

---

Electronic Thesis and Dissertation Repository

---

8-7-2014 12:00 AM

## 3D fusion of histology to multi-parametric MRI for prostate cancer imaging evaluation and lesion-targeted treatment planning

Eli Gibson

*The University of Western Ontario*

Supervisor

Dr. Aaron Fenster

*The University of Western Ontario* Joint Supervisor

Dr. Aaron Ward

*The University of Western Ontario*

Graduate Program in Biomedical Engineering

A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of Philosophy

© Eli Gibson 2014

Follow this and additional works at: <https://ir.lib.uwo.ca/etd>



Part of the [Biomedical Engineering and Bioengineering Commons](#)

---

### Recommended Citation

Gibson, Eli, "3D fusion of histology to multi-parametric MRI for prostate cancer imaging evaluation and lesion-targeted treatment planning" (2014). *Electronic Thesis and Dissertation Repository*. 2256.

<https://ir.lib.uwo.ca/etd/2256>

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact [wlsadmin@uwo.ca](mailto:wlsadmin@uwo.ca).

3D FUSION OF HISTOLOGY TO MULTI-PARAMETRIC MRI FOR PROSTATE  
CANCER IMAGING EVALUATION AND LESION-TARGETED TREATMENT  
PLANNING

(Thesis format: Integrated Article)

by

Eli Gibson

Graduate Program in Biomedical Engineering

A thesis submitted in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy

The School of Graduate and Postdoctoral Studies  
The University of Western Ontario  
London, Ontario, Canada

© Eli Gibson 2014

# Abstract

Multi-parametric magnetic resonance imaging (mpMRI) of localized prostate cancer has the potential to support detection, staging and localization of tumors, as well as selection, delivery and monitoring of treatments. Delineating prostate cancer tumors on imaging could potentially further support the clinical workflow by enabling precise monitoring of tumor burden in active-surveillance patients, optimized targeting of image-guided biopsies, and targeted delivery of treatments to decrease morbidity and improve outcomes. Evaluating the performance of mpMRI for prostate cancer imaging and delineation ideally includes comparison to an accurately registered reference standard, such as prostatectomy histology, for the locations of tumor boundaries on mpMRI. There are key gaps in knowledge regarding how to accurately register histological reference standards to imaging, and consequently further gaps in knowledge regarding the suitability of mpMRI for tasks, such as tumor delineation, that require such reference standards for evaluation.

To obtain an understanding of the magnitude of the mpMRI-histology registration problem, we quantified the position, orientation and deformation of whole-mount histology sections relative to the formalin-fixed tissue slices from which they were cut. We found that (1) modeling isotropic scaling accounted for the majority of the deformation with a further small but statistically significant improvement from modeling affine transformation, and (2) due to the depth (mean $\pm$ standard deviation (SD) 1.1 $\pm$ 0.4 mm) and orientation (mean $\pm$ SD 1.5 $\pm$ 0.9 $^\circ$ ) of the sectioning, the assumption that histology sections are cut from the front faces of tissue slices, common in previous approaches, introduced a mean error of 0.7 mm.

To determine the potential consequences of seemingly small registration errors such as described above, we investigated the impact of registration accuracy on the statistical power of imaging validation studies using a co-registered spatial reference standard (e.g. histology images) by deriving novel statistical power formulae that incorporate registration error. We illustrated, through a case study modeled on a prostate cancer imaging trial at our centre, that submillimeter differences in registration error can have a substantial impact on the required sample sizes (and therefore also the study cost) for studies aiming to detect mpMRI signal differences due to  $0.5 - 2.0 \text{ cm}^3$  prostate tumors.

With the aim of achieving highly accurate mpMRI-histology registrations without disrupting the clinical pathology workflow, we developed a three-stage method for accurately registering 2D whole-mount histology images to pre-prostatectomy mpMRI that allowed flexible placement of cuts during slicing for pathology and avoided the assumption that histology sections are cut from the front faces of tissue slices. The method comprised a 3D reconstruction of histology images, followed by 3D–3D *ex vivo*–*in vivo* and *in vivo*–*in vivo* image transformations. The 3D reconstruction method minimized fiducial registration error between cross-sections of non-disruptive histology- and *ex-vivo*-MRI-visible strand-shaped fiducials to reconstruct histology images into the coordinate system of an *ex vivo* MR image. We quantified the mean $\pm$ standard deviation target registration error of the reconstruction to be  $0.7\pm0.4 \text{ mm}$ , based on the post-reconstruction misalignment of intrinsic landmark pairs. We also compared our fiducial-based reconstruction to an alternative reconstruction based on mutual-information-based registration, an established method for multi-modality registration. We found that the



mean target registration error for the fiducial-based method (0.7 mm) was lower than that for the mutual-information-based method (1.2 mm), and that the mutual-information-based method was less robust to initialization error due to multiple sources of error, including the optimizer and the mutual information similarity metric. The second stage of the histology–mpMRI registration used interactively defined 3D–3D deformable thin-plate-spline transformations to align *ex vivo* to *in vivo* MR images to compensate for deformation due to endorectal MR coil positioning, surgical resection and formalin fixation. The third stage used interactively defined 3D–3D rigid or thin-plate-spline transformations to co-register *in vivo* mpMRI images to compensate for patient motion and image distortion. The combined mean registration error of the histology–mpMRI registration was quantified to be 2 mm using manually identified intrinsic landmark pairs.

Our data set, comprising mpMRI, target volumes contoured by four observers and co-registered contoured and graded histology images, was used to quantify the positive predictive values and variability of observer scoring of lesions following the *Prostate Imaging Reporting and Data System* (PI-RADS) guidelines, the variability of target volume contouring, and appropriate expansion margins from target volumes to achieve coverage of histologically defined cancer. The analysis of lesion scoring showed that a PI-RADS overall cancer likelihood of 5, denoting “highly likely cancer”, had a positive predictive value of 85% for Gleason 7 cancer (and 93% for lesions with volumes  $>0.5 \text{ cm}^3$  measured on mpMRI) and that PI-RADS scores were positively correlated with histological grade ( $\rho=0.6$ ). However, the analysis also showed interobserver differences in PI-RADS score of 0.6 to 1.2 (on a 5-point scale) and an agreement kappa value of only 0.30. The analysis of target volume contouring showed that target volume contours with

suitable margins can achieve near-complete histological coverage for detected lesions, despite the presence of high interobserver spatial variability in target volumes.

Prostate cancer imaging and delineation have the potential to support multiple stages in the management of localized prostate cancer. Targeted biopsy procedures with optimized targeting based on tumor delineation may help distinguish patients who need treatment from those who need active surveillance. Ongoing monitoring of tumor burden based on delineation in patients undergoing active surveillance may help identify those who need to progress to therapy early while the cancer is still curable. Preferentially targeting therapies at delineated target volumes may lower the morbidity associated with aggressive cancer treatment and improve outcomes in low-intermediate-risk patients. Measurements of the accuracy and variability of lesion scoring and target volume contouring on mpMRI will clarify its value in supporting these roles.

## Keywords

Magnetic resonance imaging, MRI, prostate cancer, registration, 3D reconstruction, focal therapy, radiology, radiation oncology, pathology, histology, histopathology.

## Co-Authorship Statement

This thesis is presented in an integrated article format, the chapters of which are based on the following publications that are either published, in press, under review, or in preparation for submission:

Chapter 2: E. Gibson, A. Fenster and A. D. Ward, "The impact of registration accuracy on imaging validation study design: a novel statistical power calculation," *Medical Image Analysis* 17, 805-815 (2013).

My contributions to this work included defining the research questions, deriving the power calculation formulae, formulating the experimental design, implementing and analyzing the Monte Carlo simulations, and drafting the manuscript. A. Fenster motivated the initial research direction. A. D. Ward contributed to defining the research questions, formulating the case study analysis, interpreting the results and drafting the manuscript. All authors helped in reviewing and editing the manuscript. The work was performed under the supervision of A. Fenster and A. D. Ward, and the project oversight of G. S. Bauman.

Chapter 3: E. Gibson, M. Gaed, J. A. Gómez, M. Moussa, S. Pautler, J. Chin, C. Crukley, G. S. Bauman, A. Fenster and A. D. Ward, "3D prostate histology image reconstruction: quantifying the impact of tissue deformation and histology section location ," *Journal of Pathology Informatics* 4, 31 (2013).

My contributions to this work included defining the research questions; identifying measurement landmarks; designing, implementing and analyzing the experiments; and drafting the manuscript. M. Gaed, J. A. Gómez, M. Moussa, and C. Crukley contributed to collecting and analyzing the histology data. S. Pautler and

J. L. Chin recruited subjects and performed prostatectomies, generating specimens for analysis. A. D. Ward contributed to defining the research questions, interpreting the results and drafting the manuscript. All authors helped in reviewing and editing the manuscript. The work was performed under the direct supervision of A. Fenster and A. D. Ward, and the project oversight of G. S. Bauman.

Chapter 4: E. Gibson, C. Crukley, M. Gaed, J. A. Gómez, M. Moussa, J. L. Chin, G. S. Bauman, A. Fenster and A. D. Ward, "Registration of prostate histology images to *ex vivo* MR images via strand-shaped fiducials," *Journal of Magnetic Resonance Imaging* 36, 1402-1412 (2012).

My contributions to this work included defining the research questions; designing and implementing the methods being evaluated; identifying measurement landmarks; designing, implementing and analyzing the experiments; and drafting the manuscript. M. Gaed, J. A. Gómez, M. Moussa, and C. Crukley contributed to collecting and analyzing the histology data. J. L. Chin recruited subjects and performed prostatectomies, generating specimens for analysis. A. D. Ward contributed to defining the research questions, designing and analyzing the experiments; interpreting the results and drafting the manuscript. All authors helped in reviewing and editing the manuscript. The work was performed under the supervision of A. Fenster and A. D. Ward, and the project oversight of G. S. Bauman.

Chapter 5: E. Gibson, M. Gaed, J. A. Gómez, M. Moussa, C. Romagnoli, S. Pautler, J. L. Chin, C. Crukley, G. S. Bauman, A. Fenster and A. D. Ward, "3D prostate histology reconstruction: an evaluation of image-based and fiducial-based algorithms," *Medical Physics* 40, 093501 (2013).

My contributions to this work included defining the research questions; designing and implementing the methods being evaluated; identifying validation landmarks; designing, implementing and analyzing the experiments; and drafting the manuscript. M. Gaed, J. A. Gómez, M. Moussa, and C. Crukley contributed to collecting and analyzing the histology data. S. Pautler and J. L. Chin recruited subjects and performed prostatectomies, generating specimens for analysis. A. D. Ward contributed to defining the research questions, designing and analyzing the experiments; interpreting the results and drafting the manuscript. All authors helped in reviewing and editing the manuscript. The work was performed under the supervision of A. Fenster and A. D. Ward, and the project oversight of G. S. Bauman.

Chapter 6: E. Gibson, C. Romagnoli, M. Bastian-Jordan, D. W. Cool, Z. Kassam, M. Gaed, M. Moussa, J. A. Gómez, S. Pautler, J. L. Chin, C. Crukley, G. S. Bauman, A. Fenster, A. D. Ward, "Prostate cancer assessment and delineation on 3 T multi-parametric MRI: interobserver and intersequence agreement," *Radiology*, (in preparation).

My contributions to this work included definition of the research questions; coordinating the observer study; designing, implementing and interpreting the analysis of the data; and drafting the manuscript. M. Gaed, J. A. Gómez, M. Moussa, and C. Crukley contributed to collection and analysis of histology data. M. Gaed contoured and graded the histology data. J. A. Gómez and M. Moussa verified the contouring and grading of the histology data. C. Romagnoli, M. Bastian-Jordan, D. W. Cool, and Z. Kassam contoured lesions on mpMRI for analysis. S. Pautler and J. L. Chin recruited subjects and performed prostatectomies, generating specimens for analysis. A. D. Ward contributed to defining the research questions, designing and analyzing the experiments; interpreting the

results and drafting the manuscript. G. S. Bauman contributed to defining the research questions. All authors helped in manuscript review and editing. The work was performed under the supervision of A. Fenster and A. D. Ward, and the project oversight of G. S. Bauman.

Chapter 7: E. Gibson, M. Gaed, J. A. Gómez, M. Moussa, C. Romagnoli, M. Bastian-Jordan, Z. Kassam, D. W. Cool, S. Pautler, J. L. Chin, C. Crukley, G. S. Bauman, A. Fenster, A. D. Ward. " Toward prostate cancer contouring guidelines on MRI: dominant lesion gross and clinical target volume coverage via accurate histology fusion," American Society for Radiation Oncology Annual Meeting; San Francisco, USA; September 2014 (accepted); in preparation for submission to Radiology.

My contributions to this work included definition of the research questions; coordinating the observer study; designing, implementing and interpreting the analysis of the data; and drafting the manuscript. C. Romagnoli, M. Bastian-Jordan, D. W. Cool, and Z. Kassam contoured lesions on mpMRI for analysis. M. Gaed, J. A. Gómez, M. Moussa, and C. Crukley contributed to collecting and analyzing the histology data. M. Gaed contoured and graded the histology data. J. A. Gómez and M. Moussa verified the contouring and grading of the histology data. S. Pautler and J. L. Chin recruited subjects and performed prostatectomies, generating specimens for analysis. A. D. Ward contributed to defining the research questions, designing and analyzing the experiments; interpreting the results and drafting the manuscript. G. S. Bauman contributed to defining the research question. All authors helped in abstract review and editing. The work was performed under the supervision of A. Fenster and A. D. Ward, and the project oversight of G. S. Bauman.

## **Acknowledgments**

The work presented in this thesis would not have been possible without the mentorship, friendship and support of many people.

My deepest gratitude goes to Dr. Aaron Ward, who has worked tirelessly to support me and to challenge me to realize my potential. Aaron's insights into the ways that communication supports all aspects of our work have reshaped the way I see the world. I am also deeply grateful to Dr. Aaron Fenster for his mentorship on research, innovation, supervision, team management and the uncountable small things that keep a team working smoothly. They have created incredible opportunities for me, and taught me to create more for myself.

I also want to thank the clinical members of my advisory committee. Dr. Glenn Bauman has been an inspiration as a leader and as a collaborator. He has modeled the leadership necessary to corral a team of independent researchers and clinicians, with a soft touch and an eye for opportunity. Dr. José Gómez-Lemus pushed me to keep my perspective broad, offering intermingled insights into medicine, music, pathology, literature, and education. My advisory committee pushed me to be ambitious, but not too ambitious, and to keep myself pointed in the right direction.

I would like to extend my thanks to the many collaborators I have had the pleasure to work with over the course of my PhD. Additionally, I would like to thank the following specific collaborators with whom I have had the opportunity to work most closely. Cathie Crukley worked at my side from the very beginning, sharing her insight into the world of pathology, and stopping stupid ideas before they gained momentum. Dr. Mena Gaed's infinite patience for contouring histology has been a blessing and an

inspiration. Dr. Cesare Romagnoli, Dr. Derek Cool, Dr. Matthew Bastian-Jordan and Dr. Zahra Kassam shared with me their insights into the clinical decision-making process during the many hours spent contouring images. Dr. Madeleine Moussa and Dr. José Gómez-Lemus engaged with us to find creative solutions that satisfied the goals of our research and the constraints of pathology diagnosis. Dr. Joe Chin and Dr. Steve Pautler were instrumental in generating our rich data set and lent their insights into the urology community to increase the impact of our work. Dr. Irina Rachinsky, Dr. Jonathan Mandel and Dr. Suha Ghoul similarly lent their expertise into the imaging community to support our work. Catherine Hildebrand and Ashley Lozanski carried the administrative load that enabled this research and kept it running smoothly. I would also like to specifically thank the other members of the CIHR Team in Image Guidance for Prostate Cancer, who have supported and guided this work.

A multitude of colleagues have supported me in my research and in my life beyond research. I have had the pleasure of being part of a community that fosters innovative research and effective communication, and I will take away important lessons and fond memories. In particular, the students and staff of the Fenster lab, the Ward lab, the Peters lab and the Drangova lab have been a second family to me. I would also like to thank Dr. Terry Peters, Dr. Grace Parraga, Dr. Charlie McKenzie, Dr. Maria Drangova, Dr. David Holdsworth, Dr. Jim Lacefield, Dr. Eugene Wong, Dr. Tamie Poepping, Dr. Jerry Battista, and Dr. Jim Johnson for their encouragement, advice and mentorship about research, teaching and leadership throughout my PhD.



I want to thank Dr. Anne Martel, Dr. Abbas Samani, Dr. Terry Peters and Dr. David D'Souza for dedicating the time and effort to evaluate this thesis and for sharing this important milestone with me.

To the Robarts coffee time crew, the magnitude of your impact on my sanity was large, the direction is debatable.

Thank you to my parents and sisters who have supported me in everything I do, and have reached out to keep me engaged over the last four years.

To my beloved Janet Allison, words cannot express the impact you have had on my life.

Many sources of funding have supported this work, including the Canadian Institutes of Health Research, Cancer Care Ontario, the Ontario Institute for Cancer Research, the National Sciences and Engineering Research Council of Canada Postgraduate Scholarship and Computer Assisted Medical Interventions CREATE Award, the Queen Elizabeth II Graduate Scholarship in Science and Technology (with generous support from Ron Yamada Ontario Graduate Scholarships), the Ontario Graduate Scholarships, and the Western Graduate Research Scholarship.

# Table of Contents

<b>Abstract</b>	<b>ii</b>
<b>Co-Authorship Statement</b>	<b>vi</b>
<b>Acknowledgments</b>	<b>x</b>
<b>Table of Contents</b>	<b>xiii</b>
<b>List of Tables</b>	<b>xviii</b>
<b>List of Figures</b>	<b>xx</b>
<b>List of Abbreviations</b>	<b>xxv</b>
<b>Chapter 1. Introduction</b>	<b>1</b>
1.1 Background .....	2
1.1.1 Prostate cancer epidemiology .....	2
1.1.2 Prostate cancer diagnosis .....	3
1.1.3 Localized prostate cancer management .....	8
1.1.4 Prostate cancer imaging .....	13
1.1.5 Histological reference standards .....	21
1.2 Problem domain .....	23
1.2.1 Key gaps in knowledge .....	23
1.2.2 Research challenges .....	23
1.2.3 Previous work addressing these challenges .....	29
1.3 Thesis research questions and objectives .....	38
1.3.1 Research questions .....	39
1.3.2 Specific research objectives .....	40
1.4 Thesis Outline .....	41
1.4.1 Chapter 2: The impact of registration accuracy on imaging validation study design: a novel statistical power calculation .....	41
1.4.2 Chapter 3: 3D prostate histology image reconstruction: quantifying the impact of tissue deformation and histology section location .....	41
1.4.3 Chapter 4: Registration of prostate histology images to <i>ex vivo</i> MR images via strand-shaped fiducials .....	42
1.4.4 Chapter 5: 3D prostate histology reconstruction: an evaluation of image- based and fiducial-based algorithms .....	43
1.4.5 Chapter 6: Prostate cancer assessment and delineation on 3 T multi- parametric MRI: interobserver and intersequence agreement .....	43
1.4.6 Chapter 7: Toward prostate cancer contouring guidelines on MRI: dominant lesion gross and clinical target volume coverage via accurate histology fusion .....	45
1.4.7 Chapter 8: Contributions of the thesis, applications and suggestions for future work .....	46

1.5 References .....	46
<b>Chapter 2. The impact of registration accuracy on imaging validation study design: a novel statistical power calculation</b>	<b>62</b>
2.1 Introduction .....	62
2.2 Statistical power calculations in imaging validation studies.....	66
2.2.1 Specification of statistical analysis and data model assumptions .....	67
2.2.2 Derivation of the statistical power formulae incorporating registration error .....	69
2.3 Simulations.....	73
2.3.1 Model accuracy under the specified assumptions.....	74
2.3.2 Model sensitivity to violations of the assumptions.....	75
2.4 Results .....	79
2.4.1 Simulation sets following assumptions.....	79
2.4.2 Simulation sets violating assumptions .....	79
2.5 Discussion .....	84
2.5.1 Simulations .....	84
2.5.2 Fractional overlap relates registration error to study outcome .....	85
2.5.3 Estimating fractional overlap .....	86
2.5.4 Case study .....	87
2.5.5 Limitations .....	92
2.6 Conclusions .....	93
2.7 References .....	93
<b>Chapter 3. 3D prostate histology image reconstruction: quantifying the impact of tissue deformation and histology section location</b>	<b>96</b>
3.1 Introduction .....	96
3.2 Materials and methods .....	101
3.2.1 Materials and imaging.....	101
3.2.2 Methods.....	103
3.2.3 Identification of landmarks and tissue slice faces.....	105
3.2.4 Tissue deformation due to histology processing (Question 1).....	107
3.2.5 Spatial misalignment of tissue sections induced by microtome cutting (Question 2) .....	109
3.2.6 Impact of reconstruction model on 3D reconstruction error (Question 3). .....	111
3.2.7 Statistical analysis .....	113
3.3 Results .....	114
3.3.1 Tissue deformation due to histology processing (Question 1).....	114
3.3.2 Spatial misalignment of tissue sections induced by microtome cutting (Question 2) .....	116
3.3.3 Impact of reconstruction model on 3D reconstruction error (Question 3). .....	119
3.4 Discussion .....	121
3.4.1 Tissue deformation due to histology processing (Question 1).....	121
3.4.2 Spatial misalignment of tissue sections induced by microtome cutting (Question 2) .....	122
3.4.3 Impact of reconstruction model on 3D reconstruction error (Question 3). .....	124
3.4.4 Limitations of cutting measurement .....	127

3.5 References .....	129
<b>Chapter 4. Registration of prostate histology images to <i>ex vivo</i> MR images via strand-shaped fiducials</b>	<b>133</b>
4.1 Introduction .....	133
4.2 Materials and methods .....	135
4.2.1 Materials .....	135
4.2.2 <i>Ex vivo</i> MR imaging .....	135
4.2.3 Digital histology imaging .....	136
4.3 Methods.....	136
4.3.1 Fiducial marking .....	137
4.3.2 2D–3D histology to MR image registration algorithm.....	140
4.3.2.1 Definitions of algorithm terms.....	140
4.3.2.2 Algorithm step 1: identify an initial plane on MR image corresponding to each histology image .....	141
4.3.2.3 Algorithm step 2: compute an affine transformation mapping each histology image to its corresponding MR image plane .....	142
4.3.2.4 Algorithm step 3: refine the fiducial correspondence using local optimization and compute affine transformation .....	142
4.3.3 Validation.....	143
4.3.4 Experiments .....	144
4.3.5 Statistical analysis .....	147
4.4 Results .....	148
4.5 Discussion .....	154
4.6 References .....	160
<b>Chapter 5. 3D prostate histology reconstruction: an evaluation of image-based and fiducial-based algorithms</b>	<b>163</b>
5.1 Introduction .....	163
5.2 Materials and methods .....	164
5.2.1 Patient selection and imaging .....	164
5.2.2 Metrics for algorithm evaluation .....	169
5.2.3 Experimental design.....	173
5.2.4 Statistical analysis .....	177
5.3 Results .....	178
5.3.1 Algorithm performance with practical initialization.....	178
5.3.2 Sensitivity of algorithm accuracy to initialization variation.....	180
5.3.3 Sources of error for mutual information .....	182
5.4 Discussion .....	184
5.4.1 Evaluating histology reconstruction methods .....	185
5.4.2 Algorithm performance with practical initialization.....	191
5.4.3 Sensitivity of algorithm accuracy to initialization variation.....	191
5.4.4 Sources of error for mutual information .....	193
5.4.5 Limitations .....	194
5.4.6 Conclusion .....	195
5.5 References .....	196

<b>Chapter 6. Prostate cancer assessment and delineation on 3 T multi-parametric MRI: interobserver and intersequence agreement</b>	<b>200</b>
6.1 Introduction .....	200
6.2 Materials and methods .....	202
6.2.1 Participants.....	202
6.2.2 Imaging .....	203
6.2.3 Lesion scoring and delineation .....	204
6.2.4 Interobserver lesion correspondence.....	206
6.2.5 Interobserver score agreement .....	207
6.2.6 Interobserver contour agreement .....	207
6.2.7 Intersequence contour agreement .....	208
6.2.8 Interaction factors .....	209
6.2.9 Statistical analysis.....	212
6.3 Results .....	213
6.3.1 Image assessment.....	215
6.3.2 Interobserver and intersequence contour agreement.....	219
6.4 Discussion .....	223
6.5 References .....	228
 <b>Chapter 7. Toward prostate cancer contouring guidelines on MRI: dominant lesion gross and clinical target volume coverage via accurate histology fusion</b>	 <b>233</b>
7.1 Introduction .....	233
7.2 Materials and methods .....	236
7.2.1 Participants.....	236
7.2.2 Imaging .....	237
7.2.3 Lesion scoring and delineation .....	238
7.2.4 Histological contouring and grading.....	239
7.2.5 Histology-mpMRI registration.....	240
7.2.6 Histology fusion.....	241
7.2.7 CTV definition .....	242
7.2.8 Histological coverage.....	242
7.2.9 CTV volumetrics.....	244
7.2.10 Statistical analysis .....	245
7.3 Results .....	245
7.4 Discussion .....	260
7.5 References .....	266
 <b>Chapter 8. Contributions of the thesis, applications and suggestions for future work</b>	 <b>269</b>
8.1 Contributions of the thesis .....	269
8.1.1 Research question 1: How does registration error impact the statistical power of imaging validation studies? .....	270
8.1.2 Research question 2: What is the performance (in terms of reconstruction error and robustness) of an extrinsic fiducial-based histology reconstruction method, and how does this compare to alternative approaches based on image-guided slicing and intensity-based image registration?.....	271

8.1.3 Research question 3: What is the variability of lesion scoring (using consensus-panel-recommended PI-RADS guidelines [1]) and contouring on mpMRI? .....	276
8.1.4 Research question 4: What are appropriate margins around observers' target volumes to achieve histological coverage of detected cancers? ...	277
8.2 Applications and future directions .....	279
8.2.1 Evaluation and refinement of existing and novel imaging .....	279
8.2.2 Use in clinical training .....	280
8.2.3 Development and evaluation of consensus guidelines for contouring.....	282
8.3 Remaining gaps in histology-imaging fusion methods.....	283
8.4 References .....	283
<b>Appendix A. Derivation of the variance of <math>R</math></b>	<b>289</b>
<b>Appendix B. Large specimen pathology processing schedule</b>	<b>292</b>
<b>Appendix C. Supplementary material to Chapter 6</b>	<b>293</b>
C.1 Manual correspondence for two <i>distinct</i> lesions .....	293
C.2 PI-RADS sequence-specific scores broken down by grade and volume .....	293
<b>Appendix D. Supplementary material to Chapter 7</b>	<b>298</b>
<b>Appendix E. Permissions</b>	<b>307</b>
E.1 Permission to reproduce previously published material in Chapter 1 .....	307
E.2 Permission to reproduce previously published material in Chapter 2 .....	310
E.3 Permission to reproduce previously published material in Chapter 4 .....	316
E.4 Permission to reproduce previously published material in Chapter 5 .....	323
<b>Curriculum Vitae</b>	<b>324</b>

## List of Tables

<b>Table 2.1:</b> Values of power simulation parameters in simulations where these parameters were held constant. ....	75
<b>Table 2.2:</b> Varied data model parameters for simulation sets .....	75
<b>Table 2.3:</b> Sensitivity of the power prediction error to variation in model parameters, under the assumptions of the derivation. ....	80
<b>Table 2.4:</b> Sensitivity of the power prediction error to variation in model parameters, for simulation sets violating the assumptions of the derivation.....	80
<b>Table 2.5:</b> Parameter ranges where the model predicted the simulated power to within 0.01 absolute error and the simulated number of subjects and MDD to within 1% relative error.....	81
<b>Table 3.1:</b> <i>Post hoc</i> analyses comparing mean TRE under varying deformation models: 95% CI of TRE for model A – model B. ....	115
<b>Table 3.2:</b> Mean $\pm$ SD TRE (mm) for four models of deformation during histological processing stages.....	115
<b>Table 3.3:</b> Descriptive statistics for the orientation, minimum depth, mean depth and maximum depth of histology sections relative to the tissue blocks from which they were cut. Sensitivity to FLE was quantified as the mean of the standard deviation of measurements in a Monte Carlo simulation with perturbed landmark positions.....	118
<b>Table 3.4:</b> <i>Post hoc</i> analyses comparing TRE/FRE after intrinsic landmark reconstruction under varying deformation models: 95% CI of TRE/FRE for model A – model B.....	120
<b>Table 3.5:</b> <i>Post hoc</i> analyses comparing TRE/FRE after intrinsic landmark reconstruction under varying deformation models: 95% CI for of the difference between reconstruction with and without the front face assumption.....	120
<b>Table 3.6:</b> Mean $\pm$ SD TRE after intrinsic landmark reconstruction under varying constraints. ....	120
<b>Table 3.7:</b> Mean $\pm$ SD FRE after intrinsic landmark reconstruction under varying constraints. ....	120
<b>Table 4.1:</b> Mean TREs of the registration methods with and without local optimization. ....	153
<b>Table 4.2:</b> Statistical analysis of accuracy of methods.....	153
<b>Table 5.1:</b> Comparison of the algorithms after practical initialization .....	180
<b>Table 5.2:</b> 95% CI on mean TRE difference for similarity metrics and search spaces across initialization algorithms. ....	180
<b>Table 5.3:</b> Sensitivity of mean TRE to initialization error quantified as 95% CI on the sensitivity coefficients of general linear models.....	181

<b>Table 6.1:</b> Subject demographics and clinical characteristics.....	203
<b>Table 6.2:</b> MRI sequence parameters.....	204
<b>Table 6.3:</b> Positive predictive values (PPVs) for any histological cancer and for Gleason 7 cancer of score assignments with PI-RADS likelihood scores of 3, 4 and 5, individually, and scores above thresholds $\geq 3$ and $\geq 4$ , broken down by lesion volume on mpMRI.....	215
<b>Table 6.4:</b> Interobserver agreement in assigning PI-RADS likelihood scores for two thresholds.....	219
<b>Table 6.5:</b> Interobserver and intersequence contour agreement, and comparison of interobserver contour agreement across sequences.....	221
<b>Table 6.6:</b> Spearman correlation coefficients* of interobserver and intersequence contour agreement with interaction factors.....	222
<b>Table 7.1:</b> Subject demographics and clinical characteristics.....	237
<b>Table 7.2:</b> mpMRI sequence details.....	238
<b>Table 7.3:</b> For each volume range and Gleason score: number of aggregate foci (subset with at least on true positive GTV). .....	247
<b>Table 7.4:</b> Per-tumor statistical analysis of the minimum margins required to achieve specified residual areas. ....	249
<b>Table 7.5:</b> Volumes of CTVs with PI margins.....	253
<b>Table B.1:</b> Our hospital's standard clinical pathology laboratory protocol for large specimens.....	292



# List of Figures

<b>Figure 1.1:</b> Illustrated Gleason patterns from the updated Gleason grading system. ....	7
<b>Figure 1.2:</b> Multiple modalities of prostate cancer imaging showing approximately the same anatomical location. A large Gleason score 4+3 cancer with tertiary grade 5 is present in the anterior prostate, alongside a small region of BPH, a common feature that mimics cancer on many modalities.....	15
<b>Figure 1.3:</b> Illustrative diagram showing positions of cuts for gross slicing a specimen where constrained slicing may be disruptive to the clinical pathology workflow. ....	33
<b>Figure 2.1:</b> Overview of the data model. ....	67
<b>Figure 2.2:</b> Simulation sets testing model accuracy under the original data model, for parameters that yielded >1% relative error in the number of subjects .....	82
<b>Figure 2.3:</b> Simulation sets testing model accuracy under the generalized data model, for parameters that yielded >1% relative error in the predicted number of subjects.....	83
<b>Figure 2.4:</b> Relationship between estimated required number of subjects and target registration error (TRE) for 4 imaging modalities .....	91
<b>Figure 3.1:</b> Overview of specimen processing, imaging and analysis. ....	97
<b>Figure 3.2:</b> Schematic representations of tissue, landmarks and measurements .....	105
<b>Figure 3.3:</b> Illustrative examples of the T1-weighted tissue slice MR (left), paraffin (middle) and histology (right) images transformed by a best-fit affine transformation aligning manually identified landmarks.....	106
<b>Figure 3.4:</b> Box plot showing the target registration errors of homologous landmarks under four deformation models for the tissue deformation due to histological processing and cutting. These results correspond to the descriptive statistics shown in the first row of Table 3.1.....	115
<b>Figure 3.5:</b> Histograms of histology section depths and orientations. The subset of tissue slices illustrated in Figure 3.7 is shown in dark gray.....	117
<b>Figure 3.6:</b> Correlation of minimum, mean and maximum histology section depths with orientations. Tissue slices corresponding to sections marked with circles are shown in Figure 3.7.....	117
<b>Figure 3.7:</b> Renderings of the spatial relationships between tissue slices, histology-visible landmarks, front face plane and histology section planes for 10 tissue slices, ordered by increasing mean depth from left to right.....	118
<b>Figure 3.8:</b> 3D reconstruction of three histology sections aligned (with a mean TRE of 0.5 mm) to an anterior view of 3D surface rendering of the corresponding intact <i>ex vivo</i> prostate gland with seminal vesicles, illustrating the potential for non-parallel, non-evenly-spaced histological tissue sections.....	124

<b>Figure 3.9:</b> Sample sizes, relative to an arbitrarily chosen baseline, for imaging validation studies of image signal differences between cancerous and background tissue for 0.2 cm <sup>3</sup> cancer foci, under assumptions that foci are spherical and reconstruction error can be modeled as a translation error distributed as a 3D Gaussian and is combined in quadrature with a 2.1 mm TRE due to registration to <i>in vivo</i> imaging. Reconstructions under differing deformation assumptions, and with or without the front face assumption are indicated .....	127
<b>Figure 4.1:</b> Specimen processing overview. ....	137
<b>Figure 4.2:</b> Fiducial configuration showing prostate surface (a and b) and a cross-section along the dotted line (c). These fiducial markers are visible on histology images (d). Insets (e and f) show enlarged views of $h_{10}$ and $h_1$ respectively.....	139
<b>Figure 4.3:</b> Co-registered T1-weighted MR images of the whole specimen (a) and a tissue block (b). Barely visible (e.g. large arrows) or invisible (e.g. small arrows) ducts on the whole specimen image are visible on the tissue block image due to Christo-Lube contrast. Labeled arrows show the appearance of surface-mounted (c) and internal (d) fiducials, respectively, on the T1W MR image.....	140
<b>Figure 4.4:</b> Homologous landmark pairs (large white arrows), used for evaluation of the registration algorithm, shown on histology images (top) and on the corresponding oblique slices on T2 (middle) and T1 (bottom) MR images. Small black arrows denote salient nearby features used to corroborate the established landmark correspondence. Salient features in these images include corpora amylaceae (labeled *), the urethra (labeled +), extrinsic fiducials (labeled °) and atrophic ducts (unlabeled). ....	149
<b>Figure 4.5:</b> Distribution of landmarks within the prostate specimens. Each landmark is shown with its AP, LR and IS position within the specimen normalized such that the extents of the resected prostate gland are from 0 to 1 for each specimen. Each one is marked as to the anatomical zone where it is located: peripheral zone (PZ), transition zone (TZ), central zone (CZ) and anterior fibromuscular stroma (AFS). The 24 landmarks with indefinite classification were omitted. ....	150
<b>Figure 4.6:</b> Three histology images registered to <i>ex vivo</i> MR images using the proposed registration method. Example validation landmarks are shown by arrows.....	151
<b>Figure 4.7:</b> Box plot of TREs for five registrations with 5–95% whiskers. ....	152
<b>Figure 4.8:</b> Bland-Altman plots comparing TREs with local optimization after initializations using the proposed method ( $TRE_p$ ) and (a) the tissue block imaging ( $TRE_{TBI}$ ) and (b) the image-guided slicing with local refinement ( $TRE_{IGSLO}$ ).....	154
<b>Figure 5.1:</b> Schematic of strand-shaped fiducial markers (1–10), with parameterized 3D fiducial space curve $m_4$ indicated with arrows, and histology showing 2D fiducial cross-sections (boxed) with $h_4$ and $h_{10}$ inset. ....	168
<b>Figure 5.2:</b> Constraining the search space in an optimization reduces the number of possible solutions to be considered, but may introduce local minima.....	169
<b>Figure 5.3:</b> Homologous landmark pairs (large white arrows), used for evaluation of the registration algorithms, shown on histology images (top) and on the	

corresponding oblique image slices on T2 (middle) and T1 (bottom) MR images. Small black arrows denote salient nearby features used to corroborate the established landmark correspondence, but not used as landmarks themselves. Salient features in these images include corpora amylacea (labeled \*), the urethra (labeled +), extrinsic fiducials (labeled °) and atrophic ducts (unlabeled). ..... 171

**Figure 5.4:** Illustrative registrations from (a) the practical initialization experiment and (b) the initialization offset by 5 mm translational error from the sensitivity experiment..... 179

**Figure 5.5:** Sensitivity of mean TRE to four types of initialization error: rotation about the (a) AP and (b) LR axes, (c) translation, and (d) scaling. .... 181

**Figure 5.6:** (a) Violin plot showing the distribution of TRE for the  $\{T^A; M_{MI}\}$  algorithm for translational initialization error; (b) scatter plot of TRE vs. the extent of tissue in the inferior direction normal to the plane  $A(\Omega)$  of the best-fit transformation. 182

**Figure 5.7:** Violin plots showing the distribution of  $D_{MI}$  for four types of initialization error: rotation about the AP (a) and LR (b) axes, translation (c), and scaling (d). ..... 184

**Figure 6.1:** Observers' process for generating the PI-RADS scores and *suspicious regions* analyzed in this study..... 206

**Figure 6.2:** Illustrative (a) T2W, (b) ADC and (c) DCE mpMRI images for 3 subjects, showing *suspicious regions* from 4 observers (yellow, magenta, cyan and red contours). The Gleason scores (GS) and mean overall likelihood scores (OL) for each lesion are shown on the left side, and sequence-specific scores are shown with each image. Panel (d) shows post-prostatectomy histology corresponding to the *distinct* lesion in the top row with insets showing the fine-scale contouring of Gleason 3+3 (yellow), 3+4 (cyan), 4+3 (green) and 4+4 (black) regions. .... 211

**Figure 6.3:** PI-RADS likelihood scores for lesions broken down by histological grade. Each vertical solid line corresponds to a *distinct* lesion, with horizontal dashes at the PI-RADS likelihood score for each observer. For each grade, Tukey box plots show the interquartile range (IQR), median (dotted line), mean (dashed line), and range of data within  $1.5 \times \text{IQR}$  of the IQR (whiskers). The dotted box denotes “false positives” where the PI-RADS likelihood scores denote likely or highly likely cancer for *non-cancerous lesions*. The dashed box denotes “higher grade false negatives” where the observer did not record a PI-RADS score for a *Gleason 7 lesion*..... 217

**Figure 6.4:** PI-RADS likelihood scores for lesions broken down by mean *suspicious region* volume. Each vertical solid line corresponds to a *distinct* lesion, with horizontal dashes at the PI-RADS likelihood score for each observer. For each volume range, Tukey box plots show the interquartile range (IQR), median (dotted line), mean (dashed line), and range of data within  $1.5 \times \text{IQR}$  of the IQR (whiskers). Note that by *a priori* agreement, a default likelihood score of 2 was assigned for all observers that did not make a *score assignment* for a *distinct* lesion recorded by at least one observer..... 218

**Figure 7.1:** Illustrative example of a contoured and graded H&E-stained whole-mount histology image (A) with magnified insets (B) and (C), showing homogenous

Gleason 3+3 (brown), intermingled Gleason 3+4 (dark green), intermingled Gleason 4+3 (purple), and prostatic intraepithelial neoplasia (light green) contours. ....	240
<b>Figure 7.2:</b> Illustrative set of true positive GTVs for a Gleason 3+4 sparse cancer. Column A shows 4 sections of midgland histology with regions of Gleason 7 cancer (green) and Gleason 6 cancer (cyan) highlighted. Columns B–D show corresponding oblique sections through the T2W, DCE and ADC images respectively, with overlaid boundaries for the corresponding GTVs (magenta), and CTVs with PI margins for 0% (solid line), 5% (dashed) and 10% (dotted) residual area for any cancer (green) and for high-grade cancer (yellow). ....	246
<b>Figure 7.3:</b> Upper bound on residual cancer (left) and high-grade cancer (right) prediction intervals for 4 observers (rows). ....	251
<b>Figure 7.4:</b> PI margins for coverage of low- and high-grade cancer for 4 observers (denoted by symbols) for 0%, 5% and 10% residual area for each CTV type. ....	252
<b>Figure 7.5:</b> PI margins for coverage of high-grade cancer for 4 observers (denoted by symbols) for 0%, 5% and 10% residual area for each CTV type. ....	252
<b>Figure 7.6:</b> Volume encompassed by CTVs with PI margins for <b>any cancer</b> for 4 observers (rows) for 0%, 5% and 10% residual cancer area (columns). ....	256
<b>Figure 7.7:</b> Volume encompassed by CTVs with PI margins for <b>high-grade cancer</b> for 4 observers (rows) for 0%, 5% and 10% residual high-grade cancer area (columns). ....	257
<b>Figure 7.8:</b> Scaling factors of the CTVs with PI margins <b>relative to their corresponding GTVs</b> for 4 observers (rows) for 0%, 5% and 10% residual cancer area (columns). Y-axis shows the cube root of the volume ratio (equivalent scaling factor in each dimension). ....	258
<b>Figure 7.9:</b> Scaling factors of the CTVs <b>relative to their corresponding aggregated foci</b> for 4 observers (rows) for 0%, 5% and 10% residual area (columns). Y-axis shows the square root of the planimetric volume ratio (equivalent scaling factor in each dimension). ....	259
<b>Figure C.1:</b> PI-RADS T2W scores broken down by histological grade. ....	294
<b>Figure C.2:</b> PI-RADS DCE scores broken down by histological grade. ....	295
<b>Figure C.3:</b> PI-RADS ADC scores broken down by histological grade. ....	295
<b>Figure C.4:</b> PI-RADS T2W scores broken down by mean <i>suspicious region</i> volume..	296
<b>Figure C.5:</b> PI-RADS DCE scores broken down by mean <i>suspicious region</i> volume..	296
<b>Figure C.6:</b> PI-RADS ADC scores broken down by mean <i>suspicious region</i> volume.	297
<b>Figure D.1:</b> Residual cancer (left) and high-grade cancer (right) for CTVs expanded by various margins from individual GTVs delineated on T2W images for 4 observers (rows). ....	298

<b>Figure D.2:</b> Residual cancer (left) and high-grade cancer (right) for CTVs expanded by various margins from individual GTVs delineated on DCE images for 4 observers (rows). .....	299
<b>Figure D.3:</b> Residual cancer (left) and high-grade cancer (right) for CTVs expanded by various margins from individual GTVs delineated on ADC images for 4 observers (rows). .....	300
<b>Figure D.4:</b> Residual cancer (left) and high-grade cancer (right) for CTVs expanded by various margins from individual composite T2W+DCE GTVs for 4 observers (rows). .....	301
<b>Figure D.5:</b> Residual cancer (left) and high-grade cancer (right) for CTVs expanded by various margins from individual composite T2W+ADC GTVs for 4 observers (rows). .....	302
<b>Figure D.6:</b> Residual cancer (left) and high-grade cancer (right) for CTVs expanded by various margins from individual composite DCE+ADC GTVs for 4 observers (rows). .....	303
<b>Figure D.7:</b> Residual cancer (left) and high-grade cancer (right) for CTVs expanded by various margins from individual composite T2W+DCE+ADC GTVs for 4 observers (rows). .....	304
<b>Figure D.8:</b> Scaling factors of the CTVs with PI margins for high-grade cancer <b>relative to their corresponding GTVs</b> for 4 observers (rows) for 0%, 5% and 10% residual high-grade cancer area (columns). Y-axis shows the cube root of the volume ratio (the equivalent scaling factor in each dimension). .....	305
<b>Figure D.9:</b> Scaling factors of the CTVs <b>relative to their corresponding aggregated high-grade foci</b> for 4 observers (rows) for 0%, 5% and 10% residual area (columns). Y-axis shows the square root of the planimetric volume ratio (the equivalent scaling factor in each dimension). .....	306

## List of Abbreviations

<b>2D</b>	Two-dimensional
<b>3D</b>	Three-dimensional
<b>ADC</b>	Apparent diffusion coefficient
<b>AFS</b>	Anterior fibromuscular stroma
<b>ANOVA</b>	Analysis of variance
<b>AP</b>	Anterior-posterior
<b>BPH</b>	Benign prostatic hyperplasia
<b>CI</b>	Confidence interval
<b>CT</b>	Computed tomography
<b>CTV</b>	Clinical target volume
<b>CZ</b>	Central zone
<b>DCE</b>	Dynamic contrast-enhanced
<b>DNA</b>	Deoxyribonucleic acid
<b>DRE</b>	Digital rectal examination
<b>DW</b>	Diffusion-weighted
<b>FLE</b>	Fiducial localization error
<b>FOV</b>	Field of view
<b>FRE</b>	Fiducial registration error
<b>FSE</b>	Fast spin echo
<b>GS</b>	Gleason score
<b>GTV</b>	Gross target volume
<b>IQR</b>	Interquartile range
<b>IS</b>	Inferior-superior
<b>k<sup>trans</sup></b>	Pharmacokinetic contrast transfer coefficient
<b>k<sup>ep</sup></b>	Pharmacokinetic rate constant
<b>LLE</b>	Landmark localization error
<b>LO</b>	Local optimization
<b>LR</b>	Left-right
<b>LRE</b>	Landmark registration error

<b>MAD</b>	Mean absolute distance
<b>MDD</b>	Minimum detectable difference
<b>MI</b>	Mutual information
<b>mpMR</b>	Multi-parametric magnetic resonance
<b>mpMRI</b>	Multi-parametric magnetic resonance imaging
<b>MR</b>	Magnetic resonance
<b>MRE</b>	Magnetic resonance elastography
<b>MRI</b>	Magnetic resonance imaging
<b>MRS</b>	Magnetic resonance spectroscopy
<b>NPV</b>	Negative predictive value
<b>OL</b>	Overall likelihood
<b>PACS</b>	Picture archiving and communication system
<b>PCA</b>	Principal component analysis
<b>PET</b>	Positron emission tomography
<b>PI</b>	Prediction interval
<b>PI-RADS</b>	Prostate imaging reporting and data system
<b>PPV</b>	Positive predictive value
<b>PSA</b>	Prostate-specific antigen
<b>PSMA</b>	Prostate-specific membrane antigen
<b>PWM</b>	Pseudo-whole-mount
<b>PZ</b>	Peripheral zone
<b>RMS</b>	Root-mean-square
<b>ROI</b>	Regions of interest
<b>SD</b>	Standard deviation
<b>SPGR</b>	Spoiled gradient echo
<b>T1</b>	Longitudinal relaxation constant, or a cancer stage
<b>T2</b>	Transverse relaxation constant, or a cancer stage
<b>T2W</b>	T2-weighted
<b>TE</b>	Echo time
<b>TLE</b>	Target localization error
<b>TNM</b>	Tumor Node Metastasis Classification of Malignant Tumours

<b>TPS</b>	Thin-plate spline
<b>TR</b>	Repetition time
<b>TRE</b>	Target registration error
<b>TRUS</b>	Transrectal ultrasound
<b>TZ</b>	Transition zone
<b>USD</b>	United States dollar
<b><math>v_e</math></b>	Pharmacokinetic contrast leakage



# Chapter 1.

## Introduction

Imaging of localized prostate cancer, and in particular multi-parametric magnetic resonance imaging (mpMRI), has the potential to support multiple stages of the clinical workflow, including detection, staging and localization of tumors, as well as selection, delivery and monitoring of treatments. Delineation of prostate cancer tumors on imaging has the potential for further improving the clinical workflow, potentially enabling precise monitoring of tumor volumes in active-surveillance patients, optimal targeting of biopsies, and targeted delivery of treatments with lower morbidity or improved outcomes for suitable patients.

Spatially accurate comparisons of prostate images and delineations to an accepted reference standard for the location and clinical significance of prostate cancers, could support evaluation and improvement of prostate cancer imaging and lesion delineation in the clinical workflow. Currently, the most readily accepted reference standard for the location and aggressiveness of prostate cancer is histological examination of prostatectomy specimens. Some applications of such a histological reference standard to investigate prostate cancer imaging and delineation require accurate registration to the images being investigated. However, there are key gaps in knowledge regarding how to accurately register histological reference standards to imaging, and how accurate the registration needs to be. Consequently, there is also a gap in knowledge regarding the suitability of imaging for tasks, such as lesion delineation, that require such reference

standards for evaluation. These gaps in knowledge will be described in further detail in Section 1.2.

The work described in this thesis focuses on (1) developing tools for accurately registering and evaluating a histological reference standard for prostate cancer imaging and lesion delineation, and (2) evaluating the accuracy and variability of delineating prostate cancer lesions on mpMRI by comparison to this accurately registered reference standard. This evaluation could potentially inform clinical guidelines on the suitability and use of mpMRI-derived target volumes for applications in localized prostate cancer diagnosis, treatment and monitoring.

## 1.1 Background

### 1.1.1 Prostate cancer epidemiology

Prostate cancer is the most commonly diagnosed non-skin cancer in Canadian men, with an estimated 23,600 new diagnoses in 2014 [1]. One in eight men will be diagnosed with prostate cancer during his lifetime [1]. It is the third most common cause of male cancer death, with an estimated 4,000 deaths in Canada in 2013. Additionally, prostate cancer has a substantial impact on patients' ongoing quality of life, because the most common prostate cancer treatments have high risks of life-changing side effects and living with untreated cancer has a high psychological burden.

The natural history of prostate cancer is highly variable: some prostate tumors become metastatic and potentially lethal, but many do not. In fact, autopsy studies of 80–90-year-old men deceased from unrelated causes have found that 70–90% had histological evidence of prostate tumors [2]. As screening tests have become more sensitive and more widespread, more of these latent non-lethal cancers have been

diagnosed and treated [3], which impacts patients' quality of life. Accordingly, prostate cancer is commonly described as over-diagnosed and over-treated [3] (for low-risk patients), while, at the same time, resulting in many deaths (of high risk patients).

Thus, key challenges in prostate cancer lie in identifying patients who do not need treatment, supporting therapy delivery for those that need aggressive treatment to extend their lives in spite of potential side effects, and developing treatment approaches for borderline patients that will survive with less aggressive treatments and will benefit from lower morbidity.

Delineation of prostate cancer may have a role in addressing each of these challenges: supporting accurate stratification of patients through optimized targeted biopsy procedures, monitoring low-risk patients through longitudinal measurements of tumor burden, and targeting curative-intent therapies for both aggressive and less aggressive treatments.

### 1.1.2 Prostate cancer diagnosis

Detection and diagnosis of prostate cancer is performed in two stages: initial detection and biopsy confirmation.

#### 1.1.2.1 Initial detection

Two techniques are currently widely used to initially detect prostate cancer in symptomatic men and as a screening tool in asymptomatic men: prostate-specific-antigen (PSA) testing and digital rectal examination (DRE).

PSA testing is a procedure where a blood sample is taken, and the level of PSA, a protein produced primarily by prostate cells, is quantified. A PSA level above 4.0 ng/mL

has been considered suspicious for cancer, and positive PSA tests result in more cancer diagnoses than other tests [4]. However, other conditions such as prostatitis and benign prostatic hyperplasia (BPH) can also increase PSA levels [5]. Furthermore, in the Prostate Cancer Prevention Trial, 15% of men with consistently lower PSA levels had a positive biopsy and 2% had a positive biopsy showing high-grade cancer [6]. The increase in detection of latent cancer has, in part, been attributed to the sensitivity of the PSA test, and its application in screening populations [1].

DRE is a procedure wherein a physician palpates the patient's prostate through the rectal wall with a gloved finger to identify tissue that has abnormally high density or asymmetry, properties which raise suspicion for prostate cancer [4]. DRE alone has a reported sensitivity of 37% and a positive predictive value (PPV) of 27%; however, DRE is typically used in conjunction with PSA testing where it yields an incremental sensitivity of 18%. In patients who would not be identified with PSA testing (i.e. PSA < 4.0 ng/mL), the PPV was 13% [7].

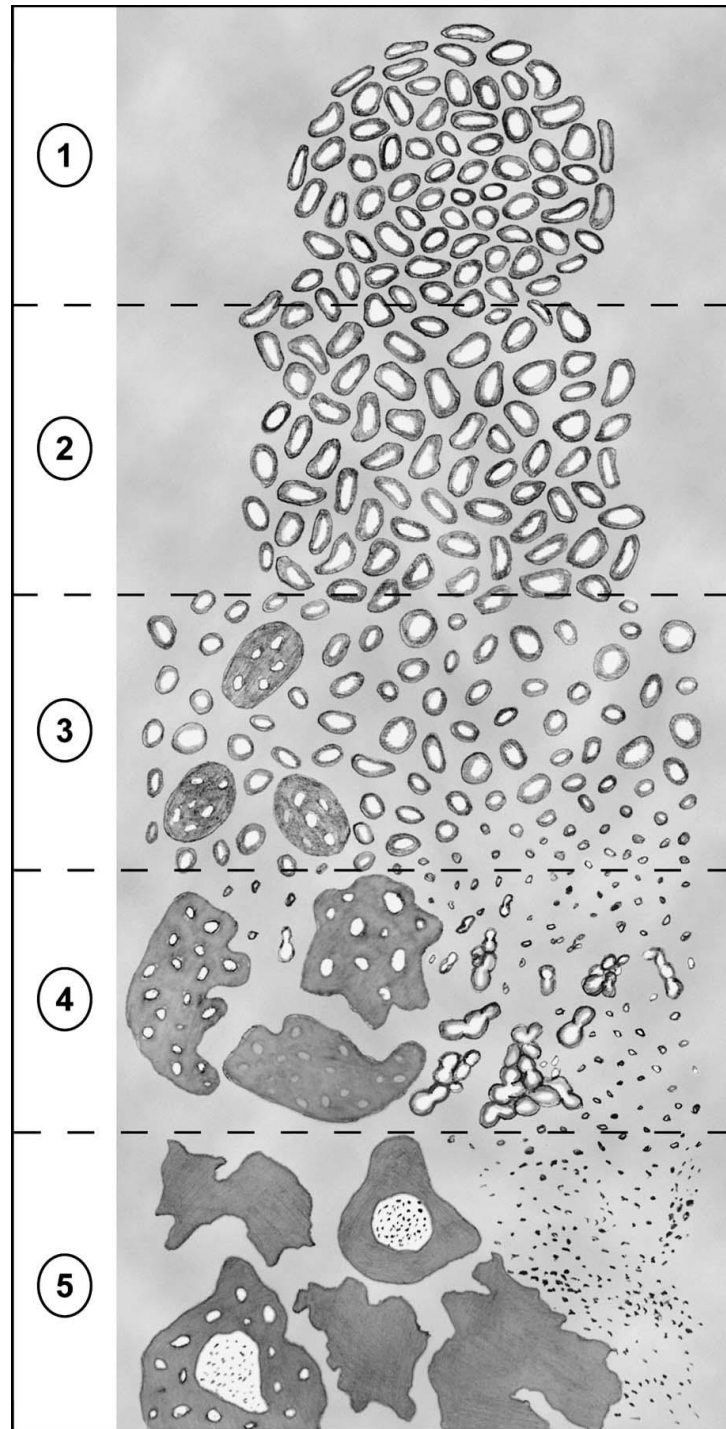
In order to detect cancers at an earlier, more treatable stage, DRE and PSA testing have been widely used to screen asymptomatic men; however, the practice remains controversial, due to uncertainty about the benefits and risks of screening [1]. The Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial found that men who underwent DRE and PSA screening had a higher incidence of prostate cancer, but the same mortality, as a control group [8]. In contrast, the European Randomized Study of Screening for Prostate Cancer trial found that the men randomized to the screening group had a higher incidence, but lower mortality, than the control group [9]. As a result of this evidence, the United States Preventive Services Task Force has recommended against

screening of asymptomatic men [10]. Correctly stratifying patients identified through high sensitivity screening could improve mortality for high-risk patients, while avoiding the overtreatment of low-risk patients.

#### 1.1.2.2 Biopsy

Patients with findings that are suspicious for prostate cancers will be referred for prostate biopsy to confirm the suspicion. Prostate biopsy is a procedure where tissue samples are taken from the prostate using a biopsy needle, through the rectal wall or the perineum, under the guidance of transrectal ultrasound. Typically 6–12 biopsy samples are taken using a systematic protocol that primarily samples the peripheral zone, where the majority of cancers are located [11, 12]. Biopsy tissue is examined microscopically to determine the number of cores containing cancer, and the extent of cancer in each core. This information is used for staging according to the "TNM Classification of Malignant Tumours" system (commonly abbreviated to the TNM system, where TNM refers to tumors nodes and metastases). For example, T1c denotes that the tumor was found in a needle biopsy performed as a result of a PSA test and T2a denotes that the tumor is in less than half of one of the prostate lobes for a palpable tumor. Cancer that is identified on biopsy is assessed according to the Gleason grading system by grading the microscopic architecture and differentiation of prostate glands on a 5 point scale [13] (illustrated in Figure 1.1), and reporting a Gleason score as the sum of the most prevalent grades. The Gleason score, clinical TNM stage and PSA levels can be used in pathologic and prognostic nomograms [14-16] to guide therapy selection. However, grading biopsy tissue remains challenging, with a 40% rate of discrepancy between biopsy and

prostatectomy Gleason grade [17, 18], due, in part, to challenges in acquiring representative samples of prostate tumors.



*Brumbaugh*

**Figure 1.1:** Illustrated Gleason patterns from the updated Gleason grading system. Reproduced with permission from Epstein et al. The 2005 ISUP Consensus Conference on Gleason Grading of Prostatic Carcinoma. The American Journal of Surgical Pathology, 29(9) 1228–1242.

