td>
<td>83.25</td>
<td>85.08</td>
</tr>
<tr>
<td>20</td>
<td>91.72</td>
<td>94.34</td>
<td>86.76</td>
<td>85.83</td>
<td>92.71</td>
<td>95.92</td>
<td>97.61</td>
</tr>
</tbody>
</table>

Appendix A. Data Tables and Figures
Table A.3: Prostate $D_{90}$ (Gy) calculated for all 20 patients, based on 5 manual (A–E) and 2 model (X and Y) sets of contours.

<table>
<thead>
<tr>
<th>Patient</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>X</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>159.94</td>
<td>138.76</td>
<td>123.91</td>
<td>154.60</td>
<td>146.23</td>
<td>157.07</td>
<td>156.96</td>
</tr>
<tr>
<td>2</td>
<td>156.12</td>
<td>149.29</td>
<td>122.88</td>
<td>156.09</td>
<td>146.59</td>
<td>130.20</td>
<td>129.50</td>
</tr>
<tr>
<td>3</td>
<td>152.32</td>
<td>152.48</td>
<td>129.65</td>
<td>149.11</td>
<td>153.20</td>
<td>150.24</td>
<td>151.40</td>
</tr>
<tr>
<td>4</td>
<td>154.46</td>
<td>137.09</td>
<td>103.46</td>
<td>152.27</td>
<td>142.55</td>
<td>137.65</td>
<td>141.53</td>
</tr>
<tr>
<td>5</td>
<td>161.01</td>
<td>143.66</td>
<td>167.18</td>
<td>156.97</td>
<td>162.40</td>
<td>140.52</td>
<td>147.33</td>
</tr>
<tr>
<td>6</td>
<td>151.55</td>
<td>146.66</td>
<td>157.86</td>
<td>150.67</td>
<td>161.47</td>
<td>170.80</td>
<td>172.06</td>
</tr>
<tr>
<td>7</td>
<td>145.56</td>
<td>130.45</td>
<td>138.71</td>
<td>140.63</td>
<td>143.26</td>
<td>149.06</td>
<td>148.41</td>
</tr>
<tr>
<td>8</td>
<td>151.54</td>
<td>128.50</td>
<td>98.22</td>
<td>141.07</td>
<td>158.40</td>
<td>156.01</td>
<td>154.54</td>
</tr>
<tr>
<td>9</td>
<td>155.68</td>
<td>151.41</td>
<td>148.35</td>
<td>137.87</td>
<td>151.50</td>
<td>150.53</td>
<td>154.92</td>
</tr>
<tr>
<td>10</td>
<td>173.16</td>
<td>160.33</td>
<td>164.78</td>
<td>162.24</td>
<td>171.57</td>
<td>163.22</td>
<td>162.98</td>
</tr>
<tr>
<td>11</td>
<td>176.29</td>
<td>173.90</td>
<td>171.88</td>
<td>169.91</td>
<td>174.62</td>
<td>174.26</td>
<td>175.18</td>
</tr>
<tr>
<td>12</td>
<td>132.39</td>
<td>140.10</td>
<td>132.59</td>
<td>136.82</td>
<td>141.03</td>
<td>129.90</td>
<td>132.61</td>
</tr>
<tr>
<td>13</td>
<td>150.65</td>
<td>148.10</td>
<td>139.57</td>
<td>145.05</td>
<td>160.39</td>
<td>156.44</td>
<td>158.05</td>
</tr>
<tr>
<td>14</td>
<td>181.77</td>
<td>178.11</td>
<td>182.73</td>
<td>176.21</td>
<td>177.93</td>
<td>173.72</td>
<td>178.60</td>
</tr>
<tr>
<td>15</td>
<td>180.30</td>
<td>174.37</td>
<td>178.72</td>
<td>177.52</td>
<td>175.95</td>
<td>137.09</td>
<td>149.55</td>
</tr>
<tr>
<td>16</td>
<td>176.03</td>
<td>163.62</td>
<td>138.48</td>
<td>157.02</td>
<td>165.71</td>
<td>151.07</td>
<td>156.07</td>
</tr>
<tr>
<td>17</td>
<td>183.10</td>
<td>174.85</td>
<td>164.10</td>
<td>173.03</td>
<td>172.85</td>
<td>149.23</td>
<td>150.95</td>
</tr>
<tr>
<td>18</td>
<td>146.82</td>
<td>144.44</td>
<td>145.83</td>
<td>159.52</td>
<td>155.31</td>
<td>161.25</td>
<td>167.93</td>
</tr>
<tr>
<td>19</td>
<td>110.86</td>
<td>117.98</td>
<td>74.92</td>
<td>100.13</td>
<td>123.25</td>
<td>133.69</td>
<td>136.25</td>
</tr>
<tr>
<td>20</td>
<td>147.24</td>
<td>154.18</td>
<td>133.79</td>
<td>134.46</td>
<td>148.98</td>
<td>158.44</td>
<td>162.30</td>
</tr>
</tbody>
</table>

Appendix A. Data Tables and Figures
### Table A.4: Ranks $R(X_{ij})$ and rank sums $R_j$ of prostate volume given in Table [A.1] for use in the Friedman statistical test and post-hoc analysis.