### 1.1.2.3 A potential role for imaging and lesion delineation

Prostate cancer screening tests are sensitive, identifying many patients with low-risk disease. Challenges remain in correctly stratifying such patients based on systematic biopsy. However, biopsies that are targeted at suspicious regions delineated on imaging may have a higher probability of sampling high-grade foci [19]. As the likelihood of sampling tissue from a lesion depends on its size and shape [20], accurate and precise delineation of intraprostatic cancer foci on imaging, and quantification of the uncertainty in delineation, could support the development of optimized planning of targeted biopsy procedures.

### 1.1.3 Localized prostate cancer management

Organ-confined prostate cancer is primarily managed with (1) expectant management without therapy, with low direct morbidity but with the psychological burden of living with untreated cancer or (2) radical therapies (radiation or surgery) that target the whole prostate with curative intent, with high rates of life-changing side effects. Emerging treatments that target the prostate cancer preferentially may offer an intermediate curative intent treatment with reduced side effects for suitable men. This section describes these treatments, and the roles that imaging and lesion delineation might play.

#### 1.1.3.1 Expectant management

Expectant management refers to multiple approaches where the patient is not given treatment: watchful waiting, where treatment is delivered if the patient becomes symptomatic; and active surveillance, where the patient is closely monitored for signs of



progression with frequent biopsy and blood tests. If progression is detected, these patients can proceed to curative-intent treatments. In patients who do not progress, there are no treatments, and, therefore, no direct side effects. However, there are indirect effects that negatively affect patients' quality of life, including anxiety and depression [21] and increased long-term erectile dysfunction rates [22]. Patients' decisions to undergo curative-intent treatment even without progression (7–13% of patients) have been attributed, in part, to these side effects [23, 24]. Expectant management is suitable for patients with low-risk cancer, defined by a Gleason score  $\leq 6$ , PSA  $< 10$  ng/ml and clinical stage T1c–T2a (i.e. the tumor occupies less than half of one lobe), constituting 50% of newly diagnosed prostate cancer [25]. In one cohort of 299 active-surveillance patients, the prostate-cancer-specific survival rate was 99.3% after 8 years of follow-up and only 35% of active-surveillance patients progressed to delayed curative-intent therapy [25]. Although differences in recurrence rates due to delayed treatment of patients who do progress have not yet been demonstrated [26, 27], delayed treatment does result in higher pathological grade at time of treatment [27]. Thus, there may be value in identifying the patients who do progress earlier. Prostate cancer imaging may have a role in ongoing monitoring of patients on active surveillance [28]. Additionally, there may be a role for prostate cancer delineation in longitudinal measurements of tumor volume [29, 30], a factor that has been found to be correlated with Gleason grade, stage, tumor progression and survival and that may have independent prognostic importance [31].

### 1.1.3.2 Prostate-focused therapies

The most common treatments for localized prostate cancer, radical prostatectomy and radical radiotherapy, treat the whole prostate to the same degree with curative intent, and with little or no adjustment for the location of the cancer within the gland.

#### 1.1.3.2.1 Radical prostatectomy

Radical prostatectomy is a procedure wherein the prostate is removed through open or minimally invasive surgery. Radical prostatectomy has the potential for cancer cure in patients with cancer that does not extend beyond the surgical margin, and is an appropriate treatment option for men with low–intermediate-risk prostate cancers, including those with organ-confined tumors containing Gleason pattern 4 or 5 [4]. However, prostatectomy patients have high rates of erectile dysfunction (34–90% [32, 33]), urinary incontinence (8–13% [32, 34]) and other complications (10–30% perioperatively and 4–10% within a year [32, 35]). Nerve-sparing radical prostatectomy, in which one or both sets neurovascular bundles lateral to the prostate are not resected, may result in lower rates of complications [34, 36]. Prostate cancer imaging and accurate delineation of tumor boundaries may support selection of nerve-sparing surgeries in patients with unilateral or bilateral cancer that is at an acceptable distance from the surgical margin.

#### 1.1.3.2.2 Radical radiation therapy

Radiation therapy refers to several procedures wherein the prostate is irradiated to damage DNA (preferentially in highly proliferative cancer cells). Examples include external beam radiation, where high energy x-rays generated outside the body are focused on the prostate from multiple angles, low-dose-rate brachytherapy where radioactive

seeds are permanently inserted into the prostate, and high-dose-rate brachytherapy where a highly radioactive emitter is moved within the prostate during the procedure and then removed. Radiation therapies may be suitable for low–intermediate-risk patients including those with PSA levels >10ng/ml and with organ-confined cancer with Gleason scores >7, particularly with dose escalation or adjuvant hormonal therapy [4]. As with radical prostatectomy, radiation therapy patients have high rates of morbidity: erectile dysfunction (8–50% [37, 38]), urinary complications (7–23% [38, 39]) and fecal incontinence (2–58% [40]).

#### 1.1.3.3 Lesion-focused therapies

Emerging treatments that preferentially target the prostate cancer lesions may allow for escalating treatment in tumors in patients to achieve better outcomes, reducing treatment in healthy tissue to achieve lower morbidity, or both.

##### 1.1.3.3.1 Whole-gland radiation with focal boost

After radical radiotherapy, local recurrence can occur within the prostate, which typically occurs at the site of the dominant lesion [41]. Advances in technologies for planning and delivering radiation have enabled radiation oncologists to deliver different doses to different regions within the prostate. Intensity-modulated-radiation-therapy-based and brachytherapy-based delivery of whole-gland radiation, with additional dose delivered to intraprostatic gross tumor volumes, have been proposed as strategies to improve outcomes [42, 43]. An alternative approach, where a reduced dose is delivered to the whole gland and an additional dose is delivered to intraprostatic lesions, with the aim to maintain control rate while reducing toxicity, is also being considered [44]. Accurate and precise delineation of lesions on imaging may support the conformal delivery of

higher dose to the dominant lesion, potentially reducing local recurrence and/or improving morbidity, while minimizing dose to surrounding organs at risk.

#### 1.1.3.3.2 Prostate-sparing and lesion-only therapies

Prostate cancer is multifocal in 67–82% of patients [45-47]; however, there is a subset of patients whose cancer is aggressive enough to require treatment but consists of a dominant lesion surrounded by primarily non-cancerous tissue. Furthermore, there is evidence that the risk of recurrence and biochemical failure, even in multifocal prostate cancers, depends primarily on the size and grade of only the largest focus [47]. Therefore, some emerging therapies aim to preserve healthy prostate tissue by leaving portions, or even the majority of the gland, entirely untreated. High-intensity focused ultrasound [48], cryotherapy [49], photodynamic therapy [50], laser ablation [51] and radiation therapy [52] are all being evaluated for focal therapy applications in the prostate. Control rates for focal therapies have not yet been well-established; however, early evidence from studies of several focal therapies suggests that they successfully reduce the side effects prevalent in radical therapies. For example, studies of cohorts undergoing cryotherapy have reported incontinence rates lower than 5% and erectile dysfunction rates of 10–35% [49]. Studies of cohorts undergoing high-intensity focused ultrasound have reported incontinence rates less than 10% and erectile dysfunction rates of 5–11% [53]. While focal therapy can be delivered as a hemi- or “hockey-stick” ablation (1/2 or 3/4 of the prostate), based on coarse localization of prostate cancer [54], conformal delivery to ablate the cancer while minimizing damage to surrounding healthy tissue and organs at risk requires that the cancer within the prostate be delineated on imaging. A recent consensus panel suggested that this delineation should be performed using mpMRI [55].

Thus, prostate cancer imaging and lesion delineation could play an important role in focal therapy planning.

#### 1.1.3.4 Summary of potential roles for imaging and lesion delineation

Imaging and lesion delineation have several potential roles in supporting management of low–intermediate-risk patients. Longitudinal imaging-based measurements (such as tumor burden) may support earlier identification of active-surveillance patients who would eventually progress to curative-intent therapies. Accurate delineation of tumors may support conformal delivery of lesion-targeted therapies for low–intermediate-risk patients to minimize damage to surrounding tissue and organs at risk, potentially further lowering side effects. Delineation of tumors may support focal boosting of radiotherapy dose to reduce local recurrence rates. Finally, imaging and lesion delineation may support the decision to deliver nerve-sparing surgeries for suitable candidates. Accurate and precise delineation of cancerous lesions on imaging could support each of these goals.

#### 1.1.4 Prostate cancer imaging

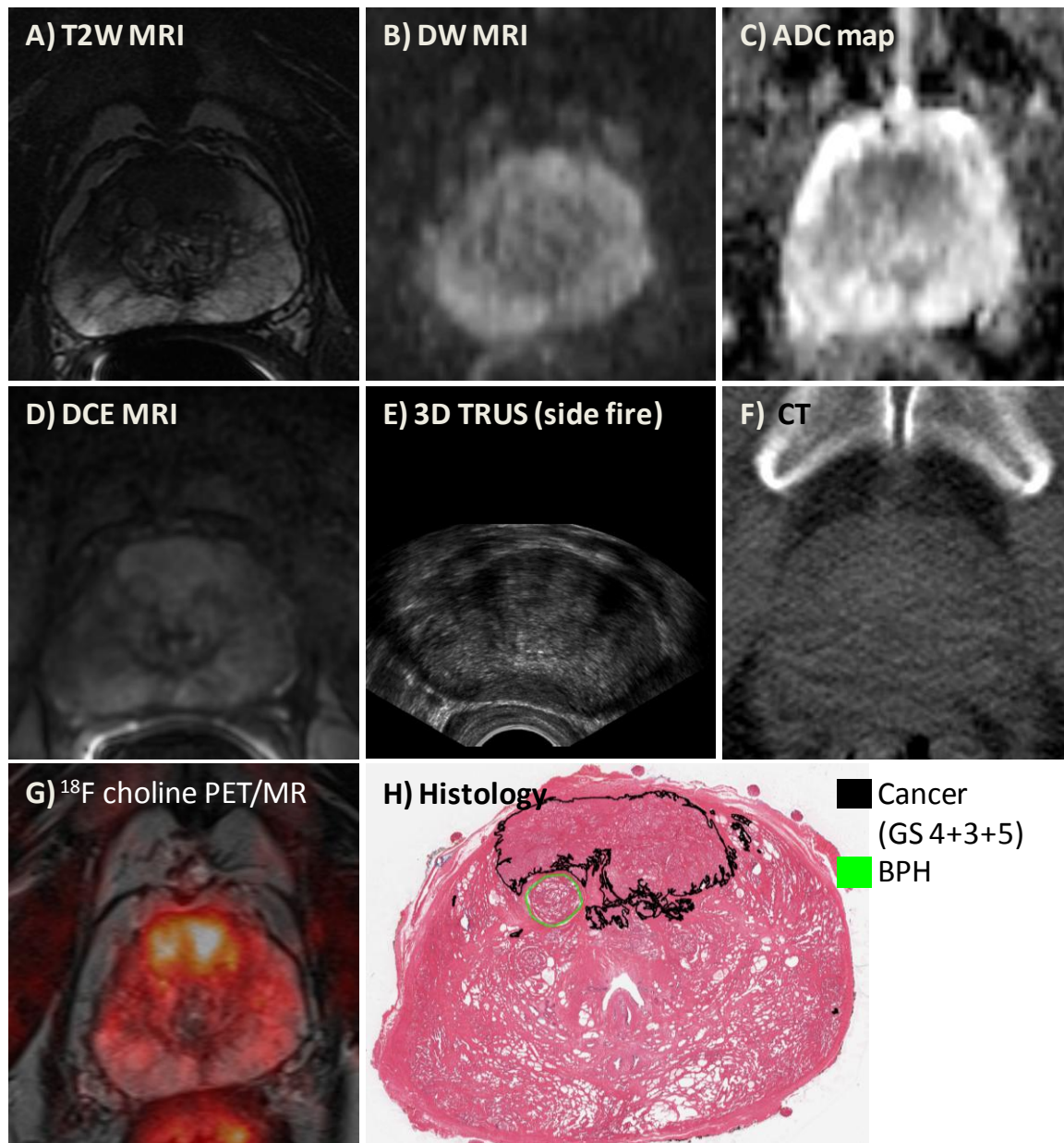
Prostate cancer imaging has the potential to extract spatial information non-invasively to support the detection, staging, grading, therapy selection, treatment planning and treatment delivery of primary and metastatic prostate cancer. Many imaging modalities can be used at different stages of the clinical workflow. mpMRI of localized prostate cancer is the focus of this thesis, and will be described in detail; however, the clinical application of ultrasound, x-ray computed tomography and nuclear imaging will

also be briefly discussed in the following sections. Figure 1.2 shows an illustrative comparison of these modalities for the same patient and anatomy.

#### 1.1.4.1 Magnetic resonance imaging

MRI has shown promise for detection [56], staging [57, 58], localization [59-61], therapy planning [43, 55, 62], and therapy monitoring [63] for localized prostate cancer, and may be suitable for lymph node staging with suitable contrast agents [64]. MRI can provide rich anatomical and pathological detail as it enables flexible control of soft tissue contrast with different acquisition sequences, and can have submillimeter resolution for some sequences.

The performance of individual MRI sequences for localized prostate cancer detection varies substantially [55], depending, in part, on the sequence used, and the location and grade of the cancer [65, 66]. Consensus guidelines [56] for detection of localized prostate cancer have been developed, recommending the use of multiple MRI sequences, comprising T2-weighted (T2W) imaging, and at least 2 functional sequences (e.g. diffusion-weighted (DW) imaging, dynamic contrast-enhanced (DCE) imaging, or MR spectroscopy). The cancer detection rate of the combination of T2W, DW and DCE imaging also varies (sensitivity: 53%–95%, specificity: 74%–96%, PPV: 51%–95%, negative predictive value [NPV]: 65%–94%) [67]. The following sections describe sequences and imaging protocols that have been investigated for imaging localized prostate cancer.



**Figure 1.2:** Multiple modalities of prostate cancer imaging showing approximately the same anatomical location. A large Gleason score 4+3 cancer with tertiary grade 5 is present in the anterior prostate, alongside a small region of BPH, a common feature that mimics cancer on many modalities.

#### 1.1.4.1.1 T2W MRI

T2W MRI is one of the basic MRI pulse sequences. It is sensitive to the transverse relaxation of precessing protons, which is different, for example, between

dense cellular tissue and fluid-filled cystic spaces. T2W MRI clearly depicts prostate zonal anatomy, due to its high spatial and contrast resolution [68]. In the peripheral zone, dense tumors appear hypointense on T2W MRI relative to the typically bright peripheral zone signal, although sparse tumors may not show such contrast [69]. In the central gland, homogenous low-intensity, ill-defined irregular boundaries, lenticular shape and invasion of the anterior fibromuscular stroma have been used to identify tumors [70]. T2W MRI has been reported to correlate with Gleason score, potentially offering prognostic information [71], although there was overlap in signals between grades, and this finding has not been widely replicated. T2W MRI interpretation is limited by the similar MR appearance of non-cancerous abnormalities, including chronic prostatitis, atrophy, scars, BPH, and post-biopsy hemorrhage [72].

#### 1.1.4.1.2 DW MRI

DW MRI is sensitive to diffusion of water molecules which decreases with the increased cellular density associated with cancer [73]. Interpretation of DW MRI for prostate cancer uses both DW MR images and post-processed apparent diffusion coefficient (ADC) maps. Prostate cancer appears hyperintense on DW MR images taken with high b-values (a parameter controlling sensitivity to diffusion [74]). Prostate cancer appears hypointense on ADC maps computed from multiple DW MR images taken with different b-values [56]. DW MRI has superior sensitivity for central gland tumors compared to other MRI sequences [65, 73]. ADC values in peripheral zone tumors have been shown to be correlated with Gleason score, potentially providing prognostic information [75-78], although there is overlap in ADC values between Gleason scores. Interpretation of DW MRI is challenged by the low resolution ( $1 \times 1 \text{ mm}^2$  to  $2 \times 2 \text{ mm}^2$  in-



plane resolution with 4–5 mm spacing) of typical acquisitions [56, 73], and frequently suffers from distortion artifacts [74, 79].

#### 1.1.4.1.3 DCE MRI

DCE MRI is sensitive to changes in vascular characteristics associated with angiogenesis due to cancer [80, 81]. DCE MRI comprises a rapid temporal sequence of T1-weighted MR images (at 1–90 s intervals) immediately prior to and for 5–10 minutes after injection of a gadolinium-chelate contrast agent. The contrast agent washes in and out more quickly in cancerous tissue than in non-cancerous tissue. Faster sequences, combined with estimates of the arterial input function, allow for precise modeling of pharmacokinetic parameters [80, 81], some of which (e.g. the contrast transfer coefficient [ $k^{\text{trans}}$ ] and the rate constant [ $k^{\text{ep}}$ ]) have been correlated with the presence of cancer [82]. DCE MRI interpretation is challenged by patient motion and by the presence of non-cancerous abnormalities that mimic cancer, including prostatitis in the peripheral zone, and highly vascularized BPH in the central gland [81]. Pharmacokinetic modeling is further challenged by the need to estimate the arterial input function [81].

#### 1.1.4.1.4 MR spectroscopy

MR spectroscopy (MRS) measures the spatial distribution of the relative concentrations of metabolites within the prostate that can be distinguished by their proton resonance frequencies, including choline, creatine, and citrate [83]. Cancerous tissue typically has elevated choline and decreased citrate compared to non-cancerous tissue. As the signals of choline and creatine are difficult to separate from each other [80], the (choline + creatine)/citrate ratio is frequently reported. MRS shows different metabolic ratios for tumor vs. normal tissue in the transition zone; however, signal strength and

interpretation remain challenging [84]. (Choline+creatine)/citrate ratios have been shown to be correlated with Gleason score [85], potentially providing prognostic information, and a retrospective study reported areas under the receiver operator characteristic curve from 0.70 to 0.78 for distinguishing low-grade from high-grade tumors using different measures based on MRS [86]. Interpretation of MRS is challenged by contamination of the signal from nearby lipids, high sensitivity to shimming, operator variability, and the presence of non-cancerous abnormalities, such as inflammation and BPH, that mimic cancer [80, 82-84].

#### 1.1.4.1.5 Emerging MRI techniques

Although MRI is showing promise for imaging prostate cancer, there remain substantial areas for improvement, particularly in detection of central gland tumors [55], and distinguishing cancer from confounders [72]. New imaging protocols such as endogenous sodium imaging [87, 88] and hyperpolarized carbon imaging [89, 90] are under development, and simultaneous hybrid imaging scanners incorporating positron emission tomography are becoming available. Additionally, new derived images computed from existing sequences are being evaluated, such as extrapolated high b-value DW images [91], diffusion compartment-model imaging [92] and alternative DCE parametric maps [93].

#### 1.1.4.2 Transrectal ultrasound

B-mode transrectal ultrasound (TRUS) imaging is commonly used to estimate prostate volume, and for image guidance during biopsy and needle-based-therapy procedures, as it enables rapid (10–20 frames per second) visualization of the prostate gland and the boundary between prostate zones. Prostate cancer in the peripheral zone,

where the majority of lesions are located, can appear hypoechoic (darker); however, the sensitivity for detecting prostate cancer on TRUS is low (59%–88%) [94-97]. Several ultrasound techniques are being investigated for prostate cancer imaging, including color and power Doppler imaging and micro-bubble-contrast-enhanced imaging [96].

#### 1.1.4.3 X-ray computed tomography

X-ray computed tomography (CT) imaging is commonly used in the assessment of metastases and in planning for whole-gland radiotherapy. CT is sensitive to differences in attenuation due to electron density, yielding good tissue-to-bone contrast, which supports detection of bone metastases [97]. CT is also used to detect lymph node metastases based on lymph node size and shape; round lymph nodes 8–10 mm in size or oval lymph nodes >10 mm in size had high specificity for metastasis [64]. Although CT is not used for detection of intraprostatic tumors due to poor soft tissue contrast [97], it is routinely used for planning whole-gland radiotherapy [4]. CT allows visualization of pelvic anatomy including nearby sensitive structures (e.g. the rectum and the bladder) and supports delineation of the prostate. Because CT and radiotherapy both involve the attenuation of electromagnetic radiation, measurements of electron density from CT can also be used in the planning of radiation therapy. Furthermore, in some external beam therapies, x-ray images can be acquired within the treatment system, supporting pre- or intraprocedural patient registration.

#### 1.1.4.4 Nuclear imaging

Nuclear imaging is commonly used in the assessment of metastases, and can also be used to detect intraprostatic cancer. Nuclear imaging measures radiation from radioactive atoms incorporated into radiotracers that preferentially accumulate at sites of

interest, such as metastatic bone tumors or intraprostatic tumors. For detecting metastases, a common form of nuclear imaging involves radiolabeled diphosphonate, which accumulates in regions of bone damage, including metastases. These bone scans have high sensitivity (95%) for osteoblastic metastases in patients with PSA levels above 20 ng/mL [97]. In prostate cancer,  $^{111}\text{In}$  capromab pendetide (marketed as ProstaScint) is a radiolabeled antibody that binds to the prostate specific membrane antigen (PSMA), which is preferentially expressed on prostate cells, including metastatic cells that originated from the prostate. For intraprostatic cancer detection, a number of radiotracers have been proposed.  $^{18}\text{F}$ -fluorodeoxyglucose, the most commonly used positron emission tomography (PET) tracer in other anatomical sites, is challenged by low metabolic glucose activity in prostate tumors and the masking effect of tracer accumulation in the bladder [98]. Several choline-based radiotracers, including  $^{11}\text{C}$ -choline and  $^{18}\text{F}$ -fluorocholine, show higher uptake in some primary and metastatic prostate cancers than in healthy tissue [99] and have less accumulation in the bladder [100]. Preliminary evidence suggests choline-based tracers may support detection of central gland tumors [101] and differentiation of tumor from confounders such as BPH and prostatitis [99].

#### 1.1.4.5 Justification for evaluating mpMRI for intraprostatic lesion delineation

Many imaging modalities are used in the prostate cancer clinical workflow; however, not all of these are suitable for delineation of intraprostatic lesions. TRUS and CT do not have the spatial and contrast resolution to support delineation of intraprostatic tumors. Nuclear imaging has been used for delineation of intraprostatic tumors [102, 103]; however, the low spatial resolution (e.g. 5 mm [104]) of clinical PET may

challenge accurate delineation of cancer boundaries. mpMRI has good soft tissue contrast showing anatomical detail, has high spatial resolution (0.5–2.0 mm in-plane [56]) and has shown promise for visualization of prostate cancer in detection, staging and coarse localization of intraprostatic cancer. Therefore, this thesis focuses on the evaluation of mpMRI for prostate cancer imaging and lesion delineation.

### 1.1.5 Histological reference standards

Evaluations of prostate cancer imaging and lesion delineation ideally include comparisons to a registered accepted reference standard. Typically, the most readily accepted reference standard for the presence, location and grade of prostate tumors is histological examination of prostate tissue (i.e., a pathologist's microscopic examination of tissue properties on ~4- $\mu$ m-thick sections of tissue).

#### 1.1.5.1 Clinical importance of histology in prostate cancer

In prostate cancer, histological examination is used at several points in the clinical workflow. Histological examination of biopsy tissue provides the first definitive diagnosis of prostate cancer, and is used as a follow-up to initial detection in screening populations, for ongoing monitoring in active surveillance, and to confirm recurrence in some post-treatment populations. Histological examination of resected prostate specimens is used to assess prognosis for patients post-prostatectomy [14], and to guide decisions about adjuvant therapy [4]. The importance of histology in the prostate cancer clinical workflow is due, in part, to the high prognostic value of the Gleason grading system. Initially developed in 1966 and updated in 2005, the Gleason grading system categorizes tissue based on its microscopic morphological appearance including

glandular differentiation and growth patterns within the stroma (Figure 1.1). It was initially evaluated in nearly 5000 men in randomized clinical trials, and shown to correlate well with survival. Since its initial development, the system has undergone multiple revisions [13], and has been adopted into international clinical guidelines [105].

#### 1.1.5.2 Biopsy vs. prostatectomy histology as a reference standard

For the validation of prostate cancer imaging, histological examination of biopsy or prostatectomy specimens has been used as a reference standard. Histology from prostatectomy specimens is the most powerful predictor of progression after prostatectomy [14], and is frequently used as the reference standard for the aggressiveness and location of prostate cancer [65, 71, 82, 84, 85, 106-109]. However, the use of prostatectomy specimens also limits the study population to patients undergoing prostatectomy, limiting the generalization of imaging findings to low-risk active surveillance and focal therapy cohorts. Histology from targeted- or template-mapping-biopsy samples, on the other hand, could be collected for such cohorts. However, biopsies are small sparse samples of tissue (~1 mm in diameter). These are commonly used for determining the presence or absence of cancer at a targeted location, but precisely determining the histological boundary of cancer from such samples would be challenging unless the sampling was impractically dense. Spatial error in biopsy targeting (3.5 mm for one MR–TRUS-fusion biopsy system as estimated by a recent analysis [20]) and the poorer reliability for assessing grade (40% rate of over- and underestimating Gleason grade with respect to prostatectomy specimens [17, 18]) would introduce further uncertainty into the analysis. For the some evaluations of lesion delineation, where the accurate spatial registration is critical to precisely evaluate the

accuracy of lesion boundaries, prostatectomy specimen histology is the more appropriate reference standard, although care must be taken in generalizing findings.

## 1.2 Problem domain

### 1.2.1 Key gaps in knowledge

The previous section identified applications in diagnosis, patient monitoring and therapy delivery where prostate cancer imaging and lesion delineation may have a role in the clinical workflow, and identified mpMRI as a suitable modality for this delineation and whole-mount histology as a suitable reference standard for evaluating the accuracy of delineation. Histology-based evaluation of the spatial accuracy of prostate cancer delineations requires that information about the histological tumor boundaries be registered into the spatial coordinates of the mpMR images being delineated. The following sections describe two key gaps in knowledge: (1) the lack of techniques for histology-imaging registration that are sufficiently accurate, robust and non-disruptive to the clinical pathology workflow; and consequently (2) the lack of knowledge about the accuracy and variability in lesion delineation on mpMRI which could lead to appropriate clinical guidelines.

### 1.2.2 Research challenges

To address these gaps in knowledge, this thesis focuses on (1) the development and evaluation of accurate, robust and non-disruptive techniques for registering a histological reference standard to mpMR images; and (2) applying these techniques for the evaluation of accuracy and variability of lesion contouring on mpMRI. This section describes key research challenges in addressing these goals.

### 1.2.2.1 Evaluation of lesion contouring

The accuracy and variability of lesion delineation on mpMRI can be measured along many dimensions. Ideally, lesion contouring could be evaluated with respect to patient outcome; however, this would require long follow-up, and may preclude paired analyses evaluating the delineations of multiple observers on one image. As a surrogate, one can evaluate spatial boundary-error measurements between the lesions on mpMRI of different observers (to assess interobserver variability), and between the lesions on mpMRI and the underlying tumor on the histological reference standard (to assess accuracy). Unlike the evaluation of staging and grading from mpMRI, where only a gross correspondence between the same tumor on *in vivo* imaging and on the reference standard must be established, the evaluation of spatial boundary errors requires densely defined information about the spatial relationship between the histology and mpMR images. This evaluation is further complicated by uncertainty in the boundary error measurements due to spatial errors in the registration of the histological reference standard to the mpMR images.

The complexity of information contained in a lesion delineation raises additional challenges. Different types of boundary errors may have different implications for different applications. A single type of boundary error measurement may not capture all of the clinically important information about the error. For example, low volumetric overlap may indicate poor overall accuracy of a contour, high boundary distance errors may indicate spicules of cancer or healthy tissue not captured by a contour, and volume differences may indicate a systematic over- or under-contouring of a focus. Thus, the



evaluation of lesion contouring requires the use of multiple metrics to yield a thorough assessment.

#### 1.2.2.2 Evaluation of registration accuracy

The accuracy of the registration of the reference standard to the mpMR images is critical to the evaluation of lesion contouring accuracy, because registration error induces uncertainty in the boundary error measurements. There can be compromises associated with achieving higher accuracy [110], including greater human interaction to guide registration algorithms to correct solutions, higher required image quality, and higher computational cost. When performing research on clinical specimens, this is particularly important as there may be compromises between disruption to the clinical pathology workflow and registration accuracy. Thus, it is important to identify the maximum acceptable level of registration error; however, the criteria for sufficiently accurate registration vary depending on the application [111]. The key determinant of the necessary target registration error (TRE) is the central question of the research that depends on the registrations. For example, a registration used to identify large cancer foci as homologous on histology and mpMRI would have less stringent registration requirements than one used to measure differences in the boundaries of mpMRI lesions and histological foci. Establishing application-specific thresholds for maximum acceptable error has been identified as a key challenge in registration [110, 111].

#### 1.2.2.3 Registration of histology to mpMRI

The registration of diagnostic prostate histology with mpMR images faces challenges due to three factors: (1) the disruptive process of acquiring histology tissue

from prostatectomy specimens, (2) the constraints on handling of clinical specimens, and (3) the properties of the resulting images.

The cutting of histology from prostatectomy specimens is a complex and disruptive process. Prostate specimens are formalin-fixed (inducing shrinkage), and are cut into 3–5-mm-thick tissue slices (resulting in the loss of the spatial relationship between slices). Any water in the slices is chemically replaced with paraffin (inducing deformation) and a single 4- $\mu$ m-thick histology section is cut from each paraffin block (resulting in further deformation and loss of 3D spatial information). Due to requirements for the archival preservation of diagnostic tissue, sections are only cut from near the faces of the tissue slices (resulting in a sparse sampling of tissue and loss of information about the 3D context of the tissue). The operator variability in the cutting of whole prostatectomy slices on a microtome introduces variability into the spatial relationships.

Reconstructing the spatial relationships between histology sections is complicated by constraints on the allowable treatment of clinical tissue. For example, the acceptable imaging modalities and the duration of imaging may be constrained to avoid disrupting the physical appearance of tissue and its chemical reactivity. The types of fiducial markers that can be used may be limited by clinical needs for patient treatment (e.g. precluding the use of fiducials *in vivo*), preservation of diagnostic tissue and avoidance of disruptions to clinical pathology workflows and diagnoses. Additionally, these constraints may vary from institution to institution, potentially limiting the reuse of existing methods.

Finally, several properties of the resulting images may introduce challenges for registration. The histology sections are sufficiently thin to be considered 2D, but the

histology images must be registered to a 3D image of a prostate specimen. Without the 3D context of the histological tissue, determining out-of-plane correspondences is challenging. Due to the non-linear deformation between the images, the section of the 3D image representing the tissue from which histology was cut may not even be planar, so that an operator may not be able to see the corresponding tissue from both images until a deformable registration is nearly complete; this complicates even the interactive registration of these images. Finally, the types of information contained in the two images are substantially different, as histology images denote the chemical affinity of tissue to chemical dyes, and MR images reflect properties of hydrogen atoms.

The combination of loss of 3D context from histology images, flexible deformation, variability in the histology acquisition processes, complex relationships relating image information, and constraints on the tools that can be used to mitigate these challenges makes the accurate registration of histology to mpMR images challenging.

There are four criteria that I have used to evaluate approaches for registering a histological reference standard to mpMRI:

- **Minimally disruptive:** The ideal method would work on either quartered or whole-mount histology collected according to usual clinical pathology protocols, in order to enable retrospective studies leveraging the large volume of existing mpMRI and histology data. Failing that, the method must not be disruptive to the pathological diagnosis: it should not result in changes to the collected histology that affect clinical diagnosis; it should not require the specimen to be out of formalin for an excessive amount of time; and it should not result in cutting tissue that is normally preserved for future diagnosis. Ideally, it would also be minimally

disruptive to the clinical pathology workflow. As workflows vary between centers, this is challenging to assess. Factors that could be considered include: minimizing delays that would impact the timeliness of the pathology report, avoiding alteration of standard cutting protocols, minimizing additional work for pathology assistants and pathologists, minimizing alteration of the specimen, and avoiding additional imaging (particularly imaging that cannot be performed in a pathology department).

- **Spatially accurate:** The 3D target registration error [112] should be evaluated using homologous point features (intrinsic landmarks or extrinsic fiducials) distributed throughout the gland on a substantial number of subjects. This error should be sufficiently low for the desired application; however, the maximum acceptable error has not yet been established for evaluation of prostate cancer imaging and delineation.
- **Robust:** The variability in performance and the failure rate of the method should be published. The method should be robust to intersubject variation in prostate anatomy and imaging appearance, and to the presence or absence of disease.
- **Widely implementable:** The method should minimize the need for specialized equipment or special expertise that may not be widely available. Software algorithms should be made available, or be easily re-implemented.

### 1.2.3 Previous work addressing these challenges

#### 1.2.3.1 Evaluation of GTV contouring

Previous evaluations of mpMRI have yielded quantifications of the sensitivity and specificity of staging and grading from different combinations of mpMR images, validated against histological information. Often this information is in the form of graded biopsies [76] that are corresponded with clinical imaging based on the origin of the biopsy. However, biopsies do not have the necessary information about the boundaries of foci to assess the accuracy of contouring. Evaluations of stage and grade that do compare against histological examination of prostatectomy specimens typically use a coarse partitioning of prostate anatomy [113, 114], or consensus between pathologists and radiologists [75] to establish correspondences between tumor foci on histology and mpMR images. These approaches may identify corresponding foci, but do not yield the spatial relationships between the histology and mpMR images required for evaluating contouring accuracy. Evaluations of the variability of contouring of the entire prostate and nearby organs at risk from imaging have been performed [115, 116]; however, these do not generally evaluate accuracy with respect to a histological reference standard and do not include measurements of cancer foci that would support focal therapy.

Two studies have recently been performed to evaluate the accuracy and variability of lesion contouring for conformal focal therapy on mpMRI [106, 117]. Rischke et al. evaluated target volumes defined on mpMRI by 5 radiation oncologists with access to the MRI report, measuring the accuracy and variability as a Dice overlap coefficient with respect to a radiologist's delineations. These measurements, while potentially important in the context of therapy, may not be representative of variability of observers not

informed by a common previous report, and rely on a reference standard with unknown accuracy and variability. Anwar et al. evaluated the histological coverage of target volumes defined by 2D margin expansion around lesions identified on mpMRI and delineated on a single 2D slice of T2W MRI. The histology-MRI registration error was reported as 2 mm in-plane, measured using the landmarks used to define the registration (i.e. a fiducial registration error, which typically is lower than the target registration error), and out of plane error was not estimated. Histological coverage was measured by breaking the histological focus and MR lesion boundaries of each focus into segments covering equal angles (from a central point that is not specified) and measuring the mean differences in distances to the central point. These measurements were aggregated across all subjects and observers, precluding evaluation of interobserver and intersubject variability. Margins of 5 mm and 8 mm were identified as covering 95% and 100% of the measurements.

To the best of my knowledge, there have been no studies to date evaluating the accuracy and variability (blinded to previous reporting on the images) of lesion contours in 3D against an accurately registered histological reference standard.

#### 1.2.3.2 Evaluation of registration accuracy

In some applications of registration, maximum acceptable registration accuracies have been identified. For example, in image-guided interventions where registration can be used to guide the tools used in the intervention, criteria have been proposed for specific interventions and imaging modalities based on clinical opinion [118] or models of what constitutes a successful intervention [119]. In registrations used to support imaging evaluation studies, some criteria have been proposed for sufficient registration

accuracy [120]; however, these criteria have not been tied to the outcomes of the studies supported by the registrations. Instead of defining criteria for registration accuracy, some approaches instead mitigate the effect of registration error on the study outcome. Examples include constraining the sizes of regions analyzed [121, 122], and qualifying study conclusions to acknowledge error as a confounder [123]. To the best of my knowledge, evaluation criteria have not been identified for the registration of prostate histology to mpMR images for evaluation of lesion contouring, and furthermore, accepted approaches for defining such criteria have not been identified.

#### 1.2.3.3 Registration of histology to mpMRI

Techniques used in the registration of clinical histology to *in vivo* images typically fall into two major categories: (1) techniques that alter the processes involved in cutting histology sections from specimens to mitigate deformation, avoid loss of spatial information or reduce variability; and (2) techniques that use assumptions about these processes and information (e.g. images or measurements) collected before, during or after the histology processing to virtually undo the deformation, reconstruct lost spatial information or account for variability. Many reconstruction methods use a combination of techniques from these two categories.

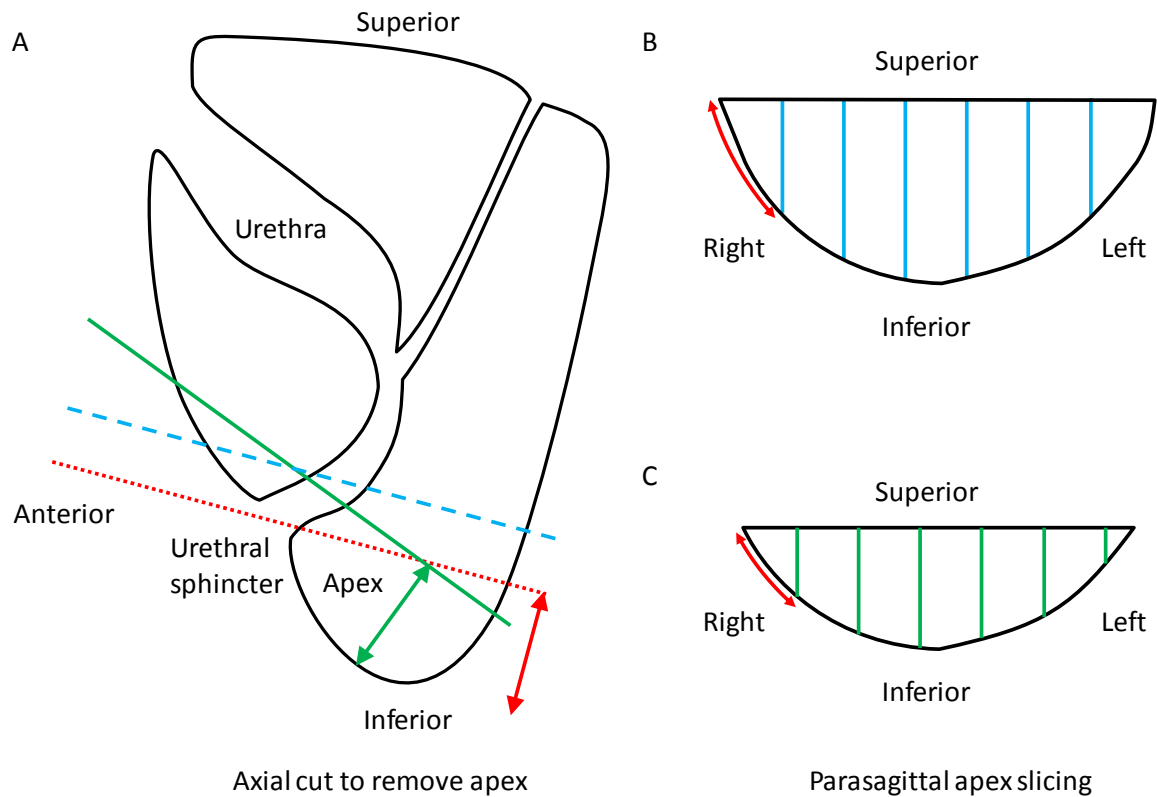
One common approach to the registration of whole-mount prostate histology to *in vivo* images is to guide the cutting of prostatectomy specimens into 3–5-mm-thick tissue slices [107, 120, 124], typically using image-guidance and specialized cutting equipment. Shah et al. [124] used a 3D printer to construct a patient-specific plastic mold based on the *in vivo* MRI, with slots to guide a custom multi-bladed knife along the *in vivo* imaging plane. Jhavar et al. [125, 126] acquired images with imaging planes

perpendicular to the posterior wall, and then cut the prostate perpendicular to the posterior wall without image guidance using special cradles and multi-bladed knives to obtain parallel slices for processing. Rouvière et al. [127] used injected fiducials (for *ex vivo* or animal studies) to define an imaging plane, used a rotating platform within the MR scanner to align these fiducials horizontally, and then embedded the specimen in wax to enable slicing parallel to the imaging plane. These methods register histology to *in vivo* images by assuming histology sections correspond to the specified slicing plane without error, and align histology within this plane using a deformable 2D registration.

Kalavagunta et al. [125] used guided slicing (similar to Jhavar et al. [125, 126]) to constrain orientation, but instead of assuming histology corresponded to a specified slice, an expert observer manually identified the corresponding slice in MRI. These methods constrain the position and orientation from which the histology is collected according to an *in vivo* imaging plane, which may be disruptive to the clinical pathology workflow when a pathologist prefers a specific orientation for diagnostic reasons. For example, the clinical pathology protocol at London Health Sciences Centre involves removing ~1 cm of the prostatic apex including the urethral sphincter for parasagittal sectioning to minimize the extent of surgical margin that is not visible on histology. For a specimen where the apex extends posterior and sufficiently inferior to the urethral sphincter (see Figure 1.3 or Walz et al. [128] Figure 11 for illustrations of such anatomy), a 1-cm apex cut perpendicular to the posterior wall may not include the sphincter. In such a case, a pathologist may prefer that the apex be removed with a 1-cm cut that is oblique to the posterior wall rather than a cut >1 cm (see Figure 1.3A) in order to minimize the amount of surgical margin that is not seen and assessed by the pathologist (see Figure 1.3B–C).



Additionally, these methods do not account for, or measure, out-of-plane error due to errors in the guidance of slicing, variability in the histology sectioning depth within the tissue slice, or error in the selection of the correct MR slice. Such underestimates of the target registration error impede the assessment of whether the registrations are sufficiently accurate for evaluating lesion contouring.



**Figure 1.3:** Illustrative diagram showing positions of cuts for gross slicing a specimen where constrained slicing may be disruptive to the clinical pathology workflow. (A) Side view of prostate. 1-cm apex cut (for parasagittal histology) constrained to be perpendicular to the posterior wall (red dotted line, red arrows show 1 cm) would not include the urethral sphincter, but a 1-cm oblique apex cut (green solid line, green arrows show 1 cm) would. To include the sphincter with a constrained cut, more tissue would have to be included in the apex histology (blue dashed line). (B) Anterior view of apex after cut along blue dashed line in (A). Pathologist makes diagnosis from histology taken from blue lines. Red arrow denotes the extent of margin that will not be assessed for extraprostatic extension and positive surgical margin. (C) Anterior view of apex after cut along green solid line in (A). Pathologist makes diagnosis from histology taken from green lines. Red arrow denotes the smaller extent of margin that will not be assessed for extraprostatic extension and positive surgical margin.

A second approach to the registration of whole-mount prostate histology to *in vivo* images uses extrinsic fiducials and/or additional imaging of the 3–5 mm tissue slices to reconstruct the spatial relationships between histology sections after they have been cut according to the clinical pathology workflow. Extrinsic fiducials have been used to orient histology images in 2D [129, 130], under the assumption that histology sections are sectioned to be parallel at constant spacing. Bart et al. [130] temporarily pierced the prostate with approximately parallel 1 mm needles running from apex to base, cut the gland axially into 3-mm-thick tissue slices with parallel cuts using a meat slicer, and interactively reconstructed histology based on the strong assumption that the needles did not deflect (their report notes that the needles did, in fact, deflect). Hughes et al. [129] adapted this method to use 3 oblique fiducial needles and collected step-sectioned parallel histology at 1-mm intervals (4.5 times more densely spaced than many clinical pathology protocols). Hughes used an automated fiducial finding algorithm, and an iterative minimization of fiducial registration error under assumptions of non-deflecting needles to align histology in plane. Alternatively, additional imaging has been used to orient histology images and account for some in-plane deformation [109, 126, 131-133]. Orczyk et al. [133] acquired photos of the tissue slices and used the posterior wall as a landmark to manually align adjacent sections into a 3D structure, blinded to 3D imaging. Jackson et al. [126] used anatomical landmarks in photos of the tissue slices to compensate for deformation during histological processing. Xu et al. [132] acquired high-resolution *ex vivo* MR images of the tissue slices for this purpose. Taylor et al. [131] and Groenendaal et al. [109] combined the parallel-needle-fiducial approach with acquiring photographs of the tissue slices to compensate for deformation during histology processing and to align

adjacent slices. Nearly all of these methods controlled the cutting of the tissue slices to be evenly spaced, with the exception of Taylor et al. [131], who measured the thicknesses of the tissue slices after cutting. As with the guided-slicing methods, these methods assumed that histology sections corresponded to the front faces of the tissue slices from which they were cut and did not account for variability in the depth of the histology sectioning within the tissue slice.

A third approach uses image content to retrospectively register histology to *in vivo* or *ex vivo* MR images. The registration of histology directly to *in vivo* images [134-138] holds the potential for minimal disruption to the clinical workflow; however, it relies on the presence of image information that may be disrupted by anatomic variability, disease processes or changes in imaging protocols.

Some methods required an expert to select a 2D MRI slice that corresponds to histology. Chappelow et al. [136] used mutual information between extracted texture features to register pseudo-whole-mount (PWM) histology (reconstructed from quartered histology) to the selected T2W MRI slice in 2D. They reported an intersection/union area ratio in the selected plane of 0.93 for 6 subjects. Kalavagunta et al. additionally required experts to contour the prostate and internal anatomy (e.g. BPH) on the PWM histology and the selected MR image slice. They used a variational minimization of overlap of these segmented regions to align histology to the selected MR image in-plane. They reported an in-plane TRE of 1.5 mm based on 103 intrinsic landmarks for 35 PWM histology sections. As Kalavagunta noted, the assumption that the expert's slice selection corresponds to the same cross-section of tissue is of "the utmost importance"; however, neither of these methods assess the validity of this assumption.

Other methods registered histology directly to the 3D MR image, with the potential to account for variability in the histology cutting depth or out-of-plane error in their initializations by identifying the section of the 3D image that corresponds to the histology that was actually cut. Unfortunately, these methods have not been evaluated using data sets and metrics that enable assessment of the robustness of the method or the expected spatial alignment of the histological reference standard. Chappelow et al. [138] leveraged the different sequences in mpMRI by first co-registering *in vivo* images, and then registering PWM histology to the mpMRI ensemble. Only the post-minimization value of the registration similarity metric was reported (for 25 subjects), which cannot be used to infer the accuracy of the registrations [112]. Patel et al. [134] used a mutual-information-based image registration, modified to weight voxels within the prostate more heavily, to deformably align histology to *in vivo* mpMRI. A mean Dice overlap of 0.83 and a mean absolute boundary distance of 0.99 mm from 2 patients were reported. Nir et al. [135] proposed a registration with particle-filtering-based optimization using an overlap metric between manual segmentations of MRI and parallel evenly-spaced histology images (that are not initially aligned in plane). They reported an in-plane mean TRE of  $2.1 \pm 1.3$  mm based on the alignment of visually corresponding landmarks (N is not reported) from 2 subjects. Prabu et al. [137] trained a statistical deformation model to regularize a free-form deformation-based registration. They reported an RMS TRE of 0.8 mm based on alignment of "key structures" from 2 patients. The small sample sizes in these analyses make it difficult to assess the robustness of these methods.

The use of *ex vivo* imaging as an intermediate registration target [139-142] allows more control over imaging, enabling the acquisition of higher resolution images that may

have more information in common with histology images. These methods address many of the challenges of 3D histology reconstruction, leaving as the remaining step a single-modality 3D *ex vivo* to 3D *in vivo* prostate MR image registration. However, these methods still rely on the presence of image information that may be disrupted.

Chappelow et al. [141] used mutual information between extracted texture features to register PWM histology to *ex vivo* MRI. They evaluated their registrations based on how well they separated the MRI intensities of cancerous and non-cancerous regions mapped from histology on data from 2 subjects. Zhan et al. [139] coarsely aligned and stacked 1.5 mm step-sectioned histology images (3 times more densely spaced than many clinical pathology protocols), and then automatically extracted 3D landmarks on both 4 T *ex vivo* MRI and reconstructed histology images. They corresponded landmarks (using shape feature vectors on the boundary and intensity features within the gland) by alternately updating correspondences to maximizing feature similarity and updating the approximating thin-plate-spline transformation. They reported a 0.88 mm TRE measured using anatomical landmarks (N is not reported) from 5 subjects. This may be a suitable approach if high-field *ex vivo* imaging and 1.5 mm step-sectioned histology is available; however, the small sample size and lack of detail regarding the evaluation makes this assessment challenging, and the use of 1.5 mm step-sectioned histology may prevent its use in clinical specimens at many centers. In a follow up to this work, Ou et al. [140] noted that the intensity features used in this approach were disrupted by the presence of cancer. To address this, Ou incorporated a subsequent registration that alternately segmented cancerous tissue and refined the registration to maximize overlap of the segmented regions; however, this introduces a bias toward aligning homogenous image

features in the registrations that may render the registrations unsuitable for evaluating prostate cancer imaging and lesion delineation. Nir et al. [142] addressed the *ex vivo* to *in vivo* registration using a biomechanical model that incorporated elasticity measurements from *in vivo* magnetic resonance elastography (MRE), yielding a  $3.1 \pm 1.4$  mean TRE (or  $3.7 \pm 1.9$  mm for a homogeneous elasticity model) based on the alignment of intrinsic landmarks (N is not reported) averaged over 6 patients. This may be a suitable approach for evaluating studies that already include MRE; however, the addition of vibration-inducing equipment and MRE acquisitions into *in vivo* protocols for registration purposes may be a barrier to its wider use.

None of the existing methods have been shown to meet all the previously described criteria for methods for registering a histological reference standard to imaging (non-disruptive, spatially accurate, robust, and widely implementable). There is a need for the development of more widely implementable, less disruptive methods with appropriate evaluation of accuracy and robustness. Consequently, there is a deficit in the evaluation of prostate cancer imaging [143] and in the evaluation of lesion delineation, which requires accurate registration.

### 1.3 Thesis research questions and objectives

To address the gaps in knowledge regarding the suitability of mpMRI for prostate cancer lesion delineation and the appropriate technology to evaluate prostate cancer imaging and lesion delineation, this thesis aims to address four research questions posed in this section through the completion of 7 research objectives. In doing so, this thesis will take steps towards addressing the 4 criteria for histology-imaging fusion methods described in Section 1.2.2.3 (non-disruptive, spatially accurate, robust, and widely

implementable) by presenting an approach to 3D histology reconstruction that (1) allows flexibility in the specimen slicing (to avoid disruption to the pathology diagnosis); (2) uses extrinsic fiducials that do not interfere with the clinical interpretation of histology (to avoid disruption to the pathology diagnosis) and are not dependent on intrinsic contrast (to improve robustness); (3) has a spatial reconstruction error of 0.7 mm and little sensitivity to initialization, which is evaluated using a 3D TRE computed from 232 homologous point landmarks in 37 whole-mount histology images from 10 patients (to appropriately quantify accuracy and robustness). Our approach has been successfully implemented at University Health Network, Toronto, demonstrating its potential portability for more widespread use.

### 1.3.1 Research questions

- How does registration error impact the statistical power of imaging validation studies? This will be answered via the completion of Objective 1 below.
- What is the performance (in terms of reconstruction error and robustness) of an extrinsic-fiducial-based histology reconstruction method, and how does this compare to alternative approaches based on image-guided slicing and intensity-based image registration? This will be answered via the completion of Objectives 2a and 2b below.
- What is the variability of lesion scoring (using consensus-panel-recommended PI-RADS guidelines [56]) and contouring on mpMRI? This will be answered via the achievement of Objectives 2c, 3a and 3b below.

- What are appropriate margins around observers' target volumes to achieve histological coverage of detected cancers? This will be answered via the achievement of Objectives 2c, 3a and 3c below.

### 1.3.2 Specific research objectives

**1 – Impact of registration error on imaging validation studies:** to establish criteria for determining the necessary accuracy of histological-reference-standard registrations for the assessment of prostate cancer imaging.

**2a – 3D histology reconstruction:** to develop methods for the 3D reconstruction of prostate histology based on strand-shaped fiducials, as well as alternative approaches based on image-guided slicing and intensity-based image registration.

**2b – 3D histology reconstruction evaluation:** to quantify the reconstruction error and robustness (variability of error, and sensitivity to initialization) of the developed methods in Objective 2a.

**2c – Histology to *in vivo* MR image registration:** to develop methods for registration of reconstructed 3D histology images to *in vivo* 3D mpMRI and quantify the registration accuracy.

**3a – mpMRI observer study:** to coordinate a multi-observer study of prostate cancer lesion scoring and delineation on mpMRI.

**3b – Variability of lesion scoring and delineation:** to quantify the interobserver and intersequence variability of lesion scoring and delineation, and factors that correlate with them, such as Gleason grade, anatomical location and lesion volume.



**3c – Margin evaluation:** to determine expansion margins around mpMRI-defined gross target volumes that would, with high likelihood, result in adequate coverage of histological cancer.

## 1.4 Thesis Outline

### 1.4.1 Chapter 2: The impact of registration accuracy on imaging validation study design: a novel statistical power calculation

The purpose of this work was to develop an approach for evaluating registration accuracy via its impact on statistical power for imaging evaluation studies that use registered reference standards (e.g. histology images). This work focused on studies to detect differences in mean image signal between normal and abnormal regions. By expressing registration error in terms of the overlap of the registered sampling regions (from reference standard images) with the true regions on the study images, the population mean and variance of sampled intensities could be expressed in terms of registration error and incorporated into the classical power calculation formula. This yielded a novel power calculation formula that can be arranged to answer three key questions affecting study design: (1) What is the maximum acceptable registration error? (2) How many subjects are needed? (3) What is the minimum detectable difference? A case study to illustrate the application of this approach was included for discussion.

### 1.4.2 Chapter 3: 3D prostate histology image reconstruction: quantifying the impact of tissue deformation and histology section location

The purpose of this work was to assess the impact of the histology cutting process on the accuracy of methods for reconstructing whole-mount histopathology images into a

3D spatial context. This work used pairs of intrinsic homologous landmarks identified on histology images and MR images of formalin-fixed prostate tissue slices before histological processing to identify the position and orientation from which the histology was originally cut. This work showed that a rigid+isotropic scale transformation accounted for the majority of in-plane deformation during histological processing, and that histology was cut from a mean depth of 1.1 mm and a mean angle of 1.5°. We showed that making the assumption, common in the literature, that histology is cut from the surface of tissue slices resulted in 0.7 mm additional error under an isotropic scaling deformation model. Although this error appears to be small, incorporating this error into the power calculations described in Chapter 2 illustrated that assumptions about the deformation model and position of histology section could have a substantial, and potentially costly, impact on the required sample sizes for imaging validation studies using a registered histological reference standard.

#### 1.4.3 Chapter 4: Registration of prostate histology images to *ex vivo* MR images via strand-shaped fiducials

The purpose of this work was to present and evaluate a method for reconstructing whole-mount histology images into the 3D context of an *ex vivo* MRI. The method used previously developed histology- and MRI-visible strand-shaped fiducials to construct a form of stereotactic frame wrapped around and within the prostate, and minimized the fiducial registration error between cross-section of the fiducials visible on histology and the parametric curves of fiducials visible on MRI. The target registration error was quantified to be 0.7 mm using 3–7 pairs of intrinsic landmarks on each of 34 histology

images. This compared favorably to alternative algorithms based on image-guided slicing (1.2 mm) and additional imaging of tissue slices (0.9 mm).

#### 1.4.4 Chapter 5: 3D prostate histology reconstruction: an evaluation of image-based and fiducial-based algorithms

The purpose of this work was to compare the accuracy and robustness of the fiducial-based reconstruction method to an alternative method based on an established multimodality image-intensity-based registration method (mutual information). The two methods were initialized using known perturbations from the optimal reconstruction (to assess robustness to initialization) and using practically achievable initializations. The image-intensity-based registration approach showed considerable sensitivity to initialization, whereas the fiducial-based approach had almost zero sensitivity to initialization. The resulting target registration errors for the practical initializations were quantified to be 0.7 mm for the fiducial-based approach and 1.2 mm for the image-registration-based approach using 232 pairs of intrinsic landmarks identified on 37 histology images and corresponding MR images. This work concluded that the fiducial-based approach was more accurate and more robust to initialization than the tested image-intensity-based registration approach.

#### 1.4.5 Chapter 6: Prostate cancer assessment and delineation on 3 T multiparametric MRI: interobserver and intersequence agreement

The purpose of this work was to evaluate mpMRI for scoring prostate cancer lesions (following the Prostate Imaging Reporting and Data System [PI-RADS] [56] detection guidelines) and for delineating suspicious lesions on different mpMRI

sequences (T2W, DCE and ADC images). Four observers scored and delineated lesions on images from each mpMRI sequence from 17 radical prostatectomy patients. We measured the positive predictive values (PPVs) of lesion scores with respect to co-registered histology, interobserver variability in lesion scoring, and the interobserver and intersequence variability in lesion delineation (using Dice overlap, mean absolute distance and absolute volume difference measures), as well as the correlation of these with interaction factors including lesion volume and histological grade. We found that a PI-RADS likelihood score of 5 (denoting highly likely cancerous) had a positive predictive value of 85% for Gleason 7 cancer, and a PPV of 93% if the delineated lesion had a volume  $>0.5 \text{ cm}^3$ . However, many false positives were observed, especially for lesions with PI-RADS likelihood scores of 3 and 4. The mean interobserver differences in the four PI-RADS scores ranged from 0.6 to 1.2 (5-point scale), with an agreement kappa of 0.30. ADC maps showed a trend towards superior interobserver contouring agreement compared to T2W or DCE images for all metrics. The observed correlation between PI-RADS scores and post-prostatectomy Gleason scores was promising; however, the interobserver variability in PI-RADS scores may impede their interpretation, and the prevalence of high PI-RADS likelihood scores for non-cancerous lesions suggests a need for further investigation of mpMRI confounders. Lesion delineation showed substantial variability, suggesting a need for standardized contouring guidelines and training.

#### 1.4.6 Chapter 7: Toward prostate cancer contouring guidelines on MRI: dominant lesion gross and clinical target volume coverage via accurate histology fusion

The purpose of this work was to evaluate a range of clinical target volume (CTV) expansion margins for lesions (or gross target volumes [GTVs]) defined on mpMR images with respect to their likelihood of yielding adequate coverage of midgland histological cancer, and to evaluate CTVs resulting from margins that yield a high likelihood of adequate coverage. Four observers delineated lesions on images from each mpMRI sequence from 25 radical prostatectomy patients. Contoured and graded histology images were co-registered with the mpMR images. CTVs were constructed with 0–30 mm margins expanded isotropically (constrained to prostate tissue) from true positive GTVs on each sequence and from composite GTVs from multiple sequences. The proportion of residual midgland cancer (and, separately, high-grade midgland cancer) remaining outside each CTV was quantified, and prediction intervals for residual area were computed for each observer and for each type of single-sequence and composite CTV. Margins yielding high likelihood (78–91%) of leaving less than 0%, 5% and 10% residual cancer were identified, and the CTV absolute volumes and relative volumes (compared to GTVs and histological cancer) were calculated. The minimal margins with high likelihood of leaving 0–10% residual cancer were lowest for composite GTVs using delineations on T2W, DCE and ADC images; however, there was substantial variation between observers, with margins of 4.5–9.0 mm, 1.5–7.5 mm and 1.5–4.5 mm for 0%, 5% and 10% residual high-grade cancer, and 6.0–11.0 mm, 5.5–10.0 mm, and 3.0–9.5 mm for 0%, 5% and 10% residual cancer of any grade. However, the

volume of the resulting CTVs from these three-sequence GTV delineations with smaller margins was not consistently lower than the CTVs of other types with larger margins. Across all observers and CTV types, 50–80% of CTVs for high-grade cancer had volumes <10 ml, potentially suitable for focal radiation boosting, and 9–51% had volumes <5 ml, potentially suitable for focal laser ablation. We concluded that lesion delineation on mpMRI with appropriate margins has the potential to generate CTVs with adequate histological coverage that are deliverable with some targeted techniques, but that there may be value in developing methods for characterizing individual lesions, in developing contouring guidelines and training programs to reduce variability, and in investigating non-isotropic margin expansions to achieve adequate histological coverage with smaller CTVs.

#### 1.4.7 Chapter 8: Contributions of the thesis, applications and suggestions for future work

This chapter summarizes the advances in knowledge related to each of the thesis' research questions, and discusses practical applications of the thesis contributions along with corresponding directions for future research.

### 1.5 References

1. Canadian Cancer Society's Advisory Committee on Cancer Statistics, "Canadian cancer statistics 2014", Toronto, ON: Canadian Cancer Society; 2014
2. G. P. Haas and W. A. Sakr, "Epidemiology of prostate cancer," *CA: A Cancer Journal for Clinicians* **47**, 273–287 (1997).
3. C. H. Bangma, S. Roemeling and F. H. Schröder, "Overdiagnosis and overtreatment of early detected prostate cancer," *World Journal of Urology* **25**, 3–9 (2007).

4. I. Thompson, J. B. Thrasher, G. Aus, A. L. Burnett, E. D. Canby-Hagino, M. S. Cookson, A. V. D'Amico, R. R. Dmochowski, D. T. Eton, J. D. Forman, S. L. Goldenberg, J. Hernandez, C. S. Higano, S. R. Kraus, J. W. Moul and C. M. Tangen, "Guideline for the management of clinically localized prostate cancer: 2007 update," *Journal of Urology* **177**, 2106–2131 (2007).
5. J. I. Izawa, L. Klotz, D. R. Siemens, W. Kassouf, A. So, J. Jordan, M. Chetner and A. E. Iansavichene, "Prostate cancer screening: Canadian guidelines 2011," *Canadian Urological Association Journal* **5**, 235–240 (2011).
6. I. M. Thompson, D. K. Pauler, P. J. Goodman, C. M. Tangen, M. S. Lucia, H. L. Parnes, L. M. Minasian, L. G. Ford, S. M. Lippman, E. D. Crawford, J. J. Crowley and C. A. C. Jr., "Prevalence of prostate cancer among men with a prostate-specific antigen level 4.0 ng per milliliter," *New England Journal of Medicine* **350**, 2239–2246 (2004).
7. F. H. Schröder, A. B. Kruger, J. Rietbergen, R. Kranse, P. van der Maas, P. Beemsterboer and R. Hoedemaeker, "Evaluation of the digital rectal examination as a screening test for prostate cancer," *Journal of the National Cancer Institute* **90**, 1817–1823 (1998).
8. G. L. Andriole, E. D. Crawford, R. L. Grubb, S. S. Buys, D. Chia, T. R. Church, M. N. Fouad, C. Isaacs, P. A. Kvale, D. J. Reding and others, "Prostate cancer screening in the randomized Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial: mortality results after 13 years of follow-up," *Journal of the National Cancer Institute*, 125–132 (2012).
9. F. H. Schröder, J. Hugosson, M. J. Roobol, T. L. Tammela, S. Ciatto, V. Nelen, M. Kwiatkowski, M. Lujan, H. Lilja, M. Zappa, L. J. Denis, F. Recker, A. Páez, L. Määttänen, C. H. Bangma, G. Aus, S. Carlsson, A. Villers, X. Rebillard, T. v. d. Kwast, P. M. Kujala, B. G. Blijenberg, U.-H. Stenman, A. Huber, K. Taari, M. Hakama, S. M. Moss, H. J. d. Koning and A. Auvinen, "Prostate-cancer mortality at 11 years of follow-up," *New England Journal of Medicine* **366**, 981–990 (2012).
10. V. A. Moyer, "Screening for prostate cancer: US Preventive Services Task Force recommendation statement," *Annals of Internal Medicine* **157**, 120–134 (2012).
11. J. E. McNeal, E. A. Redwine, F. S. Freiha and T. A. Stamey, "Zonal distribution of prostatic adenocarcinoma: correlation with histologic pattern and direction of spread," *American Journal of Surgical Pathology* **12**, 897–906 (1988).
12. M. E. Chen, D. A. Johnston, K. Tang, R. J. Babaian and P. Troncso, "Detailed mapping of prostate carcinoma foci: biopsy strategy implications," *Cancer* **89**, 1800–1809 (2000).

13. J. I. Epstein, "An update of the Gleason grading system," *Journal of Urology* **183**, 433–440 (2010).
14. R. Montironi, R. Mazzuccheli, M. Scarpelli, A. Lopez-Beltran, G. Fellegara and F. Algaba, "Gleason grading of prostate cancer in needle biopsies or radical prostatectomy specimens: contemporary approach, current clinical significance and sources of pathology discrepancies," *BJU International* **95**, 1146–1152 (2005).
15. M. W. Kattan, J. A. Eastham, A. M. Stapleton, T. M. Wheeler and P. T. Scardino, "A preoperative nomogram for disease recurrence following radical prostatectomy for prostate cancer," *Journal of the National Cancer Institute* **90**, 766–771 (1998).
16. A. W. Partin, L. A. Mangold, D. M. Lamm, P. C. Walsh, J. I. Epstein and J. D. Pearson, "Contemporary update of prostate cancer staging nomograms (Partin Tables) for the new millennium," *Urology* **58**, 843–848 (2001).
17. C. R. King, J. E. McNeal, H. Gill and J. C. Presti Jr, "Extended prostate biopsy scheme improves reliability of Gleason grading: implications for radiotherapy patients," *International Journal of Radiation Oncology\* Biology\* Physics* **59**, 386–391 (2004).
18. G. J. Kelloff, P. Choyke and D. S. Coffey, "Challenges in clinical prostate cancer: role of imaging," *American Journal of Roentgenology* **192**, 1455–1470 (2009).
19. S. Vourganti, A. Rastinehad, N. K. Yerram, J. Nix, D. Volkin, A. Hoang, B. Turkbey, G. N. Gupta, J. Kruecker, W. M. Linehan, P. L. Choyke, B. J. Wood and P. A. Pinto, "Multiparametric magnetic resonance imaging and ultrasound fusion biopsy detect prostate cancer in patients with prior negative transrectal ultrasound biopsies," *Journal of Urology* **188**, 2152–2157 (2012).
20. P. R. Martin, D. W. Cool, C. Romagnoli, A. Fenster and A. D. Ward, "Toward 3D-guided prostate biopsy target optimization: an estimation of tumor sampling probabilities," in *Proceedings of SPIE Medical Imaging, Vol. 9036* (San Diego, USA, 2014), pp. 90361C–90361C–90366.
21. E. Johansson, A. Bill-Axelson, L. Holmberg, E. Onelov, J. E. Johansson and G. Steineck, "Time, symptom burden, androgen deprivation, and self-assessed quality of life after radical prostatectomy or watchful waiting: the Randomized Scandinavian Prostate Cancer Group Study Number 4 (SPCG-4) clinical trial," *European Urology* **55**, 422–430 (2009).
22. E. Johansson, G. Steineck, L. Holmberg, J. E. Johansson, T. Nyberg, M. Ruutu and A. Bill-Axelson, "Long-term quality-of-life outcomes after radical prostatectomy or watchful waiting: the Scandinavian Prostate Cancer Group-4 randomised trial," *Lancet Oncology* **12**, 891–899 (2011).



23. M. A. Dall'Era, P. C. Albertsen, C. Bangma, P. R. Carroll, H. B. Carter, M. R. Cooperberg, S. J. Freedland, L. H. Klotz, C. Parker and M. S. Soloway, "Active surveillance for prostate cancer: a systematic review of the literature," *European Urology* **62**, 976–983 (2012).
24. D. M. Latini, S. L. Hart, S. J. Knight, J. E. Cowan, P. L. Ross, J. Duchane and P. R. Carroll, "The relationship between anxiety and time to treatment for patients with prostate cancer on surveillance," *Journal of Urology* **178**, 826–831; discussion 831–822 (2007).
25. L. Klotz, "Active surveillance with selective delayed intervention for favorable risk prostate cancer," in *Proceedings of Urologic Oncology: Seminars and Original Investigations*, (2006).
26. J. Xia, B. J. Trock, M. R. Cooperberg, R. Gulati, S. B. Zeliadt, J. L. Gore, D. W. Lin, P. R. Carroll, H. B. Carter and R. Etzioni, "Prostate Cancer Mortality following Active Surveillance versus Immediate Radical Prostatectomy," *Clinical Cancer Research* **18**, 5471–5478 (2012).
27. S. Loeb, Y. Folkvaljon, D. Robinson, D. Makarov and P. Stattin, "Does active surveillance miss the window for cure? Matched comparison of immediate versus delayed prostatectomy in a nationwide population-based cohort," in *Proceedings of American Urological Association*, (2014).
28. J. K. Mullins, D. Bonekamp, P. Landis, H. Begum, A. W. Partin, J. I. Epstein, H. B. Carter and K. J. Macura, "Multiparametric magnetic resonance imaging findings in men with low-risk prostate cancer followed using active surveillance," *BJU International* **111**, 1037–1045 (2013).
29. D. Stevens, C. Moore, H. Ahmed, C. Allen, A. Kirkham, J. Van Der Meulen and M. Emberton, "The natural history of untreated prostate MRI lesions in an active surveillance prostate cancer population – 260 patient-years," in *Proceedings of Annual Congress of the European Association of Urology*, (2012).
30. N. L. Robertson, C. M. Moore, G. Ambler, S. R. J. Bott, A. Freeman, G. Gambarota, C. Jameson, A. V. Mitra, B. Whitcher, M. Winkler, A. Kirkham, C. Allen and M. Emberton, "MAPPED study design: A 6-month randomised controlled study to evaluate the effect of dutasteride on prostate cancer volume using magnetic resonance imaging," *Contemporary Clinical Trials* **34**, 80–89 (2013).
31. T. H. van der Kwast, M. B. Amin, A. Billis, J. I. Epstein, D. Griffiths, P. A. Humphrey, R. Montironi, T. M. Wheeler, J. R. Srigley, L. Egevad and B. Delahunt, "International Society of Urological Pathology (ISUP) Consensus Conference on Handling and Staging of Radical Prostatectomy Specimens. Working group 2: T2 substaging and prostate cancer volume," *Modern Pathology* **24**, 16–25 (2011).

32. W. J. Catalona, G. F. Carvalhal, D. E. Mager and D. S. Smith, "Potency, continence and complication rates in 1,870 consecutive radical retropubic prostatectomies," *Journal of Urology* **162**, 433–438 (1999).
33. W. M. Mendenhall, R. H. Henderson, D. J. Indelicato, S. R. Keole and N. P. Mendenhall, "Erectile dysfunction after radiotherapy for prostate cancer," *American Journal of Clinical Oncology* **32**, 443–447 (2009).
34. J. A. Talcott, P. Rieker, K. J. Probert, J. A. Clark, K. I. Wishnow, K. R. Loughlin, J. P. Richie and P. W. Kantoff, "Patient-reported impotence and incontinence after nerve-sparing radical prostatectomy," *Journal of the National Cancer Institute* **89**, 1117–1123 (1997).
35. K. J. Kowalczyk, J. M. Levy, C. F. Caplan, S. R. Lipsitz, H. Y. Yu, X. Gu and J. C. Hu, "Temporal national trends of minimally invasive and retropubic radical prostatectomy outcomes from 2003 to 2007: results from the 100% Medicare sample," *European Urology* **61**, 803–809 (2012).
36. W. Y. Khoder, R. Waidelich, M. Seitz, A. J. Becker, A. Buchner, S. Trittschler and C. G. Stief, "Do we need the nerve sparing radical prostatectomy techniques (intrafascial vs. interfascial) in men with erectile dysfunction? Results of a single-centre study," *World Journal of Urology*, 1–7 (2014).
37. R. F. Sanchez-Ortiz, G. A. Broderick, E. S. Rovner, A. J. Wein, R. Whittington and S. B. Malkowicz, "Erectile function and quality of life after interstitial radiation therapy for prostate cancer," *International Journal of Impotence Research* **12 Suppl 3**, S18–24 (2000).
38. R. M. Benoit, M. J. Naslund and J. K. Cohen, "Complications after prostate brachytherapy in the Medicare population," *Urology* **55**, 91–96 (2000).
39. G. K. Zagars, A. C. von Eschenbach, D. E. Johnson and M. J. Oswald, "Stage C adenocarcinoma of the prostate. An analysis of 551 patients treated with external beam radiation," *Cancer* **60**, 1489–1499 (1987).
40. Y. Maeda, M. Hoyer, L. Lundby and C. Norton, "Faecal incontinence following radiotherapy for prostate cancer: a systematic review," *Radiotherapy and Oncology* **98**, 145–153 (2011).
41. D. Pucar, H. Hricak, A. Shukla-Dave, K. Kuroiwa, M. Drobnjak, J. Eastham, P. T. Scardino and M. J. Zelefsky, "Clinically Significant Prostate Cancer Local Recurrence After Radiation Therapy Occurs at the Site of Primary Tumor: Magnetic Resonance Imaging and Step-Section Pathology Evidence," *International Journal of Radiation Oncology\*Biophysics* **69**, 62–69 (2007).

42. I. M. Lips, U. A. van der Heide, K. Haustermans, E. N. van Lin, F. Pos, S. P. Franken, A. N. Kotte, C. H. van Gils and M. van Vulpen, "Single blind randomized phase III trial to investigate the benefit of a focal lesion ablative microboost in prostate cancer (FLAME-trial): study protocol for a randomized controlled trial," *Trials* **12**, 255 (2011).
43. G. Bauman, M. Haider, U. A. Van der Heide and C. Menard, "Boosting imaging defined dominant prostatic tumors: a systematic review," *Radiotherapy and Oncology* **107**, 274–281 (2013).
44. J. F. Williamson, J. C. Ford, M. C. Carlone and M. S. Anscher, "Focal External Beam Radiation Therapy in Management of Low-Risk Prostate Cancer: A Radiobiological Analysis," in *Proceedings of American Society for Radiation Oncology Annual Meeting*, (Miami Beach, USA, 2011).
45. V. Mouraviev, J. M. Mayes, J. F. Madden, L. Sun and T. J. Polascik, "Analysis of laterality and percentage of tumor involvement in 1386 prostatectomized specimens for selection of unilateral focal cryotherapy," *Technique in Cancer Research & Treatment* **6**, 91–95 (2007).
46. B. Djavan, M. Susani, B. Bursa, A. Basharkhah, R. Simak and M. Marberger, "Predictability and significance of multifocal prostate cancer in the radical prostatectomy specimen," *Techniques in Urology* **5**, 139–142 (1999).
47. M. Noguchi, T. A. Stamey, J. E. McNeal and R. Nolley, "Prognostic factors for multifocal prostate cancer in radical prostatectomy specimens: lack of significance of secondary cancers," *Journal of Urology* **170**, 459–463 (2003).
48. H. U. Ahmed, R. G. Hindley, L. Dickinson, A. Freeman, A. P. Kirkham, M. Sahu, R. Scott, C. Allen, J. Van der Meulen and M. Emberton, "Focal therapy for localised unifocal and multifocal prostate cancer: a prospective development study," *Lancet Oncology* **13**, 622–632 (2012).
49. E. Lecomnet, C. Moore, H. U. Ahmed and M. Emberton, "Focal therapy for prostate cancer: fact or fiction?," *Urologic Oncology* **28**, 550–556 (2010).
50. C. M. Moore, D. Pendse and M. Emberton, "Photodynamic therapy for prostate cancer – a review of current status and future promise," *Nature Clinical Practice Urology* **6**, 18–30 (2009).
51. U. Lindner, R. A. Weersink, M. A. Haider, M. R. Gertner, S. R. Davidson, M. Atri, B. C. Wilson, A. Fenster and J. Trachtenberg, "Image guided photothermal focal therapy for localized prostate cancer: phase I trial," *Journal of Urology* **182**, 1371–1377 (2009).

52. S. Langley, H. U. Ahmed, B. Al-Qaisieh, D. Bostwick, L. Dickinson, F. G. Veiga, P. Grimm, S. Machtens, F. Guedea and M. Emberton, "Report of a consensus meeting on focal low dose rate brachytherapy for prostate cancer," *BJU International* **109**, 7-16 (2012).
53. G. Bozzini, P. Colin, P. Nevoux, A. Villers, S. Mordon and N. Betrouni, "Focal therapy of prostate cancer: energies and procedures," *Urologic Oncology* **31**, 155–167 (2013).
54. M. Valerio, H. U. Ahmed, M. Emberton, N. Lawrentschuk, M. Lazzeri, R. Montironi, P. L. Nguyen, J. Trachtenberg and T. J. Polascik, "The Role of Focal Therapy in the Management of Localised Prostate Cancer: A Systematic Review," *European Urology*, (in press) (2014).
55. B. G. Muller, J. J. Fütterer, R. T. Gupta, A. Katz, A. Kirkham, J. Kurhanewicz, J. W. Moul, P. A. Pinto, A. R. Rastinehad, C. Robertson, J. de la Rosette, R. Sanchez-Salas, J. S. Jones, O. Ukimura, S. Verma, H. Wijkstra and M. Marberger, "The role of magnetic resonance imaging (MRI) in focal therapy for prostate cancer: recommendations from a consensus panel," *BJU International* **113**, 218–227 (2014).
56. J. O. Barentsz, J. Richenberg, R. Clements, P. Choyke, S. Verma, G. Villeirs, O. Rouviere, V. Logager and J. J. Fütterer, "ESUR prostate MR guidelines 2012," *European Radiology* **22**, 746–757 (2012).
57. B. Kim, R. H. Breau, D. Papadatos, D. Fergusson, S. Doucette, I. Cagiannos and C. Morash, "Diagnostic accuracy of surface coil magnetic resonance imaging at 1.5 T for local staging of elevated risk prostate cancer," *Canadian Urological Association Journal* **4**, 257–262 (2010).
58. F. Couñago, M. Recio, E. del Cerro, L. Cerezo, A. Díaz Gavela, F. J. Marcos, R. Murillo, J. M. Rodriguez Luna, I. J. Thuissard and J. L. R. Martin, "Role of 3.0 T multiparametric MRI in local staging in prostate cancer and clinical implications for radiation oncology," *Clinical and Translational Oncology*, 1–7 (2014).
59. A. Wefer, H. Hricak, D. Vigneron, F. Coakley, Y. Lu, J. Wefer, U. Mueller-Lisse, P. Carroll and J. Kurhanewicz, "Sextant localization of prostate cancer: comparison of sextant biopsy, magnetic resonance imaging and magnetic resonance spectroscopic imaging with step section histology," *Journal of Urology* **164**, 400–404 (2000).
60. M. A. Haider, T. H. van der Kwast, J. Tanguay, A. J. Evans, A.-T. Hashmi, G. Lockwood and J. Trachtenberg, "Combined T2-weighted and diffusion-weighted MRI for localization of prostate cancer," *AJR. American Journal of Roentgenology* **189**, 323–328 (2007).
61. J. Scheidler, H. Hricak, D. B. Vigneron, K. K. Yu, D. L. Sokolov, L. R. Huang, C. J. Zaloudek, S. J. Nelson, P. R. Carroll and J. Kurhanewicz, "Prostate Cancer:

- Localization with Three-dimensional Proton MR Spectroscopic Imaging—Clinicopathologic Study," *Radiology* **213**, 473–480 (1999).
62. H. Hricak, L. Wang, D. C. Wei, F. V. Coakley, O. Akin, V. E. Reuter, M. Gonen, M. W. Kattan, C. N. Onyebuchi and P. T. Scardino, "The role of preoperative endorectal magnetic resonance imaging in the decision regarding whether to preserve or resect neurovascular bundles during radical retropubic prostatectomy," *Cancer* **100**, 2655–2663 (2004).
  63. A. B. Rosenkrantz, S. M. Scionti, S. Mendrinos and S. S. Taneja, "Role of MRI in minimally invasive focal ablative therapy for prostate cancer," *AJR. American Journal of Roentgenology* **197**, W90–W96 (2011).
  64. R. A. Heesakkers, A. M. Hövels, G. J. Jager, H. van den Bosch, J. A. Witjes, H. P. Raat, J. L. Severens, E. M. Adang, C. H. van der Kaa, J. J. Fütterer and J. Barentsz, "MRI with a lymph-node-specific contrast agent as an alternative to CT scan and lymph-node dissection in patients with prostate cancer: a prospective multicohort study," *Lancet Oncology* **9**, 850–856 (2008).
  65. B. Turkbey, H. Mani, V. Shah, A. R. Rastinehad, M. Bernardo, T. Pohida, Y. Pang, D. Daar, C. Benjamin, Y. L. McKinney, H. Trivedi, C. Chua, G. Bratslavsky, J. H. Shih, W. M. Linehan, M. J. Merino, P. L. Choyke and P. A. Pinto, "Multiparametric 3T prostate magnetic resonance imaging to detect cancer: histopathological correlation using prostatectomy specimens processed in customized magnetic resonance imaging based molds," *Journal of Urology* **186**, 1818–1824 (2011).
  66. J. Thompson, N. Lawrentschuk, M. Frydenberg, L. Thompson and P. Stricker, "The role of magnetic resonance imaging in the diagnosis and management of prostate cancer," *BJU International* **112 Suppl 2**, 6–20 (2013).
  67. M. de Rooij, E. H. Hamoen, J. J. Fütterer, J. O. Barentsz and M. M. Rovers, "Accuracy of Multiparametric MRI for Prostate Cancer Detection: A Meta-Analysis," *AJR. American Journal of Roentgenology* **202**, 343–351 (2014).
  68. F. G. Claus, H. Hricak and R. R. Hattery, "Pretreatment Evaluation of Prostate Cancer: Role of MR Imaging and 1H MR Spectroscopy," *Radiographics* **24**, S167–S180 (2004).
  69. D. L. Langer, T. H. van der Kwast, A. J. Evans, L. Sun, M. J. Yaffe, J. Trachtenberg and M. A. Haider, "Intermixed Normal Tissue within Prostate Cancer: Effect on MR Imaging Measurements of Apparent Diffusion Coefficient and T2—Sparse versus Dense Cancers," *Radiology* **249**, 900–908 (2008).
  70. O. Akin, E. Sala, C. S. Moskowitz, K. Kuroiwa, N. M. Ishill, D. Pucar, P. T. Scardino and H. Hricak, "Transition Zone Prostate Cancers: Features, Detection, Localization, and Staging at Endorectal MR Imaging," *Radiology* **239**, 784–792 (2006).

71. L. Wang, Y. Mazaheri, J. Zhang, N. M. Ishill, K. Kuroiwa and H. Hricak, "Assessment of biologic aggressiveness of prostate cancer: correlation of MR signal intensity with Gleason grade after radical prostatectomy," *Radiology* **246**, 168–176 (2008).
72. A. B. Rosenkrantz and S. S. Taneja, "Radiologist, be aware: ten pitfalls that confound the interpretation of multiparametric prostate MRI," *AJR. American Journal of Roentgenology* **202**, 109–120 (2014).
73. G. Murphy, M. Haider, S. Ghai and B. Sreeharsha, "The expanding role of MRI in prostate cancer," *AJR. American Journal of Roentgenology* **201**, 1229–1238 (2013).
74. E. M. Charles-Edwards and N. M. deSouza, "Diffusion-weighted magnetic resonance imaging and its application to cancer," *Cancer Imaging* **6**, 135–143 (2006).
75. H. A. Vargas, O. Akin, T. Franiel, Y. Mazaheri, J. Zheng, C. Moskowitz, K. Udo, J. Eastham and H. Hricak, "Diffusion-weighted endorectal MR imaging at 3 T for prostate cancer: tumor detection and assessment of aggressiveness," *Radiology* **259**, 775–784 (2011).
76. B. Turkbey, V. P. Shah, Y. Pang, M. Bernardo, S. Xu, J. Kruecker, J. Locklin, A. A. Baccala, A. R. Rastinehad, M. J. Merino, J. H. Shih, B. J. Wood, P. A. Pinto and P. L. Choyke, "Is apparent diffusion coefficient associated with clinical risk scores for prostate cancers that are visible on 3-T MR images?," *Radiology* **258**, 488–495 (2011).
77. N. deSouza, S. Riches, N. Vanas, V. Morgan, S. Ashley, C. Fisher, G. Payne and C. Parker, "Diffusion-weighted magnetic resonance imaging: a potential non-invasive marker of tumour aggressiveness in localized prostate cancer," *Clinical Radiology* **63**, 774–782 (2008).
78. T. Hambrock, D. M. Somford, H. J. Huisman, I. M. van Oort, J. A. Witjes, C. A. Hulsbergen-van de Kaa, T. Scheenen and J. O. Barentsz, "Relationship between apparent diffusion coefficients at 3.0-T MR imaging and Gleason grade in peripheral zone prostate cancer," *Radiology* **259**, 453–461 (2011).
79. F. Donato Jr, D. N. Costa, Q. Yuan, N. M. Rofsky, R. E. Lenkinski and I. Pedrosa, "Geometric Distortion in Diffusion-weighted MR Imaging of the Prostate—Contributing Factors and Strategies for Improvement," *Academic Radiology* **21**, 817–823 (2014).
80. C. M. Hoeks, J. O. Barentsz, T. Hambrock, D. Yakar, D. M. Somford, S. W. Heijmink, T. W. Scheenen, P. C. Vos, H. J. Huisman and I. M. van Oort, "Prostate cancer: multiparametric MR imaging for detection, localization, and staging," *Radiology* **261**, 46–66 (2011).

81. S. Verma, B. Turkbey, N. Muradyan, A. Rajesh, F. Cornud, M. A. Haider, P. L. Choyke and M. Harisinghani, "Overview of dynamic contrast-enhanced MRI in prostate cancer diagnosis and management," *AJR. American Journal of Roentgenology* **198**, 1277–1288 (2012).
82. F. A. van Dorsten, M. van der Graaf, M. R. Engelbrecht, G. J. van Leenders, A. Verhofstad, M. Rijpkema, J. J. de la Rosette, J. O. Barentsz and A. Heerschap, "Combined quantitative dynamic contrast-enhanced MR imaging and <sup>1</sup>H MR spectroscopic imaging of human prostate cancer," *Journal of Magnetic Resonance Imaging* **20**, 279–287 (2004).
83. S. Verma, A. Rajesh, J. J. Fütterer, B. Turkbey, T. W. Scheenen, Y. Pang, P. L. Choyke and J. Kurhanewicz, "Prostate MRI and 3D MR spectroscopy: how we do it," *AJR. American Journal of Roentgenology* **194**, 1414–1426 (2010).
84. K. L. Zakian, S. Eberhardt, H. Hricak, A. Shukla-Dave, S. Kleinman, M. Muruganandham, K. Sircar, M. W. Kattan, V. E. Reuter, P. T. Scardino and J. Koutcher, "Transition Zone Prostate Cancer: Metabolic Characteristics at <sup>1</sup>H MR Spectroscopic Imaging—Initial Results," *Radiology* **229**, 241–247 (2003).
85. K. L. Zakian, K. Sircar, H. Hricak, H. N. Chen, A. Shukla-Dave, S. Eberhardt, M. Muruganandham, L. Ebor, M. W. Kattan, V. E. Reuter, P. T. Scardino and J. A. Koutcher, "Correlation of proton MR spectroscopic imaging with Gleason score based on step-section pathologic analysis after radical prostatectomy," *Radiology* **234**, 804–814 (2005).
86. T. Kobus, T. Hambrock, C. A. Hulsbergen-van de Kaa, A. J. Wright, J. O. Barentsz, A. Heerschap and T. W. Scheenen, "In vivo assessment of prostate cancer aggressiveness using magnetic resonance spectroscopic imaging at 3 T with an endorectal coil," *European Urology* **60**, 1074–1080 (2011).
87. D. Hausmann, S. Konstandin, F. Zöllner, S. Haneder, F. Wetterling, A. Nagel, D. Dinter, S. Schönberg and L. Schad, "Sodium Imaging of the Prostate at 3T," in *Proceedings of International Society for Magnetic Resonance in Medicine*, (Melbourne, Australia, 2012).
88. J. Near and R. Bartha, "Quantitative sodium MRI of the mouse prostate," *Magnetic Resonance in Medicine* **63**, 822–827 (2010).
89. M. J. Albers, R. Bok, A. P. Chen, C. H. Cunningham, M. L. Zierhut, V. Y. Zhang, S. J. Kohler, J. Tropp, R. E. Hurd, Y.-F. Yen, S. J. Nelson, D. B. Vigneron and J. Kurhanewicz, "Hyperpolarized <sup>13</sup>C Lactate, Pyruvate, and Alanine: Noninvasive Biomarkers for Prostate Cancer Detection and Grading," *Cancer Research* **68**, 8607–8615 (2008).

90. S. J. Nelson, J. Kurhanewicz, D. B. Vigneron, P. E. Z. Larson, A. L. Harzstark, M. Ferrone, M. van Criekinge, J. W. Chang, R. Bok, I. Park, G. Reed, L. Carvajal, E. J. Small, P. Munster, V. K. Weinberg, J. H. Ardenkjaer-Larsen, A. P. Chen, R. E. Hurd, L.-I. Odegardstuen, F. J. Robb, J. Tropp and J. A. Murray, "Metabolic Imaging of Patients with Prostate Cancer Using Hyperpolarized [1-13C]Pyruvate," *Science Translational Medicine* **5**, 198ra108 (2013).
91. A. B. Rosenkrantz, H. Chandarana, N. Hindman, F.-M. Deng, J. S. Babb, S. S. Taneja and C. Geppert, "Computed diffusion-weighted imaging of the prostate at 3 T: impact on image quality and tumour detection," *European Radiology* **23**, 3170–3177 (2013).
92. E. Panagiotaki, R. W. Chan, N. Dikaio, H. Ahmed, D. Atkinson, S. Punwani and D. C. Alexander, "Microstructural characterisation of normal and malignant human prostate tissue with VERDICT-MRI," in *Proceedings of International Society of Magnetic Resonance in Medicine*, (Milan, Italy, 2014).
93. W. T. Hrinivich, E. Gibson, M. Gaed, J. A. Gómez, M. Moussa, C. A. McKenzie, G. S. Bauman, A. D. Ward, A. Fenster and E. Wong, "A dimensionless dynamic contrast enhanced MRI parameter for intra-prostatic tumour target volume delineation: initial comparison with histology," in *Proceedings of SPIE Medical Imaging, Vol. 9036* (San Diego, USA, 2014), pp. 90362I–90367.
94. M. Norberg, L. Egevad, L. Holmberg, P. Sparen, B. Norlen and C. Busch, "The sextant protocol for ultrasound-guided core biopsies of the prostate underestimates the presence of cancer," *Urology* **50**, 562–566 (1997).
95. J. Sauvain, P. Palascak, D. Bourscheid, C. Chabi, A. Atassi, J. Bremon and R. Palascak, "Value of power doppler and 3D vascular sonography as a method for diagnosis and staging of prostate cancer," *European Urology* **44**, 21–31 (2003).
96. R. Clements, "The role of transrectal ultrasound in diagnosing prostate cancer," *Current Urology Reports* **3**, 194–200 (2002).
97. E. K. Outwater and J. L. Montilla-Soler, "Imaging of prostate carcinoma," *Cancer Control* **20**, 161–176 (2013).
98. P. Oehr and K. Bouchelouche, "Imaging of prostate cancer," *Current Opinion in Oncology* **19**, 259-264 (2007).
99. S. N. Reske, N. M. Blumstein, B. Neumaier, H. W. Gottfried, F. Finsterbusch, D. Kocot, P. Moller, G. Glatting and S. Perner, "Imaging prostate cancer with 11C-choline PET/CT," *Journal of Nuclear Medicine* **47**, 1249-1254 (2006).
100. E. Sutinen, M. Nurmi, A. Roivainen, M. Varpula, T. Tolvanen, P. Lehtikoinen and H. Minn, "Kinetics of [11C]choline uptake in prostate cancer: a PET study," *European Journal of Nuclear Medicine and Molecular Imaging* **31**, 317-324 (2004).



101. C. Testa, R. Schiavina, R. Lodi, E. Salizzoni, B. Corti, M. Farsad, J. Kurhanewicz, F. Manferrari, E. Brunocilla, C. Tonon and others, "Prostate Cancer: Sextant Localization with MR Imaging, MR Spectroscopy, and 11C-Choline PET/CT 1," *Radiology* **244**, 797–806 (2007).
102. M. Pinkawa, C. Attieh, M. D. Piroth, R. Holy, S. Nussen, J. Klotz, R. Hawickhorst, W. Schäfer and M. J. Eble, "Dose-escalation using intensity-modulated radiotherapy for prostate cancer – Evaluation of the dose distribution with and without 18F-choline PET-CT detected simultaneous integrated boost," *Radiotherapy and Oncology* **93**, 213–219 (2009).
103. H. Wang, H. Veas, R. Miralbell, M. Wissmeyer, C. Steiner, O. Ratib, S. Senthamizhchelvan and H. Zaidi, "18F-fluorocholine PET-guided target volume delineation techniques for partial prostate re-irradiation in local recurrent prostate cancer," *Radiotherapy and Oncology* **93**, 220–225 (2009).
104. K. Kitajima, R. C. Murphy and M. A. Nathan, "Choline PET/CT for imaging prostate cancer: an update," *Annals of Nuclear Medicine* **27**, 581–591 (2013).
105. M. B. Amin, D. J. Grignon, P. A. Humphrey and J. R. Srigley, *Gleason grading of prostate cancer: a contemporary approach*. (Lippincott Williams & Wilkins, 2004).
106. M. Anwar, A. C. Westphalen, A. J. Jung, S. M. Noworolski, J. P. Simko, J. Kurhanewicz, M. Roach, 3rd, P. R. Carroll and F. V. Coakley, "Role of endorectal MR imaging and MR spectroscopic imaging in defining treatable intraprostatic tumor foci in prostate cancer: quantitative analysis of imaging contour compared to whole-mount histopathology," *Radiotherapy and Oncology* **110**, 303–308 (2014).
107. L. H. Chen, H. Ho, R. Lazaro, C. H. Thng, J. Yuen, W. S. Ng and C. Cheng, "Optimum slicing of radical prostatectomy specimens for correlation between histopathology and medical images," *International Journal of Computer Assisted Radiology and Surgery* **5**, 471–487 (2010).
108. B. Drew, E. C. Jones, S. Reinsberg, A. C. Yung, S. L. Goldenberg and P. Kozlowski, "Device for sectioning prostatectomy specimens to facilitate comparison between histology and in vivo MRI," *Journal of Magnetic Resonance Imaging* **32**, 992–996 (2010).
109. G. Groenendaal, M. R. Moman, J. G. Korpelaar, P. J. van Diest, M. van Vulpen, M. E. Philippens and U. A. van der Heide, "Validation of functional imaging with pathology for tumor delineation in the prostate," *Radiotherapy and Oncology* **94**, 145–150 (2010).
110. D. Simon, R. V. O'Toole, M. Blackwell, F. Morgan, A. M. Digioia and T. Kanade, "Accuracy validation in image-guided orthopaedic surgery," in *Proceedings of*

- International Symposium on Medical Robotics and Computer Assisted Surgery*, (1995), pp. 185–192.
111. M. H. Loew and C. E. Rodriguez-Carranza, "Technical issues in multimodality medical image registration," in *Proceedings of IEEE Computer-Based Medical Systems*, (1998), pp. 2–7.
  112. T. Rohlfing, "Image similarity and tissue overlaps as surrogates for image registration accuracy: widely used but unreliable," *IEEE Transactions on Medical Imaging* **31**, 153–163 (2012).
  113. I. Ocak, M. Bernardo, G. Metzger, T. Barrett, P. Pinto, P. S. Albert and P. L. Choyke, "Dynamic contrast-enhanced MRI of prostate cancer at 3 T: a study of pharmacokinetic parameters," *AJR. American Journal of Roentgenology* **189**, W192–W201 (2007).
  114. K. Katahira, T. Takahara, T. C. Kwee, S. Oda, Y. Suzuki, S. Morishita, K. Kitani, Y. Hamada, M. Kitaoka and Y. Yamashita, "Ultra-high-b-value diffusion-weighted MR imaging for the detection of prostate cancer: evaluation in 201 cases with histopathological correlation," *European Radiology* **21**, 188–196 (2011).
  115. M. Debois, R. Oyen, F. Maes, G. Verswijvel, G. Gatti, H. Bosmans, M. Feron, E. Bellon, G. Kutcher, H. Van Poppel and L. Vanuytsel, "The contribution of magnetic resonance imaging to the three-dimensional treatment planning of localized prostate cancer," *International Journal of Radiation Oncology\*Biology\*Physics* **45**, 857–865 (1999).
  116. C. Lutgendorf-Caucig, I. Fotina, M. Stock, R. Potter, G. Goldner and D. Georg, "Feasibility of CBCT-based target and normal structure delineation in prostate cancer radiotherapy: multi-observer and image multi-modality study," *Radiotherapy and Oncology* **98**, 154–161 (2011).
  117. H. C. Rischke, U. Nestle, T. Fechter, C. Doll, N. Volegova-Neher, K. Henne, J. Scholber, S. Knippen, S. Kirste, A. L. Grosu and C. A. Jilg, "3 Tesla multiparametric MRI for GTV-definition of dominant intraprostatic lesions in patients with prostate cancer--an interobserver variability study," *Radiation Oncology* **8**, 183 (2013).
  118. G. Penney, A. Varnavas, N. Dastur and T. Carrell, "An image-guided surgery system to aid endovascular treatment of complex aortic aneurysms: description and initial clinical experience," in *Proceedings of Information Processing in Computer-Assisted Interventions*, Vol. 6689 (Springer Berlin / Heidelberg, 2011), pp. 13–24.
  119. W. van de Ven, G. Litjens, J. Barentsz, T. Hambroek and H. J. Huisman, "Required accuracy of MR-US registration for prostate biopsies," in *Proceedings of Prostate Cancer Imaging. Image Analysis and Image-Guided Interventions*, Vol. 6963 (Springer Berlin / Heidelberg, Toronto, Canada, 2011), pp. 92–99.

120. A. D. Ward, C. Crukley, C. A. McKenzie, J. Montreuil, E. Gibson, C. Romagnoli, J. A. Gómez, M. Moussa, J. Chin, G. Bauman and A. Fenster, "Prostate: registration of digital histopathologic images to in vivo MR images acquired by using endorectal receive coil," *Radiology* **263**, 856–864 (2012).
121. L. W. Turnbull, D. L. Buckley, L. S. Turnbull, G. P. Liney and A. J. Knowles, "Differentiation of prostatic carcinoma and benign prostatic hyperplasia: correlation between dynamic Gd-DTPA-enhanced MR imaging and histopathology," *Journal of Magnetic Resonance Imaging* **9**, 311–316 (1999).
122. M. R. Engelbrecht, H. J. Huisman, R. J. F. Laheij, G. J. Jager, G. J. L. H. van Leenders, C. A. Hulsbergen-Van De Kaa, J. J. M. C. H. de la Rosette, J. G. Blickman and J. O. Barentsz, "Discrimination of prostate cancer from normal peripheral zone and central gland tissue by using dynamic contrast-enhanced MR imaging," *Radiology* **229**, 248–254 (2003).
123. O. Rouvière, A. Raudrant, R. Ecochard, C. Colin-Pangaud, C. Pasquiou, R. Bouvier, J. M. Marechal and D. Lyonnet, "Characterization of time-enhancement curves of benign and malignant prostate tissue at dynamic MR imaging," *European Radiology* **13**, 931–942 (2003).
124. V. Shah, T. Pohida, B. Turkbey, H. Mani, M. Merino, P. A. Pinto, P. Choyke and M. Bernardo, "A method for correlating in vivo prostate magnetic resonance imaging and histopathology using individualized magnetic resonance-based molds," *Review of Scientific Instruments* **80**, 104301–104306 (2009).
125. S. G. Jhavar, C. Fisher, A. Jackson, S. A. Reinsberg, N. Dennis, A. Falconer, D. Dearnaley, S. E. Edwards, S. M. Edwards, M. O. Leach, C. Cummings, T. Christmas, A. Thompson, C. Woodhouse, S. Sandhu, C. S. Cooper and R. A. Eeles, "Processing of radical prostatectomy specimens for correlation of data from histopathological, molecular biological, and radiological studies: a new whole organ technique," *Journal of Clinical Pathology* **58**, 504–508 (2005).
126. A. S. Jackson, S. A. Reinsberg, S. A. Sohaib, E. M. Charles-Edwards, S. Jhavar, T. J. Christmas, A. C. Thompson, M. J. Bailey, C. M. Corbishley, C. Fisher, M. O. Leach and D. P. Dearnaley, "Dynamic contrast-enhanced MRI for prostate cancer localization," *British Journal of Radiology* **82**, 148–156 (2009).
127. O. Rouvière, C. Reynolds, T. Hulshizer, P. Rossman, Y. Le, J. P. Felmlee and R. L. Ehman, "MR histological correlation: a method for cutting specimens along the imaging plane in animal or ex vivo experiments," *Journal of Magnetic Resonance Imaging* **23**, 60–69 (2006).
128. J. Walz, A. L. Burnett, A. J. Costello, J. A. Eastham, M. Graefen, B. Guillonnet, M. Menon, F. Montorsi, R. P. Myers, B. Rocco and A. Villers, "A Critical Analysis

- of the Current Knowledge of Surgical Anatomy Related to Optimization of Cancer Control and Preservation of Continence and Erection in Candidates for Radical Prostatectomy," *European Urology* **57**, 179-192 (2010).
129. C. Hughes, O. Rouvière, F. Mege-Lechevallier, R. Souchon and R. Prost, "Robust alignment of prostate histology slices with quantified accuracy," *IEEE Transactions on Biomedical Engineering* **60**, 281–291 (2012).
  130. S. Bart, P. Mozer, P. Hemar, G. Lenaour, E. Comperat, R. Renaud-Penna, E. Chartier-Kastler and J. Troccaz, "MRI-histology registration in prostate cancer," in *Proceedings of Surgetica*, (2005).
  131. L. S. Taylor, B. C. Porter, G. Nadasdy, P. A. di Sant'Agnese, D. Pasternack, Z. Wu, R. B. Baggs, D. J. Rubens and K. J. Parker, "Three-dimensional registration of prostate images from histology and ultrasound," *Ultrasound in Medicine and Biology* **30**, 161–168 (2004).
  132. J. Xu, P. A. Humphrey, A. S. Kibel, A. Z. Snyder, V. R. Narra, J. J. Ackerman and S. K. Song, "Magnetic resonance diffusion characteristics of histologically defined prostate cancer in humans," *Magnetic Resonance in Medicine* **61**, 842–850 (2009).
  133. C. Orczyk, A. Mikheev, A. Rosenkrantz, J. Melamed, S. S. Taneja and H. Rusinek, "Imaging of prostate cancer: a platform for 3D co-registration of in-vivo MRI ex-vivo MRI and pathology," in *Proceedings of SPIE Medical Imaging, Vol. 8316* (2012), pp. 83162M.
  134. P. Patel, J. Chappelow, J. Tomaszewski, M. D. Feldman, M. Rosen, N. Shih and A. Madabhushi, "Spatially weighted mutual information (SWMI) for registration of digitally reconstructed ex vivo whole mount histology and in vivo prostate MRI," in *Proceedings of IEEE Engineering in Medicine & Biology Society*, (2011), pp. 6269–6272.
  135. G. Nir and S. E. Salcudean, "Registration of whole-mount histology and tomography of the prostate using particle filtering," in *Proceedings of SPIE Medical Imaging*, (2013), pp. 86760E.
  136. J. Chappelow, B. N. Bloch, N. Rofsky, E. Genega, R. Lenkinski, W. DeWolf, S. Viswanath and A. Madabhushi, "COLLINARUS: collection of image-derived non-linear attributes for registration using splines," in *Proceedings of SPIE Medical Imaging*, (2009), pp. 72592N.
  137. S. B. Prabu, R. Toth and A. Madabhushi, "A statistical deformation model (SDM) based regularizer for non-rigid image registration: application to registration of multimodal prostate MRI and histology," in *Proceedings of SPIE Medical Imaging*, (2013), pp. 86760C–86760C.

138. J. Chappelow, B. N. Bloch, N. Rofsky, E. Genega, R. Lenkinski, W. DeWolf and A. Madabhushi, "Elastic registration of multimodal prostate MRI and histology via multiattribute combined mutual information," *Medical Physics* **38**, 2005–2018 (2011).
139. Y. Zhan, Y. Ou, M. Feldman, J. Tomaszewski, C. Davatzikos and D. Shen, "Registering histologic and MR images of prostate for image-based cancer detection," *Academic Radiology* **14**, 1367–1381 (2007).
140. Y. Ou, D. Shen, M. Feldman, J. Tomaszewski and C. Davatzikos, "Non-rigid registration between histological and MR images of the prostate: a joint segmentation and registration framework," in *Proceedings of Computer Vision and Pattern Recognition Workshops*, (2009), pp. 125–132.
141. J. Chappelow, A. Madabhushi, M. Rosen, J. Tomaszewski and M. Feldman, "Multimodal image registration of ex vivo 4 Tesla MRI with whole mount histology for prostate cancer detection," in *Proceedings of SPIE Medical Imaging, Vol. 6512* (2007), pp. 65121S
142. G. Nir, R. S. Sahebjavaher, P. Kozlowski, S. D. Chang and R. Sinkus, "Model-based registration of ex vivo and in vivo MRI of the prostate using elastography," *IEEE Transactions on Medical Imaging* **32**, 1068–1080 (2013).
143. C. Meyer, B. Ma, L. P. Kunju, M. Davenport and M. Piert, "Challenges in accurate registration of 3-D medical imaging and histopathology in primary prostate cancer," *European Journal of Nuclear Medicine and Molecular Imaging* **40 Suppl 1**, S72-78 (2013).

## Chapter 2.

# **The impact of registration accuracy on imaging validation study design: a novel statistical power calculation<sup>†</sup>**

### 2.1 Introduction

Imaging evaluation with respect to an accepted reference standard imaging modality (e.g. on pathologist-contoured digital pathology images) may be a useful surrogate for clinical evaluation due to the resulting lower costs and shorter timeframes. Some aspects of these evaluations require the registration of images from the medical imaging modality being studied (henceforth, *study images*) to the images from the accepted reference standard modality (henceforth, *reference images*). Because of trade-offs associated with achieving higher accuracy [1], it is important to identify the maximum acceptable level of registration error. This threshold is application-dependent [2], and establishing application-specific thresholds for maximum acceptable error has been identified as a key challenge in the field [1, 2], but, to the best of our knowledge, criteria for acceptable registration errors have not been identified for imaging validation studies.

---

<sup>†</sup>A version of this chapter has been published: E. Gibson, A. Fenster, A. D. Ward, “The impact of registration accuracy on imaging validation study design: a novel statistical power calculation.” *Medical Image Analysis* 17:7 (2013).

In this work, we focus on developing such a criterion for a particular class of imaging validation studies: preliminary studies of the utility of an imaging modality for disease localization that test whether any observed differences in the mean image intensities (or other derived quantities) attributed to the presence of the disease (e.g. cancerous tissue) are significant, or whether they are merely due to chance. Measurement of image intensities in regions containing disease requires the localization of these regions on study images, which in turn requires the registration of the study images to the reference images, wherein the disease can be acceptably localized. An ideal (zero error) registration maps each ground truth delineation of disease features of interest onto the true region of interest (henceforth,  $R_i$  for the  $i$ -th such region, as depicted in Figure 2.1a) on the study image. A non-ideal registration ( $> \text{zero error}$ ) maps the delineation onto a region which we denote  $\tilde{R}_i$ . Registration error is ideally quantified using a measurement of distance, such as the target registration error (TRE) [3-5]. However, because our analysis depends on aggregated measurements of image intensities within regions containing disease (i.e.  $R_i$ ), but our measurements, due to non-ideal registrations, are instead aggregated measurements taken from within the sampling region  $\tilde{R}_i$ , the fidelity of the *regional overlap* between  $R_i$  and  $\tilde{R}_i$  is paramount, and is used as a surrogate measure for registration accuracy. Our measure of this fidelity is the fractional overlap,  $f_i = |\tilde{R}_i \cap R_i|/|\tilde{R}_i|$ , where  $|R|$  is the number of voxels in the region  $R$ . Fractional overlap, which ranges from 0 to 1, serves as our surrogate measure of registration accuracy throughout this chapter, with  $1 - f_i$  serving as our surrogate measure of registration error. Mapping errors that result in smaller overlap may lead to larger required number of subjects to achieve a given minimum detectable difference (MDD) on

imaging between pathologic and benign regions. This observation leads to three key questions affecting study design. **(1) What is the maximum acceptable registration error?** Specifically, for a fixed number of subjects and specified MDD, what is the maximum acceptable image registration error? **(2) How many subjects are needed?** Specifically, for a quantified image registration error and specified MDD, what is the required number of subjects? **(3) What is the minimum detectable difference?** Specifically, for a fixed number of subjects and quantified image registration error, what is the MDD?

The common concept connecting these questions is statistical power, a measure that describes the probability of a study finding a statistically significant result when there is an underlying difference to be found. Statistical power and its relationship to study designs have been actively investigated for nearly a century [6]. Calculations of statistical power depends on many factors, including the number of subjects in the study, the acceptable error rates, the size of the underlying difference, the characteristics of the population being analyzed, the statistical test [7], the types of hypotheses being tested [8, 9], and the sampling methods used [10]. This statistical power is commonly expressed in the form of a sample size calculation that relates how many subjects to recruit for a particular study design or an MDD calculation that relates how small a difference can be detected for a particular study design. Such statistical power calculations have been derived for a wide variety of study designs; however, they have not been derived for studies where the measurements rely on imperfect image registration, and the registration error is an additional factor affecting statistical power.



In the context of studies to determine whether focal disease affects intensity on study images, we propose that the acceptable registration error can be defined relative to the studies' statistical power. Thus, in this chapter, we provide a derivation that yields the relationship between image registration error, number of subjects, and MDD, where image registration is used to determine whether the presence of particular anatomy, pathology or other features of interest in the underlying tissue is reflected in a change in the mean intensity of study image voxels corresponding to the features of interest. The derivation of a statistical power calculation that incorporates uncertainty due to registration error yields a set of three equations that can be used to answer the three questions enumerated above.

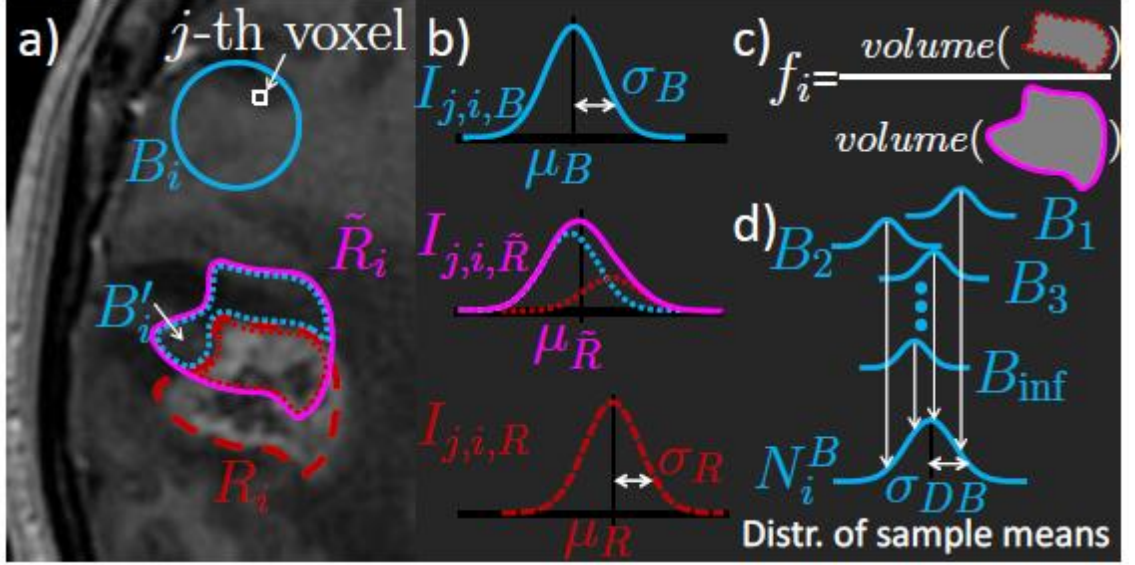
A preliminary version of this work [11] derived these relationships for a simplified image data model that cannot account for the intervoxel correlations that are pervasive in medical imaging [12]. This chapter generalizes the previous work to account for a more realistic data model, provides a more rigorous derivation of the power calculation formulae, expands the scope of the validation by measuring sensitivity to a wider range of violations of model assumptions, expands the testing of the relationships by measuring error in the number of subjects, MDD and required registration accuracy predictions, and demonstrates the application of the theory to a case study of an imaging validation project underway at our institution.

The remainder of this chapter outlines the derivation of an approximate relationship between fractional overlap and statistical power for one study design (Section 2.2), describes simulations used to validate components of the derivation (Section 2.3), presents the results of the simulations (Section 2.4) and discusses the

implications of these relationships and the application of these relationships in a case study (Section 2.5). Appendix A contains further mathematical details of the derivation.

## 2.2 Statistical power calculations in imaging validation studies

A statistical power calculation, such as the sample size calculation, relates the probability of detecting a true positive finding to other parameters of the study design under a probabilistic model of the populations being studied [13]. This relationship is based on the distribution of the test statistic, which depends on the type of statistical analysis used in the study, and the assumptions made in the data model. An overview of the data model used for this work is illustrated in Figure 2.1.



**Figure 2.1:** Overview of the data model. (a) On this axial magnetic resonance image of a brain, the true region of interest  $R_i$ , in this case a high contrast glioma for illustration purposes, is shown in red. Due to registration error, the sampling region  $\tilde{R}_i$  (shown in purple) partially overlaps the surrounding background tissue  $B'_i$  (shown in cyan). Because the background sampling region  $B_i$  is surrounded by other background tissue, we do not model registration error for this region. (b) The distributions of voxel intensities in regions  $B_i$  and  $B'_i$  (top),  $\tilde{R}_i$  (middle) and  $R_i$  (bottom). Because  $\tilde{R}_i$  contains samples from  $B'_i$  and  $R_i$ , its voxel intensity distribution  $I_{j,i,\tilde{R}}$  is a mixture of the distributions for regions  $B'_i$  ( $I_{j,i,B}$ ) and  $R_i$  ( $I_{j,i,R}$ ) (dotted lines). (c) The statistical properties of  $I_{j,i,\tilde{R}}$  are modulated by registration error, because it affects the proportion of samples from the component regions. By quantifying registration error in terms of the volume of  $\tilde{R}_i \cap R_i$  as a fraction of the volume of  $\tilde{R}_i$ , the fractional overlap, the model of its effect on the statistical properties of  $I_{j,i,\tilde{R}}$  is simplified. (d) The correlation of voxel intensities within regions  $B_i$  (or similarly  $R_i$ ) can be expressed by decomposing the total variance  $\sigma_B^2$  into two components, one reflecting the variation of the region means from region to region  $N_i^B(0, \sigma_{DB}^2)$ , and the other residual variation from voxel to voxel within a single region.

### 2.2.1 Specification of statistical analysis and data model assumptions

In this chapter, we derive and evaluate statistical power calculations for a specific statistical analysis: comparing the mean of a pool of samples drawn from  $n$  identified sampling regions  $\langle \tilde{R}_i | i = 1..n \rangle$  to the mean of another pool drawn from  $n$  background regions  $\langle B_i | i = 1..n \rangle$  using an unpaired Student's  $t$ -test on the sample intensities, with an

equal number of samples  $v$  taken from each region. In this chapter, we use a data model that makes the following assumptions:

- I. intensities of voxels containing the *features of interest*  $I_{j,i,R}$  and background  $I_{j,i,B}$  are jointly Gaussian distributed random variables that are correlated within each region and independent between regions (i.e.  $I_{j,i,R} \sim \mu_R + N_i^R(0, \sigma_{DR}^2) + N_{ij}^R(0, \sigma_{IR}^2)$  and  $I_{j,i,B} \sim \mu_B + N_i^B(0, \sigma_{DB}^2) + N_{ij}^B(0, \sigma_{IB}^2)$ , respectively, where  $N_i$  is a Gaussian random effect on the  $i$ -th region and  $N_{ij}$  is a Gaussian random effect on the  $j$ -th voxel in the  $i$ -th region);
- II. the interregion variances are equal for regions  $R_i$  and regions  $B_i$  (i.e.  $\sigma_{DR}^2 = \sigma_{DB}^2$ );
- III. when the regions  $R_i$  and  $\tilde{R}_i$  are not aligned, due to registration error, the proportions of samples that fall within  $R_i \cap \tilde{R}_i$  (termed the *fractional overlap* hereafter; see Figure 2.1c) are independently and identically distributed from a distribution  $f_i \sim F$ , and independent from the intensities;
- IV. the regions  $R_i$  and  $B_i$  are surrounded by background tissue, such that when there is registration error,  $\tilde{R}_i$  may contain a region of background tissue (denoted  $B'_i$ ), but the sampling region for background still contains only background. Intensities of voxels in regions  $B'_i$  are correlated within  $B'_i$ , but uncorrelated with voxels in any  $R_i$  or  $B_i$ ;
- V. the number of regions  $\tilde{R}_i$  is large enough that the distribution of the mean of the pooled samples approximates a Gaussian distribution; and
- VI. the number of samples,  $v$ , from each region is constant and is large enough that  $vf_i$  can be approximated as an integer.