<table>
<thead>
<tr>
<th>Patient</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>X</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>7</td>
<td>1</td>
<td>4</td>
<td>6</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>4</td>
<td>1</td>
<td>5</td>
<td>7</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>7</td>
<td>6</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>7</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>6</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>5</td>
<td>1</td>
<td>6</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>6</td>
<td>1</td>
<td>7</td>
<td>2</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>5</td>
<td>1</td>
<td>7</td>
<td>6</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>7</td>
<td>4</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>10</td>
<td>6</td>
<td>3</td>
<td>1</td>
<td>7</td>
<td>5</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>7</td>
<td>5</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>12</td>
<td>7</td>
<td>3</td>
<td>1</td>
<td>6</td>
<td>2</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>13</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>7</td>
<td>2</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>7</td>
<td>5</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>15</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>5</td>
<td>4</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>16</td>
<td>3</td>
<td>5</td>
<td>1</td>
<td>6</td>
<td>2</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>17</td>
<td>2</td>
<td>5</td>
<td>1</td>
<td>6</td>
<td>3</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>18</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>4</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>19</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>7</td>
<td>2</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>20</td>
<td>6</td>
<td>3</td>
<td>1</td>
<td>7</td>
<td>5</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>$R_j$</td>
<td>83</td>
<td>84</td>
<td>20</td>
<td>119</td>
<td>81</td>
<td>103</td>
<td>70</td>
</tr>
<tr>
<td>Patient</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>X</td>
<td>Y</td>
</tr>
<tr>
<td>---------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>1</td>
<td>7</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>5</td>
<td>1</td>
<td>7</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>7</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>2</td>
<td>1</td>
<td>6</td>
<td>5</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>2</td>
<td>7</td>
<td>4</td>
<td>5</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>7</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>9</td>
<td>7</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>10</td>
<td>7</td>
<td>2</td>
<td>5</td>
<td>4</td>
<td>6</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>11</td>
<td>5</td>
<td>7</td>
<td>6</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>12</td>
<td>3</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>7</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>13</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>7</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>14</td>
<td>6</td>
<td>4</td>
<td>7</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>15</td>
<td>6</td>
<td>3</td>
<td>7</td>
<td>4</td>
<td>5</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>16</td>
<td>7</td>
<td>6</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>17</td>
<td>7</td>
<td>6</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>18</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>4</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>19</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>20</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>

| $R_j$   | 105 | 70 | 61 | 67 | 97 | 71 | 89 |

Table A.5: Ranks $R(X_{ij})$ and rank sums $R_j$ of prostate $V_{100}$ given in Table A.2 for use in the Friedman statistical test and post-hoc analysis.
Appendix A. Data Tables and Figures

<table>
<thead>
<tr>
<th>Patient</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>X</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>5</td>
<td>1</td>
<td>6</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>6</td>
<td>1</td>
<td>2</td>
<td>7</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>2</td>
<td>1</td>
<td>6</td>
<td>5</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>2</td>
<td>7</td>
<td>4</td>
<td>6</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>7</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>9</td>
<td>7</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>10</td>
<td>7</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>6</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>11</td>
<td>7</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td>6</td>
<td>3</td>
<td>5</td>
<td>7</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>13</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>7</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>14</td>
<td>6</td>
<td>4</td>
<td>7</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>15</td>
<td>7</td>
<td>3</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>16</td>
<td>7</td>
<td>5</td>
<td>1</td>
<td>4</td>
<td>6</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>17</td>
<td>7</td>
<td>6</td>
<td>3</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>18</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>4</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>19</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>20</td>
<td>3</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>(R_j)</td>
<td>106</td>
<td>66</td>
<td>52</td>
<td>66</td>
<td>101</td>
<td>75</td>
<td>94</td>
</tr>
</tbody>
</table>

Table A.6: Ranks \(R(X_{ij})\) and rank sums \(R_j\) of prostate \(D_{90}\) given in Table A.3 for use in the Friedman statistical test and post-hoc analysis.
Figure A.1: Volumetric and dosimetric results for the BAS. Mean manual contouring values and 95% CIs (blue crosses and error bars) are shown along with values derived from model X (red circles) and Y (green squares) contours. Overplotted are the dosimetric “sub-optimal” QA thresholds (dotted lines): $V_{100} = 85\%$ and $D_{90} = 130$ Gy and 180 Gy.
Figure A.2: Volumetric and dosimetric results for the MAS. Refer to Figure A.1 caption for a description.
Figure A.3: Volumetric and dosimetric results for the AAS. Refer to Figure [A.1] caption for a description.
Appendix A. Data Tables and Figures

Figure A.4: Volumetric and dosimetric results for the BLS. Refer to Figure A.1 caption for a description.
Appendix A. Data Tables and Figures

Figure A.5: Volumetric and dosimetric results for the MLS. Refer to Figure A.1 caption for a description.
Figure A.6: Volumetric and dosimetric results for the ALS. Refer to Figure A.1 caption for a description.
Figure A.7: Volumetric and dosimetric results for the BPS. Refer to Figure A.1 caption for a description.
Figure A.8: Volumetric and dosimetric results for the MPS. Refer to Figure A.1 caption for a description.
Appendix A. Data Tables and Figures

Figure A.9: Volumetric and dosimetric results for the APS. Refer to Figure [A.1] caption for a description.