### 2.2.2 Derivation of the statistical power formulae incorporating registration error

When there is no registration error, the test statistic

$$\frac{\overline{I_{J,l,R}} - \overline{I_{J,l,B}}}{\sqrt{\frac{Var(\overline{I_{l,R}}) + Var(\overline{I_{l,B}})}{n}}}, \quad 2.1$$

which is used when the variance and intraclass correlation are unknown, has a  $t$ -distribution with  $2n - 2$  degrees of freedom, where  $\overline{I_{J,l,R}}$  and  $\overline{I_{J,l,B}}$  are the sample means and  $\overline{I_{l,R}}$  and  $\overline{I_{l,B}}$  are the region means of the  $i$ -th regions. Using the terminology introduced above in Eq. 2.1 and letting  $\mu_d = (\mu_R - \mu_B)|_{min}$  be the minimum detectable difference for a two-sample  $t$ -test, we have

$$\mu_d = T \sqrt{\frac{\sigma_{DR}^2 + \sigma_{DB}^2}{n} + \frac{\sigma_{IR}^2 + \sigma_{IB}^2}{nv}}, \quad 2.2$$

where  $T$  is a statistical threshold  $t_{\{\alpha\{2\},2n-2\}} + t_{\{\beta\{1\},2n-2\}}$ , where  $t_{\{\alpha\{2\},2n-2\}}$  and  $t_{\{\beta\{1\},2n-2\}}$  are two- and one-tailed critical values taken from the inverse cumulative distribution function of the  $t$ -distribution with  $2n - 2$  degrees of freedom, constraining type I error to  $\alpha$  and type II error to  $\beta$ . This formula holds for cases where the samples are from Gaussian distributions and there is equal variance and equal correlation within each class. The first two of the assumptions in our model are derived from this. First, to match the distributions from the classical formula, our model assumes correlated Gaussian intensity distributions, although in practice, image intensities may have varying distributions and may be truncated by the imaging systems. Second, the classical formula assumes equal intra- and interregion variances; however, in the presence of correlation

the sensitivity to unequal intraregion variances is lower than the sensitivity to unequal interregion variances [14]. Based on these observations, only the equality of the interregion variances was assumed in our model. Equation 2.2 accounts for the widening of the standard error of the mean in the presence of correlation between the samples in each class. An illustrative example of such correlation is an image whose resolution is much lower than the voxel spacing. While the variance of a randomly selected voxel from a randomly selected tumor from such images would be  $\sigma_R^2$  (i.e.  $\sigma_{DR}^2 + \sigma_{IR}^2$ ), estimating the variance using multiple measurements within a smaller number of tumor regions would yield an artificially low variance, and using this variance estimate in the calculation of the standard error of the mean  $\sigma_R/\sqrt{vn}$  would underestimate the standard error of the mean. The design factor [15] corrects for this, such that the true standard error of the mean is represented in the power calculation.

For the purpose of the derivation, we rearrange this classical power formula as

$$\mu_R - \mu_B = T \sqrt{\frac{\sigma_{IR}^2 + v\sigma_{DR}^2}{nv} + \frac{\sigma_{IB}^2 + v\sigma_{DB}^2}{nv}}, \quad 2.3$$

to isolate  $(\sigma_{IR}^2 + v\sigma_{DR}^2)/(nv)$ , the expression for the variance of the mean of  $nv$  cluster-randomized samples from the  $R_i$  regions. Note that  $\mu_d$  is the minimum detectable difference between the means of the populations being compared ( $\mu_R$  and  $\mu_B$ ).

When there is misregistration of the  $i$ -th region,  $\tilde{R}_i$  may overlap with the background region  $B'_i$  that surrounds  $R_i$  (illustrated in Figure 2.1a) such that the pool of voxels sampled from  $\tilde{R}_i$  contains samples from  $B'_i$ . If the intensity distributions  $I_{j,i,B'}$  and  $I_{j,i,R}$  differ, then samples from  $\tilde{R}_i$  do not represent the same population as samples from  $R_i$ , and Equation 2.3 does not correctly describe the statistical tests being performed. At a

high level, to derive the statistical relationships accounting for registration error, we first characterize the population of samples from  $\tilde{R}_i$ , then model the statistical comparison of samples from  $\tilde{R}_i$  and  $B$ , and finally relate this model back to the population parameters of the original data model. This is described in detail as follows.

The population of samples from  $\tilde{R}_i$  is a mixture of samples from  $R_i$  and samples from  $B'_i$ , as illustrated in Figure 2.1b. The proportion of samples that contain the feature of interest can be quantified as the fractional overlap  $f_i = |\tilde{R}_i \cap R_i|/|\tilde{R}_i|$  (illustrated in Figure 2.1c). Ignoring partial volumes,  $\tilde{R}_i$  will have  $\text{round}(vf_i)$  samples from  $R_i$  and  $v - \text{round}(vf_i)$  samples from  $B'_i$ . Each of the samples is a Gaussian random variable with a distribution that depends on whether the sample is from  $R_i$  or from  $B'_i$ .

Noting that Equation 2.3 incorporates expressions for the mean of  $I_{j,i,R}$  and the variance of the mean of  $nv$  cluster-randomized samples from  $I_{j,i,R}$ , we derive these same parameters for the population of samples from  $\tilde{R}_i$ . Because  $I_{j,i,\tilde{R}}$  is a mixture of two component distributions, the mean of the population  $\mu_{\tilde{R}}$  is a weighted mean of the means of the component distributions weighted by the number of samples from each component. Since the number of samples depends on the fractional overlap, we marginalize this weighted mean over the distribution of fractional overlap:

$$\mu_{\tilde{R}} = \int \frac{\sum_i v f_i \mu_R + \sum_i v (1 - f_i) \mu_B}{nv} \prod_i p(F_i = f_i) df = (\mu_R - \mu_B) \mu_F - \mu_B \quad 2.4$$

The variance of the mean of  $nv$  cluster-randomized samples from  $\tilde{R}$  can be expressed as

$$\sigma_{\tilde{R},nv}^2 = \frac{\mu_d^2 \sigma_F^2 + \sigma_D^2 (\mu_F^2 + \sigma_F^2) + \left( \frac{\sigma_{IR}^2}{v} - \sigma_{B,2}^2 \right) \mu_F + \sigma_{B,1}^2}{n} \quad 2.5$$

where  $\sigma_D^2 = \sigma_{DR}^2 + \sigma_{DB}^2$ ,  $\sigma_{B,1}^2 = \frac{\sigma_{IB}^2}{v} + \sigma_{DB}^2$  and  $\sigma_{B,2}^2 = \frac{\sigma_{IB}^2}{v} + 2\sigma_{DB}^2$ . The derivation of this expression is given in Appendix A.

To model the statistical comparison of samples from regions  $\tilde{R}_i$  and  $B_i$ , we incorporate  $\mu_{\tilde{R}}$  and  $\sigma_{\tilde{R},nv}^2$  into the classical formula in Equation 2.3, by substituting  $\mu_{\tilde{R}}$  for  $\mu_R$ , and substituting  $\sigma_{\tilde{R},nv}^2$  (the variance of the mean of  $nv$  cluster-randomized samples from regions  $\tilde{R}_i$ ) for  $\frac{\sigma_{IR}^2 + v\sigma_{DR}^2}{nv}$  (the variance of the mean of  $nv$  cluster-randomized samples from regions  $R_i$ ). Although the classical formula in Equation 2.3 is an exact statistical model, the formula resulting from these substitutions describes an approximate relationship, as the population of samples from  $\tilde{R}_i$  is not, in general, Gaussian, and so the distribution of the test statistic is not, in general, a  $t$ -distribution.

This relationship can be related to the population parameters of  $R_i$  and  $B_i$  by noting that  $\mu_{\tilde{R}}$  is a function of the difference in population means  $\mu_d = \mu_R - \mu_B$ . This expression can be solved for  $\mu_F$ ,  $n$ , and  $\mu_d$ , as given in the power calculation formulae in equations 2.6, 2.7 and 2.8, respectively:

$$\mu_F = \frac{\sigma_{B,2}^2 - \sigma_{IR}^2/v \pm \sqrt{\left(\frac{\sigma_{IR}^2}{v} - \sigma_{B,2}^2\right)^2 - 4\left((\sigma_D^2 + \mu_d^2)\sigma_F^2 + 2\sigma_{B,1}^2\right)\left(\sigma_D^2 - \frac{\mu_d^2 n}{T^2}\right)}}{2\left(\sigma_D^2 - \frac{\mu_d^2 n}{T^2}\right)} \quad 2.6$$

$$n = T^2 \left( \frac{\sigma_D^2(\mu_F^2 + \sigma_F^2) + (\sigma_{IR}^2/v - \sigma_{B,2}^2)\mu_F + 2\sigma_{B,1}^2}{\mu_F^2 \mu_d^2} + \frac{\sigma_F^2}{\mu_F^2} \right) \quad 2.7$$

$$\mu_d = \sqrt{\frac{\sigma_D^2(\mu_F^2 + \sigma_F^2) + (\sigma_{IR}^2/v - \sigma_{B,2}^2)\mu_F + 2\sigma_{B,1}^2}{n\mu_F^2/T^2 - \sigma_F^2}} \quad 2.8$$



Note that the right-hand side of Equation 2.7 is a function of  $n$ , because  $T = t_{\{\alpha\{2\},2n-2\}} + t_{\{\beta\{1\},2n-2\}}$ . This equation must be solved approximately by numerical methods.

## 2.3 Simulations

We performed Monte Carlo simulations to assess the accuracy of the derived statistical model, and the sensitivity of the model to assumption violations. In each iteration of a simulation, we modeled an experiment comparing the sample means of voxels from  $n$  regions  $\tilde{R}_i$  and  $n$  regions  $B_i$  under the alternative hypothesis that the population mean intensities  $\mu_R$  and  $\mu_B$  differed by  $\mu_d$ , with two-sample  $t$ -tests of the null hypothesis that the population mean intensities were equal. By running multiple simulation iterations, a 95% confidence interval on the power of this experiment was estimated. Each sampled region  $\tilde{R}_i$  was modeled by sampling a fractional overlap  $f_i$  from a specified distribution, and sampling  $\text{round}(vf_i)$  intensity samples from  $I_{j,i,R}$ , the distribution of intensities of regions  $R_i$ , and  $v - \text{round}(vf_i)$  intensity samples from  $I_{j,i,B}$ , the distribution of intensities of regions  $B'_i$ . Each sampled region  $B_i$  was modeled by sampling  $v$  intensity samples from  $I_{j,i,B}$ . The difference in the means of  $I_{j,i,B}$  and  $I_{j,i,R}$  was set to be the  $\mu_d$  predicted by the model, so that if the model were perfect, the probability of a negative  $t$ -test result from the simulation would match the model's type II error parameter  $\beta$ . In each simulation, 4,000,000 simulation iterations were run so that the width of the 95% confidence interval on the power of the simulated study was 0.001 (e.g. [0.7995,0.8005]). For each simulation, the model was evaluated by (1) comparing the simulation's power to the power predicted by the model (with all other parameters

matching the simulation), (2) comparing the simulation's number of subjects  $n$  and underlying population difference  $\mu_R - \mu_B$  to the number of subjects and  $\mu_d$  predicted by the model for the observed power (with all other parameters matching the simulation). To measure each different aspect of the model, a set of multiple simulations with controlled variation of an underlying parameter (henceforth referred to as a *simulation set*) was run. Simulation sets were run with controlled variation of  $v$ , of the population distributions of  $f_i$ ,  $I_{j,i,B}$ , and  $I_{j,i,R}$ , and of the experimental design parameters  $n$ ,  $\alpha$  and  $\beta$ . Note that the population distributions for these experiments were not necessarily constrained by the model assumptions, in order to allow the testing of violations of these assumptions.

### 2.3.1 Model accuracy under the specified assumptions

To test the model accuracy under the assumptions specified in Section 2.2.1, we ran simulation sets using the data model specified in the derivation and parameters that did not violate any specified assumptions. Voxel intensity distributions were drawn from

$$I_{j,i,R} \sim \mu_R + N_t^R(0, \sigma_{DR}^2) + N_{ij}^R(0, \sigma_{IR}^2), \text{ and} \quad 2.9$$

$$I_{j,i,B} \sim \mu_B + N_t^B(0, \sigma_{DB}^2) + N_{ij}^B(0, \sigma_{IB}^2), \quad 2.10$$

and the model parameters  $\sigma_{DR}^2$ ,  $\sigma_{DB}^2$ ,  $\sigma_{IR}^2$ , and  $\sigma_{IB}^2$  were varied. A variety of distributions of fractional overlap were tested by using the Pearson family of distributions. These distributions facilitate the independent manipulation of mean  $\mu_T$ , variance  $\sigma_T^2$ , skew  $\gamma_T$  and kurtosis  $k_T$ , and include the normal distribution, giving a well-understood reference point when varying distribution parameters. Because fractional overlap is a proportion, the Pearson distributions were truncated to  $[0,1]$  and the mean and variance of fractional overlap did not in general match the specified  $\mu_T$  and  $\sigma_T^2$ . The  $\mu_F$  and  $\sigma_F^2$  used in the

model were estimated based on 1,000,000 samples from the distributions.  $n$  and  $v$  were varied over a range of values high enough to satisfy assumptions 5 and 6. Error rate control parameters  $\alpha$  and  $\beta$  were also varied. Each of the parameters was varied independently in one set of simulations, with the other parameters held constant with values listed in Table 2.1. The ranges of the parameters varied in these simulation sets are described in Table 2.2.

**Table 2.1:** Values of power simulation parameters in simulations where these parameters were held constant.

$\alpha$	$\beta$	$\sigma_{DR}^2 = \sigma_{DB}^2$	$\sigma_{IR}^2$	$\sigma_{IB}^2$	$n$	$v$	$\mu_T$	$\sigma_T$	$\gamma_T$	$k_T$
0.05	0.2	0.5	1	1	30	30	{0.5,0.7,0.9}	0.2	0	3

**Table 2.2:** Varied data model parameters for simulation sets, denoted as *min* to *max*.

$\alpha$	$\beta$	$\sigma_{DR}^2 = \sigma_{DB}^2$	$\sigma_{IR}^2$	$\sigma_{IB}^2$	$n$	$v$	$\mu_T$	$\sigma_T$	$\gamma_T$	$k_T$
0.01 to 0.2	0.1 to 0.3	0.1 to 2	0.1 to 2	0.1 to 2	30 to 80	30 to 80	0.1 to 1	0 to 1	-1 to 1	1.5 to 10

### 2.3.2 Model sensitivity to violations of the assumptions

To test the sensitivity of the derived model to violations of the assumptions in Section 2.2.1, we constructed a more general data model, and ran sets of simulations varying parameters independently, with ranges of values including extremes where the model assumptions were violated.

#### 2.3.2.1 Generalized data model

The more general data model included a separate intensity distribution  $I_{j,i,B'}$  for voxels from background regions  $B'_i$  surrounding  $R_i$ , and used more flexible voxel intensity distribution models.

In practice, imaging intensity distributions in normal and abnormal regions may vary based on the anatomy, the pathology, the imaging modality, and the digitization processes (such as truncation of the intensity ranges stored in the images), and would not in general be distributed as a Gaussian. Due to the intractable number of possible intensity distributions, these effects were not explored directly; instead, the sensitivity of the power formulae to the intensity distributions was assessed by replacing the Gaussian distributions with the more general Pearson distributions, a flexible class of distributions that can model skew and kurtosis and encompasses many common distributions including unbounded distributions such as the Gaussian distribution, the  $t$ -distribution and the gamma distribution, as well as bounded distributions such as the uniform distribution and the beta distribution. The generalized data model is characterized by the following intensity distributions:

$$I_{j,i,R} \sim \mu_R + P_i^R(\sigma_{DR}, \gamma_{DR}, k_{DR}) + P_{ij}^R(\sigma_{IR}, \gamma_{IR}, k_{IR}) + C_{RB} + C_{RR} + C_{RB'}, \quad 2.11$$

$$I_{j,i,B} \sim \mu_B + P_i^B(\sigma_{DB}, \gamma_{DB}, k_{DB}) + P_{ij}^B(\sigma_{IB}, \gamma_{IB}, k_{IB}) + C_{RB} + C_{BB}, \text{ and} \quad 2.12$$

$$I_{j,i,B'} \sim \mu_B + P_i^{B'}(\sigma_{DB}, \gamma_{DB}, k_{DB}) + P_{ij}^{B'}(\sigma_{IB}, \gamma_{IB}, k_{IB}) + C_{RB'}, \quad 2.13$$

where  $P_i(\sigma, \gamma, k)$  and  $P_{ij}(\sigma, \gamma, k)$  denote Pearson distributed random variables sampled once per region and once per voxel, respectively, with standard deviation  $\sigma$ , skew  $\gamma$  and kurtosis  $k$ , and  $C_{RB}$ ,  $C_{RR}$ ,  $C_{BB}$  and  $C_{RB'}$  were zero-mean Gaussian random variables with variances  $\sigma_{RB}^2$ ,  $\sigma_{RR}^2$ ,  $\sigma_{BB}^2$ , and  $\sigma_{RB'}^2$ , respectively. Note that distributions  $P_{ij}^B$  and  $P_{ij}^{B'}$  (and distributions  $P_i^B$  and  $P_i^{B'}$ ) share the same population parameters as they both represent background tissue; however, for a given  $i$  and  $j$ , separate independent samples were taken from these two distributions.  $C_{RB}$  was sampled once per pair of regions  $R_i$  and  $B_i$  and added to both, to introduce correlated (matched) pairs.  $C_{RR}$  and  $C_{BB}$  were sampled once

per simulation iteration and added to all  $R_i$  and  $B_i$ , respectively, introducing correlation within region types (or, effectively, variability in  $\mu_R$  and  $\mu_B$  from iteration to iteration).  $C_{RB'}$  was sampled once per region  $R_i$  and added to  $R_i$  and the surrounding background region  $B'_i$ , introducing correlation between regions and their surroundings.

### 2.3.2.2 Simulation sets

Two types of simulation sets were run to assess the sensitivity of the model to violations of the assumptions. First, parameters from the original data model were varied to violate the assumptions of the derivation. The parameters  $n$  and  $v$  were independently varied through a range (2–30) that potentially violated the assumptions that these values were sufficiently large. The ratio  $\sigma_{DR}^2/\sigma_{DB}^2$  was varied from 0.25 to 4 to violate the assumption of equality of interregion variances. Because the sensitivity of the  $t$ -test to non-Gaussian populations is known to increase with lower  $n$  [16], the variations described in Table 2.2 (apart from  $n$ ) were repeated with  $n = 20$  and  $n = 10$ . Second, the additional parameters introduced in the generalized data model representing skew ( $\gamma_{DB}$ ,  $\gamma_{DR}$ ,  $\gamma_{IB}$  and  $\gamma_{IR}$ ) and kurtosis ( $k_{DB}$ ,  $k_{DR}$ ,  $k_{IB}$  and  $k_{IR}$ ) of voxel intensity distributions and correlations between regions ( $\sigma_{RB}^2$ ,  $\sigma_{RR}^2$ ,  $\sigma_{BB}^2$ , and  $\sigma_{RB'}^2$ ) were varied. Skew parameters were varied from -1 to 1, and set to 0 when not being varied. Kurtosis parameters were varied from 1.5 to 10, and set to 3 (the kurtosis of the normal distribution) when not being varied. Correlation variances were varied from 0.1 to 2, and set to 0 when not being varied. Other parameters were set according to Table 2.1. As above, the variations were repeated with  $n = 20$  and  $n = 10$ .

### 2.3.2.3 Statistical analysis

Statistical analysis was performed in SPSS 21 (IBM SPSS, 2010, Chicago, IL). For the simulation sets following the assumptions of the derivation, the accuracy of the model was assessed based on the maximum relative error in the model's prediction of the simulated number of subjects and MDD, and the maximum absolute error in power across all the levels of the varied parameter in the simulation set. The sensitivity of the model to variation of model parameters was assessed using a linear regression with the power as the dependent variable and the varied parameter as the independent variable, averaged over the three levels of  $\mu_T$  (except in the case of varying  $\mu_T$ , where there was no averaging). For the regression assessing sensitivity to the parameter  $\beta$ , the null hypothesis was that the linear coefficient of  $\beta$  was -1, as the power of an ideal model should be  $1 - \beta$ . For the other regressions, the null hypothesis, corresponding to no sensitivity to parameter variation, was that the linear coefficient of the varied parameter was 0. For the simulations violating the assumptions of the derivation, the sensitivity of the model to these violations was assessed by (1) using separate linear regression for  $n = 30$ ,  $n = 20$ , and  $n = 10$ , with the power as the dependent variable and the varied parameter as the independent variable, averaged over three levels of  $\mu_T$ , and (2) identifying the ranges of each parameter where the model predicted the number of subjects and MDD to within 1% relative error and the power to within 0.01 absolute error. The simulation set varying  $n$  was assessed using a single linear regression.

## 2.4 Results

### 2.4.1 Simulation sets following assumptions

The simulations were run for all simulation sets that followed the assumptions of the derivation. Table 2.3 shows the linear sensitivities of the model's predicted power to the varied parameter for the simulation sets described in Table 2.2 that met two criteria: (1) regression detected a significant effect ( $p < 0.05$ ) and (2) the magnitude of the effect was  $\geq 0.0005$ . For the remaining parameters, ( $\sigma_{DR}^2 = \sigma_{DB}^2$ ,  $\sigma_{IR}^2$ ,  $\sigma_{IB}^2$ ,  $n$ ,  $v$ ,  $\sigma_T$ ,  $\gamma_T$ , and  $k_T$ ), the test failed to detect a significant effect or detected an effect  $< 0.0005$ . Under the assumptions of the derivation, the model predicted the power to within 0.003 of the simulation power, predicted the number of subjects to within 0.8% relative error from the simulation number of subjects, and predicted the MDD to within 0.5% relative error from the simulation MDD for all parameter levels for all simulations sets.

### 2.4.2 Simulation sets violating assumptions

Simulations were run for all the simulation sets that violated the assumptions of the derivation. Table 2.4 shows the linear sensitivities of the model's predicted power to the varied parameter for the simulation sets violating the model assumptions that met two criteria for any of the three levels of  $n$ : (1) regression detected a significant effect ( $p < 0.05$ ) and (2) the magnitude of the effect was  $\geq 0.0005$ . For the remaining parameters, ( $\sigma_{IR}^2$ ,  $\sigma_{IB}^2$ ,  $\gamma_{IR}$ ,  $\gamma_{IB}$ ,  $k_{IR}$ ,  $k_{IB}$ ,  $k_{DR}$ ,  $n$ ,  $v$ ,  $k_T$ ,  $\sigma_{RB}^2$ ,  $\sigma_{RR}^2$ ,  $\sigma_{BB}^2$ , and  $\sigma_{RB'}^2$ ), the test failed to detect a significant effect or detected an effect  $< 0.0005$ . Simulation sets for parameters where the model erred in the prediction of the number of subjects by more than 1% for any of the parameter levels with any  $\mu_T$  or any  $n$  are shown in Figure 2.2 for

the original data model, and shown in Figure 2.3 for the generalized data model. In these figures, the y-axes indicate the relative percentage difference between the number of subjects predicted by the model and the simulations. A value of 0% indicates that the model exactly predicted the simulation results. Values of  $-x\%$  and  $+x\%$  indicate that the model under- and overestimated the number of subjects, respectively. Parameter ranges where the model predicted the simulated power to within 0.01 absolute error and the simulated number of subjects and MDD to within 1% relative error are given in Table 2.5.

**Table 2.3:** Sensitivity of the power prediction error to variation in model parameters, under the assumptions of the derivation. The linear coefficients of the varied parameter are shown only for simulation sets that met two criteria: (1) regression detected a significant effect ( $p < 0.05$ ) and (2) the magnitude of the effect was  $\geq 0.0005$ . Parameters  $\sigma_{DR}^2 = \sigma_{DB}^2$ ,  $\sigma_{IR}^2$ ,  $\sigma_{IB}^2$ ,  $n$ ,  $v$ ,  $\sigma_T$ ,  $\gamma_T$ , and  $k_T$ , where the test failed to find a significant effect or the magnitude of the effect was  $< 0.0005$ , were omitted from the table. An ideal model would yield a linear coefficient of 0, except for the simulation set varying  $\beta$ , which should yield a linear coefficient of -1.

	$\alpha$	$\beta$	$\mu_T$
<b>Linear coefficient (1/unit of parameter variation)</b>	0.015	-0.999	0.003

**Table 2.4:** Sensitivity of the power prediction error to variation in model parameters, for simulation sets violating the assumptions of the derivation. The linear coefficients and the intercepts are shown only for simulation sets that met two criteria for any level of  $n$ : (1) regression detected a significant effect ( $p < 0.05$ ) and (2) the magnitude of the effect was  $> 0.0005$ . Parameters  $\sigma_{IR}^2$ ,  $\sigma_{IB}^2$ ,  $\gamma_{IR}$ ,  $\gamma_{IB}$ ,  $k_{IR}$ ,  $k_{IB}$ ,  $k_{DR}$ ,  $n$ ,  $v$ ,  $k_T$ ,  $\sigma_{RB}^2$ ,  $\sigma_{RR}^2$ ,  $\sigma_{BB}^2$ , and  $\sigma_{RB'}^2$ , where the test failed to find a significant effect or the magnitude of the effect was  $< 0.0005$  for all levels of  $n$ , were omitted from the table. An ideal model would yield a linear coefficient of 0, except for the simulation set varying  $\beta$ , which should yield a linear coefficient of -1. Simulation sets that yielded effects  $< 0.0005$  are denoted with a \*.

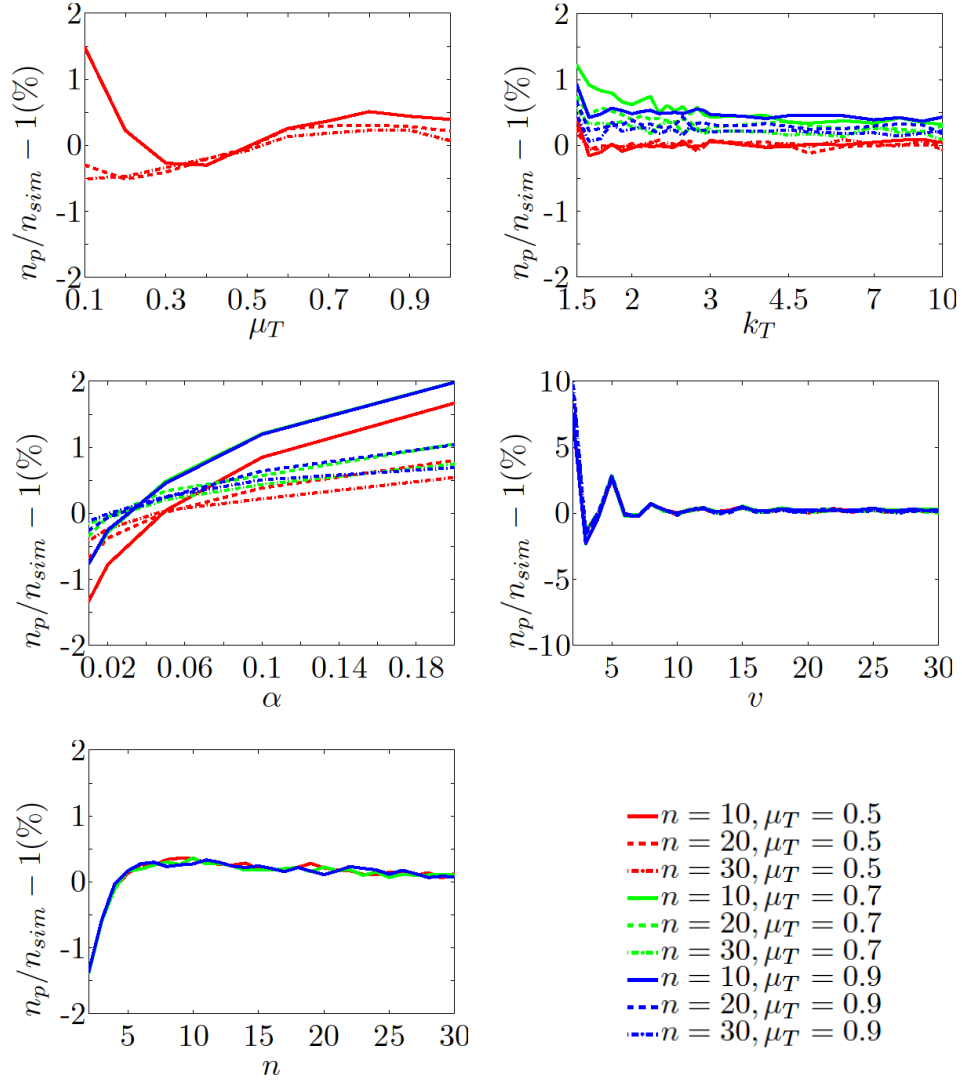
	$n$	$\alpha$	$\beta$	$\sigma_{DR}^2/\sigma_{DB}^2$	$\gamma_{DR}$	$\gamma_{DB}$	$k_{DB}$	$\mu_T$	$\sigma_T$	$\gamma_T$
<b>Linear coefficient</b>	<b>30</b>	0.015	-0.999	0.001	0.006	-0.015	*	0.003	*	*
	<b>20</b>	0.027	-0.993	0.002	0.008	-0.019	0.001	0.004	*	-0.001
	<b>10</b>	0.054	-0.987	0.003	0.01	-0.026	0.001	*	0.001	-0.001



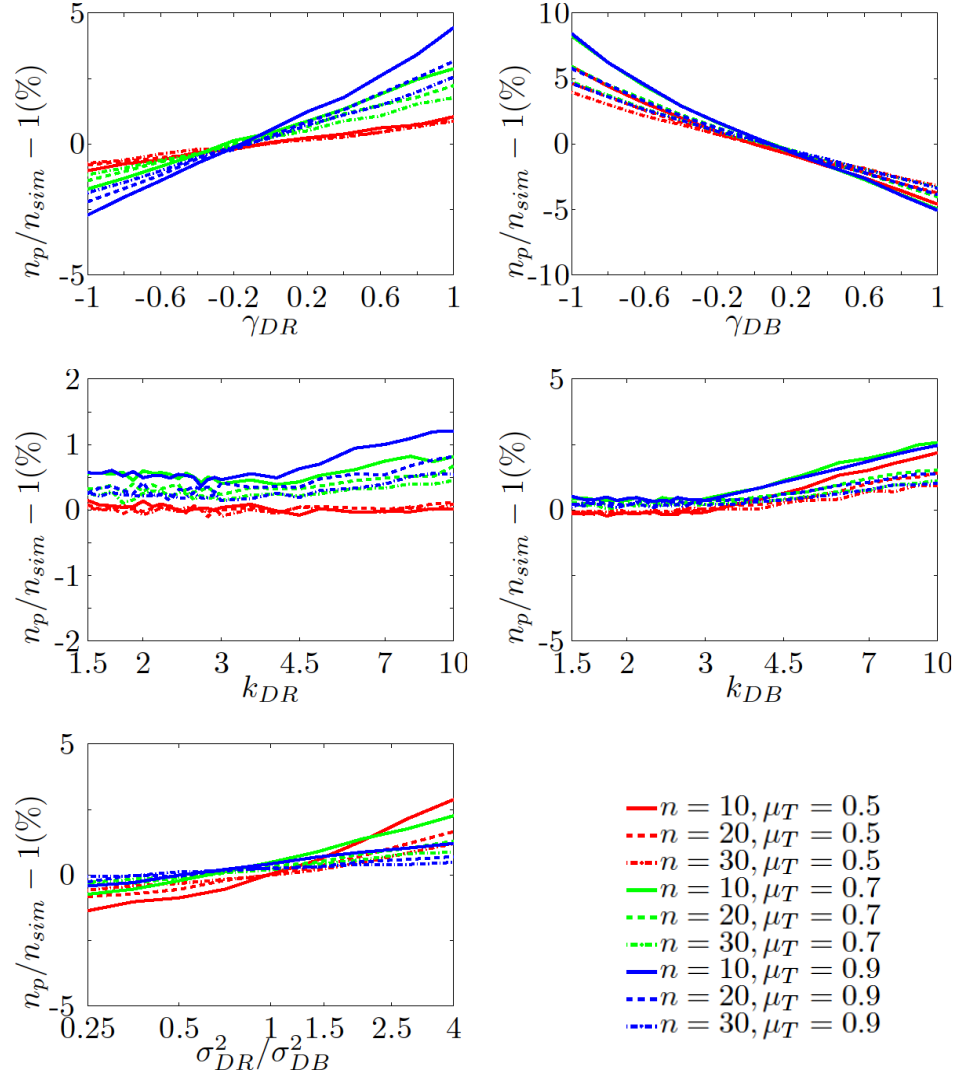
**Table 2.5:** Parameter ranges where the model predicted the simulated power to within 0.01 absolute error and the simulated number of subjects and MDD to within 1% relative error. Ranges are denoted *min* to *max* if the whole range of rested parameter values fit these criteria. Otherwise, *min* and/or *max* were replaced with the identified lower and/or upper bound that did fit the criteria. Parameters where errors were below these thresholds throughout the tested range were omitted ( $\sigma_{DR}^2 = \sigma_{DB}^2$ ,  $\sigma_{IR}^2$ ,  $\sigma_{IB}^2$ ,  $\gamma_{IR}$ ,  $\gamma_{IB}$ ,  $k_{IR}$ ,  $k_{IB}$ ,  $\sigma_T$ ,  $\gamma_T$ ,  $\sigma_{RR}^2$ ,  $\sigma_{BB}^2$ ,  $\sigma_{RB}^2$ ,  $\sigma_{RB'}^2$ ).

$n$	$\alpha$	$\sigma_{DR}^2/\sigma_{DB}^2$	$\gamma_{DR}$	$\gamma_{DB}$	$k_{DR}$	$k_{DB}$	$\nu$	$\mu_T$	$k_T$
30	<i>min</i> to <i>max</i>	<i>min</i> to 2.8	−0.4 to 0.2	−0.2 to 0.2	<i>min</i> to <i>max</i>	<i>min</i> to 8	6 to <i>max</i>	<i>min</i> to <i>max</i>	<i>min</i> to <i>max</i>
20	<i>min</i> to 0.1	<i>min</i> to 2	−0.4 to 0.2	0 to 0.2	<i>min</i> to <i>max</i>	<i>min</i> to 6	6 to <i>max</i>	<i>min</i> to <i>max</i>	<i>min</i> to <i>max</i>
10	0.02 to 0.05	0.5 to 1.4	−0.4 to 0	0 to 0.2	<i>min</i> to 6	<i>min</i> to 4	6 to <i>max</i>	0.2 to <i>max</i>	1.6 to <i>max</i>

As  $n$  was decreased, for the simulation sets that otherwise followed the assumptions, the model error increased for all simulation sets. For  $n = 10$ , the model predicted the power to within 0.008 of the simulation power, predicted the number of subjects to within 2% relative error from the number of subjects in the simulation, and predicted the minimum detectable difference to within 1.1% relative error from the simulation minimum detectable difference for all parameter levels for all simulation sets.



**Figure 2.2:** Simulation sets testing model accuracy under the original data model, for parameters that yielded  $>1\%$  relative error in the number of subjects at any level of  $n \in \{10, 20, 30\}$  or  $\mu_T \in \{0.5, 0.7, 0.9\}$ . The remaining parameters ( $\sigma_{IR}^2$ ,  $\sigma_{IB}^2$ ,  $\sigma_T$ ,  $\gamma_T$ ,  $\sigma_{DR}^2 = \sigma_{DB}^2$ , and  $\beta$ ) showed  $<1\%$  error in the predicted number of subjects for all simulation sets. Each curve represents the relative error in the predicted number of subjects for a particular  $n$  and  $\mu_T$  for varying parameter levels. An ideal model would yield a horizontal line at 0% for all levels of  $n$  and  $\mu_T$ . Multiple lines with different slopes, or different shapes suggest that there is an interaction between the sensitivity to the parameter being varied and  $n$  or  $\mu_T$ . Negative relative errors mean the model underestimated the number of subjects, potentially leading to an underpowered study. Positive relative errors mean the model overestimated the number of subjects, resulting in potentially inefficient use of resources.



**Figure 2.3:** Simulation sets testing model accuracy under the generalized data model, for parameters that yielded  $>1\%$  relative error in the predicted number of subjects at any level of  $n \in \{10, 20, 30\}$  or  $\mu_T \in \{0.5, 0.7, 0.9\}$ . The remaining parameters ( $\sigma_T$ ,  $\gamma_T$ ,  $\sigma_{IR}^2$ ,  $\gamma_{IR}$ ,  $k_{IR}$ ,  $\sigma_{IB}^2$ ,  $\gamma_{IB}$ ,  $k_{IB}$ ,  $\sigma_{RR}^2$ ,  $\sigma_{BB}^2$ ,  $\sigma_{RB}^2$ ,  $\sigma_{RB'}^2$ ,  $\alpha$ ,  $k_T$ ,  $\mu_T$ , and  $v$ ) showed  $<1\%$  error in the predicted number of subjects for all simulation sets. Each curve represents the relative error in the predicted number of subjects for a particular  $n$  and  $\mu_T$  for varying parameter levels. An ideal model would yield a horizontal line at 0% for all levels of  $n$  and  $\mu_T$ . Multiple lines with different slopes, or different shapes suggest that there is an interaction between the sensitivity to the parameter being varied and  $n$  or  $\mu_T$ . Negative relative errors mean the model underestimated the number of subjects, potentially leading to an underpowered study. Positive relative errors mean the model overestimated the number of subjects, resulting in potentially inefficient use of resources.

## 2.5 Discussion

In this work, we derived a set of novel statistical power formulae that directly model registration error for a statistical test of the difference in signal intensity between foreground and background regions, accounting for correlation within these regions.

This work provides a derivation of a statistical power calculation incorporating image registration uncertainty and addressing three central questions in the design of imaging validation studies. (1) **Eq. 2.6:** What is the maximum acceptable registration error? (2) **Eq. 2.7:** How many subjects are needed? (3) **Eq. 2.8:** What is the MDD between normal and pathologic image regions? Specifically, we derived an approximate relationship between an arbitrary distribution of fractional overlap and the statistics of a study design, elucidating a relationship between registration error, number of subjects and statistical power, yielding a set of three equations that are central to the design of imaging validation studies.

These relationships could be used in several applications. During study design, Eq. 2.7 or 2.8 could be used to evaluate or control the power, after estimating imaging properties and registration errors, while Eq. 2.6 could be used to guide the selection of registration algorithms under the constraints of a study design. During algorithm development, Eq. 2.6 could be used to assess whether an algorithm has sufficient accuracy for a particular application.

### 2.5.1 Simulations

We ran Monte Carlo simulations to test the fidelity of our model both when our assumptions were met and when some of them were violated. When the assumptions of the derivation were met, the model predicted the number of subjects and MDD to within

1% error of the simulation and the power to within 0.01 absolute error of the simulation for all of the simulation sets. When the assumptions of the derivation were met, the source of the residual error is the deviation of the distribution of voxel intensities in the sampling region from the Gaussian distribution assumed by the classical power calculation. The simulations suggest that this source of error is negligible over the range of parameters simulated.

When the assumptions were violated, the model still predicted the number of subjects, MDD and power of the simulation to within 11% over all tested parameter values. The model was most sensitive to the combination of a skewed distribution of regional mean intensities and low  $n$ . The parameter ranges shown in Table 2.5 and Figures 2.2 and 2.3 may suggest study parameter ranges for which this model is appropriate. For example, the model's accuracy decreases as the number of samples  $v$  decreases below 6 and as the number of subjects  $n$  decreases below 4, suggesting that the model may not be appropriate for studies where the pathological regions occupy fewer than 6 voxels, or the study includes fewer than 4 subjects. In addition to serving to evaluate the statistical model, Figure 2.3 may also serve as a guideline for compensating for known deviations from the assumed model. For example, if it is known that the distribution of the background voxels has a positive skew of 1, a conservative study design could recommend an additional 4% beyond the number of subjects predicted by the model.

### 2.5.2 Fractional overlap relates registration error to study outcome

The post-registration misalignment of many homologous landmarks not used to define the registration (often quantified as the TRE) is considered to be the ideal general

metric for measuring registration error [4, 5]. In the study design addressed in this work, where image signals averaged within foreground and background regions are compared, the test statistic, and thus the study outcome, is invariant to some types of registration error. Specifically, two different registrations that map a point on the study image to different points that are both within the region  $R_i$  will yield the same test statistic. Because of this invariance, the statistical power of such a study is more directly tied to a surrogate measure for registration error that shares this invariance. Furthermore, as fractional overlap for a sampling region is linearly related to the average signal inside the sampling region, it is readily integrated into the derivation, which is a desirable feature for a surrogate measure.

### 2.5.3 Estimating fractional overlap

The presented power formulae depend on estimates of two population parameters of the distribution of fractional overlap for the registration methods and data used in a study. The statistical properties of fractional overlap have not yet been well characterized in the general case, and methods for estimating its population parameters have not yet been standardized. Thus, the practical use of the formulae requires careful consideration to be made in selecting an approach to this parameter estimation that is appropriate to the problem at hand. One possible approach [11] would estimate  $\mu_F$  and  $\sigma_F^2$  from a quantified target registration error [3], under assumptions of having translational registration errors distributed as a 3D Gaussian, and having spherical regions of known volumes. Depending on the study, these assumptions may be overly strong. An alternative approach to estimating the mean fractional overlap that avoids the assumption of spherical regions with known volumes would be to (1) characterize the distribution of tumor shapes (e.g.,

as a sample of tumor shapes taken from pilot data), (2) characterize the distribution of registration errors (e.g., as a distribution of per-region translations modeled as a 3D Gaussian distribution whose covariance is defined based on a quantified TRE), (3) repeatedly sample from these distributions and measure the fractional overlap for each sample, and (4) estimate the population mean and standard deviation of fractional overlap as the sample mean and standard deviation. The characterization of the statistical properties of these approaches, and the development of generalized approaches to estimating fractional overlap, would support the application of the power calculation formulae.

#### 2.5.4 Case study

The application of these power calculations can be illustrated through a sample size calculation for a case study, modeled closely after a prostate cancer imaging validation study currently taking place at our institution. In this study, patients scheduled for radical prostatectomy (surgical removal of the prostate) undergo *in vivo* imaging before surgery using multi-parametric magnetic resonance (MR) imaging (structural, diffusion weighted and dynamic contrast-enhanced (DCE) sequences). These images could be processed to yield derived images (T2 maps from structural imaging, the average diffusion coefficient (ADC) from diffusion weighted MRI, and both the contrast transfer coefficient  $K^{trans}$  and the contrast leakage  $v_e$  from DCE MRI). Following surgery, the prostate specimens are processed for whole-mount histology [17] and are digitized, annotated by a pathologist to identify cancer of different grades [18], and reconstructed and registered to the *in vivo* images. The contemporary per-patient cost for this study underway in our center is approximately \$10,000 USD, so correctly estimating

the required number of subjects is important. Although these data can be applied to answer many questions, in this illustration we calculate the number of subjects for a hypothesis testing whether these modalities show significant signal differences between cancer and benign regions in the prostatic peripheral zone, known to harbor the majority of prostate cancer [19].

To predict the required number of subjects for such a study, we can use Equation 2.7. In order to use this equation, we need to specify three study design parameters: (1)  $\alpha$ , the acceptable false positive error rate (chosen as 0.05, in our case study);  $\beta$ , the acceptable false negative error rate (chosen as 0.2 in our case study); and  $\mu_d$ , the minimum magnitude of signal difference we need to distinguish with these error rates. For our case study, we estimated the differences in mean intensities from the literature: Langer et al. [20] reported medians and ranges (from which means and variances can be estimated [21]) for the intensity in tumor and benign tissue in the prostatic peripheral zone for T2, ADC,  $K^{trans}$  and  $v_e$  images. We also need to estimate 6 model parameters:

- I.  $v$ , the average number of samples per tumor, which can be calculated as the average tumor volume divided by the volume of an image voxel. For this case study, we assume an even distribution of tumors with 0.5 cc, 1 cc and 2 cc volumes.
- II.  $\sigma_R^2$ , the variance of tumor voxel intensities. For this case study, we used the variance estimated from statistics reported by Langer et al. [20] for tumor peripheral zone tissue for T2, ADC,  $K^{trans}$  and  $v_e$ .



- III.  $\sigma_B^2$ , the variance of background voxels. For this case study, we used the variance estimated from statistics reported by Langer et al. [20] for benign peripheral zone tissue for T2, ADC,  $K^{trans}$  and  $v_e$ .
- IV.  $\rho$ , the interclass correlation coefficient relating the relative contributions of intra- and interregion variances to the total variance of voxel intensities. This parameter is typically assumed to be equal for both classes, and so can represent  $\sigma_{DB}^2/(\sigma_{DB}^2 + \sigma_{IB}^2)$  or  $\sigma_{DR}^2/(\sigma_{DR}^2 + \sigma_{IR}^2)$ . This can be estimated from the literature, or calculated from pilot data as  $s_D^2/(s_D^2 + s_I^2)$ , where  $s_D^2$  is the variance of the mean intensities of tumor regions and  $s_I^2$  is the variance of tumor voxels after subtracting the mean intensity for each tumor. For this case study, without pilot data or information from the literature specifying  $\rho$ , we used a conservative estimate of  $\rho = 1$ .
- V.  $\mu_F$ , the mean fractional overlap of our tumor sampling regions with the underlying tumor for our registration algorithm. For this case study, we estimated  $\mu_F$  using the first approach described in Section 2.5.3, i.e., estimating  $\mu_F$  from a TRE estimate under the assumptions of spherical tumors and 3D Gaussian translational error, and using the same tumor volumes as used for estimating  $v$ . For this case study, we explored multiple levels of TRE.
- VI.  $\sigma_F^2$ , the variance of fractional overlap of our registration algorithm. For the case study, this was also estimated using the first approach described in Section 2.5.3.

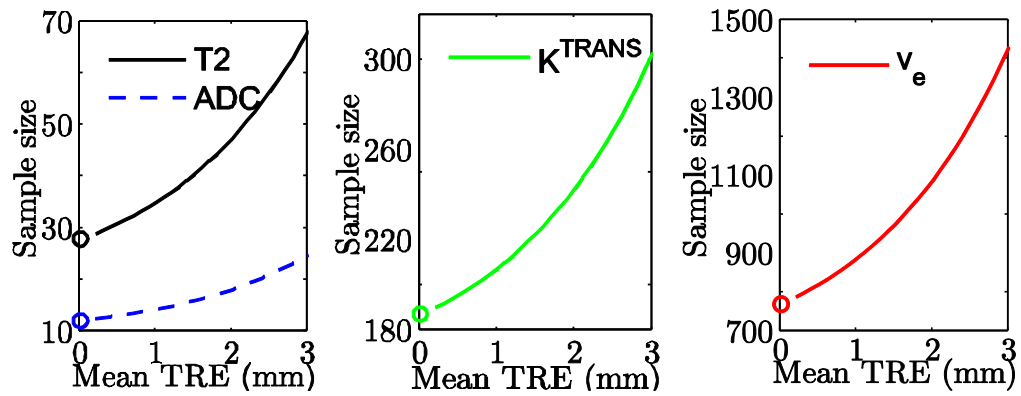
From these parameters, we can calculate 4 derived parameters: (1)  $\sigma_{IR}^2 = (1 - \rho)\sigma_R^2$ , (2)  $\sigma_D^2 = \rho(\sigma_R^2 + \sigma_B^2)$ , (3)  $\sigma_{B,1}^2 = \sigma_B^2(1 + (v - 1)\rho)/v$ , and (4)  $\sigma_{B,2}^2 = \sigma_B^2(1 + (2v - 1)\rho)/v$ . The number of subjects can be calculated by substituting these parameters and  $T = t_{\{\alpha\{2\}, 2n-2\}} + t_{\{\beta\{1\}, 2n-2\}}$  into Equation 2.7 and solving numerically

for the fixed point of  $n$  [22]. For the case study, this was performed by expressing the right hand side of Equation 2.7 as  $g(n)$  and finding the zero-crossing of  $n - g(n)$  using an iterative solver implemented as *fzero* in Matlab R2011b (The Mathworks Inc., Natick, MA).

Because it may be necessary to estimate  $\rho$  in the absence of pilot data or reported values, an intuitive understanding of  $\rho$  is important. The effect of  $\rho$  can be understood by comparing the variance of the mean of cluster-randomized samples in the absence of correlation  $(\sigma_{DR}^2 + \sigma_{IR}^2)/(nv)$ , and in the presence of correlation,  $(\sigma_{DR}^2 + \sigma_{IR}^2)/\left(n \frac{v}{1+(v-1)\rho}\right)$ .  $\rho$  has an effect equivalent to changing the number of independent samples that can be collected from each region. Since the power depends on the total number of independent samples, the number of subjects needed is inversely proportional to the number of independent samples per region. Assuming  $\rho = 1$  (i.e. effectively 1 sample per region) would yield a conservative estimate of the necessary number of subjects. As initial data is collected during the study and better estimates for  $\rho$  can be made, the study may reassess the required number of subjects. For example,  $\rho = 0.5$  (i.e. effectively ~2 samples per region) would halve the required number of subjects.

If we had an ideal zero-error registration, we could apply the classical power calculation in Equation 2.2 by solving for  $n$ , substituting in this study's parameter estimates for  $\alpha$ ,  $\beta$ ,  $v$ ,  $\sigma_R^2$ ,  $\sigma_B^2$  and  $\rho$ , and numerically solving for the fixed point of  $n$ . This would yield an estimated number of subjects for each modality: 28 for T2, 13 for ADC, 188 for  $K^{trans}$ , and 772 for  $v_e$ . In practice, since histology-*in vivo* MR image registrations have error, Equation 2.7 predicts the functional relationship between number of subjects and registration error using this study's parameter estimates (shown in Figure

2.4). In this relationship, the  $x$ -axis is target registration error, from which we calculated the mean and variance of fractional overlap [11]. For a target registration error of 1 mm, we would predict a required number of subjects of 36 for T2, 15 for ADC, 208 for  $K^{trans}$ , and 889 for  $v_e$ . For a target registration error of 2.5 mm, we would predict a required number of subjects of 48 for T2, 19 for ADC, 243 for  $K^{trans}$ , and 1090 for  $v_e$ , in which case, the classical power calculation underestimated the number of subjects by as much as 40%.



**Figure 2.4:** Relationship between estimated required number of subjects and target registration error (TRE) for 4 imaging modalities (note the differing y-axis scales illustrating the variable sensitivity of the required number of subjects to TRE for the different modalities). An estimate of the required number of subjects using the classical power formula (marked with circles) will underestimate the required number of subjects when there is registration error. This relationship also illustrates the high potential impact for improving registration error in the context of imaging validation studies. Based on contemporary per-patient cost for a study underway in our center, the per-patient cost is more than \$10,000 USD, so reducing the required number of subjects can have a substantial impact on the overall study cost.

In practice, several methods have been proposed for reconstruction of prostate histology to *in vivo* MR imaging. The required number of subjects (and thus the study cost) varies as a function of the registration error. Figure 2.4 shows this relationship, demonstrating the value in improving registration accuracy.

### 2.5.5 Limitations

The novel power calculations derived in this work should be considered in the context of two key limitations. First, the classical power calculation on which this work builds assumes that the samples being compared are drawn from a Gaussian distribution. In the presence of registration error, one sample may no longer be Gaussian, causing the distribution of the  $t$  statistic to deviate from a  $t$ -distribution. Thus, this work relies on the robustness of the  $t$ -test to non-Gaussian distributions. Second, the derivations in this work are exact under strong assumptions. Because statistical power calculations inherently depend on the particular statistical analysis being performed, the derived power calculations can be directly applied only in the context of the described statistical design. The approach used in this derivation, however, may facilitate derivations of power calculations suitable for other statistical designs. The specific data models assumed in the derivation may not perfectly describe data for a given problem; however, the analysis of sensitivity to violations of these assumptions should allow users to assess whether the model is appropriate for a given experiment. Third, due to the assumptions of the classical power formula from which the novel power calculation formulae were derived, our model assumes that the interregion variances  $\sigma_{DR}^2$  and  $\sigma_{DB}^2$  are equal. Depending on the study, this assumption may be violated if, for example, the variability of the pathology from subject to subject results in more variable appearance between pathological regions than between benign regions. Prediction errors due to such violations are shown in Figure 2.3. Fourth, the power calculation formulae depend on the estimation of the mean and standard deviation of fractional overlap, which are required for the application of the formulae. The statistics of fractional overlap, and their

dependence on registration method and data, are not as well characterized as those of other registration metrics such as TRE. The two approaches to this estimation discussed in the case study, one using strong assumptions about region shapes and the distribution of TRE, and one incorporating region shape information from pilot data, may warrant further investigation.

## 2.6 Conclusions

In this work, we derived novel power calculations relating registration error to the minimum detectable difference and the number of subjects. These power calculation formulae enable imaging researchers to answer three central questions in the design of certain imaging validation studies: (1) What is the maximum acceptable registration error? (2) How many subjects are needed? (3) What is the minimum detectable difference between normal and pathologic image regions? These formulae were accurate to within 1% error when the model's assumptions were followed, and the sensitivity of the model to violations of the assumptions was characterized. The case study presented in this work highlighted the importance of adjusting the number of subjects to account for registration error, showing that the classical power calculation underestimated the required number of subjects by as much as 40% for a study design modeled closely after an ongoing imaging validation project in our center.

## 2.7 References

1. D. Simon, R. V. O'Toole, M. Blackwell, F. Morgan, A. M. Digioia and T. Kanade, "Accuracy validation in image-guided orthopaedic surgery," in *Proceedings of International Symposium on Medical Robotics and Computer Assisted Surgery*, (1995), pp. 185–192.

2. M. H. Loew and C. E. Rodriguez-Carranza, "Technical issues in multimodality medical image registration," in *Proceedings of IEEE Computer-Based Medical Systems*, (1998), pp. 2–7.
3. J. M. Fitzpatrick, D. L. G. Hill, Y. Shyr, J. West, C. Studholme and C. R. J. Maurer, "Visual assessment of the accuracy of retrospective registration of MR and CT images of the brain," *IEEE Transactions on Medical Imaging* **17**, 571–585 (1998).
4. J. West, J. M. Fitzpatrick, M. Y. Wang, B. M. Dawant, C. R. Maurer Jr, R. M. Kessler, R. J. Maciunas, C. Barillot, D. Lemoine, A. Collignon and others, "Comparison and evaluation of retrospective intermodality brain image registration techniques," *Journal of Computer Assisted Tomography* **21**, 554–568 (1997).
5. T. Rohlfing, "Image similarity and tissue overlaps as surrogates for image registration accuracy: widely used but unreliable," *IEEE Transactions on Medical Imaging* **31**, 153–163 (2012).
6. J. Neyman and E. S. Pearson, "On the use and interpretation of certain test criteria for purposes of statistical inference: part I," *Biometrika*, 175–240 (1928).
7. M. W. Lipsey, *Design sensitivity: statistical power for experimental research*. (Sage Publications, 1990).
8. R. Hayes and S. Bennett, "Simple sample size calculation for cluster-randomized trials," *International Journal of Epidemiology* **28**, 319–326 (1999).
9. K. J. Friston, A. P. Holmes and K. J. Worsley, "How many subjects constitute a study?," *Neuroimage* **10**, 1–5 (1999).
10. A. Donner and N. Klar, *Design and analysis of cluster randomization trials in health research*. (London: Arnold, 2000).
11. E. Gibson, A. Fenster and A. D. Ward, "Registration accuracy: how good is good enough? A statistical power calculation incorporating image registration uncertainty," in *Proceedings of Medical Image Computing and Computer Assisted Intervention, Vol. 7511* (Springer, Nice, France, 2012), pp. 643–650.
12. M. Brett, W. Penny and S. Kiebel, "Introduction to random field theory," *Human Brain Function*, 867–879 (2003).
13. A. E. Mace, *Sample-size determination*. (New York: Reinhold, 1964).
14. E. J. H. Korendijk, C. J. M. Maas, M. Moerbeek and P. G. M. Van der Heijden, "The influence of misspecification of the heteroscedasticity on multilevel regression parameter and standard error estimates," *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences* **4**, 67–72 (2008).

15. L. Kish, *Survey sampling*. (New York: Wiley, 1965).
16. P. Diehr and T. Lumley, "The importance of the normality assumption in large public health data sets," *Annual Review of Public Health* **23**, 151–169 (2002).
17. R. Montironi, T. van der Kwast, L. Boccon-Gibod and A. V. Bono, "Handling and pathology reporting of radical prostatectomy specimens," *European Urology* **44**, 626–636 (2003).
18. D. Gleason, "Histologic grading and staging of prostatic carcinoma," *American Journal of Surgical Pathology* **5**, 193 (1981).
19. J. Haffner, E. Potiron, S. Bouyé, P. Puech, X. Leroy, L. Lemaitre and A. Villers, "Peripheral zone prostate cancers: location and intraprostatic patterns of spread at histopathology," *The Prostate* **69**, 276–282 (2009).
20. D. L. Langer, T. H. van der Kwast, A. J. Evans, A. Plotkin, J. Trachtenberg, B. C. Wilson and M. A. Haider, "Prostate tissue composition and MR measurements: investigating the relationships between ADC, T2, K(trans), v(e), and corresponding histologic features," *Radiology* **255**, 485–494 (2010).
21. S. P. Hozo, B. Djulbegovic and I. Hozo, "Estimating the mean and variance from the median, range, and the size of a sample," *BMC Medical Research Methodology* **5**, 13 (2005).
22. D. M. Murray and P. J. Hannan, "Planning for the appropriate analysis in school-based drug-use prevention studies," *Journal of Consulting and Clinical Psychology* **58**, 458 (1990).

## Chapter 3.

# **3D prostate histology image reconstruction: quantifying the impact of tissue deformation and histology section location<sup>†</sup>**

### 3.1 Introduction

Determining the 3D spatial relationship between corresponding histology and *in vivo* images, for the purpose of evaluating prostate cancer imaging, is often performed in two steps: (1) a reconstruction of histology images to a 3D *ex vivo* spatial context, and (2) an alignment of reconstructed histology to *in vivo* images.

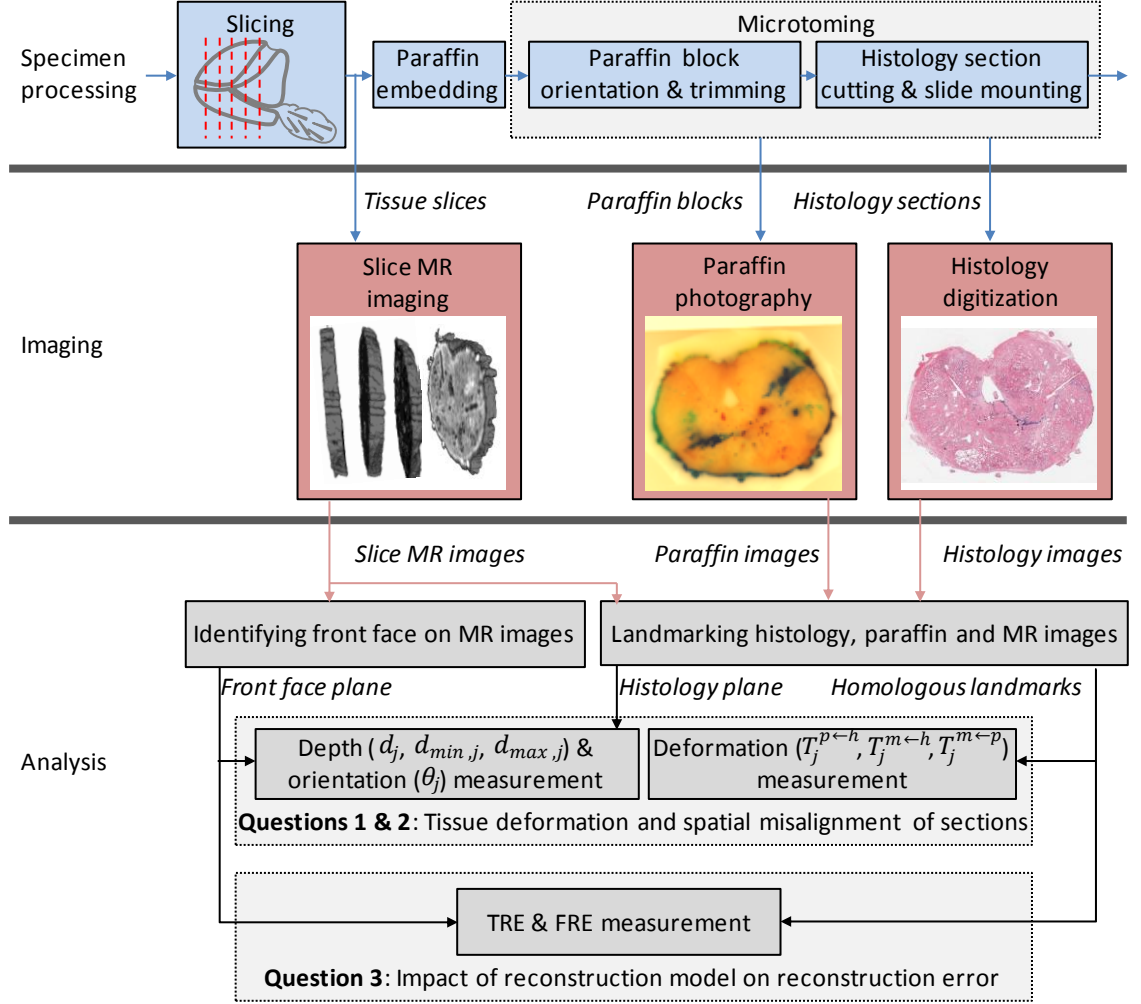
The challenges in 3D histology reconstruction can be illustrated in the context of the process of collecting histology from radical prostatectomy specimens, which typically proceeds as follows (shown in the first row of Figure 3.1). After surgery, the prostate is fixed in a formalin solution and then cut into 3–5-mm-thick tissue slices at the pathology bench. These tissue slices proceed through a series of chemical baths to replace water in the tissue with paraffin, and the slices are embedded in a block of translucent paraffin. This block is mounted to a microtome by hand, aligned by eye to square the tissue face (as seen through the translucent paraffin) with the microtome blade, and tissue is cut until a full cross-section can be collected. Once a sufficient depth has been reached, the

---

<sup>†</sup> A version of this chapter has been published: E. Gibson, M. Gaed, J. A. Gómez, M. Moussa, S. Pautler, J. L. Chin, C. Crukley, G. S. Bauman, A. Fenster, A. D. Ward, “3D prostate histology image reconstruction: Quantifying the impact of tissue deformation and histology section location.” J Pathol Inform 4:31 (2013).



operator cuts a 4  $\mu\text{m}$  histological section, allows it to expand on a water bath to flatten the section, and mounts it on a glass slide.



**Figure 3.1:** Overview of specimen processing, imaging and analysis.

The 3D reconstruction of histology consists of retroactively determining the positions of the cutting and the deformation of the tissue to determine the original 3D spatial relationships of histological tissue, a process that remains an active area of research [1-4]. 2D to 3D deformable reconstructions for clinical prostate specimens have many degrees of freedom and sparse out-of-plane information content. A common approach to mitigate these challenges is to make simplifying assumptions about the

spatial relationship of histological tissue to the corresponding tissue in the specimen. Some assumptions made in existing approaches for 3D histology reconstruction are enumerated below. The focus of this chapter is on testing the strength of the first two assumptions in this list.

- The deformation of the histological tissue after coarse slicing fits a specified constrained deformation model, such as the rigid [1], rigid + isotropic scaling (referred to as *similarity* throughout this chapter) [5, 6], affine [2, 7], or thin-plate-spline (TPS) [8, 9] deformation models. This is referred to as the ***deformation model assumption*** throughout this chapter.
- Each histological section corresponds to the front face of the 3–5-mm-thick tissue slice from which it was taken [3, 5-7, 9-11]. This is referred to as the ***front face assumption*** throughout this chapter.
- Each histological section corresponds to a planar surface in the specimen [1-3, 7, 10, 11].
- Each histological section corresponds to a surface in the specimen defined prospectively by carefully controlling the position and orientation of the cuts made during specimen slicing [3, 10, 12, 13].
- The histological sections correspond to parallel, evenly spaced surfaces in the specimen (typically justified based on controlled cutting of tissue slices) which are determined retrospectively using additional imaging (typically photographs of the faces of tissue slices) during or after the slicing of specimens into thick tissue slices [5, 7, 9, 11].

By constraining the allowable spatial relationship between each histology section and the tissue from which it was cut, these assumptions may simplify the registration problem by decreasing the degrees of freedom but may also impact the accuracy of 3D reconstruction methods. In the context of imaging validation studies that evaluate imaging modalities by comparison to a 3D reconstructed histological reference, reconstruction accuracy affects the statistical power (i.e. the probability of a study finding an existing statistically significant effect) of studies that apply the reconstruction methods [14, 15]. Because of this relationship, 3D histology image reconstructions with greater error create a requirement for more patients to be enrolled in the study, which can have a substantial impact on the cost of the study (see discussion for an illustrative case study) or run the risk of improperly evaluating the imaging modality if the study is underpowered for the error inherent in the technique. Thus, it is important to consider these simplifying assumptions, and their impact on reconstruction error, in the development and/or selection of reconstruction methods for such studies.

The strengths of these assumptions depend on their fidelity to the processes the tissue undergo throughout the preparation of histological sections. For example, the correspondence of the histological sections to the front faces of tissue slices depends in part on the skill and experience of the microtome operator determining the paraffin block face orientation and cutting depth. This task is complicated by the fact that the tissue face is hidden behind a translucent layer of paraffin, challenging the assessment of tissue face orientation until the tissue has already been exposed by the microtome blade and thus sectioned. The strengths of these assumptions, and their impact on the reconstruction error, have not, to the best of our knowledge, been quantified in the literature. This

complicates the selection of appropriate assumptions for reconstruction method developers, and also complicates the selection of reconstruction methods (that may incorporate such assumptions) for study designers.

In this work, our objective was to quantify the spatial relationships between histological sections, paraffin-embedded blocks, and the corresponding tissue slices from which the sections were taken to answer three questions, referred to according to the following enumeration throughout this chapter.

- **Question 1:** How does prostate tissue deform during histology processing?  
Specifically, with what accuracy can rigid, similarity, affine or TPS deformation models align homologous landmarks on histology sections (Figure 3.2f, labeled  $h_{i,j}$ ), on the cut paraffin-embedded tissue blocks (Figure 3.2e, labeled  $p_{i,j}$ ) and on MR images of formalin-fixed tissue slices (Figure 3.2c-d, labeled  $m_{i,j}$ ). This question constitutes a test of the strength of the *deformation model assumption*.
- **Question 2:** What spatial misalignment of the tissue sections is induced by microtome cutting? Specifically, relative to the front faces of the tissue slices, from what depth (Figure 3.2c, labeled  $d_j$ ) and at what orientation (Figure 3.2c, labeled  $\theta_j$ ) are histology sections taken? This question constitutes a test of the strength of the *front face assumption*.
- **Question 3:** How does the choice of reconstruction model affect the accuracy of histology reconstructions? A reconstruction model is defined in this chapter by a choice of one of four deformation models (rigid, affine, similarity, TPS), and a choice of whether or not to make the front face assumption. This question resolves to the following two more specific questions: for all possible

reconstruction models as defined above, (1) what are the target registration errors (TRE) and (2) the fiducial registration errors (FRE) of least-squares best-fit landmark-based reconstructions?

These questions constitute an evaluation of the *impact on registration error* of making the *deformation model assumption* and the *front face assumption*. The answers to these questions were quantified using homologous landmarks manually identified on histology images, paraffin block face images, and MR images of the tissue slices, and the resulting errors were evaluated in the context of a type of imaging validation study that relies on histology image reconstruction.

## 3.2 Materials and methods

### 3.2.1 Materials and imaging

As part of an ongoing prospective imaging validation study, we obtained prostate specimens from 6 subjects after radical prostatectomy with the following inclusion criteria: (1) male, (2) age 18 years or older, and (3) clinical prostate cancer stage T1 or T2 with histological confirmation from biopsy. The exclusion criteria were: (1) prior therapy for prostate cancer, (2) use of 5-alpha reductase inhibitors within 6 months of the study start, (3) inability to comply with preoperative imaging, (4) allergy to contrast agents, (5) sickle cell or other anemias, (6) hip prosthesis, (7) sources of artifact within the pelvis, and (8) contraindications to MRI. This research was approved by our institutional human subjects research ethics board, and informed consent was obtained from each subject.

An overview of the processing, imaging and measurement of these data is shown in Figure 3.1. After resection, fixation (10% buffered formalin for 48 hours), and marking with fiducial strands [3], the prostatic apex was removed and the mid-gland was gross-

sectioned into 4.4-mm-thick tissue slices (3–5 per specimen, 21 total). MR images of these tissue slices were acquired using a Discovery MR750 (GE Healthcare, Waukesha, WI, USA) at 3 T using an endorectal coil (Prostate eCoil, Medrad, Inc., Warrendale, PA, USA). Tissue slices were immobilized in tissue processing and embedding cassettes and immersed in Christo-Lube (Lubrication Technology Inc., Franklin Furnace, OH, USA) to provide a black background and minimize boundary artifacts on imaging. Imaging used a T1-weighted protocol (3D SPGR, TR 6.5 ms, TE 2.5 ms, bandwidth  $\pm 31.25$  kHz, 8 averages, FOV 14×14×6.2 cm, slice thickness 0.4 mm, 256×192 matrix, 312 slices, flip angle 15°, 25 min) and a T2-weighted protocol (3D FSE, TR 2000 ms, TE 151.5 ms, bandwidth  $\pm 125$  kHz, 3 averages, FOV 14×14×6.2 cm, slice thickness 0.4 mm, 320×192 matrix, 312 slices, 25 min). These images are referred to as *tissue slice MR images* throughout this chapter.

Following MR imaging, formalin-fixed tissue slices were decalcified in a hydrochloric acid/chelating agent solution (Cal-Ex Decalcifier, Fisher Chemical, Ottawa, Canada) overnight and then dehydrated and embedded in paraffin using a series of chemical baths of formalin, ethanol, xylene and paraplast under our hospital's standard clinical pathology laboratory protocol for large blocks (wherein the duration of ethanol, xylene and paraplast are lengthened). The full processing schedule is given in Appendix B.

The embedded blocks were sectioned by one of fifteen clinical histotechnologists in our hospital's clinical pathology laboratory. Each block was mounted by hand on the chuck of a microtome (RM2245 or RM2255, Leica Biosystems, Nussloch, Germany), and the operator attempted to align the front face of the tissue, as seen through the semi-

transparent paraffin covering, with the cutting axis by manually adjusting mechanical control knobs on the microtome. Sections were repeatedly cut until a full cross-section of the tissue block was reached. A final 4  $\mu\text{m}$  section was cut from the block, floated on a hot water bath to remove distortion and mounted on a positively-charged glass slide. All sections were stained with hematoxylin and eosin.

After the clinical pathology assessment was complete, stained histology sections were digitized on a ScanScope GL (Aperio Technologies, Vista, CA, USA) bright field slide scanning system with a 0.5  $\mu\text{m}$  pixel size. These images are referred to as *histology images* throughout this chapter.

Photographs of the exposed face of paraffin-embedded tissue blocks were acquired using a Pentax K200D with an SMC Pentax D FA 100 mm F2.8 macro lens (Pentax Imaging Company, Denver, CO, USA). The camera was attached to the camera-mount column of the photography table to ensure the optical axis was perpendicular to the tissue blocks. Labels containing a 4 mm long scale marker were affixed to the cut surfaces, and used to calibrate the pixel size of the images. These images are referred to as *paraffin images* throughout.

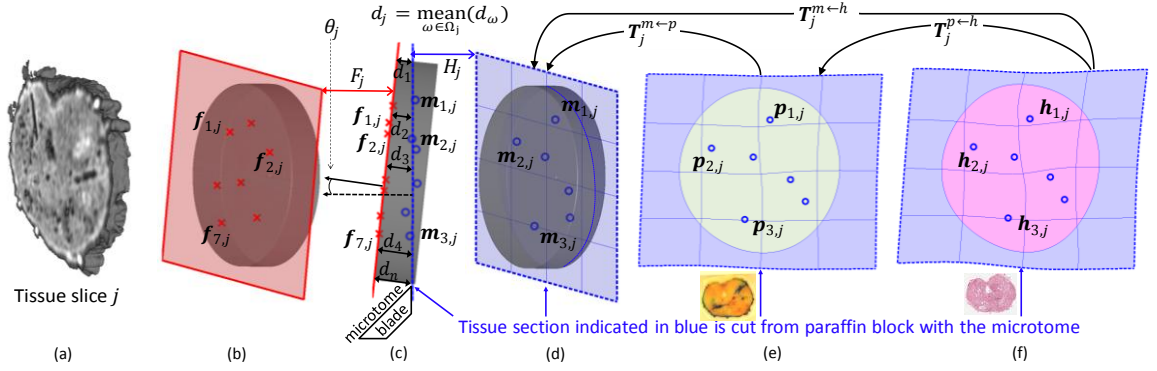
Three sections were excluded from our analysis, because we identified an insufficient number of homologous landmarks on the three imaging modalities (histology, paraffin, and tissue slice MR images). Five sets of homologous landmarks were necessary for our evaluation; see discussion for details.

### 3.2.2 Methods

Our method is illustrated in the “Analysis” portion of Figure 3.1, and is summarized at a high level as follows. To characterize prostate tissue deformation due to

histology processing (Question 1), we determined the class of deformation that best mapped tissue on histology sections to the homologous tissue on the paraffin block faces and on the formalin-fixed tissue slices. Specifically, we assessed which of the four evaluated classes of transformation (rigid, similarity, affine or TPS) best mapped homologous landmarks on histology images (Figure 3.2f, labeled  $\mathbf{h}_{i,j}$ ), on paraffin images (Figure 3.2e, labeled  $\mathbf{p}_{i,j}$ ) and on tissue slice MR images (Figure 3.2c-d, labeled  $\mathbf{m}_{i,j}$ ). To measure the spatial misalignment of tissue sections induced by microtome cutting (Question 2), we characterized the locations from which histology sections were taken from within tissue slices, by estimating the depth (Figure 3.2c, labeled  $d_j$ ) and the orientation (Figure 3.2c, labeled  $\theta_j$ ) relative to the front face of the tissue slice from which the histology sections were taken. To assess the impact of the choice reconstruction model (i.e. choice of deformation model, plus choice of whether or not to make the front face assumption) on 3D reconstruction error (Question 3), we estimated two reconstruction error measures, the TRE and the FRE, using different reconstruction models. The impact of these assumptions will depend, in part, on the reconstruction algorithm that is used. We used least-squares best-fit alignment of manually identified homologous intrinsic fiducials for these measurements. This approach is parameter-free, has an analytic solution for each considered reconstruction model and has an accuracy that depends only on the number and placement of fiducials and not on image properties. In the four subsections that follow, we describe the details of the selection of these landmarks, as well as the details of the methods used to address each of the three central questions of this work.



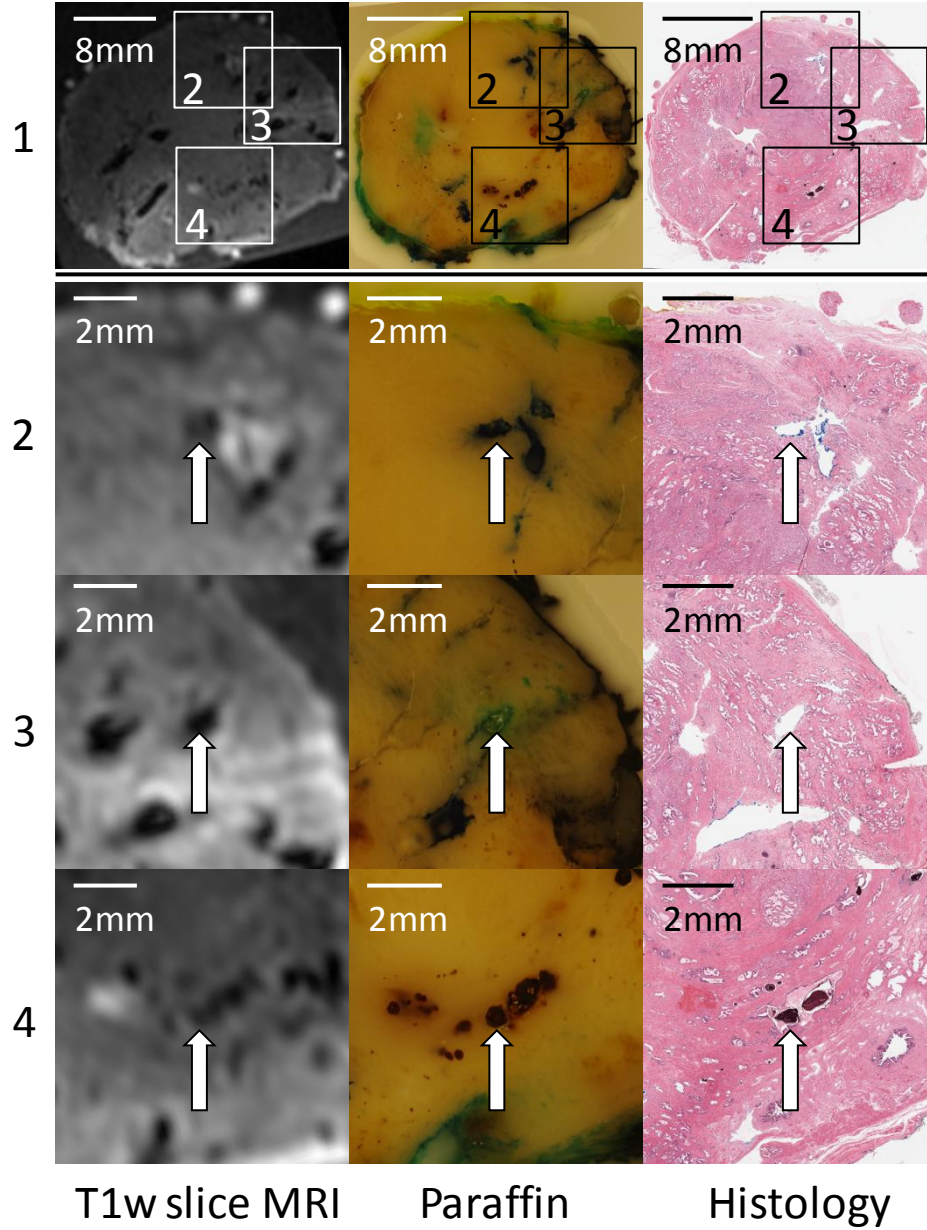


**Figure 3.2:** Schematic representations of tissue, landmarks and measurements, including (a) a surface rendering of a tissue slice MR image; (b) a schematic rendering of the tissue slice with front face fiducials  $f_{i,j}$ , and the best-fit front face plane  $F_j$ ; (c) a projected side view of the tissue slice as oriented over the microtome blade with front face fiducials  $f_{i,j}$ , the best fit front face plane  $F_j$ , histology-visible landmarks  $m_{i,j}$ , the best-fit histology section plane  $H_j$ , the orientation  $\theta_j$  and the depth measurement  $d_j$ ; (d) a schematic rendering of the tissue slice with histology-visible landmarks and best-fit histology section plane  $H_j$ ; (e) a schematic rendering of the paraffin block face after histological sectioning showing homologous landmarks  $p_{i,j}$  and (f) a schematic rendering of the corresponding histology section with homologous landmarks  $h_{i,j}$ .

### 3.2.3 Identification of landmarks and tissue slice faces

All of these measurements rely on identifying homologous landmarks in histology images, paraffin images and tissue slice MR images. For each tissue slice, we identified 7–15 distinct landmarks (162 in total), comprising the centers of atrophic ducts, cysts and corpora amylacea with diameters less than 1 mm. Illustrative examples of these images and identified homologous landmarks are shown in Figure 3.3. The positions of the landmarks (2D for histology and paraffin images, 3D for tissue slice MR images) on these modalities are denoted  $\mathbf{h}_{i,j}$ ,  $\mathbf{p}_{i,j}$ , and  $\mathbf{m}_{i,j}$ , respectively, for  $i$ -th landmark on the  $j$ -th tissue slice. Landmarks were interactively localized using 3D Slicer (Surgical Planning Lab, Harvard Medical School, Boston, USA), which required that the histology images be loaded into random-access memory; as the full resolution images typically occupy 15–

20 GB, we downsampled them to a  $30 \times 30 \mu\text{m}$  voxel size for landmark identification, yielding images 10–40 MB in size.



**Figure 3.3:** Illustrative examples of the T1-weighted tissue slice MR (left), paraffin (middle) and histology (right) images transformed by a best-fit affine transformation aligning manually identified landmarks. The three highlighted regions in row 1 are shown magnified in rows 2, 3 and 4, with corresponding landmarks denoted by arrows.

We estimated the front face of the  $j$ -th tissue slice by first manually identifying seven 3D points  $\{\mathbf{f}_{i,j} | i = 1..7\}$  evenly distributed across the face of the  $j$ -th tissue slice on the tissue slice MR image (approximately in the configuration shown in Figure 3.2b), and then computing the least-squares best-fit plane  $F_j$  to these points.

Because variability in landmark localization introduces uncertainty into spatial relationships measured in this work, the fiducial localization error (FLE) was estimated on histology (FLE<sub>h</sub>), paraffin (FLE<sub>p</sub>) and tissue slice MR (FLE<sub>m</sub>) images. Based on previous measurements using these histology images and MR images of intact prostate specimens using the same protocols [16], FLE<sub>h</sub> and FLE<sub>m</sub> were taken to be 0.05 mm and 0.16 mm respectively. These measurements quantified FLE as an unbiased estimator of the standard deviation of repeated localizations of landmarks. FLE<sub>p</sub> was estimated to be 0.05 mm, the same as FLE<sub>h</sub>, because the types of fiducials and the pixel sizes were similar (30  $\mu\text{m}$  for histology images, 18  $\mu\text{m}$  for paraffin images).

### 3.2.4 Tissue deformation due to histology processing (Question 1)

To assess the fidelity of the different deformation models, we quantified deformation between formalin-fixed tissue slices, paraffin blocks and histology sections under each of the models. Thus, we assessed the deformation due to three histology processes: (1) paraffin processing and embedding (denoted with superscript  $m \leftarrow p$ , corresponding to the transformation  $T_j^{m \leftarrow p}$  from the paraffin images  $[p]$  to the tissue slice MR images  $[m]$ ), (2) histological sectioning and mounting (denoted with superscript  $p \leftarrow h$ , corresponding to the transformation  $T_j^{p \leftarrow h}$  from the histology images  $[h]$  to the paraffin images  $[p]$ ), and (3) the combination of both processes (denoted with superscript

$m \leftarrow h$ , corresponding to the transformation  $T_j^{m \leftarrow h}$  from the histology images  $[h]$  to the tissue slice MR images  $[m]$ ).

Deformation models were compared using the mean TRE of homologous landmark pairs identified on images before and after each process after landmark-based registrations constrained by 4 deformation models of increasing flexibility: rigid, similarity (rigid + isotropic scaling), affine (rigid + scaling + skewing), and non-linear TPS [17]. The mean TRE was estimated as the misalignment between homologous landmarks after transformation by a least-squares best-fit transformation constrained by the deformation model, and was calculated using a leave-one-out cross-validation:

$$TRE^{t \leftarrow s, k} = \text{mean}_{j \in \{1..J\}, v \in \{1..I_j\}} \|T_{\hat{P}_{v,j}, \hat{Q}_{v,j}}^{t \leftarrow s, k}(\mathbf{s}_{v,j}) - \mathbf{t}_{v,j}\|, \quad 3.1$$

where  $T_{\hat{P}, \hat{Q}}^{t \leftarrow s, k}$  is the transformation of type  $k \in \{\text{rigid}, \text{similarity}, \text{affine}, \text{TPS}\}$  that best maps the vector of image landmarks  $\hat{Q}_{v,j} = \langle \mathbf{s}_{i,j} | i \in \{1..I_j\}, i \neq v \rangle$  from the  $j$ -th slice on the source modality  $s \in \{h, p\}$  to the vector of image landmarks  $\hat{P}_{v,j} = \langle \mathbf{t}_{i,j} | i \in \{1..I_j\}, i \neq v \rangle$  from the  $j$ -th slice on the target modality  $t \in \{p, m\}$ ,  $J = 21$  tissue slices, and  $v$  denotes which landmark is left out. For example,  $T_{\hat{P}_{3,2}, \hat{Q}_{3,2}}^{m \leftarrow h, \text{rigid}}$  is the best-fit rigid transformation for the second tissue slice that maps from histology to tissue slice MR image coordinates that is fit to all but the third identified fiducial.

The sensitivity of these measurements to FLE depends on the spatial configuration and number of landmarks identified for each tissue slice. For example, the best-fit transformation, and hence the leave-one-out TRE, of a section with few landmarks centrally clustered near the urethra and one landmark near the prostate boundary could be more sensitive to a misplaced landmark than that of a section with many widely-spaced landmarks. Although relationships between TRE and FLE with

respect to spatial distribution of the landmarks have been characterized for rigid transformations [18], to the best of our knowledge there is no closed form solution for calculating this sensitivity for a leave-one-out TRE for all four deformation models. Thus, we assessed this sensitivity instead by Monte Carlo simulation. For each tissue slice, landmarks on histology, paraffin, and tissue MR images were modeled as  $\mathbf{h}_{i,j} + \mathbf{G}^{2D}$ ,  $\mathbf{p}_{i,j} + \mathbf{G}^{2D}$ , and  $\mathbf{m}_{i,j} + \mathbf{G}^{3D}$ , respectively, where  $\mathbf{G}^{2D}$  is a 2D Gaussian random variable sampled for each landmark with  $x$  and  $y$  components distributed as  $N\left(0, \frac{\text{FLE}_h^2}{2}\right)$ , and  $\mathbf{G}^{3D}$  is a 3D Gaussian random variable sampled for each landmark with  $x$ ,  $y$  and  $z$  components distributed as  $N\left(0, \frac{\text{FLE}_m^2}{3}\right)$ . For each tissue slice, the TRE measurements were calculated for 5,000 sets of perturbed landmarks, and the standard deviation of these measurements was calculated. The sensitivity of the TRE measurements to FLE was quantified as the average of these standard deviations across all tissue slices. The number of samples was chosen such that the standard error of the standard deviation would be 1% of the standard deviation itself.

### 3.2.5 Spatial misalignment of tissue sections induced by microtome cutting (Question 2)

To assess the strength of the front face assumption, we quantified the depth and orientation of histology sections relative to the front face of the tissue slices from which they were cut. The depth and orientation of each histology section were both estimated based on the spatial relationship between two planes: the best-fit plane  $F_j$  through the points identified on the front face of the tissue slice in the tissue slice MR image and the plane  $H_j$ , an estimate of the tissue from which histology section was cut, computed as the

best-fit plane through the landmark points  $\{\mathbf{m}_{i,j} | i = 1..I_j\}$  in the tissue slice MR image corresponding to homologous landmarks visible on the histology image. The orientation  $\theta_j$  of the  $j$ -th histology section within the corresponding tissue slice was measured as the angle between the normal of plane  $F_j$  and plane  $H_j$ : specifically,  $\theta_j = \cos^{-1}(|\mathbf{n}(F_j) \cdot \mathbf{n}(H_j)|)$ , where  $\mathbf{n}(P)$  is the 3D unit normal of plane  $P$ . The depth  $d_j$  of the  $j$ -th histology section from within the tissue slice was measured as the minimum, average and maximum distances from the tissue points  $\Omega_j$  (the intersection of plane  $H_j$  with tissue identified on tissue slice MR) to the front face plane  $F_j$ : specifically,  $d_{min,j} = \min_{\omega \in \Omega_j} (D(F_j, \omega))$ ,  $d_j = \text{mean}_{\omega \in \Omega_j} (D(F_j, \omega))$  and  $d_{max,j} = \max_{\omega \in \Omega_j} (D(F_j, \omega))$ , where  $D(P, \omega)$  is the distance from 3D point  $\omega$  to the plane  $P$ . Tissue points on the tissue slice MR image were identified by a threshold-based segmentation of the T1-weighted tissue slice MR image using a manually selected threshold, followed by manual editing.

The sensitivity of these measurements to FLE also depends on the spatial configuration and number of landmarks identified for each tissue slice. For example, the estimated orientation of a section with few landmarks centrally clustered near the urethra could be more sensitive to a misplaced landmark than that of a section with many widely-spaced landmarks. Because, to the best of our knowledge, there is no closed form solution for calculating this sensitivity with respect to spatial distribution of the landmarks, we assessed it instead by Monte Carlo simulation. For each tissue slice, histology-visible landmarks on tissue slice MR images were modeled as  $\mathbf{m}_{i,j} + \mathbf{G}^{3D}$ , and front face landmarks were modeled as  $\mathbf{f}_{i,j} + \mathbf{G}^{3D}$ . As was done for Question 1, the sensitivity of the depth and orientation measurements to FLE was measured as the mean,

across all tissue slices, of the standard deviation of each measure across 5,000 sets of perturbed landmarks.

### 3.2.6 Impact of reconstruction model on 3D reconstruction error (Question 3)

For a reconstruction algorithm that uses a particular reconstruction model (i.e. a specified deformation model with or without the front face assumption), reconstruction accuracy may decrease if the true spatial relationships between histology sections and the tissue slices from which they were cut are different from the assumed constraints. While the impact will depend on the 3D reconstruction methods used, it can be explored by examining reconstructions based on the least-squares best-fit transformation of identified homologous intrinsic landmarks under various reconstruction models. This reconstruction approach was chosen because the reconstructions are parameter-free, they can be solved analytically avoiding reconstruction errors due to local optima, and their accuracy depends only on the number and placement of the fiducials and not on image properties.

The impact of the reconstruction model was quantified using the TRE (calculated in a leave-one-out manner). This is analogous to the TRE described in Equation 3.1, but with an expanded set of transformation types  $\vec{k}$  that includes the deformation types  $\{rigid, similarity, affine, TPS\}$  both with and without an additional constraint imposed by the front face assumption. Because the TPS transformation is an interpolating spline (i.e. source fiducials used to define the transformation are mapped exactly to target fiducials) and target fiducials may lie at a non-zero depth from the front face, the front face assumption cannot be directly applied. However, a transformation that does satisfy the front face assumption can be realized by first projecting the target fiducials used to

define the transformation onto the front face, and then defining a TPS transformation from source fiducials to the projected target fiducials. For rigid, similarity and affine transformations, the constrained least-squares fitting (constrained by the front face assumption) of transformed source fiducials to target fiducials is mathematically equivalent to the unconstrained least-squares fitting of transformed source fiducials to the projected target fiducials. Thus, for the reconstructions where the front face assumption was made, the target fiducials were projected onto the front face for all four deformation models. The sensitivity of TRE to FLE was quantified as for the TRE in Question 1.

In addition to quantifying the reconstruction error for these particular reconstructions, we can also calculate the lower bound on reconstruction error as measured by the identified landmarks for any possible reconstruction algorithm constrained by a particular reconstruction model. This lower bound is quantified as the FRE,

$$\text{FRE}^{t \leftarrow s, \vec{k}} = \text{mean}_{j \in \{1..J\}, i \in \{1..I_j\}} \left\| T_{P_j, Q_j}^{t \leftarrow s, \vec{k}}(\mathbf{s}_{i,j}) - \mathbf{t}_{i,j} \right\|, \quad 3.2$$

where  $T_{P_j, Q_j}^{t \leftarrow s, \vec{k}}$  is the transformation of type  $\vec{k}$  (the expanded set of transformation types described in the previous paragraph) that best maps the vector of image landmarks  $Q_j = \langle \mathbf{s}_{i,j} | i \in \{1..I_j\} \rangle$  from the  $j$ -th slice on the source modality  $s \in \{h, p\}$  to the vector of image landmarks  $P_j = \langle \mathbf{t}_{i,j} | i \in \{1..I_j\} \rangle$  from the  $j$ -th slice on the target modality  $t \in \{p, m\}$  and  $J = 21$  tissue slices. Note that unlike the TRE, the FRE includes all fiducials when fitting the transformation  $T_{P_j, Q_j}^{t \leftarrow s, \vec{k}}$ , and represents a lower bound on the TRE as measured using the identified intrinsic landmarks. Because the TPS



transformation is an interpolating spline (i.e. source fiducials used to define the transformation are mapped exactly to target fiducials) the FRE of an unconstrained TPS transformation is 0, by construction, for any configuration of fiducials. The sensitivity of FRE to FLE was quantified as for the TRE in Question 1.

### 3.2.7 Statistical analysis

Statistical analyses were performed in SPSS 20 (IBM, Chicago, USA). The depth and orientation measurements were characterized with descriptive statistics (mean and standard deviation) and 95% confidence intervals (CI) on the means were computed. Correlations of the depth measurements with the orientation were assessed using pairwise Spearman correlations.

The TRE measurements quantifying deformation during paraffin processing and embedding, histological sectioning and mounting, and the combination of both processes were characterized with descriptive statistics. We assessed differences in mean TRE between the deformation models using separate 1-way repeated-measures analysis of variance (ANOVA) tests with Greenhouse–Geisser correction for asphericity with the deformation model as the factor. Pairwise *post hoc* analysis of adjacent levels (i.e. rigid vs. similarity, similarity vs. affine and affine vs. TPS deformation models) was performed by constructing 95% CI on the differences in mean TRE.

To assess the impact of reconstruction assumptions, we assessed the differences in mean TRE using a 2-way repeated-measures ANOVA with a Greenhouse–Geisser correction for asphericity, with the two assumptions (the deformation model assumption and the front face assumption) as factors. Pairwise *post hoc* analysis was performed by constructing 95% CI on the difference in FRE and TRE due to the front face assumption

under each deformation model, and on the pairwise differences due to the deformation model between adjacent levels with and without the front face assumption. Note that the FRE of a reconstruction under stricter assumptions is mathematically guaranteed to be equal to or higher than the FRE under relaxed assumption. For example, the rigid deformation assumption is stricter than the affine deformation assumption, and the front face assumption is stricter than eliminating that assumption.

### 3.3 Results

#### 3.3.1 Tissue deformation due to histology processing (Question 1)

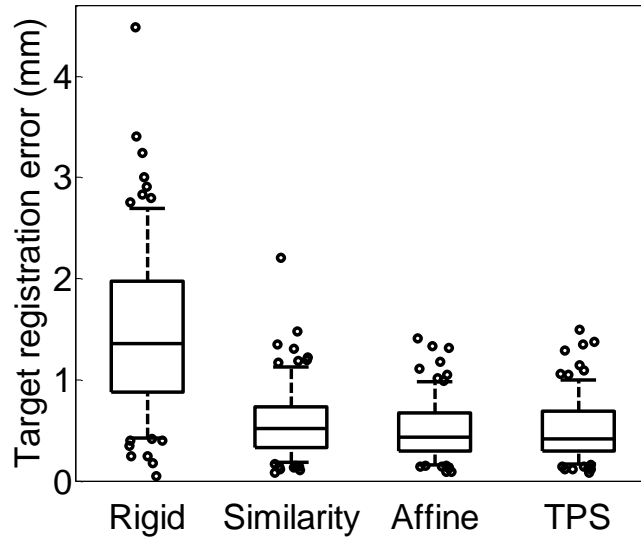
The key finding regarding tissue deformation was that modeling isotropic scaling as in the similarity deformation model improved the mean TRE by 0.8–1.0 mm, while modeling skew or thin-plate-spline deformation improved the mean TRE by less than 0.1 mm (bolded intervals in the first row of Table 3.1). The mean and standard deviation of TRE for the three histology processes under the four deformation models are shown in Table 3.2. For the combined deformation from tissue slice to histology section (shown in the first row of Tables 3.1 and 3.2 and as a box plot in Figure 3.4), the similarity model had a significantly lower mean TRE than the rigid model and the affine model has a significantly lower mean TRE than the similarity model (by 0.9 mm and 0.06 mm differences respectively), but *post hoc* analyses failed to show a statistically significant difference between the affine and TPS deformation models (0.005 mm difference).

**Table 3.1:** *Post hoc* analyses comparing mean TRE under varying deformation models: 95% CI of TRE for model A – model B. Key findings are shown in bold.

<i>Deformation model A</i>	<i>Rigid</i>	<i>Similarity</i>	<i>Affine</i>
<i>Deformation model B</i>	<i>Similarity</i>	<i>Affine</i>	<i>Thin-plate spline</i>
Tissue slice MR to histology images	<b>[0.78,0.98]</b>	<b>[0.03,0.10]</b>	<b>[-0.02,0.02]</b>
Tissue slice MR to paraffin images	[1.07,1.31]	[-0.04,0.01]	[-0.03,0.01]
Paraffin to histology images	[0.12,0.20]	[0.02,0.07]	[-0.01,0.02]

**Table 3.2:** Mean $\pm$ SD TRE (mm) for four models of deformation during histological processing stages. Statistical comparisons (performed between adjacent columns) where the statistical tests failed to detect a significant difference are connected by lines.

<i>First image</i>	<i>Second image</i>	<i>Rigid</i>	<i>Similarity</i>	<i>Affine</i>	<i>Thin-plate spline</i>
Tissue slice MR	Histology	1.44 $\pm$ 0.73	0.56 $\pm$ 0.31	0.50 $\pm$ 0.27	— 0.50 $\pm$ 0.28
Tissue slice MR	Paraffin	1.71 $\pm$ 0.82	0.52 $\pm$ 0.26	— 0.54 $\pm$ 0.26	— 0.54 $\pm$ 0.28
Paraffin	Histology	0.42 $\pm$ 0.27	0.26 $\pm$ 0.18	0.22 $\pm$ 0.14	— 0.21 $\pm$ 0.14



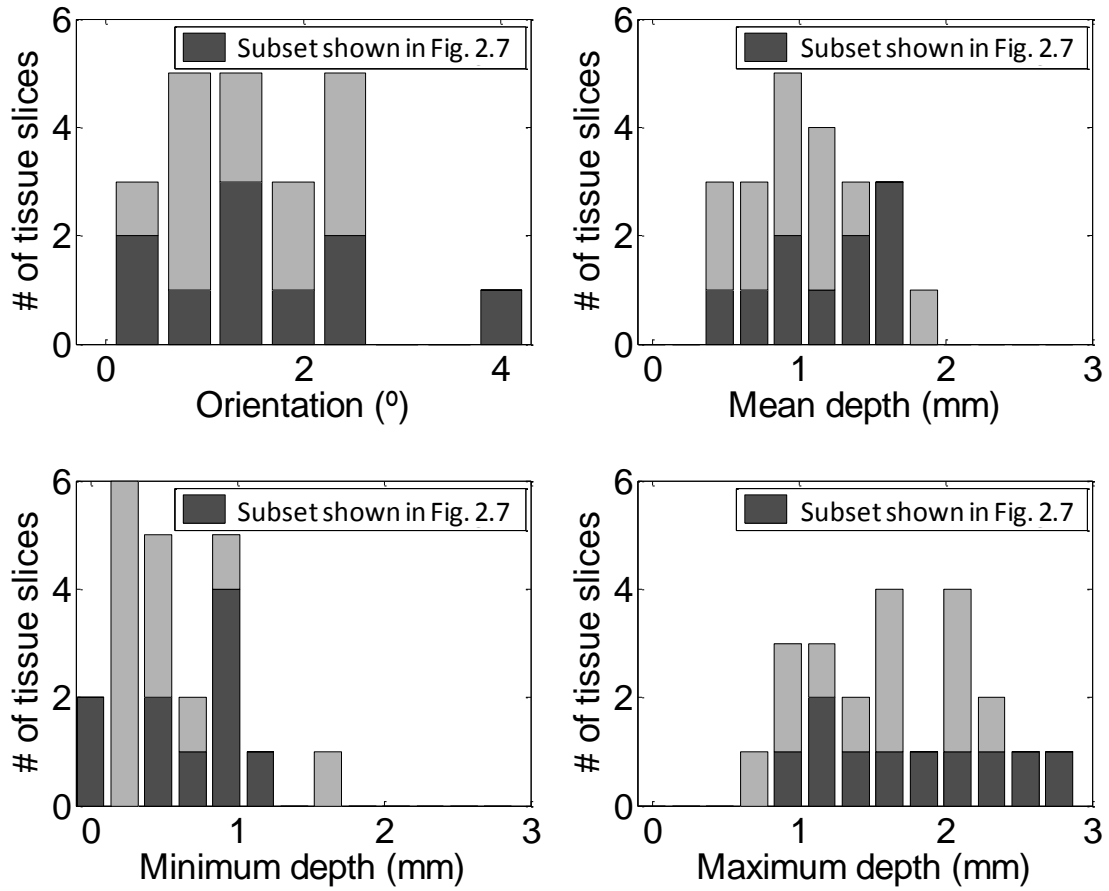
**Figure 3.4:** Box plot showing the target registration errors of homologous landmarks under four deformation models for the tissue deformation due to histological processing and cutting. These results correspond to the descriptive statistics shown in the first row of Table 3.1.

For the intermediate deformation due to paraffin embedding (second row of Tables 3.1 and 3.2), the rigid model had a significantly higher mean TRE than the similarity model (by a 1.2 mm difference), but *post hoc* analyses failed to show a

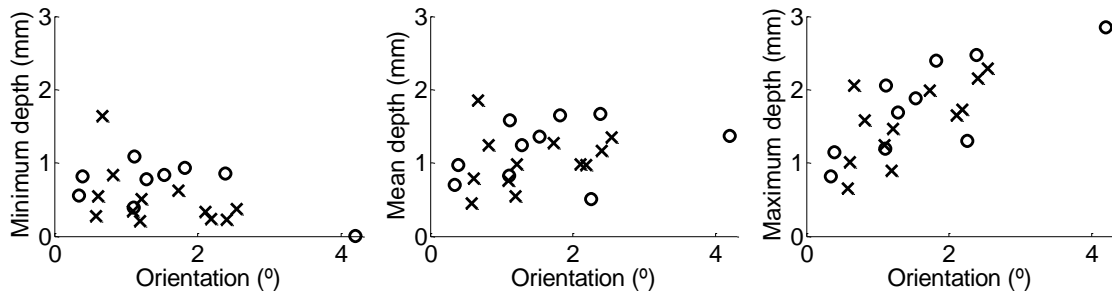
statistically significant difference between the other models (differences  $< 0.03$  mm). For the intermediate deformation due to histological sectioning (third row of Tables 3.1 and 3.2), the affine model had a significantly lower mean TRE (by 0.05 mm) than the similarity model and the similarity model had a significantly lower mean TRE (by 0.2 mm) than the rigid model. *Post hoc* analyses failed to show a statistically significant difference between the affine and TPS deformation models (0.009 mm difference). The sensitivities of the TRE to the observed FLE ranged from 0.05 to 0.13 mm.

### 3.3.2 Spatial misalignment of tissue sections induced by microtome cutting (Question 2)

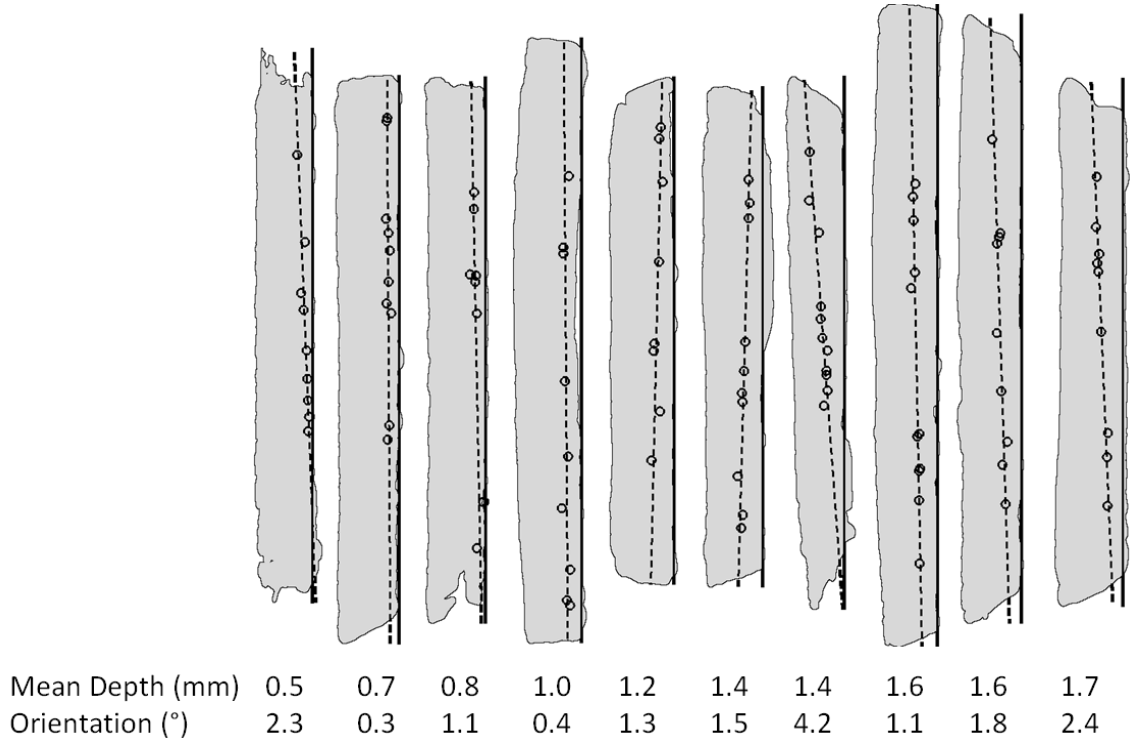
The key finding regarding the spatial misalignment of tissue sections was that the 95% confidence interval on the mean of orientation was  $1.1$  to  $1.9^\circ$  and the 95% confidence interval on the mean of mean depth was 0.9 to 1.3 mm (bolded intervals in the second column of Table 3.3). The distributions of depth and orientation measurements are shown in Figure 3.5, and the correlation plots of orientation with the minimum (Spearman  $r = -0.4$ ), mean (Spearman  $r = 0.4$ ) and maximum (Spearman  $r = 0.75$ ) section depth measures are shown in Figure 3.6. A subset of the tissue slices, chosen to illustrate the range of depths and orientations, are shown in Figure 3.7 with the front face  $F_j$  and the best fit plane to  $H_j$  superimposed. The standard deviation, 95% CI on the mean, and sensitivity to FLE for the orientation, minimum section depth, mean section depth, and maximum section depth are shown in Table 3.3.



**Figure 3.5:** Histograms of histology section depths and orientations. The subset of tissue slices illustrated in Figure 3.7 is shown in dark gray.



**Figure 3.6:** Correlation of minimum, mean and maximum histology section depths with orientations. Tissue slices corresponding to sections marked with circles are shown in Figure 3.7.



**Figure 3.7:** Renderings of the spatial relationships between tissue slices, histology-visible landmarks, front face plane and histology section planes for 10 tissue slices, ordered by increasing mean depth from left to right. Each tissue slice is shown as a silhouette projected along  $n(F_j) \times n(H_j)$ , the cross-product of the front face and histology section plane normals. With this projection, the front face plane  $F_j$  and histology section plane  $H_j$  can be represented as solid and dashed lines respectively. The projected histology-visible landmarks  $m_{i,j}$  are shown as circles.

**Table 3.3:** Descriptive statistics for the orientation, minimum depth, mean depth and maximum depth of histology sections relative to the tissue blocks from which they were cut. Sensitivity to FLE was quantified as the mean of the standard deviation of measurements in a Monte Carlo simulation with perturbed landmark positions. Key findings are shown in bold.

	<i>Standard deviation 95% CI on mean Sensitivity to FLE</i>		
<b>Orientation (°)</b>	0.9	<b>[1.1,1.9]</b>	0.30
<b>Minimum depth (mm)</b>	0.4	[0.4,0.7]	0.12
<b>Mean depth (mm)</b>	0.4	<b>[0.9,1.3]</b>	0.05
<b>Maximum depth (mm)</b>	0.6	[1.4,1.9]	0.13

### 3.3.3 Impact of reconstruction model on 3D reconstruction error (Question 3)

The two key findings regarding the impact of the reconstruction model on 3D reconstruction error were as follows. (1) Modeling isotropic scaling (as in the similarity deformation model) improved the mean TRE by 0.5 to 0.7 mm if the front face assumption was made and by 0.8 to 1.0 mm if the front face assumption was not made, but modeling skew or thin-plate-spline deformation improved mean TRE by less than 0.1 mm (bolded intervals in Table 3.4). (2) Under a similarity deformation model, the front face assumption increased the mean TRE by 0.6 mm to 0.8 mm (bolded interval in Table 3.5). The mean and standard deviation of TRE and FRE for the intrinsic landmark-based reconstructions are shown in Tables 3.6 and 3.7, respectively. The 95% CI for the difference in TRE and FRE due to deformation model and due to the front face assumption are shown in Tables 3.4 and 3.5, respectively. The sensitivities of the TRE to the observed FLE ranged from 0.10 to 0.13 mm. The sensitivities of the FLE to the observed FLE were 0 mm (by construction) for the reconstruction model comprising the thin-plate-spline deformation model without the front face assumption and 0.07 to 0.09 mm for the remaining reconstruction models.

**Table 3.4:** *Post hoc* analyses comparing TRE/FRE after intrinsic landmark reconstruction under varying deformation models: 95% CI of TRE/FRE for model A – model B. Key findings are shown in bold.

<i>Deformation model A</i>	<i>Rigid</i>	<i>Similarity</i>	<i>Affine</i>
<i>Deformation model B</i>	<i>Similarity</i>	<i>Affine</i>	<i>Thin plate spline</i>
<b>95% CI of mean TRE (mm) with front face assumption</b>	<b>[0.54,0.70]</b>	<b>[0.01,0.06]</b>	<b>[-0.01,0.01]</b>
<b>95% CI of mean TRE (mm) without front face assumption</b>	<b>[0.78,0.98]</b>	<b>[0.03,0.10]</b>	<b>[-0.02,0.02]</b>
<b>95% CI of mean FRE (mm) with front face assumption</b>	[0.47,0.62]	[0.03,0.06]	[0.04,0.07]
<b>95% CI of mean FRE (mm) without front face assumption</b>	[0.75,0.94]	[0.07,0.13]	[0.29,0.35] <sup>†</sup>

†. The FRE after an unconstrained thin-plate-spline transformation is 0 mm by construction.

**Table 3.5:** *Post hoc* analyses comparing TRE/FRE after intrinsic landmark reconstruction under varying deformation models: 95% CI for of the difference between reconstruction with and without the front face assumption.

	<i>Rigid</i>	<i>Similarity</i>	<i>Affine</i>	<i>Thin-plate spline</i>
<b>95% CI of mean TRE (mm)</b>	[0.36,0.46]	<b>[0.60,0.75]</b>	[0.63,0.77]	[0.62,0.77]
<b>95% CI of mean FRE (mm)</b>	[0.40,0.51]	[0.68,0.83]	[0.74,0.89]	[1.00,1.15] <sup>†</sup>

†. The FRE after an unconstrained thin-plate-spline transformation is 0 mm by construction.

**Table 3.6:** Mean±SD TRE after intrinsic landmark reconstruction under varying constraints. Statistical comparisons (performed between adjacent columns and rows) where the statistical tests failed to detect a significant difference are connected by lines.

	<i>Rigid</i>	<i>Similarity</i>	<i>Affine</i>	<i>Thin-plate spline</i>
<b>With front face assumption</b>	1.85±0.71	1.23±0.48	1.20±0.47	— 1.20±0.47
<b>Without front face assumption</b>	1.44±0.73	0.56±0.31	0.50±0.27	— 0.50±0.28

**Table 3.7:** Mean±SD FRE after intrinsic landmark reconstruction under varying constraints. All statistical comparisons (performed between adjacent columns and rows) showed significant differences.

	<i>Rigid</i>	<i>Similarity</i>	<i>Affine</i>	<i>Thin-plate spline</i>
<b>With front face assumption</b>	1.72±0.66	1.18±0.48	1.13±0.48	1.08±0.50
<b>Without front face assumption</b>	1.27±0.66	0.42±0.25	0.32±0.19	0±0 <sup>†</sup>

†. The FRE after an unconstrained thin plate spline transformation is 0 by construction.



## 3.4 Discussion

*In vivo* prostate imaging is increasingly being validated against 3D reconstructed histology images [19-22]. Many algorithms for 3D reconstruction limit the degrees of freedom by making simplifying assumptions about the cutting of histology sections from the prostate gland which may affect the accuracy of reconstruction. This work explored two such assumptions: the deformation model assumption that histology sections have been deformed under a specified deformation model relative to the fixed tissue, and the front face assumption that histology sections correspond to the front face of the tissue slice from which it was cut. Operator variability in sectioning could lead to histology sections that are not taken coincident with or parallel to the front face of the tissue slice, and the cumulative deformation of the histology section due to dehydration, cutting, water-bath expansion and slide-mounting processes may not be accurately modeled by the chosen transformation. In this work, we quantified the spatial relationship between histology images and the formalin-fixed tissue slices from which they were taken and evaluated the impact of the reconstruction model assumptions on 3D reconstruction error.

### 3.4.1 Tissue deformation due to histology processing (Question 1)

Modeling the deformation due to paraffin processing and histological sectioning as affine deformation yielded the lowest mean TRE (0.5 mm), although the difference between the affine and similarity models was 0.06 mm, and our analysis failed to show a statistically significant difference between the affine and TPS models. The 0.5 mm TRE under the affine and TPS deformation model is larger than the 0.2 mm FLE, suggesting that there is some submillimeter-scale non-affine deformation that occurs, but that is not

well-captured by the interpolation of the TPS deformation model with the landmark configurations identified in this work.

The analysis of deformation from tissue block MR to paraffin images suggests that most of the deformation during the paraffin processing is characterized by isotropic scaling, which is consistent with the dehydration that occurs during this process. The analysis of deformation from paraffin to histology images suggests that some further affine deformation occurs during histological sectioning, which is consistent with anisotropic cutting forces that are applied during sectioning. The larger mean TRE across all deformation models and the larger change in TRE with isotropic scaling for the paraffin processing compared to the microtome sectioning suggests that paraffin processing is the source of most of the observed deformation. Notably, the mean TRE of the combined processes under the rigid deformation model is less than that of the paraffin processing alone, which is consistent with expansion on the water bath partially cancelling out contraction due to dehydration during paraffin processing.

To perform these analyses, at least five sets of homologous landmarks were required; in particular, four sets of landmarks are required for a 2D–3D TPS transformation to define a non-affine transformation and a fifth is needed to enable the leave-one-out evaluation. Three histology sections were omitted from the analysis because fewer than five sets of landmarks were identifiable.

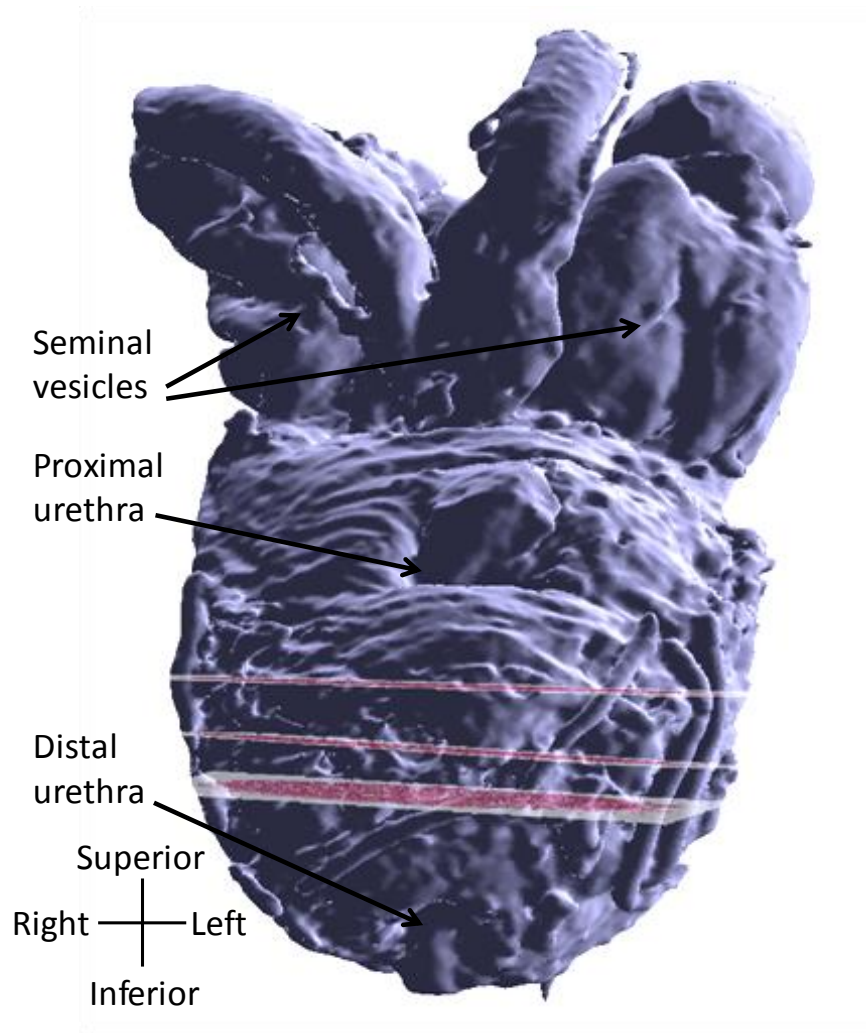
### 3.4.2 Spatial misalignment of tissue sections induced by microtome cutting (Question 2)

The histology sections were taken at a mean depth of 1.1 mm, and were taken at an average angle of  $1.5^\circ$  relative to the front face. To illustrate these values, a histology

section cut with the mean depth and the mean orientation from a hypothetical tissue slice 30 mm in diameter (typical for our sample of tissue slices), would be 0.7 mm from the front face at the closest point and 1.5 mm from the front face at the furthest point.

The standard deviations of the orientation ( $0.9^\circ$ ) and depth measurements (0.4 mm) are greater than would be expected due to the 0.16 mm FLE alone, suggesting that there is operator variability in the alignment of the tissue block face with the microtome blade and in the depth of cutting. The variability in the minimum section depth suggests that the variability in depth is not directly caused by variability in tissue block alignment; if the variability in depth of cutting were the result of variability in tissue block alignment followed by consistently cutting until a full cross-section of tissue were barely reached, we would expect the minimum depth to have low variability. We speculate that the continued cutting beyond the best-fit front face plane could be due to concavity of the tissue front face that can be introduced during paraffin embedding, which would require a deeper cut to achieve a full face. This continued cutting could also be due to the practice of removing the paraffin block from the microtome to cool the cutting surface with ice, leading to variability in the orientation when the block is replaced.

The impact of the observed variability in depth and orientation on the relative spatial relationship of histology sections in a 3D reconstruction can be seen in Figure 3.8, where the tissue slices were sliced to be parallel at an even spacing (by embedding the specimen in an agar gel and cutting it on a rotary slicer), but after an alignment of tissue slice MR images (and accompanying registered histology sections) with an MR image of the intact *ex vivo* specimen with a 0.5 mm TRE, the non-parallel and uneven spacing of the three midgland histology images can be seen.



**Figure 3.8:** 3D reconstruction of three histology sections aligned (with a mean TRE of 0.5 mm) to an anterior view of 3D surface rendering of the corresponding intact *ex vivo* prostate gland with seminal vesicles, illustrating the potential for non-parallel, non-evenly-spaced histological tissue sections.

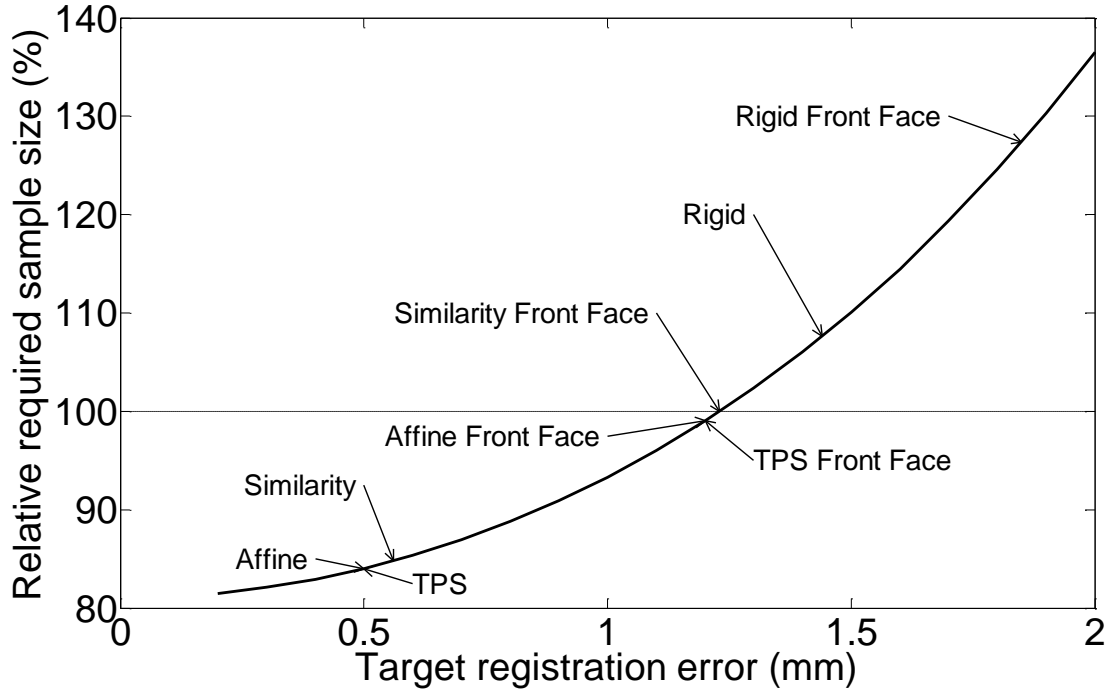
### 3.4.3 Impact of reconstruction model on 3D reconstruction error (Question 3)

If a reconstruction algorithm used a reconstruction model wherein histology sections corresponded to the front faces of tissue slices, the lower bound of the achievable mean TRE for any of the tested deformation models would be 1.1 mm (for the affine and TPS deformation models), suggesting that to achieve submillimeter

reconstruction error, the front face assumption should not be made. For a reconstruction model unconstrained by the front face assumption using an affine deformation model, the lower bound of the mean TRE is 0.3 mm (the corresponding lower bound of the mean TRE for the TPS model is 0 mm by construction and therefore does not provide for an informative comparison). The improvement in TRE for modeling isotropic scaling (from rigid to similarity deformation models) was 0.78 to 0.98 mm when the front face assumption was not made, but the improvement from the similarity deformation model to the more flexible affine and TPS deformation models was less than 0.1 mm, suggesting that a similarity transformation may be sufficient. Note that this reconstruction error is only a component of the overall registration error from the histology images to *in vivo* images; Groenendaal et al. [5] and Orczyk et al. [23] presented two methods for the registration of reconstructed *ex vivo* images to *in vivo* images with reported errors of 2.1 mm and 1.6 mm, which would, under the assumption that these errors are independent, be added in quadrature with reconstruction error.

It is important to interpret these reconstruction errors in the context of the application in which the reconstructions could be used. A recent model (described in Chapter 2) quantifying the impact of registration error on the statistical power (and thus the required sample size) of imaging validation studies [14, 15] can be used to relate the differences in reconstruction error in these experiments to an appropriate application. This can be illustrated through the scenario of an imaging validation study testing for differences between the mean imaging signal of tumors and normal tissue regions, under the assumptions that the tumors are spherical foci of the smallest clinically significant volume ( $0.2 \text{ cm}^3$ ) [24] and that reconstruction error can be modeled as an isotropic

Gaussian. For this scenario, our reconstruction is combined with a registration of reconstructed histology to *in vivo* images with mean TRE 2.1 mm (as reported by Groenendaal et al. [5]). In this scenario, we can compare the required sample sizes for the imaging validation study over a range of reconstruction errors as compared to an arbitrarily chosen baseline. In this illustration, we use landmark-based reconstruction under the similarity deformation model and the front face assumption with a mean TRE of 1.23 mm as the baseline. The relative required sample sizes for reconstruction errors ranging from 0.3 – 2.0 mm mean TRE are shown in Figure 3.9, with each of the assessed combinations of assumptions marked. Between the worst and best performing sets of assumptions, there is a 1.5-fold difference in required sample size. Based on per-patient costs of \$10,000, from an ongoing imaging validation study in our centre with 66 subjects, a 1.5-fold reduction in sample size could yield savings of \$220,000 for the same statistical power.



**Figure 3.9:** Sample sizes, relative to an arbitrarily chosen baseline, for imaging validation studies of image signal differences between cancerous and background tissue for  $0.2 \text{ cm}^3$  cancer foci, under assumptions that foci are spherical and reconstruction error can be modeled as a translation error distributed as a 3D Gaussian and is combined in quadrature with a 2.1 mm TRE due to registration to *in vivo* imaging. Reconstructions under differing deformation assumptions, and with or without the front face assumption are indicated, with the reconstruction using a similarity transform and the front face assumption arbitrarily chosen as the 100% baseline reference.

#### 3.4.4 Limitations of cutting measurement

The conclusions of this work should be considered in the context of the limitations of the performed experiments. This study had four notable limitations. First, the process of sectioning tissue for histological examination varies between laboratories and many aspects of the processing can affect the sectioning distortion, and possibly the amount of trimming before a full cross-section is successfully cut. Examples include tissue type, embedding medium, water-bath duration and temperature [21], knife quality and angle [25], and possibly operator skill. It is also unclear how these results would generalize when prostate tissue slices are cut into quarters before paraffin processing, an

approach adopted in many clinical laboratories. While this study used histological sections sectioned by multiple histotechnologists, the use of a consistent processing protocol in a single clinical laboratory prevented us from assessing the impact of these other factors on the identified spatial relationships. A second limitation of the study is that the fidelity of the TPS deformation to the true underlying deformation is limited in part by the number of homologous landmarks identifiable on tissue slice MR, histology and paraffin block face images. It is not clear if the 7–15 landmarks identified per section are sufficient to characterize the unknown underlying deformation. If 7–15 landmarks are too few, then the TPS model may not capture the underlying deformation (resulting in a higher reported TRE) even if the deformation could be well-described by a TPS model. A third limitation of the study is that only one non-linear deformation model was assessed in this study, although there are an infinite number of such models. This work does, however, suggest an upper bound of less than 0.5 mm for the possible improvement that could be derived from better non-linear deformation models. Fourth, our assessment of the impact of the reconstruction errors on the statistical power of imaging validation studies examines the histology image reconstruction in isolation; if the reconstruction were followed by additional processing, such as 3D image registration, the impact of the reconstruction errors would be challenging to isolate, and was not assessed by this study.

In conclusion, this work addressed three questions, as follows. (1) How does prostate tissue deform during histology processing? (2) What spatial misalignment of the tissue sections is induced by microtome cutting? (3) How does the choice of reconstruction model affect the accuracy of histology reconstructions? The key conclusions from these investigations are that for accurate 3D reconstruction of whole-



mount histology, the reconstruction model should not assume that histology corresponds to the front face of the tissue slices from which it was cut as such an assumption yields a higher mean TRE by 0.6 to 0.8 mm, and should use a similarity deformation model as the mean TRE under this model is 0.5 to 0.7 mm lower than that of a rigid deformation model and within 0.1 mm of the affine and thin-plate-spline deformation models with more degrees of freedom. The mean TRE of 0.56 mm was measured for the least-squares best-fit fiducial-based reconstruction using a similarity deformation model without the front face assumption. Additionally, our characterization of the misalignment of histology sections revealed a mean section depth of 1.1 mm (with maximum depths as high as 2.8 mm) and a mean section orientation of  $1.5^\circ$  (with orientations as high as  $4.2^\circ$ ), which may support commensurate heuristics in 3D reconstruction. Finally, in the context of imaging validation studies testing for imaging signal differences between cancerous and background tissue for the smallest clinically significant prostate cancer foci by correlation with reconstructed histology images, the range of reconstruction errors seen in this work would result in a 1.5-fold difference in the required sample size for such a study under our modeling assumptions, potentially translating to a difference of hundreds of thousands of dollars.

### 3.5 References

1. C. Hughes, O. Rouvière, F. Mege-Lechevallier, R. Souchon and R. Prost, "Robust alignment of prostate histology slices with quantified accuracy," *IEEE Transactions on Biomedical Engineering* **60**, 281–291 (2012).
2. E. Gibson, C. Crukley, M. Gaed, J. A. Gómez, M. Moussa, J. L. Chin, G. S. Bauman, A. Fenster and A. D. Ward, "Registration of prostate histology images to ex vivo MR images via strand-shaped fiducials," *Journal of Magnetic Resonance Imaging* **36**, 1402–1412 (2012).

3. A. D. Ward, C. Crukley, C. A. McKenzie, J. Montreuil, E. Gibson, C. Romagnoli, J. A. Gómez, M. Moussa, J. Chin, G. Bauman and A. Fenster, "Prostate: registration of digital histopathologic images to in vivo MR images acquired by using endorectal receive coil," *Radiology* **263**, 856–864 (2012).
4. J. Chappelow, B. N. Bloch, N. Rofsky, E. Genega, R. Lenkinski, W. DeWolf and A. Madabhushi, "Elastic registration of multimodal prostate MRI and histology via multiattribute combined mutual information," *Medical Physics* **38**, 2005–2018 (2011).
5. G. Groenendaal, M. R. Moman, J. G. Korporaal, P. J. van Diest, M. van Vulpen, M. E. Philippens and U. A. van der Heide, "Validation of functional imaging with pathology for tumor delineation in the prostate," *Radiotherapy and Oncology* **94**, 145–150 (2010).
6. B. Castaneda, K. Hoyt, M. Zhang, D. Pasternack, L. Baxter, P. Nigwekar, A. di Sant'Agnese, J. Joseph, J. Strang, D. J. Rubens and K. J. Parker, "P1C-9 prostate cancer detection based on three dimensional sonoelastography," in *Proceedings of IEEE Ultrasonics Symposium*, (New York, USA, 2007), pp. 1353–1356.
7. L. S. Taylor, B. C. Porter, G. Nadasdy, P. A. di Sant'Agnese, D. Pasternack, Z. Wu, R. B. Baggs, D. J. Rubens and K. J. Parker, "Three-dimensional registration of prostate images from histology and ultrasound," *Ultrasound in Medicine and Biology* **30**, 161–168 (2004).
8. Y. Zhan, Y. Ou, M. Feldman, J. Tomaszewski, C. Davatzikos and D. Shen, "Registering histologic and MR images of prostate for image-based cancer detection," *Academic Radiology* **14**, 1367–1381 (2007).
9. H. Park, M. R. Piert, A. Khan, R. Shah, H. Hussain, J. Siddiqui, T. L. Chenevert and C. R. Meyer, "Registration methodology for histological sections and in vivo imaging of human prostate," *Academic Radiology* **15**, 1027–1039 (2008).
10. V. Shah, T. Pohida, B. Turkbey, H. Mani, M. Merino, P. A. Pinto, P. Choyke and M. Bernardo, "A method for correlating in vivo prostate magnetic resonance imaging and histopathology using individualized magnetic resonance-based molds," *Review of Scientific Instruments* **80**, 104301–104306 (2009).
11. A. S. Jackson, S. A. Reinsberg, S. A. Sohaib, E. M. Charles-Edwards, S. Jhavar, T. J. Christmas, A. C. Thompson, M. J. Bailey, C. M. Corbishley, C. Fisher, M. O. Leach and D. P. Dearnaley, "Dynamic contrast-enhanced MRI for prostate cancer localization," *British Journal of Radiology* **82**, 148–156 (2009).
12. L. H. Chen, H. Ho, R. Lazaro, C. H. Thng, J. Yuen, W. S. Ng and C. Cheng, "Optimum slicing of radical prostatectomy specimens for correlation between

- histopathology and medical images," *International Journal of Computer Assisted Radiology and Surgery* **5**, 471–487 (2010).
13. M. S. Breen, T. L. Lancaster, R. S. Lazebnik, S. G. Nour, J. S. Lewin and D. L. Wilson, "Three-dimensional method for comparing in vivo interventional MR images of thermally ablated tissue with tissue response," *Journal of Magnetic Resonance Imaging* **18**, 90–102 (2003).
  14. E. Gibson, A. Fenster and A. D. Ward, "Registration accuracy: how good is good enough? A statistical power calculation incorporating image registration uncertainty," in *Proceedings of Medical Image Computing and Computer Assisted Intervention*, Vol. 7511 (Springer, Nice, France, 2012), pp. 643–650.
  15. E. Gibson, A. Fenster and A. D. Ward, "The impact of registration accuracy on imaging validation study design: a novel statistical power calculation," *Medical Image Analysis* **17**, 805–815 (2013).
  16. J. M. Fitzpatrick, J. B. West and C. R. Maurer, Jr., "Predicting error in rigid-body point-based registration," *IEEE Transactions on Medical Imaging* **17**, 694–702 (1998).
  17. F. L. Bookstein, "Principal warps: thin-plate splines and the decomposition of deformations," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11**, 567–585 (1989).
  18. J. M. Fitzpatrick and J. B. West, "The distribution of target registration error in rigid-body point-based registration," *IEEE Transactions on Medical Imaging* **20**, 917–927 (2001).
  19. R. Garcia-Parra, D. Wood, R. B. Shah, J. Siddiqui, H. Hussain, H. Park, T. Desmond, C. Meyer and M. Piert, "Investigation on tumor hypoxia in resectable primary prostate cancer as demonstrated by 18F-FAZA PET/CT utilizing multimodality fusion techniques," *European Journal of Nuclear Medicine and Molecular Imaging* **38**, 1816–1823 (2011).
  20. E. Mena, B. Turkbey, H. Mani, S. Adler, V. A. Valera, M. Bernardo, V. Shah, T. Pohida, Y. McKinney, G. Kwarteng, D. Daar, M. L. Lindenberg, P. Eclarinal, R. Wade, W. M. Linehan, M. J. Merino, P. A. Pinto, P. L. Choyke and K. A. Kurdziel, "11C-Acetate PET/CT in localized prostate cancer: a study with MRI and histopathologic correlation," *Journal of Nuclear Medicine* **53**, 538–545 (2012).
  21. M. Piert, H. Park, A. Khan, J. Siddiqui, H. Hussain, T. Chenevert, D. Wood, T. Johnson, R. B. Shah and C. Meyer, "Detection of aggressive primary prostate cancer with 11C-choline PET/CT using multimodality fusion techniques," *Journal of Nuclear Medicine* **50**, 1585–1593 (2009).

22. G. Groenendaal, A. Borren, M. R. Moman, E. Monninkhof, P. J. van Diest, M. E. Philippens, M. van Vulpen and U. A. van der Heide, "Pathologic validation of a model based on diffusion-weighted imaging and dynamic contrast-enhanced magnetic resonance imaging for tumor delineation in the prostate peripheral zone," *International Journal of Radiation Oncology, Biology, Physics* **82**, e537–544 (2012).
23. C. Orczyk, A. Mikheev, A. Rosenkrantz, J. Melamed, S. S. Taneja and H. Rusinek, "Imaging of prostate cancer: a platform for 3D co-registration of in-vivo MRI ex-vivo MRI and pathology," in *Proceedings of SPIE Medical Imaging, Vol. 8316* (2012), pp. 83162M.
24. J. I. Epstein, P. C. Walsh, M. Carmichael and C. B. Brendler, "Pathologic and clinical findings to predict tumor extent of nonpalpable (stage T1c) prostate cancer," *Journal of the American Medical Association* **271**, 368–374 (1994).
25. H. Yaegashi, T. Takahashi and M. Kawasaki, "Microcomputer-aided reconstruction: a system designed for the study of 3-D microstructure in histology and histopathology," *Journal of Microscopy* **146**, 55–65 (1987).

## Chapter 4.

# Registration of prostate histology images to *ex vivo* MR images via strand-shaped fiducials<sup>†</sup>

### 4.1 Introduction

The registration of histology to *in vivo* images faces several challenges. First, it is a multi-modality registration (e.g. from histology to MRI). Second, deformations induced by the imaging (e.g. due to the MRI endorectal receive coil), resection, and fixation processes are best modeled by non-rigid transformations. Finally, each section is taken from a variable position and orientation within a separate 3–5-mm-thick tissue block, so adjacent histology images are sparsely and irregularly spaced, and non-parallel. The variability and sparseness of histology image spacing introduces substantial challenges to the reconstruction of a 3D histology image for 3D registration. Differences in image content and scale, and substantial non-linear deformations introduce challenges to direct 2D to 3D registration.

Acquiring *ex vivo* prostate 3D images can address these challenges. *In vivo* to *ex vivo* image registration can compensate for tissue distortion due to surgical resection and fixation. Histology to *ex vivo* image registration can then compensate for slicing variability and the reduced distortion due to tissue shrinkage and sectioning during

---

<sup>†</sup> A version of this chapter has been published: E. Gibson, C. Crukley, M. Gaed, J. A. Gómez, M. Moussa, J. L. Chin, G. S. Bauman, A. Fenster, A. D. Ward, “Registration of prostate histology images to *ex vivo* MR images via strand-shape fiducials.” J Magn Reson Imaging 36:6 (2012).

histoprocessing. Many previous methods (described in detail in Section 1.2.3.3) approach histology to *ex vivo* image registration by guiding the slicing of the *ex vivo* specimen, in order to approximately constrain the positions and orientations of the tissue slices from which each section was cut, or by retrospectively registering histology-*ex vivo* MR images using manual or automated image-based registration methods (with additional images optionally acquired during specimen preparation). None of the existing work meets the criteria for histology-imaging registration discussed earlier in Section 1.2.2.3: (1) permitting specimen slicing and sectioning according to normal clinical pathology protocols, (2) providing registrations whose accuracies are robust to varying appearance of the prostate on imaging and pathology, (3) providing a quantitative evaluation of the registration error using a 3D target registration error (TRE), and supporting wide implementation by avoiding the use of specialized equipment.

In this study, we present and evaluate a method for registration of histology to *ex vivo* prostate MR images, developed in our laboratory, that takes steps towards meeting these criteria. The method utilizes strand-shaped fiducial markers that allow the determination of the location and orientation of each section without constraining the slicing of the specimen. Using non-anatomical fiducials provides robustness to variations in the appearance of the prostate on MR and histology images. Our registration method was evaluated using intrinsic anatomical landmarks and compared to previous methods. We also demonstrated that a proposed local registration optimization can be used to improve registrations provided by previous methods.

## 4.2 Materials and methods

This study was conducted with the approval of the Human Subjects Research Ethics Board of our institution and the informed consent of all subjects.

### 4.2.1 Materials

We obtained nine prostate specimens after radical prostatectomy with the following inclusion criteria: age 18 years or older, and histologically confirmed clinical prostate cancer stage T1 or T2. The exclusion criteria were: prior therapy for prostate cancer, use of 5-alpha reductase inhibitors within 6 months of the study start, inability to comply with preoperative imaging, allergy to contrast agents, sickle cell or other anemias, hip prosthesis, sources of artifact within the pelvis, and contraindications to MRI.

### 4.2.2 *Ex vivo* MR imaging

MR images were acquired for all specimens using a Discovery MR750 (GE Healthcare, Waukesha, WI, USA) at 3 T. After resection, formalin fixation (10% buffered formalin for 48 hours), and marking with fiducial strands [1], specimens were immobilized in a syringe filled with Christo-Lube (Lubrication Technology Inc., Franklin Furnace, OH, USA) to provide a black background and minimize boundary artifacts. Using an endorectal coil (Prostate eCoil, Medrad, Inc., Warrendale, PA, USA) placed flush with the syringe, specimens were imaged using a T1-weighted (3D SPGR, TR 6.5 ms, TE 2.5 ms, bandwidth  $\pm 31.25$  kHz, 8 averages, FOV 140 $\times$ 140 $\times$ 62 mm, slice thickness 0.4 mm, slice spacing 0.2 mm, 256 $\times$ 192 matrix, 312 slices, flip angle 15°, 25 min) protocol used in the registration and validation and a T2-weighted (3D FSE, TR

2000 ms, TE 151.5 ms, bandwidth  $\pm 125$  kHz, 3 averages, FOV 140×140×62 mm, slice thickness 0.4 mm, slice spacing 0.2 mm, 320×192 matrix, 312 slices, 25 min) used only in the validation. Each specimen was then coarsely sliced into whole-mount tissue blocks 4.4 mm thick, as required by clinical protocol. As part of another study from which our specimens were obtained, each specimen was sliced using image guidance to yield cuts coincident with *in vivo* MR imaging planes [1]. For each specimen, all tissue blocks were separated and immobilized in tissue processing cassettes, immersed in Christo-Lube, and imaged using the same coils. MR protocols were as above but with a larger FOV to cover the separated tissue blocks. These images are referred to as *block* MR images hereafter.

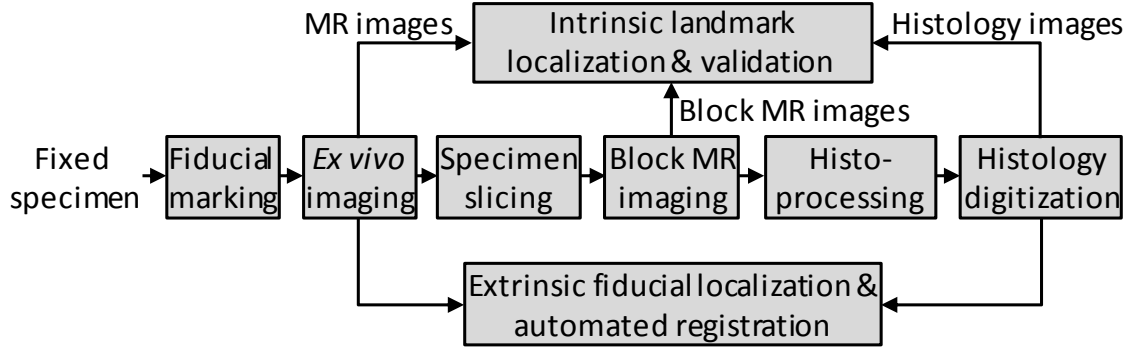
#### 4.2.3 Digital histology imaging

After standard whole-mount paraffin embedding, a 4- $\mu$ m-thick section was cut from each midgland tissue block and stained with hematoxylin and eosin. The resulting 34 slides were digitized on a ScanScope GL (Aperio Technologies, Vista, CA, USA) bright field slide scanning system with a 0.5  $\mu$ m resolution, and downsampled to a 30  $\mu$ m resolution.

### 4.3 Methods

The overall process of our method is outlined in Figure 4.1.



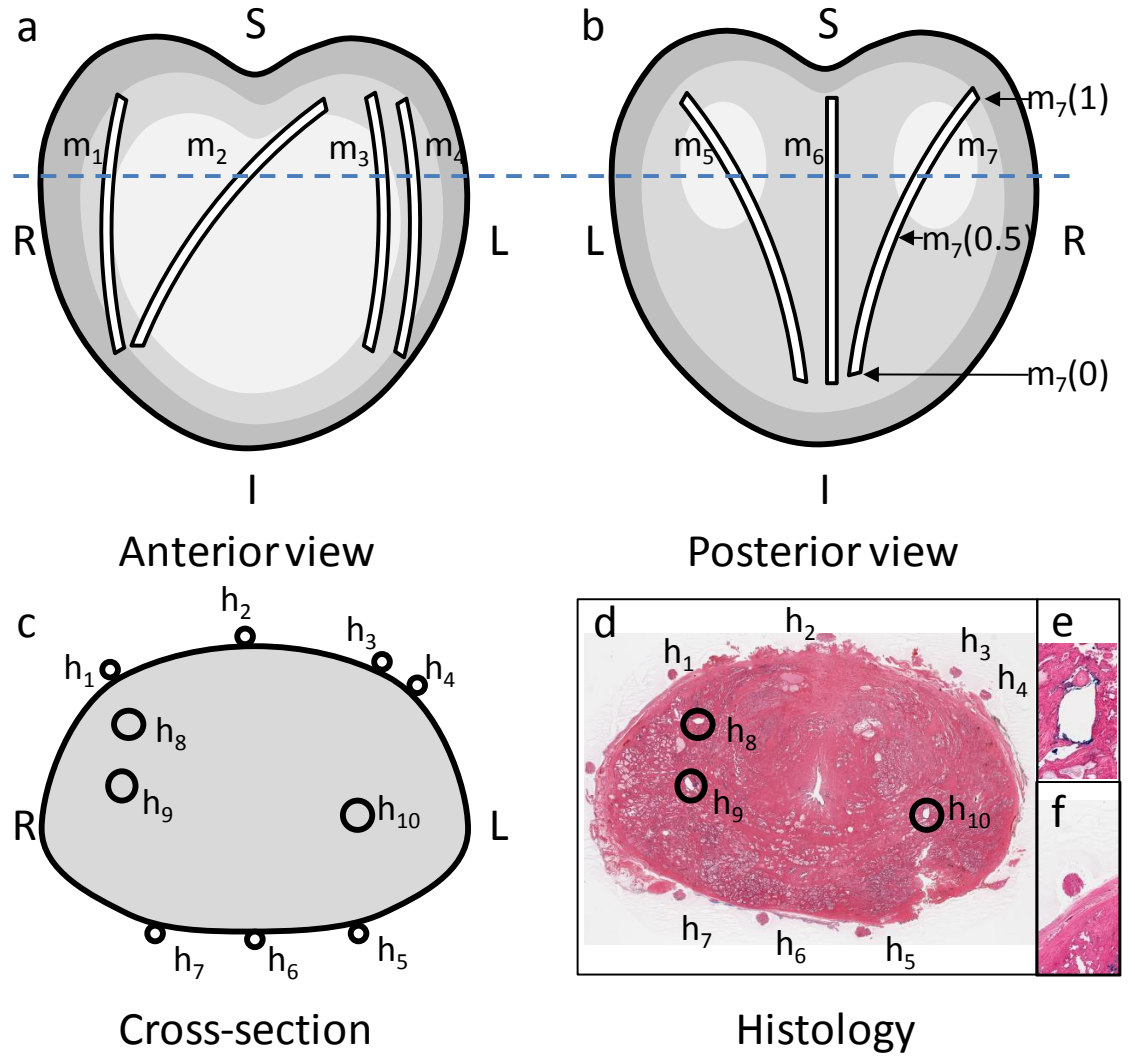


**Figure 4.1:** Specimen processing overview.

#### 4.3.1 Fiducial marking

Two sets of strand-shaped fiducials visible on T1-weighted MR and histology images (details in [1]) were added to all specimens, running from apex to base: 7 surface-mounted fiducials and 3 internal fiducials. External fiducials consisted of cylindrical strands of lamb kidney cortex (16-gauge biopsy,  $31 \pm 6$  mm mean  $\pm$  standard deviation (SD) length) infused with a 1:40 solution of Magnevist (Bayer AG, Germany) gadolinium contrast and 10% buffered formalin. They were rigidly affixed to the surface of the specimen with Loctite 411 toughened, heat-resistant, ethyl cyanoacrylate adhesive (Henkel Inc., Germany) along the entire length of the fiducial. Internal fiducials consisted of cotton thread infused with a 1:40 solution of Magnevist and blue Tissue Marking Dye (Triangle Biomedical Sciences Inc., Durham, NC, USA). Internal fiducials were inserted by (1) introducing an 18-gauge cannula and stylet with a Quincke-type point (BD Medical Inc, Franklin Lakes, NJ, USA), which is designed for clean tissue separation without damage, (2) removing the stylet, (3) inserting the thread through the cannula and (4) removing the cannula leaving the thread in place. The resiliency to deformation of formalin-fixed prostate tissue and the flexibility of the cotton thread minimize the potential for distortion of the prostate during the insertion of the internal fiducial.

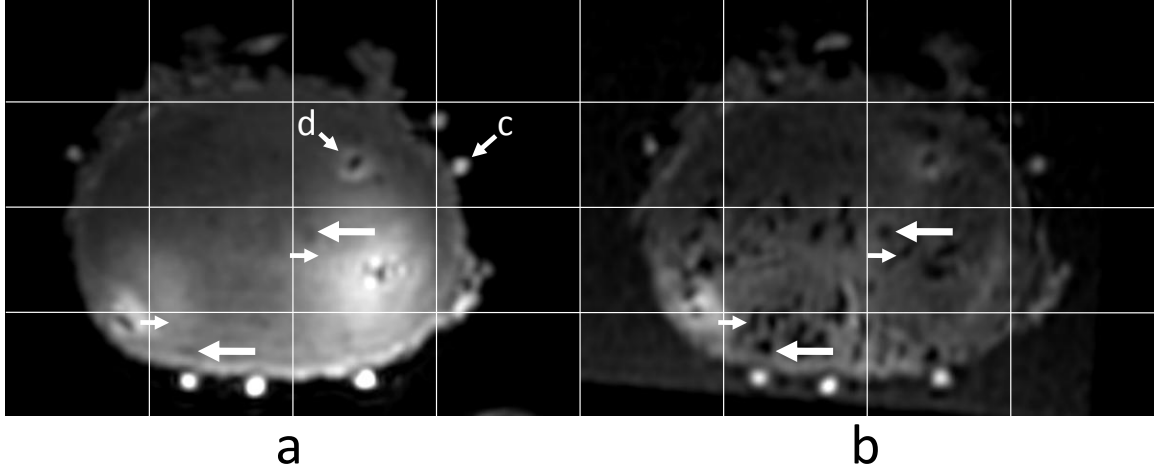
Given two potential planes in the MR image, each fiducial would intersect the two planes at two different points in the planes. Thus, the intersections of all fiducials on any given plane form a set of points with a particular configuration. The layout of fiducials was designed to maximize the difference in the configurations of points for the possible cutting planes. This facilitates the registration of each histology image to the correct plane in the MR image by matching the configuration of fiducial cross-sections. The layout of the fiducials is illustrated in Figure 4.2. Neighboring pairs of surface-mounted fiducials were placed at approximate  $45^\circ$  angles to each other, with 4 roughly inferior-superior ( $\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_4, \mathbf{m}_6$ ) and 3 oblique ( $\mathbf{m}_3, \mathbf{m}_5, \mathbf{m}_7$ ) fiducials. This makes the configuration of fiducial cross-sections on an intersecting plane sensitive to the inferior-superior position and orientation of the plane.



**Figure 4.2:** Fiducial configuration showing prostate surface (a and b) and a cross-section along the dotted line (c). These fiducial markers are visible on histology images (d). Insets (e and f) show enlarged views of  $h_{10}$  and  $h_1$  respectively.

The centerlines of fiducial tracks were manually localized on the T1-weighted MR images at  $\sim 1$  mm intervals, using 3D Slicer (Surgical Planning Lab, Harvard Medical School, Boston, USA), and interpolated to  $\sim 0.2$  mm intervals using cubic splines. Fiducials, shown in Figures 4.2e-f and 4.3c-d, were readily identifiable on histology and MR images. In some histology images, some fiducial cross-sections were absent due to

the section not intersecting the fiducial or due to detachment during processing. These sections were registered using only the remaining fiducial cross-sections.



**Figure 4.3:** Co-registered T1-weighted MR images of the whole specimen (a) and a tissue block (b). Barely visible (e.g. large arrows) or invisible (e.g. small arrows) ducts on the whole specimen image are visible on the tissue block image due to Christo-Lube contrast. Labeled arrows show the appearance of surface-mounted (c) and internal (d) fiducials, respectively, on the T1W MR image.

#### 4.3.2 2D–3D histology to MR image registration algorithm

##### 4.3.2.1 Definitions of algorithm terms

On MR images, the  $i$ -th fiducial track forms a 3D parametric curve  $\mathbf{m}_i (s_i \in [0,1])$ , as illustrated on  $\mathbf{m}_7$  in Figure 4.2b. On the histology image, the cross-section of the  $i$ -th fiducial gives a 2D point  $\mathbf{h}_i$  (Figure 4.2d), which corresponds to some 3D point along the corresponding fiducial track. For the 10 fiducial cross-section points in each histology image, any possible correspondence can be encoded as a 10D vector  $\mathbf{s} = [s_1, s_2, \dots, s_{10}]$ . For example,  $\mathbf{s} = [0.4, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5]$  denotes that the first fiducial cross-section on the histology image corresponds to the point 40% along the length of the first fiducial track on the MR image, measured from the apex, and all other

cross-sections correspond to the midpoints of their respective tracks. Each  $\mathbf{s}$  determines a least-squares best-fit 2D–3D affine transformation  $\mathbf{A}_s$  mapping fiducial cross-sections on the histology image to their corresponding points on the MR image.

#### 4.3.2.2 Algorithm step 1: identify an initial plane on MR image

corresponding to each histology image

Finding the MR image plane corresponding to a histology image involves finding the correspondence vector  $\bar{\mathbf{s}}$ , defined as

$$\bar{\mathbf{s}} = \operatorname{argmin}_s \sum_{i=1}^{10} \|\mathbf{A}_s \mathbf{h}_i - \mathbf{m}_i(s_i)\|^2, \quad 4.1$$

yielding the smallest squared residual error after the affine transformation  $\mathbf{A}_{\bar{\mathbf{s}}}$  which transforms the histology image to its corresponding MR image plane.

An exhaustive search of the 10D correspondence space to find this vector would be computationally prohibitive, and a greedy minimization may tend to find local minima far from the global minimum. For an efficient, broad search of the space of correspondences, we manually selected a triplet of fiducials  $[i_1, i_2, i_3]$  corresponding to three dimensions of  $\mathbf{s}$ , defining a 3D correspondence vector  $\mathbf{t} = [s_{i_1}, s_{i_2}, s_{i_3}]$ , and the unique 2D–3D affine transformation  $\mathbf{A}_t$  that maps  $[\mathbf{h}_{i_1}, \mathbf{h}_{i_2}, \mathbf{h}_{i_3}]$  to  $[\mathbf{m}_{i_1}(s_{i_1}), \mathbf{m}_{i_2}(s_{i_2}), \mathbf{m}_{i_3}(s_{i_3})]$ . We performed an exhaustive search of this reduced space of 3D correspondence vectors. This yielded a 3D correspondence vector  $\bar{\mathbf{t}}$  with the minimal residual computed from all 10 fiducials,

$$\bar{\mathbf{t}} = \operatorname{argmin}_t \sum_{i=1}^{10} \|\mathbf{A}_t \mathbf{h}_i - C(\mathbf{A}_t \mathbf{h}_i, \mathbf{m}_i)\|^2, \quad 4.2$$

where  $C(\mathbf{p}, \mathbf{m})$  is the closest point to  $\mathbf{p}$  on parametric curve  $\mathbf{m}$ .

The fiducial triplets were manually selected to include widely spaced fiducials, including at least one diagonal fiducial. When all fiducials were present, the triplet  $[2,4,7]$  (i.e.,  $\mathbf{m}_2, \mathbf{m}_4$ , and  $\mathbf{m}_7$  in Figure 4.2(a-b)) was used. The plane defined by the points  $\{\mathbf{m}_i(\bar{t}_i) | i \in [i_1, i_2, i_3]\}$  was taken to be the MR image plane corresponding to the histology image.

#### 4.3.2.3 Algorithm step 2: compute an affine transformation mapping each histology image to its corresponding MR image plane

The next step was to compute a 10D correspondence vector  $\check{\mathbf{s}}$ , defining an affine transformation  $\mathbf{A}_{\check{\mathbf{s}}}$  (which utilizes the spatial information provided by all 10 fiducial markers) mapping the histology image onto the MR image.

We construct  $\check{\mathbf{s}}$  based on the plane defined in step 1, such that each fiducial cross-section  $\mathbf{h}_i$  corresponds to the closest point on  $\mathbf{m}_i$  to the defined plane.  $\check{s}_i$  is defined as,

$$\check{s}_i = \underset{s_i}{\operatorname{argmin}} \left| \left( \mathbf{m}_{i_2}(\bar{t}_2) - \mathbf{m}_{i_1}(\bar{t}_1) \right) \times \left( \mathbf{m}_{i_3}(\bar{t}_3) - \mathbf{m}_{i_1}(\bar{t}_1) \right) \cdot \left( \mathbf{m}_i(s_i) - \mathbf{m}_{i_1}(\bar{t}_1) \right) \right|. \quad 4.3$$

#### 4.3.2.4 Algorithm step 3: refine the fiducial correspondence using local optimization and compute affine transformation

$\mathbf{A}_{\check{\mathbf{s}}}$  is intended to yield a near-optimal histology to MR image alignment. To find the optimal correspondence vector  $\bar{\mathbf{s}}$ , we used a Nelder-Mead greedy simplex minimization [2] of the squared residual FRE in the 10D correspondence space (Equation 4.1), initialized with  $\check{\mathbf{s}}$ . This yields the correspondence vector at the nearest optimum in the 10D correspondence space, giving a corresponding affine transformation.

### 4.3.3 Validation

TREs for all specimens were used to evaluate the registration [3]. TREs were calculated as the 3D post-registration misalignments of small anatomical landmarks identified on histology and MR images.

Identifying homologous landmarks on histology and MR images is challenging [4] due to a large 3D search space, a lack of 3D context on histology images and few identifiably homologous structures. We developed a protocol for identifying such landmarks on histology and whole-gland *ex vivo* MR images by constraining the search using an approximate alignment informed by "block MR" images of the sliced tissue blocks, acquired using the same MR imaging protocol described above. The smaller search space and higher contrast features (as shown in Figure 4.3) due to infiltration of Christo-Lube into ducts exposed by specimen slicing facilitate aligning histology to block MR images; similar image content and 3D context facilitate aligning block to whole-gland MR images. Landmarks included the centers of atrophic ducts, cysts and corpora amylaceae (or parts thereof) with diameters less than 1 *mm*, which are large enough to be resolved on the *ex vivo* MR images (voxel size 0.27×0.27×0.20 mm). Homology of these landmarks was evaluated based on spatial relationships with nearby salient features. Potential landmarks without nearby salient features, or for which the nearby salient features did not uniquely constrain the correspondence, were not included as landmarks. Landmarks were identified by one observer under advisement from 2 genitourinary pathologists and 1 radiologist with expertise in prostate imaging.

Measurement of TREs based on localized landmarks incorporates error from the registration and error in localizing the landmarks. The target localization error (TLE) of

landmarks on each modality was quantified as an unbiased estimator of the standard deviation of repeated localizations of the same landmark [3], at least one day apart:

$$\text{TLE} = \sqrt{\frac{1}{J} \sum_{j=1}^J \frac{1}{K-1} \sum_{k=1}^K \left\| \mathbf{p}_{j,k} - \frac{1}{K} \sum_{k=1}^K \mathbf{p}_{j,k} \right\|^2}, \quad 4.4$$

where  $\mathbf{p}_{j,k}$  is the  $k$ -th localization of the  $j$ -th fiducial,  $J = 16$  landmarks, and  $K = 7$  repeated localizations of each landmark. The TLE on histology images ( $\text{TLE}_h$ ) used landmarks pre-specified on MR images, while the TLE on MR images ( $\text{TLE}_m$ ) used landmarks pre-specified on histology images. Landmarks for this assessment comprised all landmarks from 3 blocks from 3 randomly chosen specimens for calculating  $\text{TLE}_h$ , and a separate 3 blocks for calculating  $\text{TLE}_m$ .

#### 4.3.4 Experiments

We conducted four experiments: (1) to evaluate the proposed method, (2) to compare the accuracy of the proposed method, where the plane selection and registration is informed by fiducial correspondence, to a previous approach, where the plane selection is informed by image-guided slicing, (3) to determine the reduction in TRE provided by the internal fiducials, and (4) to determine the reduction in TRE provided by the augmentation of the proposed method to use an additional set of images as an intermediate registration target.

##### 4.3.4.1 Experiment 1: To evaluate the proposed method

To assess the accuracy of the proposed method, TREs were calculated after steps 2 and 3 of the algorithm. We documented the fiducial cross-sections that were absent on histology images and correlated the number of remaining fiducials with the mean TRE.



To assess the anisotropy of the TRE, we analyzed the post-registration misalignment vector for each landmark pair, denoted  $\mathbf{TRE}_j$  for the  $j$ -th landmark pair, in the MR image coordinate space. We computed the 3D principal component analysis (PCA) of these vectors using MATLAB (The Mathworks Inc., Natick, MA) yielding the directions of maximum variance, and the variance of error along these directions. An isotropic TRE would yield 3 equal variances. To examine anisotropy related to the histology images, which are not all parallel, we also decomposed each  $\mathbf{TRE}_j$  into in-plane and out-of-plane components, and computed an aggregate in-plane and out-of-plane variance:

$$in-plane = Var(\{\|\mathbf{TRE}_j - (\mathbf{TRE}_j \cdot \mathbf{n}_j)\mathbf{n}_j\| | j = 1..N\}), \text{ and} \quad 4.5$$

$$out-of-plane = Var(\{\mathbf{TRE}_j \cdot \mathbf{n}_j | j = 1..N\}), \quad 4.6$$

where  $\mathbf{n}_j$  is the unit normal of the corresponding histology plane, and  $N = 184$  fiducials.

For an isotropic TRE, in-plane variance, being the sum of two equal 1-dimensional variances, would be twice the size of the out-of-plane variance.

4.3.4.2 Experiment 2: To compare the proposed method to a previous method based on image-guided slicing [1]

The proposed method was compared to a previous method [1] to evaluate which had a lower mean TRE. Because the specimens used in this study were obtained from another study where specimens were sliced using image-guided slicing defined in the previous method, slicing planes from the previous method were known. In the previous work, these slicing planes were taken to be the MR image plane corresponding to the histology image. To directly compare the previous method to the proposed method, algorithm steps 2 and 3 were applied to these slicing planes to yield affine registrations. This was accomplished by defining  $\{\mathbf{m}_i(\bar{t}_i) | i \in [i_1, i_2, i_3]\}$  to contain three non-collinear

points on the slicing plane for algorithm step 2 for the previous method. A TRE was calculated after step 2, to compare the previous method against the proposed method, and after step 3, to evaluate the effect of local optimization on the previous method.

#### 4.3.4.3 Experiment 3: To compare the accuracy of the method with and without the use of internal fiducials

To assess the effect of using only surface-mounted fiducials (to reduce the processing time required and render the method suitable to pathology environments prohibiting internal fiducials), we computed a TRE for the proposed algorithm, excluding the internal fiducials, with and without the local optimization.

#### 4.3.4.4 Experiment 4: To evaluate the effect of incorporating information from block MR images

Since each section is cut from a tissue block, each histology image corresponds to the prostate region corresponding to that block. However, in algorithm step 1, the entire lengths of the fiducial strands, including parts outside of the corresponding block, are searched. In principle, step 1 could associate a histology image with a plane outside the block from which it came. To overcome this potential problem, previous approaches [4-11] have used additional images as intermediate registration targets. Testing this approach, we *replaced algorithm step 1* with a procedure, based on [11], wherein the plane search described in step 1 was run for each histology image using the *corresponding* block MR image taken prior to sectioning. As in [11], least-squares best-fit rigid transformations of the fiducial markers was used to place these planes within the whole-gland MR image. We subsequently carried out algorithm steps 2 and 3 as above, computed a TRE, and compared it to that yielded by the proposed method.

#### 4.3.5 Statistical analysis

Statistical analyses were performed in Prism 5.04 (Graphpad Software, Inc., San Diego, USA). To compensate for the positive skew in the measured TREs, data were transformed using a square-root function. To compare the mean TREs of the different methods, we used a repeated-measures ANOVA of the transformed TREs from each method, followed by *post hoc* analyses with Bonferroni multiple-comparison correction. After the square-root transformation, the TRE distribution for the proposed method without internal fiducials remained non-normal (D'Agostino & Pearson omnibus normality test;  $p < 0.0001$ ). To confirm the findings involving this method, Wilcoxon matched-pairs signed-rank tests were performed.

To assess relative accuracy, pairwise *post hoc* analyses tested for a difference in the mean TRE between the proposed method and three alternative methods: image-guided slicing, tissue block imaging, and the proposed method without internal fiducials. To evaluate whether the local optimization improved the previous image-guided slicing method [1], *post hoc* analysis tested for a difference in the mean TRE of the previous method with and without local optimization. Confidence intervals on these differences were estimated using the same analysis on the untransformed data; these intervals must be interpreted with the non-normality of the underlying data in mind. To assess the interchangeability of initialization methods, we generated two Bland-Altman plots comparing the proposed method to the image-guided method with local optimization, and the tissue block imaging method with local optimization.

Additionally, the accuracy of the proposed method inside and outside the peripheral zone was compared using an unpaired t-test omitting landmarks with indefinite

classification, and the accuracy of the proposed method with and without internal fiducials omitting sections with missing fiducial cross-sections was performed using a paired t-test. Finally, to assess the robustness of the proposed method to missing fiducial cross-sections, a Spearman correlation test was performed between the TRE for each landmark and the number of fiducial cross-sections on the section containing the landmark.

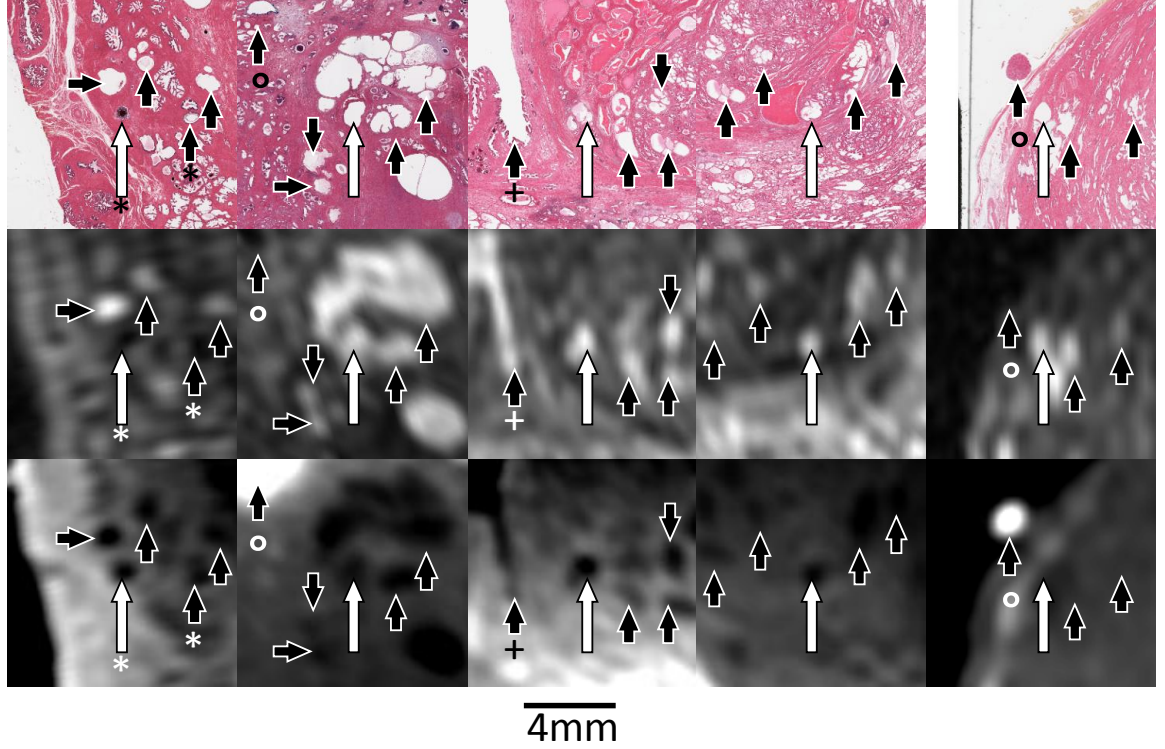
## 4.4 Results

The proposed method required ~3 hours per specimen including fiducial application (90 minutes), imaging (80 minutes), fiducial localization (12 minutes) and registration computation ( $59 \pm 11$  seconds). Our application of previous methods using block MR images [11] and image-guided slicing [1] required ~5 and ~11 hours per specimen, respectively, for all processing steps.

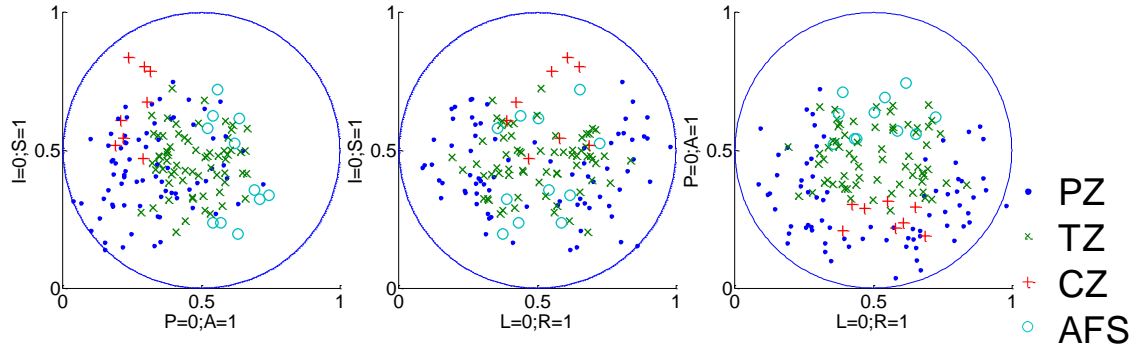
The landmark localization protocol yielded 3–7 homologous landmark pairs on each of 3–5 sections per specimen, totaling 184 landmarks on 34 histology images. Figure 4.4 shows histology and MR images of 5 homologous landmark pairs and nearby salient features, confirmed by a genitourinary pathologist (J.A.G.) and a radiologist specializing in prostate imaging (C.R.). These landmarks were located in the peripheral zone (75/184), transitional zone (66/184), central zone (8/184) and anterior fibromuscular stroma (11/184); 24/184 landmarks were not definitively categorized. The spatial distribution of the landmarks is illustrated in Figure 4.5, showing the landmarks evenly distributed throughout the posterior midgland. Few landmarks were near the apex and base, because only midgland sections were included in this analysis, and few landmarks were near the anterior due to a lack of salient features in the anterior fibromuscular

stroma. The TLE on histology images was 50  $\mu\text{m}$ , relative to a pixel size of  $30 \times 30 \mu\text{m}^2$ .

The TLE on MR images was 0.16 mm, relative to a voxel size of  $0.27 \times 0.27 \times 0.2 \text{ mm}^3$ .

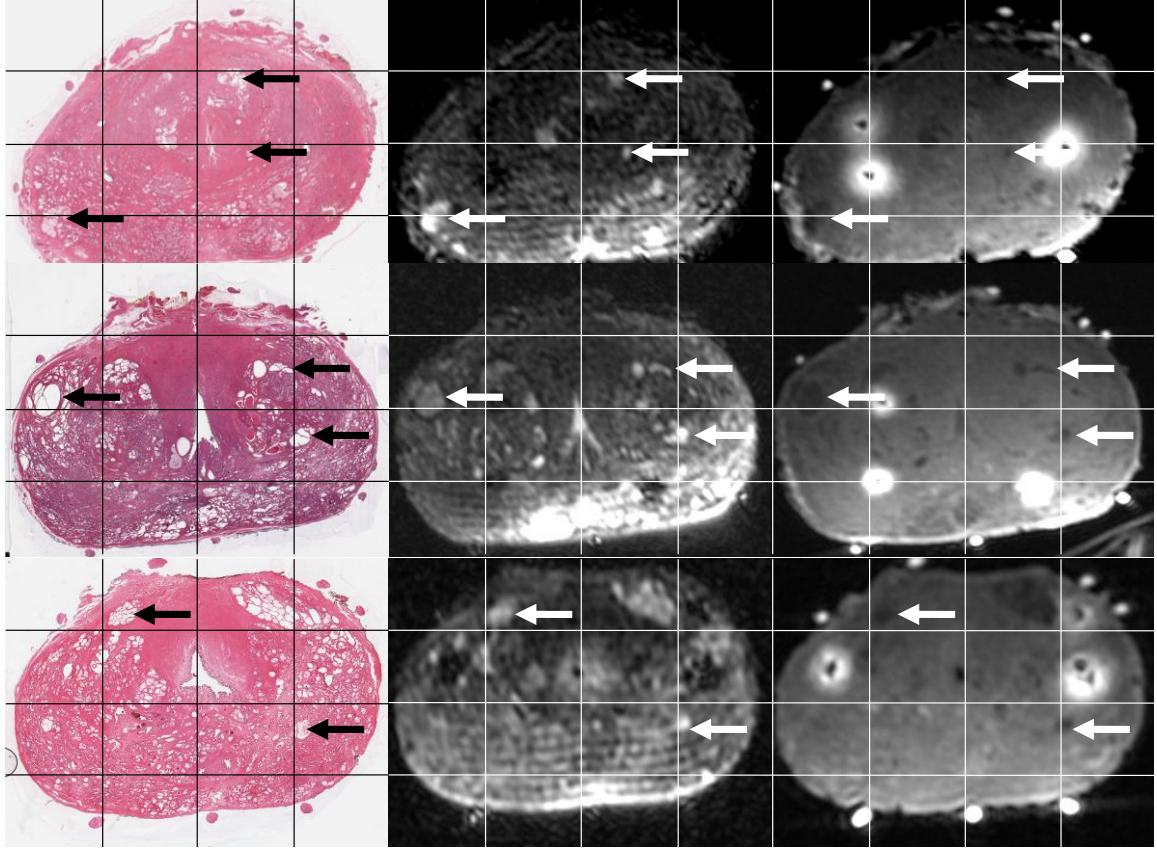


**Figure 4.4:** Homologous landmark pairs (large white arrows), used for evaluation of the registration algorithm, shown on histology images (top) and on the corresponding oblique slices on T2 (middle) and T1 (bottom) MR images. Small black arrows denote salient nearby features used to corroborate the established landmark correspondence. Salient features in these images include corpora amylaceae (labeled \*), the urethra (labeled +), extrinsic fiducials (labeled °) and atrophic ducts (unlabeled).



**Figure 4.5:** Distribution of landmarks within the prostate specimens. Each landmark is shown with its AP, LR and IS position within the specimen normalized such that the extents of the resected prostate gland are from 0 to 1 for each specimen. Each one is marked as to the anatomical zone where it is located: peripheral zone (PZ), transition zone (TZ), central zone (CZ) and anterior fibromuscular stroma (AFS). The 24 landmarks with indefinite classification were omitted.

Three histology and MR images co-registered using the proposed method are shown in Figure 4.6. The proposed method yielded a mean TRE of  $0.71 \pm 0.38$  mm. An unpaired t-test failed to show a significant difference between the mean TRE inside and outside the peripheral zone ( $p=0.75$ , 95% CI, -0.14 to 0.11 mm). This TRE is anisotropic, with variances of 0.33, 0.13 and  $0.09 \text{ mm}^2$ ; the principal direction of variation is  $[0.45, -0.12, -0.88]$ , which is not aligned to the MR image axes. The out-of-plane variance is  $0.39 \text{ mm}^2$ , larger than the largest variance of the 3D PCA. The in-plane variance is  $0.24 \text{ mm}^2$ , a factor of 3 lower than the  $0.78 \text{ mm}^2$  that would be expected for an isotropic TRE.

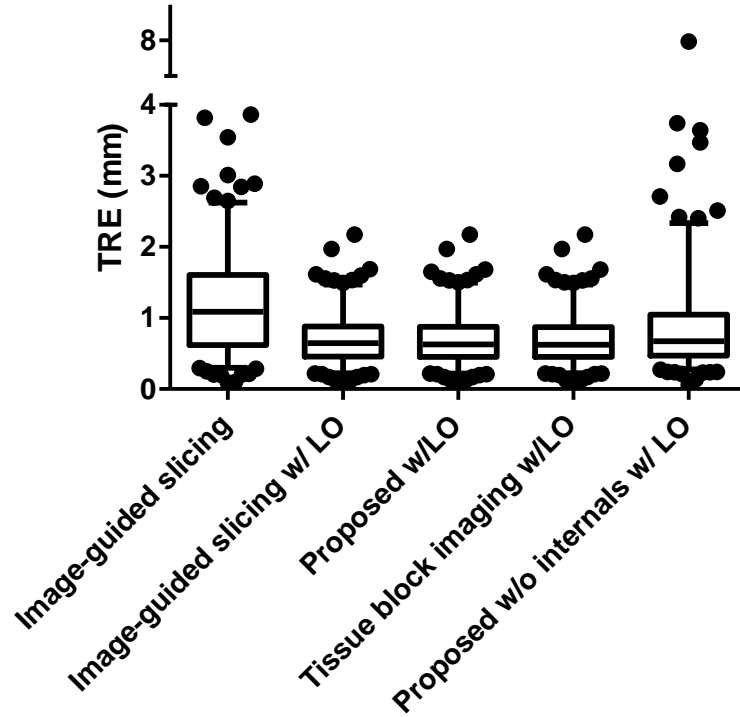


**Figure 4.6:** Three histology images registered to *ex vivo* MR images using the proposed registration method. Example validation landmarks are shown by arrows.

The proportion of fiducial cross-sections absent on histology images was 20/340, 7 due to the section not passing through the fiducial, and 13 due to detachment during specimen cutting or processing. As a result, 7/34, 5/34 and 1/34 sections were registered using 9, 8 and 7 fiducial cross-sections, respectively. The Spearman correlation test failed to show a correlation between TRE and the number of fiducial cross-sections used ( $p=0.25$ , 95% CI for  $r$ : -0.23 to 0.06).

The mean and standard deviation of the TREs before and after local optimization are shown in **Table 4.1** for four registrations: the proposed method, the proposed method without internal fiducials, the method using tissue block imaging, and the method using image-guided slicing. The TRE data are shown in Figure 4.7 for the five registrations

examined in the discussed experiments, and the *post hoc* analyses are summarized in Table 4.2.



**Figure 4.7:** Box plot of TREs for five registrations with 5–95% whiskers.



**Table 4.1:** Mean TREs of the registration methods with and without local optimization.

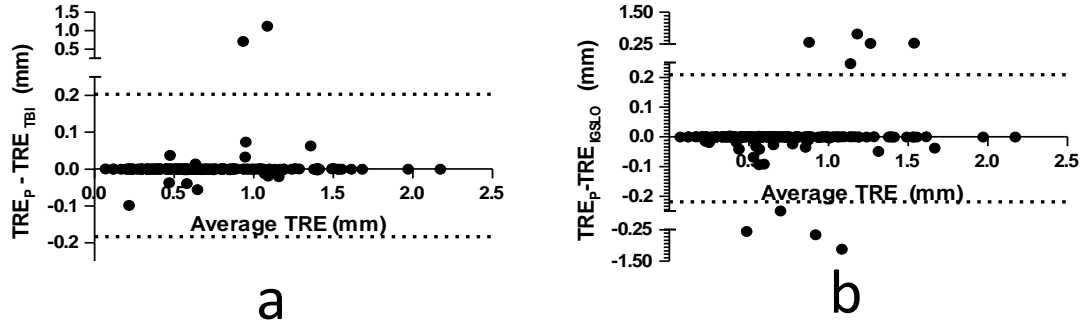
	<i>Proposed method</i>	<i>Image-guided slicing method</i>	<i>Tissue block imaging method</i>	<i>Proposed method without internal fiducials</i>
<b>Mean±SD TRE (mm) without local optimization</b>	0.76±0.43	1.21±0.74	0.86±0.50	0.95±0.84
<b>Mean±SD TRE (mm) with local optimization</b>	0.71±0.38	0.71±0.37	0.70±0.36	0.92±0.82

**Table 4.2:** Statistical analysis of accuracy of methods. Each row represents a paired t-test of TREs of landmarks between two methods.

<i>Method A</i>	<i>Method B</i>	<i>p-value</i> <i>h<sub>0</sub>:</i> $\sqrt{TRE_A} - \sqrt{TRE_B} = 0$	<i>95% CI (mm)</i> <i>of</i> $TRE_A - TRE_B$
<b>Image-guided slicing</b>	<b>Proposed w/ LO</b>	< 0.0001	[0.38,0.63]
<b>Image-guided slicing</b>	<b>Image-guided slicing w/ LO</b>	< 0.0001	[0.38,0.62]
<b>Image-guided slicing w/ LO</b>	<b>Proposed w/ LO</b>	> 0.05	[-0.12,0.13]
<b>Proposed w/ LO w/o internal</b>	<b>Proposed w/ LO</b>	< 0.0001	[0.09,0.34]
<b>Tissue block imaging w/ LO</b>	<b>Proposed w/ LO</b>	> 0.05	[-0.13,0.11]

The proposed registration had a lower mean TRE ( $p < 0.0001$ ) than that using image-guided slicing. Applying local optimization to the method using image-guided slicing improves the TRE ( $p < 0.0001$ ). The *post hoc* analysis failed to show a statistically significant difference between the mean TRE of the proposed registration and the method using image-guided slicing with local optimization ( $p > 0.05$ , 95% CI, -0.12 to 0.13 mm).

A Bland-Altman comparison of these methods is shown in Figure 4.8a.



**Figure 4.8:** Bland-Altman plots comparing TREs with local optimization after initializations using the proposed method ( $TRE_P$ ) and (a) the tissue block imaging ( $TRE_{TBI}$ ) and (b) the image-guided slicing with local refinement ( $TRE_{IGSLO}$ ). Each point represents one landmark pair.

The mean TRE for the proposed method with internal fiducials was lower than that without internal fiducials ( $p < 0.0001$  using ANOVA *post hoc* analysis,  $p = 0.0006$  using non-parametric confirmation). Notably, for the method without internal fiducials, all 9 of the outliers seen in Figure 4.7 are from sections with missing fiducial cross-sections. Omitting sections with missing fiducial cross-sections, the difference in mean TRE with (0.67 mm) and without (0.75 mm) internal fiducials was less significant ( $p = 0.024$ , 95% CI, 0.01 to 0.14 mm using t-test;  $p = 0.15$  using non-parametric confirmation).

*Post hoc* analysis comparing the proposed method to the method incorporating block MR images failed to show a statistically significant difference between the mean TREs ( $p > 0.05$ , 95% CI, -0.13 to 0.11 mm). A Bland-Altman comparison of these methods is shown in Figure 4.8b.

## 4.5 Discussion

In this chapter, we proposed a method for registration of histology to *ex vivo* MR images with submillimeter accuracy that reduces the processing time over a previously

adopted method. Our registration uses a controlled spatial configuration of non-anatomical fiducials to address the challenges of 2D to 3D multimodality registration. We register histology to *ex vivo* MR images acquired using a similar arrangement to that used in clinical *in vivo* prostate MRI (3 T magnet, T1- and T2-weighted sequences, endorectal receive coil). Thus, our method addresses many challenges of the full registration of histology to *in vivo* MR images, leaving as the remaining step (outside the scope of the present work) a single-modality 3D *ex vivo-in vivo* prostate MR image registration, a task addressed in previous work [4] using standard methods. Furthermore, the histology to *ex vivo* image registration can be used directly in validation studies of *ex vivo* imaging (e.g. [12]).

While some disruption of clinical workflow is inevitable when using an intermediate registration to *ex vivo* imaging, our method reduces the extra-clinical processing time required to collect data for registration from 11 hours using a previous workflow [1] to 3 hours; a reduction of 70%. This 3-hour processing time could be integrated into existing clinical workflows with no delay to the clinical diagnosis by scheduling the processing around specimen batch processing schedules, which was not generally possible with the 11-hour workflow.

The maximum acceptable registration error for registration-based validations depends on application-specific information regarding (a) the size of the regions of interest to be registered, (b) the imaging properties of these regions, (c) the type of analysis and (d) the power of the analysis. Large registration error may constrain the size of region that can be included in an analysis [13, 14], or may underestimate the differences between benign and cancerous tissues on imaging [15]. When the registration

error is unknown, studies must err on the side of caution, excluding more small regions than necessary, or searching for smaller effects than necessary. The measurement of TRE is challenging [4] as it requires the identification of homologous landmarks in histology images and other medical images, and it is therefore frequently omitted in analyses of these registration techniques [4, 16, 17]. In prostate histology to *ex vivo* MR image registration, few efforts have been made to report TRE: Kimm et al. [18] reported a 2D TRE of 0.86 mm based on 3 sections from a single specimen including non-anatomical landmarks in their measurement, Zhan et al. [19] and Ou et al. [20] reported TREs of 0.82 mm and 0.79 mm using anatomical landmarks with no further indication of the number or type of landmarks used, or whether the TRE was 2D or 3D. The proposed landmark localization framework enabled the localization of 184 landmarks to robustly quantify the 3D TRE of the proposed registration for 34 sections from 9 subjects. The TLE is sufficiently small that localization variability does not dominate the measurement of TRE using these landmarks. While these TLEs are large relative to the differences in mean TRE between methods, the use of repeated-measures ANOVA comparisons and paired *post hoc* analyses mitigates the effect of TLE in these comparisons.

The mean $\pm$ SD TRE for the proposed method was found to be 0.71 $\pm$ 0.38 mm, more accurate than the previously adopted method based on image-guided slicing 1.21 $\pm$ 0.74 mm. The size of the smallest clinically relevant cancer focus is a topic of controversy, with thresholds from 0.2 cc [21] – 0.5 cc [22]. The mean TRE is sufficient to achieve an 85% overlap of a hypothetical spherical 0.2 cc tumor. If used as a component of a histology image to *in vivo* image registration, the mean error of the combined registration would be added in quadrature, assuming independent errors. For example, if

the *ex vivo* to *in vivo* registration also had a mean TRE of 0.71 mm, the mean combined error would yield a 75% overlap of a 0.2 cc tumor, or an 85% overlap of a 0.5 mm tumor.

The anisotropic variance of the TRE, highest in the direction perpendicular to the histology plane, could have several causes: out-of-plane deformation of the tissue during slicing or histology processing, poor sensitivity of the configuration of fiducial cross-sections to the plane selection, or higher out-of-plane error in locating the landmarks. This last effect is consistent with the landmark location procedure: the landmarks on MR images were typically placed at the center of an anatomical feature, and on histology images, the center of an anatomical feature can only be accurately identified in the histology image plane.

The proposed local optimization can also be leveraged to improve previous registration methods to achieve similar accuracy. The difference in mean TRE between the proposed method and the image-guided slicing registration with local optimization is less than 0.13 mm, with 95% confidence. The high agreement between landmark TREs seen in the Bland-Altman plot in Figure 4.8a suggests that this improved image-guided slicing registration can provide equivalent registration accuracy to the proposed method in specific research contexts where histology interrogation of particular MR imaging planes is required. In all other research contexts, the proposed method is applicable and provides a substantially faster, less clinically disruptive solution. The proposed method augmented with block MR images also has high agreement between landmark TREs; however, there is no clear research context where this method would be required, suggesting the additional complexity and delay is not warranted.

While the rigid fixation of external fiducial markers to the prostate surface by the cyanoacrylate adhesive inhibits fiducial markers from moving along the surface, it is possible for cross-sections of the fiducial markers to detach entirely, which occurred for 4% of fiducial marker cross-sections. An additional 2% of fiducial marker cross-sections were not visible on histology images due to the sections not intersecting with the fiducial markers. The proposed method was robust to the absence of 1–3 of the fiducial cross-sections on histology images. There is, however, a theoretical limit to this robustness, as the method will not function with fewer than 4 fiducials as the residual fiducial registration error being minimized would be uniformly zero.

The proposed registration method can be used with or without internal fiducials, although we observed a trade-off between method simplification (by eliminating internal fiducials) and accuracy. Registrations of sections with no absent fiducials were more accurate, suggesting that the proposed method without internal fiducials is less robust to loss of fiducials. Elimination of internal fiducials may still be warranted if required by an institution's pathology workflow and our reported accuracy is sufficient for the study. In this case, careful handling to ensure slicing through the fiducials and to avoid the loss of surface-mounted fiducials may increase the registration accuracy.

An ideal error measurement of an image registration algorithm informs as to the post-registration misalignment of anatomically homologous points. The most straightforward approach to obtaining this information is to identify homologous point landmark pairs in the images to be registered and compute an aggregate of the post-registration Euclidean distances between them (e.g. yielding a mean TRE). The use of non-point structural landmarks (e.g. via the introduction of additional strand-shaped

fiducial markers to be used only for validation) complicates the error calculation due to the need to establish an accurate point correspondence across these more complex structures, which may be an ill-posed problem. In our application, this would involve the computation of a correspondence between the points defined by the cross-sections of the strand-shaped fiducial markers on the histology images and the space curves visible in the MR images. Because the location of the histological section is unknown, it is ambiguous which point along the curve on the MR image corresponds to the point on the histology image. The approach used in our study avoids this ambiguity by identifying homologous point landmarks in the histology and MR images.

One limitation of the method is that the registration assumes there is an affine transformation between histology and post-fixation *ex vivo* images. This assumption is supported by preliminary work [11] and the work described in Chapter 3 using the same whole-mount histology processing protocol. However, if this assumption is violated, the registration could be adapted to incorporate more flexible non-linear transformations. A second limitation is that all specimens were sliced according to the constraints of the ongoing study from which specimens were drawn. While this allowed a comparison of registration accuracy on the same specimens, it prevented a measurement of the range of acceptable variation in slicing orientations recoverable by the proposed registration and local optimization. We hypothesize that the method is robust to a wide range of slicing angles as long as the slicing plane passes through the applied fiducial strands.

In conclusion, the proposed method for histology image – *ex vivo* MR prostate image registration yields submillimeter registration error of 0.71 mm while reducing the processing time by 8 hours (70% reduction) relative to a currently adopted method. It

does not require guidance of specimen slicing, does not require block face or tissue block imaging as an intermediate registration target, and does not depend on the stable appearance of images that may be disrupted by the disease process.

## 4.6 References

1. A. D. Ward, C. Crukley, C. A. McKenzie, J. Montreuil, E. Gibson, C. Romagnoli, J. A. Gómez, M. Moussa, J. Chin, G. Bauman and A. Fenster, "Prostate: registration of digital histopathologic images to in vivo MR images acquired by using endorectal receive coil," *Radiology* **263**, 856–864 (2012).
2. J. C. Lagarias, J. A. Reeds, M. H. Wright and P. E. Wright, "Convergence properties of the Nelder–Mead simplex method in low dimensions," *SIAM Journal on Optimization* **9**, 112–147 (1998).
3. J. M. Fitzpatrick, J. B. West and C. R. Maurer, Jr., "Predicting error in rigid-body point-based registration," *IEEE Transactions on Medical Imaging* **17**, 694–702 (1998).
4. H. Park, M. R. Piert, A. Khan, R. Shah, H. Hussain, J. Siddiqui, T. L. Chenevert and C. R. Meyer, "Registration methodology for histological sections and in vivo imaging of human prostate," *Academic Radiology* **15**, 1027–1039 (2008).
5. G. Groenendaal, M. R. Moman, J. G. Korporaal, P. J. van Diest, M. van Vulpen, M. E. Philippens and U. A. van der Heide, "Validation of functional imaging with pathology for tumor delineation in the prostate," *Radiotherapy and Oncology* **94**, 145–150 (2010).
6. A. S. Jackson, S. A. Reinsberg, S. A. Sohaib, E. M. Charles-Edwards, S. Jhavar, T. J. Christmas, A. C. Thompson, M. J. Bailey, C. M. Corbishley, C. Fisher, M. O. Leach and D. P. Dearnaley, "Dynamic contrast-enhanced MRI for prostate cancer localization," *British Journal of Radiology* **82**, 148–156 (2009).
7. L. S. Taylor, B. C. Porter, G. Nadasdy, P. A. di Sant'Agnese, D. Pasternack, Z. Wu, R. B. Baggs, D. J. Rubens and K. J. Parker, "Three-dimensional registration of prostate images from histology and ultrasound," *Ultrasound in Medicine and Biology* **30**, 161–168 (2004).
8. M. Breen, R. Lazebnik and D. Wilson, "Three-dimensional registration of magnetic resonance image data to histological sections with model-based evaluation," *Annals of Biomedical Engineering* **33**, 1100–1112 (2005).



9. E. Bardinnet, Ourselin, S., Dormont, D., Malandain, G., Tandé, D., Parain, K., Ayache, N., Yelnik, J., "Co-registration of histological, optical and MR data of the human brain," in *Proceedings of Medical Image Computing and Computer Assisted Intervention*, (2002), pp. 548–555.
10. J. Xu, P. A. Humphrey, A. S. Kibel, A. Z. Snyder, V. R. Narra, J. J. Ackerman and S. K. Song, "Magnetic resonance diffusion characteristics of histologically defined prostate cancer in humans," *Magnetic Resonance in Medicine* **61**, 842–850 (2009).
11. E. Gibson, C. Crukley, J. A. Gómez, M. Moussa, G. Bauman, A. Fenster and A. D. Ward, "Tissue block MRI for slice orientation-independent registration of digital histology images to *ex vivo* MRI of the prostate," in *Proceedings of International Symposium on Biomedical Imaging*, (Chicago, USA, 2011), pp. 566–569.
12. A. Madabhushi, M. D. Feldman, D. N. Metaxas, J. Tomaszewski and D. Chute, "Automated detection of prostatic adenocarcinoma from high-resolution *ex vivo* MRI," *IEEE Transactions on Medical Imaging* **24**, 1611–1625 (2005).
13. L. W. Turnbull, D. L. Buckley, L. S. Turnbull, G. P. Liney and A. J. Knowles, "Differentiation of prostatic carcinoma and benign prostatic hyperplasia: correlation between dynamic Gd-DTPA-enhanced MR imaging and histopathology," *Journal of Magnetic Resonance Imaging* **9**, 311–316 (1999).
14. M. R. Engelbrecht, H. J. Huisman, R. J. F. Laheij, G. J. Jager, G. J. L. H. van Leenders, C. A. Hulsbergen-Van De Kaa, J. J. M. C. H. de la Rosette, J. G. Blickman and J. O. Barentsz, "Discrimination of prostate cancer from normal peripheral zone and central gland tissue by using dynamic contrast-enhanced MR imaging," *Radiology* **229**, 248–254 (2003).
15. O. Rouvière, A. Raudrant, R. Ecochard, C. Colin-Pangaud, C. Pasquiou, R. Bouvier, J. M. Marechal and D. Lyonnet, "Characterization of time-enhancement curves of benign and malignant prostate tissue at dynamic MR imaging," *European Radiology* **13**, 931–942 (2003).
16. Y. Mazaheri, L. Bokacheva, D.-J. Kroon, O. Akin, H. Hricak, D. Chamudot, S. Fine and J. A. Koutcher, "Semi-automatic deformable registration of prostate MR images to pathological slices," *Journal of Magnetic Resonance Imaging* **32**, 1149–1157 (2010).
17. V. Shah, T. Pohida, B. Turkbey, H. Mani, M. Merino, P. A. Pinto, P. Choyke and M. Bernardo, "A method for correlating *in vivo* prostate magnetic resonance imaging and histopathology using individualized magnetic resonance-based molds," *Review of Scientific Instruments* **80**, 104301–104306 (2009).
18. S. Y. Kimm, J. H. Lee, D. G. Nishimura, T. V. Tarin, B. S. Hu, K. Jensen and J. D. Brooks, "MR-histology Correlation in *Ex-vivo* Human Prostate Specimens," in

- Proceedings of International Society for Magnetic Resonance in Medicine*, (Hawaii, USA, 2009).
19. Y. Zhan, Y. Ou, M. Feldman, J. Tomaszewski, C. Davatzikos and D. Shen, "Registering histologic and MR images of prostate for image-based cancer detection," *Academic Radiology* **14**, 1367–1381 (2007).
  20. Y. Ou, D. Shen, M. Feldman, J. Tomaszewski and C. Davatzikos, "Non-rigid registration between histological and MR images of the prostate: a joint segmentation and registration framework," in *Proceedings of Computer Vision and Pattern Recognition Workshops*, (2009), pp. 125–132.
  21. J. I. Epstein, P. C. Walsh, M. Carmichael and C. B. Brendler, "Pathologic and clinical findings to predict tumor extent of nonpalpable (stage T1c) prostate cancer," *Journal of the American Medical Association* **271**, 368–374 (1994).
  22. T. A. Stamey, F. S. Freiha, J. E. McNeal, E. A. Redwine, A. S. Whittemore and H. P. Schmid, "Localized prostate cancer: relationship of tumor volume to clinical significance for treatment of prostate cancer," *Cancer* **71**, 933–938 (1993).

## Chapter 5.

# **3D prostate histology reconstruction: an evaluation of image-based and fiducial-based algorithms<sup>†</sup>**

### 5.1 Introduction

Reconstruction methods based on the registration of 2D digital histology images to *in vivo* [1-6] or *ex vivo* [7-9] 3D images (referred to as *registration-based reconstruction methods* throughout this chapter) hold the potential for automated 3D histology image reconstruction with fewer assumptions about the relative positions of the histology sections and less disruption of pathologists' tissue slicing workflows. Furthermore, the reconstructions generated by these methods result in histology images that are inherently registered to a volumetric 3D image, further facilitating the co-registration with clinical images. Some such registration algorithms use only intrinsic image information and allow existing specimen handling processes to be used with minimal alteration beyond the additional imaging; however, this type of registration algorithm relies on the presence of intrinsic image features that may vary due to anatomical variation and, as noted by Ou et al. [7], by the presence of cancer foci, the very tissue these methods aim to align. To mitigate this dependence, other registration algorithms use image information derived from extrinsic fiducial markers applied to

---

<sup>†</sup> A version of this chapter has been published: E. Gibson, M. Gaed, J. A. Gómez, M. Moussa, C. Romagnoli, S. Pautler, J. L. Chin, C. Crukley, G. S. Bauman, A. Fenster and A. D. Ward, "3D prostate histology reconstruction: an evaluation of image-based and fiducial-based algorithms," *Medical Physics* **40**, 093501 (2013).

specimens before imaging, which increases robustness to image variation, but requires additional specimen handling.

This work explores this trade-off for registration-based reconstruction methods between intrinsic information-based and extrinsic fiducial-based registration algorithms via a direct comparison of the registration errors of four algorithms on the same data set quantified as a 3D TRE using homologous landmarks. This comparison comprised three experiments measuring the registration errors of these algorithms, measuring the sensitivity of the performance of each algorithm to initialization, and identifying the sources of any substantial registration errors that were observed. This work builds on our laboratory's previous image-guided slicing [10] method based on an altered pathology workflow using image guidance to prospectively control the slicing of the specimen, and on our subsequent extrinsic-fiducial-registration-based reconstruction method (Chapter 4) [11] that eliminated image-guided slicing to reduce this workflow disruption. In this work, we investigated an intrinsic image-information-based registration that could further reduce workflow disruption by eliminating the need for extrinsic fiducial markers and compared it to our previous extrinsic-fiducial-registration-based reconstruction method. To the best of our knowledge, this represents the first direct comparison between intrinsic-image-registration-based and extrinsic-fiducial-registration-based prostate histology reconstruction methods on the same data set with error quantified as a 3D TRE.

## 5.2 Materials and methods

### 5.2.1 Patient selection and imaging

We obtained twelve radical prostatectomy specimens from patients with prostate cancer being treated by one of three collaborating urologists and surgical oncologists.

The inclusion criteria were: (1) age 18 or older, (2) having histologically confirmed clinical prostate cancer of stage T1 or T2 on a previous biopsy, and (3) suitable for and consenting to radical prostatectomy. The exclusion criteria were: (1) prior therapy for prostate cancer, (2) use of 5-alpha reductase inhibitors within 6 months of the study start, (3) inability to comply with preoperative imaging, (4) allergy to contrast agents, (5) sickle cell or other anemias, (6) sources of artifact within the pelvis such as hip and penile prostheses, and (7) contraindications to MRI such as electronic implants, metal in the orbit, cerebral aneurysm clips, claustrophobia and morbid obesity. The subjects' ages ranged from 47 to 69 years. The specimens had volumes (estimated from transrectal ultrasound at biopsy) ranging from 19 to 49 cm<sup>3</sup> and biopsy Gleason scores of 6 or 7. This study was approved by the University of Western Ontario Human Subjects Research Ethics Board, and informed consent was obtained from each subject.

Prostatectomy specimens were formalin fixed in 10% buffered formalin for 48 hours and marked with strand-shaped fiducial markers approximately 1 mm in diameter on the exterior of the specimen and internal to the specimen (illustrated in Figure 5.1) that are visible on histology and MR images [11]. After adding fiducial markers, the intact specimens were imaged *ex vivo* in a 3 T GE Discovery MR750 (GE Healthcare, Waukesha, WI, USA) with an endorectal coil (Prostate eCoil, Medrad, Inc., Warrendale, PA, USA) using a T1-weighted protocol (3D SPGR, TR 6.5 ms, TE 2.5 ms, bandwidth  $\pm 31.25$  kHz, 8 averages, FOV 140×140×62 mm, image slice thickness 0.4 mm, image slice spacing 0.2 mm, 256×192 matrix, 312 image slices, flip angle 15°, 25 minutes) and a T2-weighted protocol (3D FSE, TR 2000 ms, TE 151.5 ms, bandwidth  $\pm 125$  kHz, 3 averages, FOV 140×140×62 mm, image slice thickness 0.4 mm, image slice spacing 0.2

mm, 320×192 matrix, 312 image slices, 25 minutes). During MR imaging, specimens were immobilized in a syringe filled with Christo-Lube (Lubrication Technology Inc., Franklin Furnace, OH, USA) to provide low signal in the background and minimize boundary artifacts.

As part of another study from which our specimens were obtained, removal of the prostatic apex was performed using image guidance to yield cuts coincident with *in vivo* MR imaging planes obtained preoperatively, and the midgland was cut into  $4.4 \pm 0.2$ -mm-thick tissue slices using a rotary slicer to yield parallel cuts [10]. After gross slicing, the resulting tissue slices were immobilized in tissue processing cassettes, immersed in Christo-Lube and imaged using the same imaging protocol. To distinguish these images from the whole-gland images described above, these images are referred to as *tissue slice MR* images hereafter.

After standard whole-mount paraffin embedding, a 4  $\mu$ m histology section was cut from each midgland tissue slice and stained with hematoxylin and eosin. After staining, histology sections were digitized on a ScanScope GL (Aperio Technologies, Vista, CA, USA) bright field slide scanner, and downsampled to a 30  $\mu$ m pixel size.

#### 5.2.1.1 3D reconstruction methods

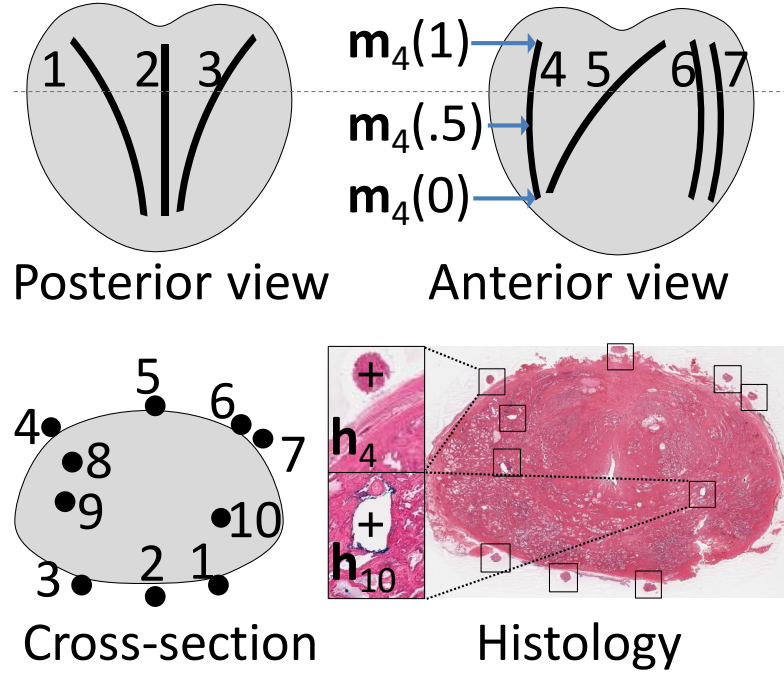
We evaluated four 3D registration-based reconstruction methods corresponding to four registration algorithms. Specifically, for each histology image independently, each algorithm computed a 2D–3D affine transformation from each histology image to the space of the corresponding whole-gland MR image by minimizing a similarity metric  $M$  over a constrained search space of linear transformations  $T$ , yielding a transformation

$$\tilde{A}_{i,\{T;M\}} = \underset{\hat{A} \in T}{\operatorname{argmin}} M(H_i, W_i, \hat{A}), \quad 5.1$$

where  $H_i$  is the  $i$ -th scalar histology image,  $W_i$  is the whole-gland MR image from the corresponding specimen, and  $\hat{A}: \mathbb{R}^2 \rightarrow \mathbb{R}^3$  maps the histology image plane  $\Omega$  into a plane  $\hat{A}(\Omega)$  in the whole-gland MR image space. The minimization was performed using a local iterative Nelder–Mead simplex algorithm implemented in Matlab R2011b (The Mathworks Inc., Natick, MA) as *fminsearch*. This minimization was initialized using an affine transformation, the construction of which differed in different experiments. The differences between the algorithms were the choice of the similarity metric  $M$  [mutual information ( $M_{MI}$ ) or fiducial registration error ( $M_{FRE}$ )] and the search space  $T$  [affine transformations ( $T^A$ ) or *fiducial-constrained* affine transformations ( $T^F$ )]. Each algorithm is correspondingly denoted  $\{T; M\}$ .

Two similarity metrics were used: mutual information ( $M_{MI}$ ) and fiducial registration error ( $M_{FRE}$ ).  $M_{MI}$  measures the negative normalized mutual information [12] (MI) of a 2D histogram relating image intensities from  $H_i$  and the oblique image slice through  $W_i$  defined by the plane  $\hat{A}(\Omega)$ , resampled at the coordinates of the transformed histology image. Three parameters that may affect the performance of the metric were varied in this work: the number of joint 2D intensity histogram bins, the color channel used for  $H_i$ , and the choice of MRI sequence. As these parameters are varied in the experiments, their selection will be described in Section 5.2.3. To avoid using fiducial information in  $M_{MI}$ , voxels containing fiducials were excluded from the histogram calculation.  $M_{FRE}$  measures the post-transformation misalignment between strand-shaped fiducial markers identified on  $H_i$  and  $W_i$ . Internal and external fiducial strands were attached to the specimen before imaging as shown in Figure 5.1. The  $j$ -th fiducial strand

on  $W_i$  was semi-automatically localized as a 3D parametric curve  $\mathbf{m}_{i,j}(s_{i,j} \in [0,1])$ . The center of the cross-section of the  $j$ -th fiducial marker on  $H_i$  was manually localized as a 2D point  $\mathbf{h}_{i,j}$ .  $M_{FRE}$  is the sum of squared 3D distances for each fiducial between  $\hat{A}(\mathbf{h}_{i,j})$  and the closest point on  $\mathbf{m}_{i,j}$  to the plane  $\hat{A}(\Omega)$ .

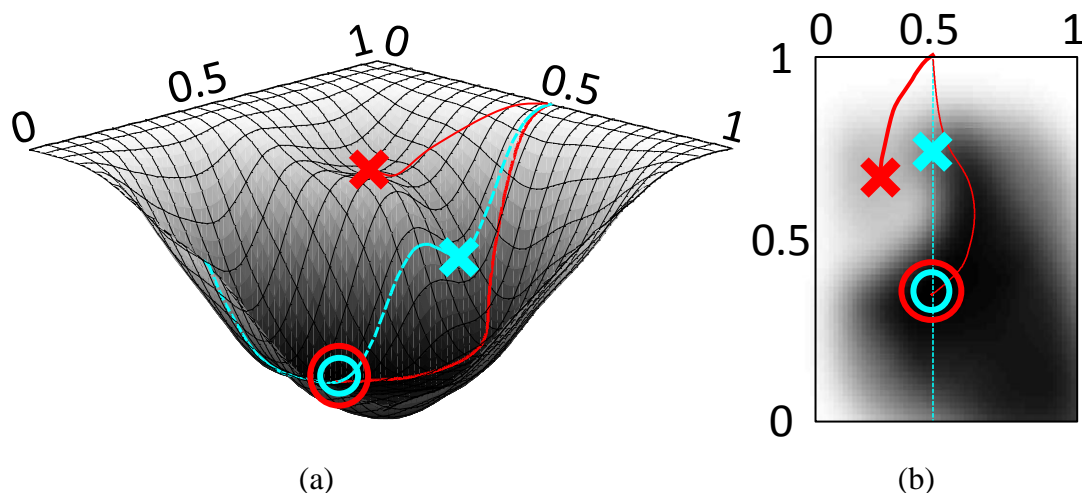


**Figure 5.1:** Schematic of strand-shaped fiducial markers (1–10), with parameterized 3D fiducial space curve  $\mathbf{m}_4$  indicated with arrows, and histology showing 2D fiducial cross-sections (boxed) with  $\mathbf{h}_4$  and  $\mathbf{h}_{10}$  inset.

Two search spaces were used: affine transformations ( $T^A$ ) and *fiducial-constrained* affine transformations ( $T^F$ ).  $T^A$  is the space of 2D–3D affine transformations, realized by optimizing over the uppermost 9 matrix elements of a  $4 \times 3$  homogeneous transformation matrix.  $T^F$  is a subspace of  $T^A$  used in previous work [11] that includes only transformations that can be realized from a best-fit affine mapping of the vector of 2D fiducial points  $\langle \mathbf{h}_{i,1}, \mathbf{h}_{i,2}, \dots, \mathbf{h}_{i,10} \rangle$  to a vector of 3D fiducial points  $\langle \mathbf{m}_{i,1}(s_{i,1}), \mathbf{m}_{i,2}(s_{i,2}), \dots, \mathbf{m}_{i,10}(s_{i,10}) \rangle$ . This space was realized by optimizing over the



vector of curve parameters,  $\langle \widehat{s}_{l,1}, \widehat{s}_{l,2}, \dots, \widehat{s}_{l,10} \rangle$ . Constraining the search space in this way is a trade-off between eliminating solutions (and potentially local minima) known to be non-optimal, and potentially introducing new local minima not present in the initial search space. This trade-off is illustrated in Figure 5.2.



**Figure 5.2:** Constraining the search space in an optimization reduces the number of possible solutions to be considered, but may introduce local minima. On this illustrative metric in a 2D search space (surface view (a) and 2D view (b), with darker intensity representing lower values), the red curves show two monotonically decreasing paths an unconstrained optimizer might take, one ending at the global optimum (red circle) and one ending at a local minimum (red x). If it is known *a priori* that the optimal solution for the metric lies on the cyan dashed line, an optimization could search just these solutions, avoiding solutions and local minima known to be suboptimal (off the cyan dashed line); however, this constrained space may have local optima (e.g. the cyan x) not present in the unconstrained space.

## 5.2.2 Metrics for algorithm evaluation

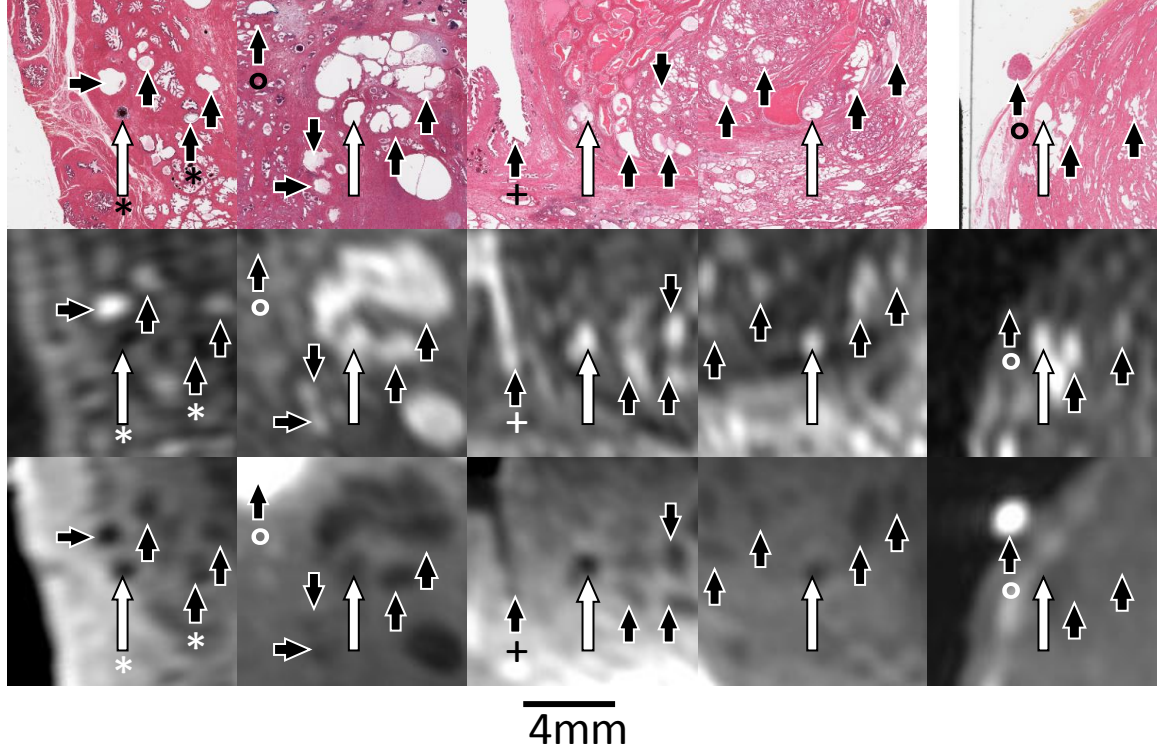
### 5.2.2.1 Target registration error

We quantified the registration error of each algorithm using the mean TRE [13], i.e. the mean post-registration 3D Euclidean distance between corresponding pairs of homologous landmarks on the whole-gland MR images and on the reconstructed

histological images, averaged over 232 landmarks from 37 sections taken from 10 specimens.

These point landmarks (3–16 landmarks per section) were interactively identified using 3D Slicer by one observer under advisement from 2 genitourinary pathologists (J.A.G. and M.M.) and 1 radiologist (C.R.) with expertise in prostate imaging.

Landmarks included the centers of atrophic ducts, cysts, and corpora amylacea, and were selected based on salient shape characteristics and the presence of nearby salient features that uniquely constrained the correspondence. Illustrative examples of these landmarks are shown in Figure 5.3.



**Figure 5.3:** Homologous landmark pairs (large white arrows), used for evaluation of the registration algorithms, shown on histology images (top) and on the corresponding oblique image slices on T2 (middle) and T1 (bottom) MR images. Small black arrows denote salient nearby features used to corroborate the established landmark correspondence, but not used as landmarks themselves. Salient features in these images include corpora amylacea (labeled \*), the urethra (labeled +), extrinsic fiducials (labeled °) and atrophic ducts (unlabeled). Figure reproduced here with permission from the Journal of Magnetic Resonance Imaging [11].

#### 5.2.2.2 Sensitivity to initialization

The sensitivity of the algorithms' mean TRE to initialization error was evaluated by initializing the algorithms at sets of different transformations, including the “ideal” section-specific least squares best-fit 2D–3D affine transformation aligning homologous landmarks on histology and MR images  $A_i$ , and also at four sets of transformations that deviated from  $A_i$  by controlled rotations, translations and scaling factors.  $A_i$  is computed as,

$$A_i = \underset{\hat{A} \in T^A}{\operatorname{argmin}} \sum_{k=1}^{N_i} \|p_{i,k} - \hat{A}(q_{i,k})\|^2, \quad 5.2$$

where  $q_{i,k}$  is the  $k$ -th landmark in the  $i$ -th histology image,  $p_{i,k}$  is the homologous landmark in the corresponding whole-gland MR image, and  $N_i$  is the number of landmarks identified in the  $i$ -th histology image. Sensitivity was quantified as the linear coefficient of a general linear model relating post-registration TRE to the absolute rotation, translation or scaling change of the initialization from  $A_i$ . Because identification of  $A_i$  is sensitive to error in the placement of the landmarks, sections with fewer than 5 identified landmarks were excluded from this calculation; thus, a total of 28 sections, taken from 10 specimens, were used.

### 5.2.2.3 Quantifying contribution of the sources of post-registration TRE

The landmark localization error (LLE) (measuring the repeatability of the landmark placement) was quantified as an unbiased estimator of the standard deviation of seven repeated localizations (at least one day apart) of 16 landmarks identified on MR and histology [11]. This metric was measured independently for histology and MR images.

The landmark registration error (LRE) (measuring the combined effect of error in localizing landmarks and the distance from the correct homology to the closest point in the search space) was quantified as the mean residual misalignment of homologous landmarks after applying the best-fit transformation from the search space,

$$LRE = \frac{1}{\sum_{i=1}^{S'} N_i} \sum_{i=1}^{S'} \min_{\hat{A} \in T} \sum_{k=1}^{N_i} \|p_{i,k} - \hat{A}(q_{i,k})\|, \quad 5.3$$

where  $S'$  is the number of sections. To mitigate under-estimation of LRE due to overfitting, sections with fewer than 5 identified landmarks were excluded from this calculation; in total,  $\sum_{i=1}^{S'} N_i = 198$  landmarks from  $S' = 28$  sections were used.

We explored the relative impacts of two other sources of error (the similarity metric not having a globally optimal value at the best transformation in the search space, and the optimizer not converging to the globally optimal value) by examining the similarity metric values after minimization. Ideally, the similarity metric value at  $A_i$  would be the global minimum of the similarity metric function (i.e.  $\min_{\hat{A} \in T} M(H_i, W_i, \hat{A}) = M(H_i, W_i, A_i)$ ) and the minimization would converge to this transformation ( $\tilde{A}_{i,\{T;M\}} = A_i$ ); in this case, the difference in metric values  $D_{i,\{T;M\}} = M(H_i, W_i, \tilde{A}_{i,\{T;M\}}) - M(H_i, W_i, A_i) = 0$ . If  $D_{i,\{T;M\}} > 0$ , the optimizer did not find the global minimum of the similarity metric function. If  $D_{i,\{T;M\}} < 0$ , the global minimum of the similarity function was not at  $A_i$ . We quantified the relative contributions of these sources of error as the proportion of minimizations that resulted in  $D_{i,\{T;M\}} > 0$  to those that resulted in  $D_{i,\{T;M\}} < 0$ .

### 5.2.3 Experimental design

We conducted three experiments in this work. In the first experiment (Section 5.2.3.1), the registration errors of four algorithms for the registration of 2D prostate histology images to 3D *ex vivo* MR images were compared by measuring the post-registration TRE of intrinsic homologous point landmarks when initialized with a straightforward initialization involving aligning the histology images in 2D, stacking them at even intervals and aligning the stack to the *ex vivo* MR image. The algorithms

differed in the optimized similarity metric (mutual information or fiducial registration error) and the search space examined (affine transformations or a subset of affine transformations constrained by the spatial configuration of fiducial markers). In the second experiment (Section 5.2.3.2), the sensitivities of TRE to initializations that deviated from the best-fit transformation for these four algorithms were compared by varying the initialization and recording the resulting TREs. In the third experiment (Section 5.2.3.3), the causes of substantial registration errors were explored by quantifying contributions to error from five possible sources: (1) the selection of parameters of the similarity metric, (2) the incorrect localization of the homologous landmarks, (3) the correct transformation not being part of the search space, (4) the similarity metric not having a globally optimal value at the best transformation in the search space, and (5) the optimizer not converging to the globally optimal value.

The first two experiments required the selection of a single configuration of the three MI parameters described in Section 5.2.1.1. Rather than resorting to *ad hoc* parameter selection, we used the nine histology images from two randomly selected specimens from our data set to guide the selection of the MI parameters. For these nine histology images,  $M_{MI}$  at the best-fit affine registration (identified using homologous intrinsic landmarks) was minimized over all parameter combinations (histopathology image color channel chosen from red, green, blue, and the mean of all three; MRI sequence chosen from the T1-weighted and T2 weighted sequences; and number  $N$  of joint intensity histogram bins chosen from  $\{6, 16, 32, 64, 128\}$ ). This procedure yielded one set of parameters giving the lowest  $M_{MI}$  (the blue channel of the histology image, the T1-weighted MRI sequence and  $N = 6$ ); this set is henceforth referred to as the *selected*

*MI parameter set.* The two specimens used to determine the selected MI parameter set were not used in any of our experiments; thus, for all three experiments, our results were calculated based only on the remaining 10 specimens in our data set.

### 5.2.3.1 Algorithm performance with practical initialization

To assess the algorithm performance in practical applications, we initialized each algorithm using a transformation that can be practically achieved by researchers able to control, and collect data during the coarse slicing of the specimen. With this initialization, we executed each algorithm, and computed the post-registration TRE. To define the initialization, extrinsic fiducials were stacked into a 3D volume by (1) identifying the vector  $\langle \mathbf{f}_{i,1}, \mathbf{f}_{i,2}, \dots, \mathbf{f}_{i,10} \rangle$  of 2D points of extrinsic fiducial cross-sections on the front face of the  $i$ -th tissue slice in a standard 2D coordinate system, (2) stacking these vectors at 4.4 mm intervals (i.e. the quantified gross slicing thickness [10]) to yield a 3D point set, (3) computing a rigid transform  $\tilde{R}_i$  mapping the 3D point set to the fiducial marker curves  $\langle \mathbf{m}_{i,1}, \mathbf{m}_{i,2}, \dots, \mathbf{m}_{i,10} \rangle$  in the whole-gland MR image using the iterative closest points algorithm [14], and (4) computing the best-fit affine initialization transformation  $\tilde{A}_i$  mapping  $\langle \mathbf{h}_{i,1}, \mathbf{h}_{i,2}, \dots, \mathbf{h}_{i,10} \rangle$  to  $\langle \tilde{R}_i \mathbf{f}_{i,1}, \tilde{R}_i \mathbf{f}_{i,2}, \dots, \tilde{R}_i \mathbf{f}_{i,10} \rangle$  for each section  $i$ . In this work, the vector  $\langle \mathbf{f}_{i,1}, \mathbf{f}_{i,2}, \dots, \mathbf{f}_{i,10} \rangle$  was determined by interactively identifying the inferior-most point of each fiducial on tissue slice MR images that were manually registered to whole-gland MR images, and projecting these points onto the plane at the front face of the superior-most tissue slice; however, an analogous vector  $\langle \mathbf{f}_{i,1}, \mathbf{f}_{i,2}, \dots, \mathbf{f}_{i,10} \rangle$  could also be identified by photography of the specimen in a common coordinate system after each cut during gross sectioning, as performed by Breen et al. [15]. For this experiment,  $M_{MI}$  used the selected MI parameter set.

### 5.2.3.2 Sensitivity of algorithm accuracy to initialization variation

Due to the variability in imaging data sets and initialization methods, initialization errors may deviate from those in the experiment described above. Accordingly, the sensitivity of the algorithms' registration error to initialization error was assessed. Each algorithm was initialized with the best-fit affine transformation  $A_i$  and also with four sets of transformations that deviated from  $A_i$  by controlled rotations, translations or scaling factors, and the post-registration TRE was calculated. The first and second sets of transformations varied the initialization by rotations about the AP axis and about the LR axis, respectively, with rotations in both sets ranging from  $-15^\circ$  to  $15^\circ$  in  $3^\circ$  increments. The AP and LR axes were determined by manually aligning the whole-gland MR images in 3D Slicer. The third set of transformations varied the initialization by translations in the inferior direction normal to the plane  $A_i(\Omega)$ , ranging from -10 mm to 10 mm of translation in 1 mm increments. The fourth set of transformations varied the initialization by scaling factors from -10% to 10% in 2.5% increments. For each type of deviation from  $A_i$ , a linear coefficient of sensitivity was calculated as described in Section 5.2.2.2. For this experiment,  $M_{MI}$  used the selected MI parameter set.

### 5.2.3.3 Sources of error for mutual information

Based on preliminary observations of the results of these experiments, we were motivated to explore the relative contributions of four causes of error to the observed TRE. To quantify variability in landmark localization, we computed the LLE for histology and MR images. To quantify error caused by the combined effect of landmark localization and search space we computed the LRE for  $T^A$ . To quantify the relative prevalence of errors in the optimizer and in the similarity metric, we computed the



difference between the similarity metric value at the post-optimization transformation and at the landmark-based best fit transformation ( $D_{i,\{T^A, M_{MI}\}}$ ) as discussed in Section 5.2.2.3.

To explore the impact of parameter selection on the observed errors, the accuracy and sensitivity to initialization experiments (described in Sections 5.2.3.1 and 5.2.3.2) were run for  $\{T^A; M_{MI}\}$  with all combinations of the  $M_{MI}$  parameters described in Section 5.2.3. (i.e. different histology color channels, MRI sequences and histogram sizes).

## 5.2.4 Statistical analysis

Statistical analyses were performed in SPSS 19 (SPSS, Inc., Chicago, USA). ANOVA tests used repeated measures ANOVA with Greenhouse–Geisser corrections for asphericity, and used a threshold of  $p = 0.05$  for significance. For experiment 1, the improvement in mean TRE from the pre-registration initialization to the post-registration transformation was assessed using a 1-way ANOVA and pairwise *post hoc* analyses between the TRE at initialization compared to the TRE after each of the algorithms. Differences in the mean TRE due to the search space ( $T^A$  or  $T^F$ ) and the similarity metric ( $M_{MI}$  or  $M_{FRE}$ ) were assessed using a 2-way ANOVA and pairwise *post hoc* analyses. For experiment 2, the sensitivities of the algorithms to the four types of initialization errors (rotational about AP and LR axes, translational, and scaling) were assessed using four separate 2-way ANOVAs. In each ANOVA, the factors were the choice of algorithm and the rotational, translational or scaling deviations of the initialization from  $A$ . In *post hoc* analyses for each algorithm, we modeled sensitivity as a general linear model relating post-registration TRE to the absolute rotation, translation or scale of the initialization from  $A_i$ , estimating the linear coefficient as a measure of sensitivity. The sensitivity of the mean TRE of algorithms using  $M_{MI}$  to the choice of MRI sequence, histology image

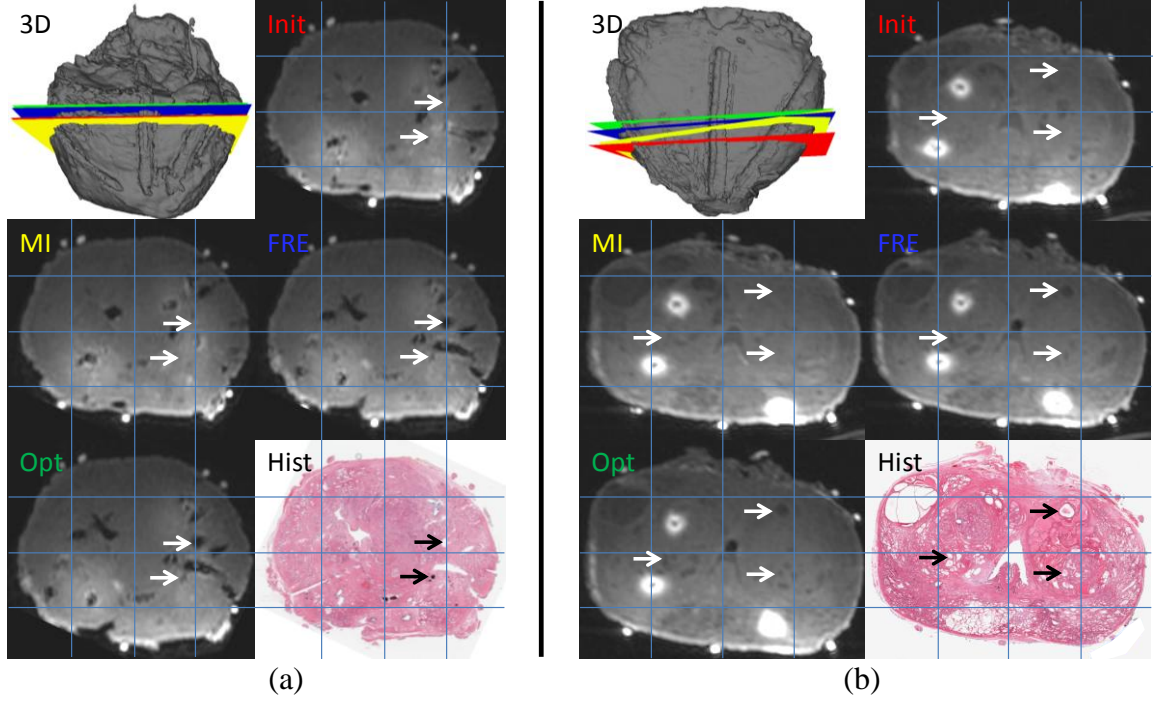
channel and number of histogram bins was assessed using a 3-way ANOVA and *post hoc* analyses. The relative contributions of sources of error were estimated with point estimates for the LLE, LRE, and the proportion of  $D_{i,\{T^A;M_{MI}\}}$  above and below zero for the practical initializations and each set of perturbed initializations for each set of  $M_{MI}$  parameters.

## 5.3 Results

### 5.3.1 Algorithm performance with practical initialization

The key finding of this experiment was that the choice of similarity metric affected the mean TRE, with the landmark-based similarity metric outperforming the intensity-based similarity metric: algorithms using  $M_{FRE}$  had a mean TRE of  $0.7 \pm 0.4$  mm, an improvement of 0.5–0.7 mm from the initialization (with mean TRE  $1.3 \pm 0.6$  mm), whereas algorithms using  $M_{MI}$  had a mean TRE of  $1.2 \pm 0.7$  mm, an improvement of 0.1–0.2 mm from the initialization.

Results for individual algorithms, and computational times are shown in Table 5.1. An illustrative registration from the practical initializations is shown in Figure 5.4a. All algorithms showed a statistically significant improvement in mean TRE over initialization. The 2-way ANOVA of TRE showed a significant difference due to choice of metric but failed to find a difference due to the choice of search space and failed to find an interaction between the two factors. Pairwise *post hoc* analyses (shown in Table 5.2) showed that (1) algorithms using  $M_{FRE}$  yielded a lower mean TRE than those using  $M_{MI}$  for algorithms using both search spaces, and (2) the difference in mean TREs due to the choice of search space was less than 0.12 mm with 95% confidence.



**Figure 5.4:** Illustrative registrations from (a) the practical initialization experiment and (b) the initialization offset by 5 mm translational error from the sensitivity experiment, showing a posterior view of a surface rendering of the prostate with sections (represented by colored planes) reconstructed via 4 different transformations: the initialization (red), the transformation given by  $\{T^A; M_{MI}\}$  (yellow), the transformation given by  $\{T^A; M_{FRE}\}$  (blue), and the best-fit transformation (green). Corresponding oblique image slices of the T1-weighted MR image are shown in the “Init”, “MI”, “FRE”, and “Opt” panels, respectively. The H&E-stained histology image is shown in the “Hist” panel.

**Table 5.1:** Comparison of the algorithms after practical initialization: mean  $\pm$  standard deviation of TRE, 95% confidence interval (CI) on the improvement over the initialization  $\ddot{A}$  (i.e.,  $TRE_{\ddot{A}} - TRE_{\{T,M\}}$ ), and mean computational time per section of the four registration algorithms, (columns 2 to 5).

	$\{T^A; M_{FRE}\}$	$\{T^A; M_{MI}\}$	$\{T^F; M_{FRE}\}$	$\{T^F; M_{MI}\}$	$\ddot{A}$
<b>TRE (mm)</b>	$0.7 \pm 0.4$	$1.2 \pm 0.7$	$0.7 \pm 0.4$	$1.2 \pm 0.7$	$1.3 \pm 0.6$
<b><math>TRE_{\ddot{A}} - TRE_{\{T,M\}}</math> (mm)</b>	[0.55,0.70]	[0.05,0.22]	[0.53,0.69]	[0.06,0.23]	
<b>Computation time (s)</b>	2	32	9	38	

**Table 5.2:** 95% CI on mean TRE difference for similarity metrics and search spaces across initialization algorithms.

<i>Factor held constant</i>	$M_{FRE}$	$M_{MI}$	$T^A$	$T^F$
<i>Pairwise difference analyzed</i>	$T^F - T^A$	$T^F - T^A$	$M_{MI} - M_{FRE}$	$M_{MI} - M_{FRE}$
<b>95% CI on mean TRE difference</b>	[-0.03,0.00]	[-0.10,0.12]	[0.40,0.57]	[0.38,0.55]

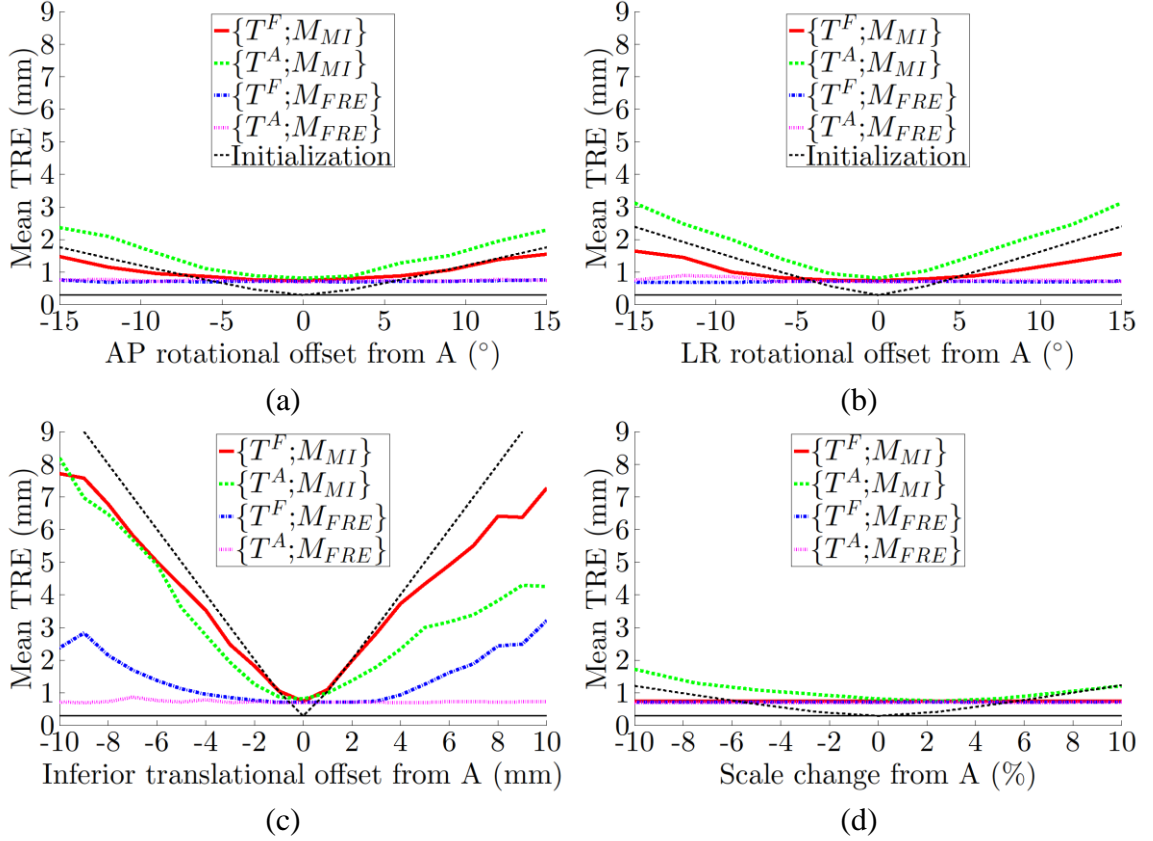
### 5.3.2 Sensitivity of algorithm accuracy to initialization variation

The key finding from this experiment was that the landmark-based algorithm using an affine transformation ( $\{T^A; M_{FRE}\}$ ) showed negligible sensitivity to initialization error, while the other three algorithms showed greater sensitivity.

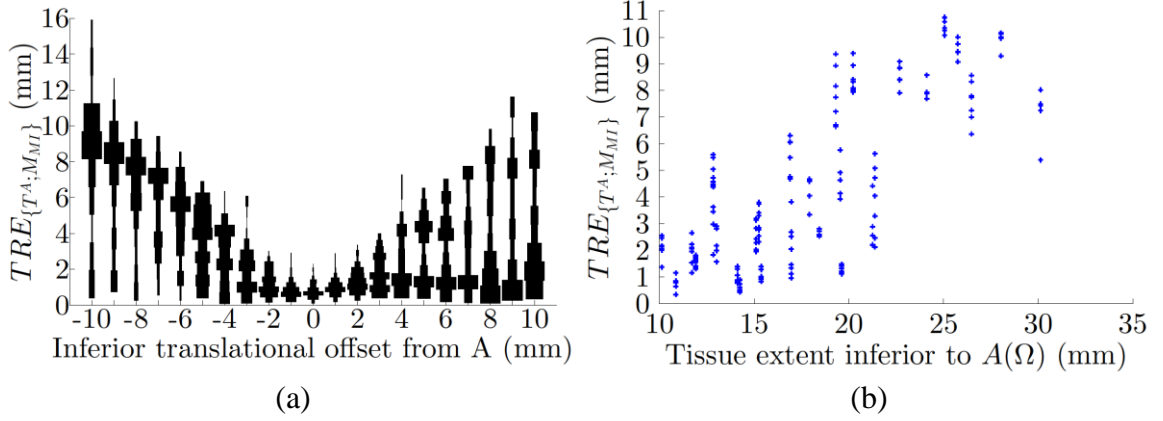
95% CI on the sensitivity coefficients for each type of initialization error for each algorithm are shown in Table 5.3, and the graphical relationships are shown in Figure 5.5. An illustrative registration from the sensitivity experiment is shown in Figure 5.4b. The ANOVA showed significant effects of sensitivity to initialization. The  $\{T^A; M_{MI}\}$  algorithm shows an asymmetry in the sensitivity to translational offset; the distributions of TRE for translational offsets (Figure 5.6a) are bimodal for extreme inferior translational offsets, and this variation was correlated ( $\rho = 0.77$ ) with the extent of tissue in the inferior direction normal to the plane  $A(\Omega)$  of the best-fit histological section (Figure 5.6b).

**Table 5.3:** Sensitivity of mean TRE to initialization error quantified as 95% CI on the sensitivity coefficients of general linear models.

<i>Perturbation</i>	$\{T^A; M_{FRE}\}$	$\{T^A; M_{MI}\}$	$\{T^F; M_{FRE}\}$	$\{T^F; M_{MI}\}$
<b>AP rotation (mm/<math>^\circ</math> rotation AP)</b>	[0.00,0.01]	[0.15,0.17]	[0.00,0.00]	[0.06,0.07]
<b>LR rotation (mm/<math>^\circ</math> rotation LR)</b>	[0.00,0.01]	[0.10,0.12]	[0.00,0.01]	[0.05,0.06]
<b>Translation (mm/mm translation)</b>	[0.00,0.00]	[0.57,0.61]	[0.22,0.26]	[0.70,0.74]
<b>Scaling (mm/% scale)</b>	[0.00,0.00]	[0.06,0.08]	[0.00,0.00]	[0.00,0.00]



**Figure 5.5:** Sensitivity of mean TRE to four types of initialization error: rotation about the (a) AP and (b) LR axes, (c) translation, and (d) scaling.



**Figure 5.6:** (a) Violin plot showing the distribution of TRE for the  $\{T^A; M_{MI}\}$  algorithm for translational initialization error; (b) scatter plot of TRE vs. the extent of tissue in the inferior direction normal to the plane  $A(\Omega)$  of the best-fit transformation.

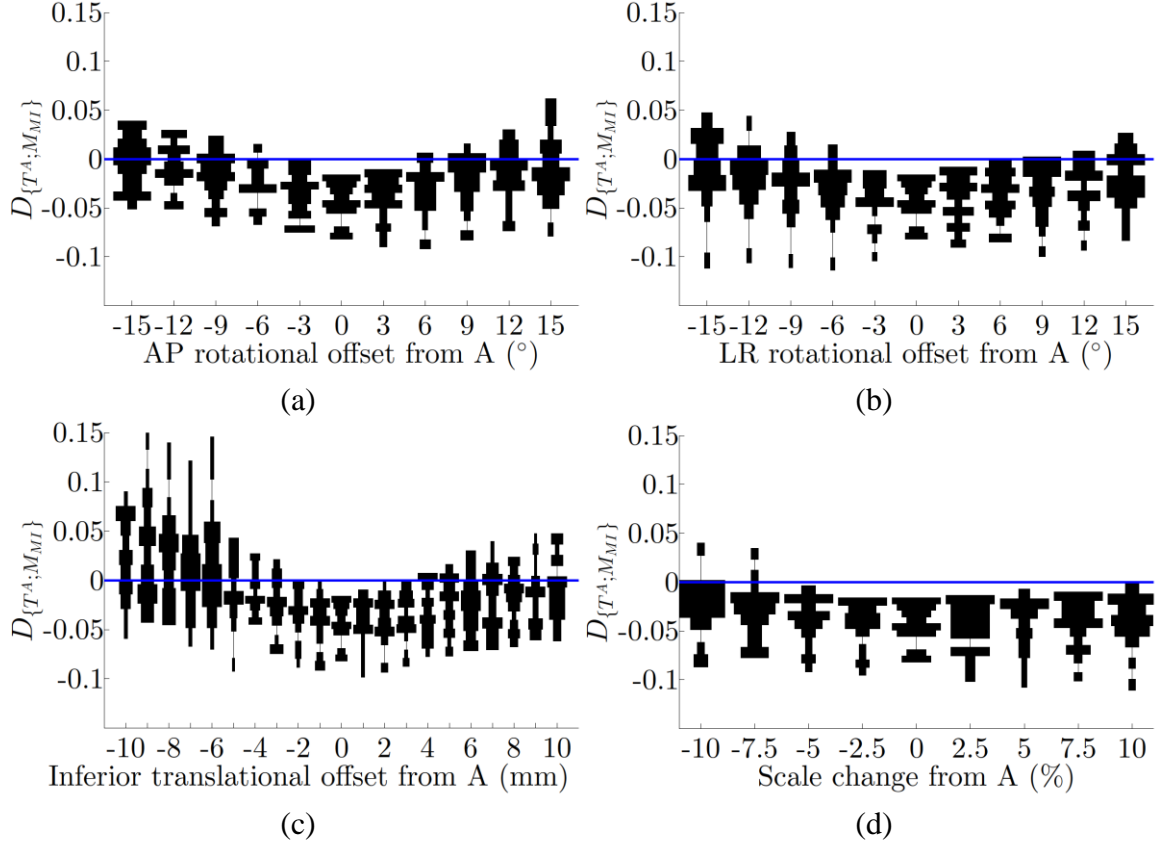
### 5.3.3 Sources of error for mutual information

The key findings for this experiment were (1) the observed mutual information-based registration errors were not fully explained by landmark localization and landmark registration errors, and (2) in more than 80% of cases, the mutual information metric value at the ideal registration was not the globally optimal value, suggesting that the mutual information metric may not be ideally suited to these cases. More specifically, (1) the landmark localization error values were 0.05 mm and 0.16 mm on histology and MR images, respectively, and the landmark registration error was 0.29 mm, insufficient to account for the observed registration errors with mutual information, and (2) across the 40 sets of mutual information parameters tested, the proportion of optimizations that resulted in metric values less than those at the best fit transformation ( $D_{\{T^A; M_{MI}\}} < 0$ , showing evidence of errors due to the similarity metric) ranged from 81% to 98%.

Figure 5.7 shows the distribution of  $D_{\{T^A; M_{MI}\}}$  over all histology sections for each perturbed initialization for  $M_{\{T^A; M_{MI}\}}$  using the selected MI parameter set. For the perturbed initializations, with the  $M_{MI}$  parameters used in experiments 1 and 2, we

observed that 18% of the optimizations in our sample resulted in  $D_{\{T^A; M_{MI}\}} > 0$  and 82% resulted in  $D_{\{T^A; M_{MI}\}} < 0$ . Of the optimizations that resulted in  $D_{\{T^A; M_{MI}\}} > 0$ , the majority (10% of all optimizations) resulted from initializations that were translated from  $A_i$  by  $\geq 5$  mm.

The lowest mean TRE for all of the parameter combinations was  $1.2 \pm 0.7$  mm, using the selected MI parameter set. Mean TRE was sensitive to MRI sequence selection, histology channel selection and number of histogram bins, and there were significant interactions between the factors. Sensitivity to MRI sequence selection was the most significant effect where all algorithms using the T2-weighted MRI sequence had higher mean TRE than all those using the T1-weighted MRI sequence (95% CI on the difference of estimated marginal means: [0.78, 0.97] mm). Within the algorithms using the T1-weighted MRI sequence, a *post hoc* 2-way ANOVA showed that mean TRE was sensitive to the number of histogram bins, but failed to detect sensitivity to the choice of histology channel or interaction between the factors. Among the algorithms using the T1-weighted MRI sequence, the 95% CI on the difference between the parameter combination with the lowest mean TRE (the selected MI parameter set using the blue histology channel, the T1-weighted MRI sequence and  $N = 6$ ) and the highest mean TRE (using the blue histology channel, the T1-weighted MRI sequence and  $N = 16$ ) was [0.07, 0.22] mm. On our sample of 10 specimens, no MI parameter combination yielded a lower mean TRE than the selected MI parameter set that was used in experiments 1 and 2.



**Figure 5.7:** Violin plots showing the distribution of  $D_{MI}$  for four types of initialization error: rotation about the AP (a) and LR (b) axes, translation (c), and scaling (d). Each data point represents one run of the  $\{T^A; M_{MI}\}$  algorithm;  $D_{\{T^A; M_{MI}\}} > 0$  (above the blue line) shows that the optimizer did not find the global minimum, while  $D_{\{T^A; M_{MI}\}} < 0$  (below the blue line) shows that the similarity metric had values lower than that of the best-fit transformation.

## 5.4 Discussion

In this chapter, we have performed a comparison of accuracy and sensitivity to initialization for four registration-based 3D histology reconstruction algorithms on the same data set, comparing registration algorithms that use intrinsic image information with algorithms that use extrinsic fiducial information. Specifically, we explored the effects of using  $M_{MI}$ , a similarity metric based on the mutual information of intrinsic image information, compared to using  $M_{FRE}$ , a similarity metric introduced in previous work [11] based on the fiducial registration error of extrinsic fiducials. We also explored



the impact of using the transformation spaces  $T^A$ , comprising the unconstrained set of affine transforms, compared to using  $T^F$ , a transformation space used in previous work [11] comprising a subset of affine transforms constrained to align extrinsic fiducials.

### 5.4.1 Evaluating histology reconstruction methods

Two criteria that are important in the evaluation of methods for the 3D reconstruction of histology from clinical specimens are the disruption to the clinical workflow and the accuracy of the reconstruction.

#### 5.4.1.1 Disruption to the clinical workflow

Because the specimens are used clinically, care must be taken to minimize unnecessary disruption to the clinical use of the data for the sake of research. The clinical workflow can be disrupted in several ways, including altering which tissue is cut for histological examination, altering the appearance of histology sections or altering the timeline of processing by introducing additional processing steps.

Several existing reconstruction methods control which tissue is cut for histological examination by constraining the slicing of the specimen into 3–5-mm-thick tissue slices, so that histology is taken from tissue that approximately corresponds to *in vivo* imaging planes [10, 16-18]. The spatial relationships of these histology sections may differ from those normally used for diagnosis by a pathologist.

Retrospective methods, including registration-based reconstruction methods that use registration algorithms based on extrinsic fiducial information [11, 19, 20] or intrinsic image information (as described in this work and elsewhere [4-9]), as well as other

retrospective methods [21, 22], avoid the disruption of controlling which tissue is cut for histological examination by retrospectively determining the spatial relationships between histology sections regardless of how they were cut, enabling the slicing of histology according to the pathologists' workflow. Many approaches achieve this by adding extrinsic fiducials to the prostate specimen; however, care must be taken to avoid disrupting the appearance of the histology by adding features that would not typically appear on histology or removing/moving tissue when fiducials are introduced within the prostate. By using intrinsic image information, the mutual information-based registration algorithm described in this work and other such methods [2-9, 23] avoid this challenge.

The use of additional *ex vivo* imaging for registration, as seen in the methods presented in this work and other existing work, may disrupt the processing timeline of the pathology laboratory, which some methods [21] avoid by relying on known geometrical properties of their extrinsic fiducials.

#### 5.4.1.2 Accuracy of the reconstruction

The accuracy of the reconstruction (and the accuracy of the following registration to *in vivo* imaging) may constrain the types of questions that can be asked of the data and may impact the power of some statistical analyses of the data (Chapter 2) [24]. A comparison of the accuracy of reconstruction methods is challenging, due to the variety of methods for quantifying accuracy.

Some methods do not report reconstruction errors as point-to-point distances, instead using surrogates such as overlap of the prostate gland [8, 25], prostate boundary distances [2, 22], qualitative assessment of tumor alignment [6, 18], post-optimization similarity metric values [6], or image intensity properties [9].

Even reconstruction errors reported as distance must be interpreted carefully, considering the modalities involved in the reconstruction, the method of measurement, the dimensionality of the measurement and the errors that may be omitted from the measurement. Previous work in our laboratory using a guided-slicing approach measured reconstruction error from histology to *in vivo* images comprising a 2D in-plane mean error of  $1.1 \pm 0.7$  mm based on intrinsic landmarks and a root-mean-square error in the position of the slicing plane of 0.9 mm based on physical measurements. However, these measurements were made under an assumption that histology corresponded to the front faces of the thick tissue slices (referred to as the *front face assumption* hereafter), and as the mean distance of whole-mount prostate histology sections from the front face has been estimated to be  $\sim 1$  mm (Chapter 3) [26] these errors may underestimate the total error accordingly. Park et al. measured reconstruction errors of 3.74 and 2.26 mm based on a distance between medial axes of foci contoured on histology and *in vivo* images. Direct comparison with these errors is challenging, as they represent reconstruction to align with *in vivo* images, where there is more deformation.

Groenendaal et al. reported an error of  $2.2 \pm 0.5$  mm from histology to *in vivo* images, but broke the error down into a histology—fixed tissue image TRE component of  $0.7 \pm 0.3$  mm based on extrinsic landmarks which similarly held the front face assumption, and a fixed tissue to *in vivo* image error component of  $2.1 \pm 0.5$  mm. Orczyk et al. broke the error into a histology-fresh tissue slice image FRE (where the landmarks used for error calculation were also used to compute the transformation, typically underestimating the TRE) component of  $0.9 \pm 0.3$  mm measured using intrinsic anatomical landmarks that did not measure out-of-plane error due to the 3D reconstruction and held the front face

assumption, and a reconstructed specimen—*in vivo* image FRE of  $1.6 \pm 0.3$  mm. Kimm et al. reported 2D displacements (interpretable as a surrogate for a distance-based reconstruction error) from histology to fixed tissue of  $1.2 \pm 0.6$  mm and  $0.9 \pm 0.4$  mm for two methods measured using the misalignment of intrinsic landmarks using a metric that is invariant to rotational and translational error and that may underestimate error relative to a 2D TRE, and that held the front face assumption. Nir et al. reported a mean  $\pm$ SD 2D TRE of  $1.5 \pm 0.9$  mm using intrinsic fiducials. Prabu et al. reported a mean 2D TRE of 1.4 mm using intrinsic landmarks. Two pieces of related work by Zhan et al. and Ou et al. used intrinsic landmarks to measure TREs of 0.8 mm and  $0.6\text{--}1.0 \pm 0.43\text{--}0.79$  mm (range is for landmarks chosen by 2 observers), respectively. The identity and number of the intrinsic landmarks are not specified, and it is not clear if the calculated TREs are 2D or 3D. Recent work by Hughes et al. reported a 2D TRE of  $0.6 \pm 0.5$  mm using one extrinsic landmark per section and a 3-fiducial FRE of  $0.1 \pm 0.1$  mm on a beef liver surrogate and a 3-fiducial FRE of  $0.2 \pm 0.1$  mm on prostate specimens.

To the best of our knowledge, no existing publications have directly compared the accuracy of intrinsic image registration-based algorithms to extrinsic fiducial registration-based reconstruction algorithms on the same data set and measured the error as a 3D TRE based on homologous landmarks localized in 3D on the 3D images (to avoid underestimation due to the front face assumption). Our work addresses this gap in knowledge by comparing our previously developed fiducial-based algorithm to an intrinsic image-based algorithm using mutual information, measuring the 3D TRE using 232 homologous intrinsic point landmarks localized in 3D on the *ex vivo* MR images. The fiducial-based reconstructions yielded a 3D TRE of  $0.7 \pm 0.4$  mm, among the highest

accuracies of current fiducial-based methods. The methods based on mutual information of image intensities yielded a 3D TRE of  $1.2 \pm 0.7$  mm, suggesting that more complex intensity-based methods, such as those of Zhan et al. and Ou et al, may be necessary to eliminate the use of fiducial markers while achieving accuracies as low as the most accurate fiducial-based methods.

There is no TRE value that is universally accepted as necessary and sufficient for all applications of 3D prostate histology reconstruction. The necessary TRE depends principally on the central question of the research study served by the 3D histology reconstruction algorithm. More specifically, as it has recently been shown that TRE can affect the statistical power and the number of subjects needed to answer a research question (Chapter 2) [24], answering one particular research question (e.g. localization of small, high-grade prostate cancer on imaging) could require a 3D histology algorithm with a submillimeter TRE, whereas a different question (e.g. coarse identification of larger prostate tumors) could be answerable using an algorithm with a multi-millimeter TRE. Determination of the necessary and sufficient TRE can be done in detail in the context of a typical application of 3D histology reconstruction methods: studies evaluating imaging modalities by correspondence with a histological reference standard. In such studies, error in the alignment of ground truth information to the images being validated introduces uncertainty that prohibits certain analyses or reduces the statistical power. The recent introduction of a sample size calculation that incorporates quantified registration error (Chapter 2) [24] provides an avenue to quantify the impact of the algorithms' mean TRE in imaging validation studies, assuming reconstruction error dominates registration error. The sample size calculation addresses imaging validation

studies to detect a difference in the mean signal intensity between regions of interest (ROI) on images from a modality under study based on ROI defined on registered images from a reference standard modality. It augments the classical power calculation formulae by modeling the impact of the mean and variance of registration accuracy on the study power (and thus the required sample size), showing that higher registration accuracy leads to higher study power or a lower required sample size. Because it also provides a mapping from TRE to the registration error parameters used in the calculation (under the assumption of spherical regions of interest of specified volume, and 3D Gaussian-distributed target registration error), we can use this relationship to assess the impact of mean TRE from different algorithms.

The potential impact of small improvements in TRE can be elucidated by applying this sample size calculation in the context of a simplified case study. For a hypothetical imaging validation study of signal differences between tumor and benign tissue for the smallest clinically significant prostate cancer focus (0.2 cc [27]), the sample size calculation predicts (assuming equal image intensity variance of tumor and benign tissue) that an algorithm with a mean TRE of 1.2 mm ( $\{T^A; M_{MI}\}$ ) would need 7% fewer subjects than would be needed with a mean TRE of 1.3 mm (pre-optimization), while an algorithm with a mean TRE of 0.7 mm ( $\{T^A; M_{FRE}\}$ ) would need 27% fewer subjects. Considering the high per-subject cost of imaging validation studies (e.g. a contemporary cost of more than \$10,000 USD per subject for the study on which this chapter is based), we observe that using a more accurate reconstruction can have a substantial impact on the overall cost of a study. For example, in the ongoing imaging validation study of 66 subjects at our center, the above sample size reduction of 7% resulting from a 0.1 mm

improvement in TRE could translate to a need for 4 fewer subjects, resulting in a savings of approximately \$40,000 USD.

#### 5.4.2 Algorithm performance with practical initialization

When initialized with the described practical initialization (with a TRE of 1.3 mm), all algorithms significantly improved the mean TRE. Since the choice of practical initialization was inspired by several existing histology reconstruction methods that identify the front faces of tissue slices via guided sectioning or tissue slice imaging, these results suggest that the accuracy of existing reconstruction methods could be improved by applying these algorithms. Although  $\{T^A; M_{MI}\}$  showed a small improvement over initialization without the use of fiducial markers, the size of the difference (95% CI: 0.05 to 0.22 mm) should be considered when assessing its value for refining an image-guided-slicing reconstruction method.

Our experiments failed to detect a significant difference in mean TRE between algorithms due to the choice of search space used. As the difference in mean TRE was less than 0.12 mm, and the mean computational times of all four algorithms was less than 40 seconds per section, constraining the search space to the fiducials may not warrant the added complexity.

#### 5.4.3 Sensitivity of algorithm accuracy to initialization variation

The sensitivity of algorithms using  $M_{FRE}$  to initialization error was lower than that of algorithms using  $M_{MI}$ .  $\{T^A; M_{FRE}\}$  showed no material change in mean TRE across the range of initializations tested. This suggests that this algorithm can be used without the need for accurate initialization. The algorithms using  $M_{MI}$  were more sensitive to

rotational, translational and, in the case of  $\{T^A; M_{MI}\}$ , scaling initialization, suggesting that algorithms using  $M_{MI}$  should be initialized close to the correct transformation. The algorithm using  $\{T^A; M_{MI}\}$  was less sensitive to translation in the inferior direction, particularly for the more inferior sections. The initial translation in the inferior direction combined with the narrowing of the apical region resulted in initializations with poor overlap of tissue. We speculate that superior translation (compensating for the initialization error) could rapidly improve the overlap (and the metric value), thus directing the optimizer in that direction. This effect would be less pronounced for more superior sections and for initialization errors in the superior direction because the cross-section of the tissue in the midgland and base would not change as quickly with superior/inferior translation. As in the previous experiment, constraining the search space to the fiducial markers does not appear to improve the sensitivity of algorithms using either similarity metric.

The range of initializations was chosen to model realistic use cases. The range of rotational offsets tested reflects variation seen in unguided specimen slicing [28]. The broad range of translational offsets was chosen because the variation of this property in the usual clinical setting was unknown. The range of scaling factors was chosen by dividing a reported scaling factor between *in vivo* tissue and fixed tissue (0.96) by reported scaling factors between *in vivo* and histology (0.87 to 0.96) [29], yielding a scaling factor from histology to fixed tissue (1.00 to 1.10). The tested range was extended to be from 0.90 to 1.10 for symmetry. The use of ranges that span realistic use cases suggests that these results may be generalizable to practical use cases.



#### 5.4.4 Sources of error for mutual information

The high sensitivity of the mean TRE of the  $\{T^A; M_{MI}\}$  algorithm to initialization error, and the marginal improvement over the initialization motivated a deeper exploration of the causes of these errors for this algorithm.

The four sources of post-registration TRE described in Section 5.2.2.3 are interrelated, and cannot be measured directly without knowledge of the true image homology (for error due to landmark placement and erroneous search space) and the complex interactions between the search space, the similarity metric, the data and the optimizer (for errors due to the similarity metric and the optimizer). However, quantifying aspects of these errors may elucidate their relative contributions. The upper bound for the sum of the LRE (reflecting variability in landmark placement) and LLE (reflecting manual landmark alignment that cannot be accounted for by affine deformation), which occurs when they are perfectly correlated, was only 0.48 mm, suggesting the majority of the residual error after registration with  $\{T^A; M_{MI}\}$  is due to other sources. Measurements of  $D_{i,\{T^A; M_{MI}\}}$  show that errors due to the optimizer converging to a local minimum, as well as errors due to the optimizer converging to a global or local minimum whose metric value is less than that at the best-fit transformation, are present. Thus, it appears that both the image similarity metric and the optimizer may be sources of error contributing to our measured TRE. We observe from Figure 5.7 that for optimizations initialized close to the best-fit transformation, the image similarity metric seems to be the dominant contributing factor to the TRE, whereas for initializations far from the best-fit transformation, we observe more frequently that the

optimizer converges to a local optimum. This brings into question the suitability of using the mutual information of these modalities for this task.

The variation of MI parameters suggests that these results were not disrupted by error due to a suboptimal selection of parameters. While the TRE was sensitive to parameter variation, and in particular to the choice of MRI sequence used, the reported TRE values in our experiment measuring registration accuracy after practical initialization were the lowest for any tested combination of MI parameters, and the evidence of errors due to the similarity metric was present for all tested parameters.

#### 5.4.5 Limitations

The conclusions of this work should be considered in the context of the scope of our experiments. Our experiments tested a single image-based similarity metric, mutual information, against a single fiducial-based similarity metric. Mutual information was tested as it is a commonly chosen metric for intermodality registrations; however, the poor performance of mutual information in this application cannot be generalized to other intensity-based metrics. This work tested a range of values for several parameters of mutual information; however, mutual information has other tunable parameters that were not varied, including spacing of histogram bins and sampling patterns. An exhaustive search of the parameter space of  $M_{MI}$  was beyond the scope of this work. It should also be noted that while sensitivities to rotational, translational and scaling initialization error were assessed independently, the interactions between these sensitivities and the sensitivities to anisotropic scaling were not investigated. Finally, the specimens, drawn from another study, were cut into tissue slices following an image-guided slicing

approach; thus, the variability of histology section orientations relative to the prostate gland may be less than seen in typical clinical workflows.

#### 5.4.6 Conclusion

This work compares the accuracy and sensitivity to initialization of 3D prostate histology registration-based reconstruction algorithms. Minimizing a mutual information-based image similarity metric over affine transformations yielded a mean TRE of 1.2 mm when initialized using a practical initialization (pre-optimization mean TRE 1.3 mm), and was sensitive to the initialization error. Errors due to both the optimizer and the mutual information-based similarity metric contributed to the residual error, with the latter more common when optimizations were initialized close to the best-fit transformation. In contrast, minimizing a fiducial registration error-based metric over affine transformations yielded a mean TRE of 0.7 mm, with no material change in mean TRE across initializations representing the full range of clinical use cases. Thus, for registration-based reconstruction methods, we recommend the use of the extrinsic-fiducial-based registration algorithm using affine transformations. If the application of fiducials is prohibited and assuming MR imaging sequences equivalent to those used in this study, the MI-based algorithm could be used to achieve a 0.1 to 0.2 mm improvement in registration error, with the caveat that accurate initialization may be critical to its performance. For an imaging validation study to detect a difference in the image intensity of tumor and benign tissue for the smallest clinically significant prostate cancer focus (0.2 cc [27]), the mutual information-based algorithm could reduce the number of subjects needed by 7% compared to an approach based on stacking aligned histology

images on the front faces of tissue slices at a fixed spacing, while the fiducial-registration-error-based algorithm could reduce the number of subjects needed by 27%.

## 5.5 References

1. J. Chappelow, B. N. Bloch, N. Rofsky, E. Genega, R. Lenkinski, W. DeWolf and A. Madabhushi, "Elastic registration of multimodal prostate MRI and histology via multiattribute combined mutual information," *Medical Physics* **38**, 2005–2018 (2011).
2. P. Patel, J. Chappelow, J. Tomaszewski, M. D. Feldman, M. Rosen, N. Shih and A. Madabhushi, "Spatially weighted mutual information (SWMI) for registration of digitally reconstructed ex vivo whole mount histology and in vivo prostate MRI," in *Proceedings of IEEE Engineering in Medicine & Biology Society*, (2011), pp. 6269–6272.
3. G. Nir and S. E. Salcudean, "Registration of whole-mount histology and tomography of the prostate using particle filtering," in *Proceedings of SPIE Medical Imaging*, (2013), pp. 86760E.
4. J. Chappelow, S. Viswanath, J. Monaco, M. Rosen, J. Tomaszewski, M. Feldman and A. Madabhushi, "Improving supervised classification accuracy using non-rigid multimodal image registration: detecting prostate cancer," in *Proceedings of SPIE Medical Imaging*, (2008), pp. 69150V.
5. S. B. Prabu, R. Toth and A. Madabhushi, "A statistical deformation model (SDM) based regularizer for non-rigid image registration: application to registration of multimodal prostate MRI and histology," in *Proceedings of SPIE Medical Imaging*, (2013), pp. 86760C–86760C.
6. J. Chappelow, B. N. Bloch, N. Rofsky, E. Genega, R. Lenkinski, W. DeWolf, S. Viswanath and A. Madabhushi, "COLLINARUS: collection of image-derived non-linear attributes for registration using splines," in *Proceedings of SPIE Medical Imaging*, (2009), pp. 72592N.
7. Y. Ou, D. Shen, M. Feldman, J. Tomaszewski and C. Davatzikos, "Non-rigid registration between histological and MR images of the prostate: a joint segmentation and registration framework," in *Proceedings of Computer Vision and Pattern Recognition Workshops*, (2009), pp. 125–132.
8. Y. Zhan, Y. Ou, M. Feldman, J. Tomaszewski, C. Davatzikos and D. Shen, "Registering histologic and MR images of prostate for image-based cancer detection," *Academic Radiology* **14**, 1367–1381 (2007).

9. J. Chappelow, A. Madabhushi, M. Rosen, J. Tomaszewski and M. Feldman, "Multimodal image registration of ex vivo 4 Tesla MRI with whole mount histology for prostate cancer detection," in *Proceedings of SPIE Medical Imaging*, Vol. 6512 (2007), pp. 65121S
10. A. D. Ward, C. Crukley, C. A. McKenzie, J. Montreuil, E. Gibson, C. Romagnoli, J. A. Gómez, M. Moussa, J. Chin, G. Bauman and A. Fenster, "Prostate: registration of digital histopathologic images to in vivo MR images acquired by using endorectal receive coil," *Radiology* **263**, 856–864 (2012).
11. E. Gibson, C. Crukley, M. Gaed, J. A. Gómez, M. Moussa, J. L. Chin, G. S. Bauman, A. Fenster and A. D. Ward, "Registration of prostate histology images to ex vivo MR images via strand-shaped fiducials," *Journal of Magnetic Resonance Imaging* **36**, 1402–1412 (2012).
12. F. Maes, A. Collignon, D. Vandermeulen, G. Marchal and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Transactions on Medical Imaging* **16**, 187–198 (1997).
13. J. M. Fitzpatrick and J. B. West, "The distribution of target registration error in rigid-body point-based registration," *IEEE Transactions on Medical Imaging* **20**, 917–927 (2001).
14. P. J. Besl and N. D. McKay, "A method for registration of 3-D shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **14**, 239–256 (1992).
15. M. S. Breen, T. L. Lancaster, R. S. Lazebnik, S. G. Nour, J. S. Lewin and D. L. Wilson, "Three-dimensional method for comparing in vivo interventional MR images of thermally ablated tissue with tissue response," *Journal of Magnetic Resonance Imaging* **18**, 90–102 (2003).
16. A. S. Jackson, S. A. Reinsberg, S. A. Sohaib, E. M. Charles-Edwards, S. Jhavar, T. J. Christmas, A. C. Thompson, M. J. Bailey, C. M. Corbishley, C. Fisher, M. O. Leach and D. P. Dearnaley, "Dynamic contrast-enhanced MRI for prostate cancer localization," *British Journal of Radiology* **82**, 148–156 (2009).
17. L. H. Chen, H. Ho, R. Lazaro, C. H. Thng, J. Yuen, W. S. Ng and C. Cheng, "Optimum slicing of radical prostatectomy specimens for correlation between histopathology and medical images," *International Journal of Computer Assisted Radiology and Surgery* **5**, 471–487 (2010).
18. V. Shah, T. Pohida, B. Turkbey, H. Mani, M. Merino, P. A. Pinto, P. Choyke and M. Bernardo, "A method for correlating in vivo prostate magnetic resonance imaging and histopathology using individualized magnetic resonance-based molds," *Review of Scientific Instruments* **80**, 104301–104306 (2009).

19. G. Groenendaal, M. R. Moman, J. G. Korporaal, P. J. van Diest, M. van Vulpen, M. E. Philippens and U. A. van der Heide, "Validation of functional imaging with pathology for tumor delineation in the prostate," *Radiotherapy and Oncology* **94**, 145–150 (2010).
20. S. Y. Kimm, T. V. Tarin, J. H. Lee, B. Hu, K. Jensen, D. Nishimura and J. D. Brooks, "Methods for registration of magnetic resonance images of ex vivo prostate specimens with histology," *Journal of Magnetic Resonance Imaging* **36**, 206–212 (2012).
21. C. Hughes, O. Rouvière, F. Mege-Lechevallier, R. Souchon and R. Prost, "Robust alignment of prostate histology slices with quantified accuracy," *IEEE Transactions on Biomedical Engineering* **60**, 281–291 (2012).
22. S. Bart, P. Mozer, P. Hemar, G. Lenaour, E. Comperat, R. Renaud-Penna, E. Chartier-Kastler and J. Troccaz, "MRI-histology registration in prostate cancer," in *Proceedings of Surgetica*, (2005).
23. C. Orczyk, A. Mikheev, A. Rosenkrantz, J. Melamed, S. S. Taneja and H. Rusinek, "Imaging of prostate cancer: a platform for 3D co-registration of in-vivo MRI ex-vivo MRI and pathology," in *Proceedings of SPIE Medical Imaging, Vol. 8316* (2012), pp. 83162M.
24. E. Gibson, A. Fenster and A. D. Ward, "Registration accuracy: how good is good enough? A statistical power calculation incorporating image registration uncertainty," in *Proceedings of Medical Image Computing and Computer Assisted Intervention, Vol. 7511* (Springer, Nice, France, 2012), pp. 643–650.
25. L. S. Taylor, B. C. Porter, G. Nadasdy, P. A. di Sant'Agnese, D. Pasternack, Z. Wu, R. B. Baggs, D. J. Rubens and K. J. Parker, "Three-dimensional registration of prostate images from histology and ultrasound," *Ultrasound in Medicine and Biology* **30**, 161–168 (2004).
26. E. Gibson, J. A. Gomez, M. Moussa, C. Crukley, G. Bauman, A. Fenster and A. D. Ward, "3D reconstruction of prostate histology based on quantified tissue cutting and deformation parameters," in *Proceedings of SPIE Medical Imaging, Vol. 8317* (2012), pp. 83170N
27. J. I. Epstein, P. C. Walsh, M. Carmichael and C. B. Brendler, "Pathologic and clinical findings to predict tumor extent of nonpalpable (stage T1c) prostate cancer," *Journal of the American Medical Association* **271**, 368–374 (1994).
28. G. J. Jager, E. T. Ruijter, C. A. van de Kaa, J. J. de la Rosette, G. O. Oosterhof, J. R. Thornbury and J. O. Barentsz, "Local staging of prostate cancer with endorectal MR imaging: correlation with histopathology," *American Journal of Roentgenology* **166**, 845–852 (1996).

29. A. R. Schned, K. J. Wheeler, C. A. Hodorowski, J. A. Heaney, M. S. Ernstoff, R. J. Amdur and R. D. Harris, "Tissue-shrinkage correction factor in the calculation of prostate cancer volume," *American Journal of Surgical Pathology* **20**, 1501–1506 (1996).

## Chapter 6.

# **Prostate cancer assessment and delineation on 3 T multi-parametric MRI: interobserver and intersequence agreement<sup>†</sup>**

### 6.1 Introduction

Accurate delineation of suspected intraprostatic cancer foci on imaging is important for many aspects of treatment planning for men with localized prostate cancer. For example, the use of targeted biopsy procedures may enhance the probability of sampling high-grade foci [1], thus improving the ability to stratify men between intervention or careful monitoring with active surveillance. Estimations of cancer volume may have prognostic importance complementary to Gleason grade [2]. For many patients with intermediate-grade localized prostate cancer, focal therapy may be an appropriate treatment option with lower risk of morbidity than radical prostatectomy or radiotherapy [3, 4]. Focal therapies, including cryotherapy [3], focal laser ablation [5] and high-intensity focused ultrasound [6], normally aim to treat prostate cancer under image guidance by focusing treatment on cancerous lesions identified and delineated on pre-

---

<sup>†</sup> A version of this chapter is in preparation for submission: E. Gibson, C. Romagnoli, M. Bastian-Jordan, D. W. Cool, Z. Kassam, M. Gaed, M. Moussa, J. A. Gómez, S. Pautler, J. L. Chin, C. Crukley, G. S. Bauman, A. Fenster, A. D. Ward, “Prostate cancer assessment and delineation on 3 T multi-parametric MRI: interobserver and intersequence agreement” Radiology (in preparation).



treatment images. Accurate and precise delineation of these lesions on imaging is of critical importance because untreated aggressive prostate cancer increases the risk of recurrence, while treatment of healthy tissue and surrounding organs at risk increases the risk of morbidity. Furthermore, for men with more advanced cancers, treatment intensification using focal radiotherapy boosts to intraprostatic gross tumor volumes has been proposed as a strategy to improve outcomes [7, 8].

mpMRI has shown promise for detection [9, 10], staging [11] and sextant localization [12-14] of prostate cancer. mpMRI can include T2W [9, 15], DW [16], DCE [11, 17] and MR spectroscopy [18, 19] images, as well as parametric maps derived from these images, such as ADC [20, 21] maps. The European Society of Urogenital Radiology has recently published clinical guidelines [10] for prostate mpMRI, including a structured reporting scoring system (PI-RADS) which includes, for each lesion, scores for each sequence and a score for likelihood of cancer. Such guidelines are aimed at *scoring* lesions and localizing them to standardized prostate regions. However, the use of mpMRI for *delineation* (identifying the boundaries of suspicious regions on imaging) of prostate cancer lesions has not yet been widely studied [22, 23].

In this study, we investigated the variability of prostate cancer scoring and delineation on images from three 3 T mpMRI sequences: T2W, ADC and DCE images. Specifically, for lesions scored to be equivocal, likely cancerous or highly likely cancerous (PI-RADS likelihood scores 3 – 5), we addressed the following four questions. (1) What is the agreement in PI-RADS likelihood and sequence-specific scores (referred to hereafter as *interobserver score agreement*)? (2) What is the average agreement in delineated suspicious regions (measured using the Dice overlap coefficient, the mean

absolute distance between boundaries, and the absolute volume difference) between observers on images from each of these sequences (referred to hereafter as *interobserver contour agreement*)? (3) What is the average agreement in delineated suspicious regions between images from the different sequences (referred to hereafter as *intersequence contour agreement*)? (4) What are the interactions of *interobserver score agreement*, *interobserver contour agreement* and *intersequence contour agreement* with the histological grade, anatomical location and volume of the lesions?

## 6.2 Materials and methods

### 6.2.1 Participants

The study was approved by the human subjects research ethics board of our institution, and written informed consent was obtained from all subjects prior to enrollment. 107 patients were screened for this study between March 2010 and December 2011, selected from patients being treated for prostate cancer by three collaborating urologists and surgical oncologists (including S.P. and J.L.C., with 13 and 30 years of experience, respectively). The inclusion criteria were: (1) age 18 or older, (2) clinical stage T1 or T2 prostate cancer histologically confirmed on previous biopsy, and (3) suitability for and willingness to undergo radical prostatectomy. The exclusion criteria (with the numbers excluded in parentheses) were: (1) prior therapy for prostate cancer (n=8); (2) use of 5-alpha reductase inhibitors within 6 months of the study start (n=6); (3) inability to comply with preoperative imaging (n=1); (4) allergy to contrast agents (n=0); (5) sickle cell or other anemias (n=0); (6) insufficient renal function (n=1); (7) sources of artifact within the pelvis, such as hip and penile prostheses (n=1); (8) contraindications to MRI such as electronic implants, metal in the orbit, cerebral aneurysm clips,

claustrophobia and morbid obesity (n=11); and (9) prostate sizes exceeding the dimensions of 5.1 cm × 7.6 cm (2 inch × 3 inch) whole-mount histology slides (n=15). In total, 17 subjects were included in this study. The demographics and clinical characteristics of these subjects are given in Table 6.1.

**Table 6.1:** Subject demographics and clinical characteristics.

<b>Patient age median, range (y)</b>	63, 45–68
<b>PSA level median, range (ng/mL)</b>	4.8, 1.3–11.2
<b>Prostate volume estimated from transrectal ultrasound median, range (cm<sup>3</sup>)</b>	35, 19–49
<b>Pathologic stage (number of subjects)</b>	
<b>T2c</b>	12
<b>T3a</b>	5
<b>Patient overall Gleason score (number of subjects)</b>	
<b>3+3</b>	5
<b>3+4</b>	11
<b>4+3</b>	1

### 6.2.2 Imaging

Subjects were imaged before radical prostatectomy in a 3 T GE Discovery MR750 (GE Healthcare, Waukesha, WI, USA) with an endorectal coil (Prostate eCoil, Medrad, Inc., Warrendale, PA, USA). The imaging protocol included T2W 2D fast spin echo, DCE 3D spoiled gradient-recalled echo, and DW 2D echo-planar sequences, with parameters shown in Table 6.2. All phases of the DCE images, without computed pharmacokinetic maps, were available for interpretation during scoring and delineation tasks. DW images were post-processed on the MRI system console to generate ADC images, and both were available for interpretation during scoring and delineation tasks.

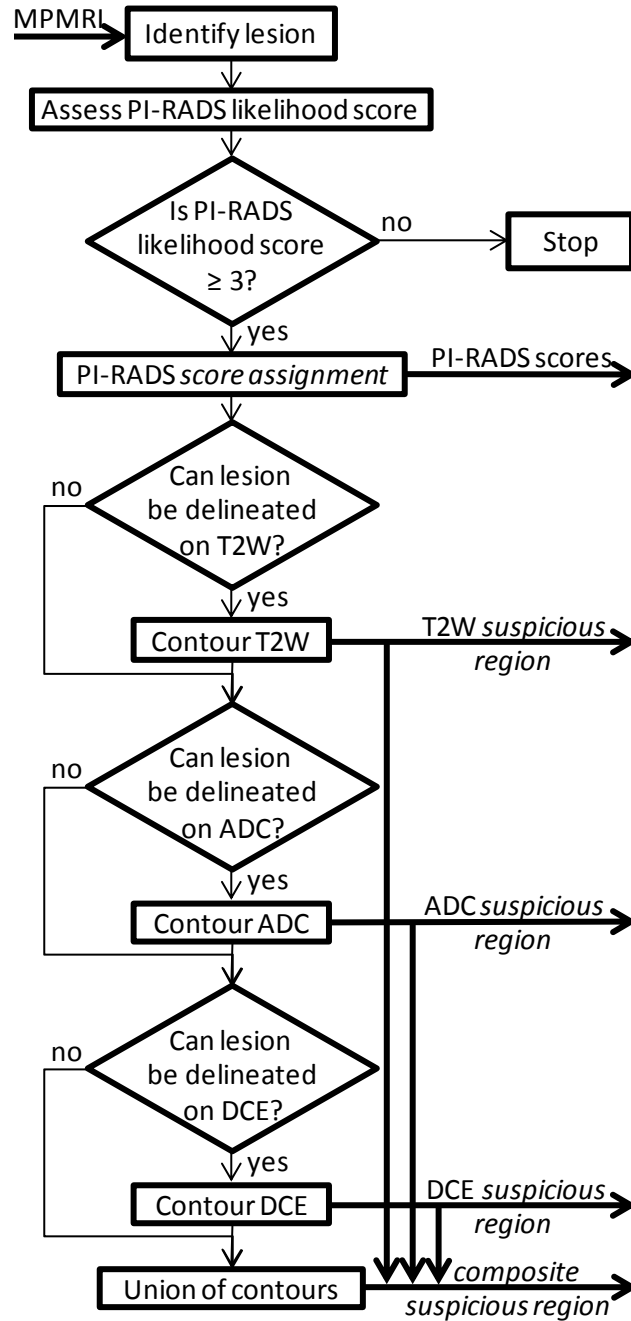
**Table 6.2:** MRI sequence parameters.

<i>Sequence</i>	<i>T2W</i>	<i>DCE</i>	<i>DW</i>
<b>Repetition time (ms)</b>	4000–13000	5.6–5.9	4000
<b>Echo time (ms)</b>	156–164	2.1–2.2	70–77
<b>Bandwidth (kHz)</b>	31.25	31.25	125.00
<b>Number of excitations</b>	2	1–2	3
<b>Field of view (cm)</b>	14	14	14
<b>Slice thickness (mm)</b>	2.2	2.8	3.3–3.6
<b>Slice spacing (mm)</b>	2.2	1.4	3.3–3.6
<b>Matrix</b>	320 × 192	256 × 192	128 × 256
<b>Number of slices</b>	40	42	20–34
<b>Flip angle (°)</b>	90	15	90
<b>Temporal spacing (s)</b>	N/A	90	N/A

### 6.2.3 Lesion scoring and delineation

Images from each subject were independently read, blinded to histopathological reports, by four observers who were aware of the inclusion criteria for the study: a radiology resident (D.W.C.), a radiology fellow (M.B.-J.) and two radiologists (C.R. and Z.K.) with 4.5, 2, 5.5, and 2 years of experience in prostate MRI assessment, respectively. For each lesion, observers followed the process illustrated in Figure 6.1, and described below, for lesion scoring and delineation. Images were assessed following the PI-RADS guidelines [10] using ClearCanvas Workstation 7.1 (ClearCanvas Inc., Toronto, Canada). These guidelines specify recording, for each lesion, the location within a 27-segment template, a 5-point likelihood score [denoting that clinically significant cancer is (1) highly unlikely, (2) unlikely, (3) equivocal, (4) likely or (5) highly likely] as well as separate 5-point sequence-specific scores for each sequence assessed (i.e. three scores; one for each of the T2W, ADC and DCE sequences in our study) where higher scores indicate image features associated with cancer. In this study, this record

(henceforth referred to as a *score assignment*) was made only for lesions (henceforth referred to as *scored* lesions) having a PI-RADS likelihood score of 3 (“equivocal”) or greater. By *a priori* agreement, if a *scored* lesion recorded by at least one observer has no corresponding *score assignment* recorded by another observer, a default PI-RADS likelihood score of 2, denoting “unlikely cancerous”, was assigned for the latter observer. On each sequence (i.e. T2W, DCE and ADC), each observer delineated a *suspicious region* corresponding to each *scored* lesion that the observer felt could be contoured. In this chapter, the term *composite suspicious region* refers to the union of the *suspicious regions* (from different sequences) for a *scored* lesion recorded by an observer. Delineations of *suspicious regions* were performed in ITK-SNAP 2.4 ([www.itksnap.org](http://www.itksnap.org)) using either the paintbrush or polygon mode (at the observer’s discretion), yielding a binary label map depicting each *suspicious region*. For DCE images, *suspicious regions* were delineated on one phase chosen by the observer. If one or more images was considered by an observer to be non-diagnostic due to imaging artifacts, that observer could decline to record the corresponding PI-RADS scores and/or to delineate the corresponding *suspicious regions* for that image; these missing data were excluded from the measurements of agreement.



**Figure 6.1:** Observers' process for generating the PI-RADS scores and *suspicious regions* analyzed in this study. Thin arrows denote process flow. Thick arrows denote data (images, *suspicious region* contours, and PI-RADS scores).

#### 6.2.4 Interobserver lesion correspondence

To calculate *interobserver contour* and *score agreements*, correspondences of *scored* lesions between different observers were defined. Between observers, two *scored*

lesions were taken to correspond if their *composite suspicious regions* overlapped. This correspondence approach was intended to yield aggregate lesions which were spatially distinct, and therefore each set of corresponding *scored* lesions is referred to in aggregate as a *distinct* lesion in this chapter. For two cases, the overlap of *composite suspicious regions* did not yield unambiguous lesion correspondence. In these cases, *suspicious regions* were manually corresponded (as described in detail in Appendix B).

### 6.2.5 Interobserver score agreement

The *interobserver score agreement* was quantified as the average of the absolute differences in PI-RADS likelihood and sequence-specific scores between pairs of observers for corresponding *scored* lesions. We further quantified agreement in PI-RADS likelihood scores using Schouten's weighted kappa [24] and using the proportion of *distinct* lesions with four-observer consensus on classifying *scored* lesions with respect to 2 thresholds:  $<3$  (highly unlikely or unlikely cancer) vs.  $\geq 3$  (equivocal, likely or highly likely cancerous) and  $<4$  (unlikely cancer or equivocal) vs.  $\geq 4$  (likely or highly likely cancer). The interactions of PI-RADS scores and *interobserver score agreement* with histological grade, anatomical location (i.e. peripheral zone or non-peripheral zone) and volume were quantified using correlation coefficients.

### 6.2.6 Interobserver contour agreement

For each of the sequences, the *interobserver contour agreement* for corresponding *suspicious regions* was measured using three pairwise metrics: (1) the Dice overlap coefficient, (2) the symmetric mean absolute distance (MAD) and (3) the absolute difference between volumes. When *suspicious regions* corresponded across more than

two observers, the agreement was averaged across all pairwise comparisons. The Dice overlap was computed as the volume of the overlap of two *suspicious regions* divided by the average volume of the two *suspicious regions*. The symmetric MAD was computed by (a) resampling two *suspicious regions* onto a 3D isotropic grid with a voxel side length of 0.25 mm (approximately the smallest voxel dimension in the data set) using nearest neighbor interpolation; (b) computing the average 3D Euclidean distance from each voxel on the boundary of the first resampled *suspicious region* to the nearest voxel on the boundary of the second resampled *suspicious region*, and vice versa from the second resampled *suspicious region* to the first; and (c) averaging the two average distances. Correlations of the *interobserver contour agreements* with the histological grade, anatomical location, lesion volume as measured on MRI, mean PI-RADS likelihood score and the mean PI-RADS sequence-specific score (for the same sequence) were assessed.

### 6.2.7 Intersequence contour agreement

For each pair of sequences, the *intersequence contour agreement* for *suspicious regions* for the same *scored* lesion was measured using the same three pairwise metrics as described in the previous section. When *suspicious regions* for the same *scored* lesion were identified on a pair of sequences by more than one observer, the metrics were averaged across the observers. To account for different imaging grids for different sequences, distortion and patient motion, each metric in this analysis was computed by (1) interactively rigidly registering the MR images to a common reference frame using 3D Slicer 3.6 (Surgical Planning Lab, Harvard Medical School, Boston, USA), (2) resampling the *suspicious regions* onto a common 0.25 mm isotropic grid using nearest-



neighbor interpolation and (3) computing the three pairwise metrics on the resampled *suspicious regions*. The mean intersequence target registration error was estimated to be 1 mm using intrinsic landmarks (comprising small atrophic cysts in regions of benign hyperplasia or in the peripheral zone) on images from 3 subjects (4–6 landmark pairs per image pair, 41 landmark pairs in total). Correlations of the *intersequence contour agreements* with the histological grade, anatomical location, lesion volume as measured on MRI, and mean likelihood score were assessed.

*Score assignment* was performed on all sequences together; however, delineation of *suspicious regions* was performed on each sequence separately, and each observer made a subjective judgment of whether or not he/she could adequately delineate on each sequence a *suspicious region* corresponding to each *scored* lesion. For each sequence, we therefore also calculated the ratio of the number of delineated *suspicious regions* to the number of *scored* lesions. The ideal value of this ratio is 1, where every *scored* lesion was also contoured.

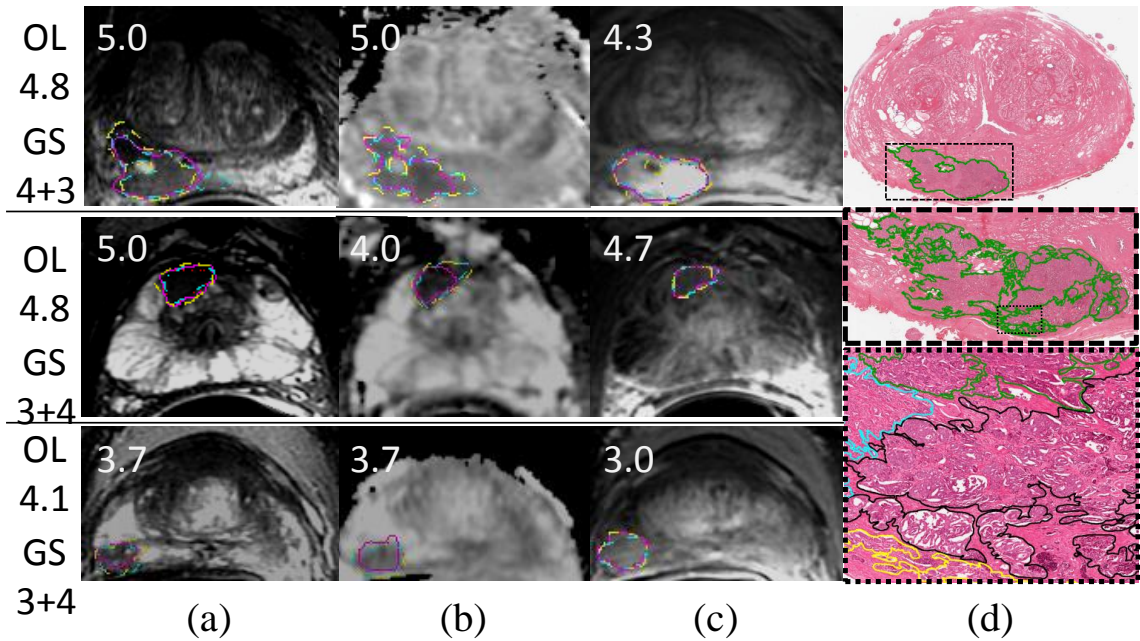
## 6.2.8 Interaction factors

### 6.2.8.1 Histological grade

To determine the histological grade for each *distinct* lesion, digitized histology images from the prostatectomy specimens were examined, and the resulting histological grades were assigned to *distinct* lesions based on their corresponding *composite suspicious regions* on MR images. Histology images were independently contoured and graded in ImageScope 10 (Aperio Technologies, Vista, CA, USA) using a pen-enabled 21.5 inch widescreen display (Cintiq 22HD, Wacom Co., Ltd., Otone, Japan) at a 20x

equivalent scale (illustrative contours are shown in Figure 6.2) by a clinician (M.G., with 2.5 years of experience in contouring prostate cancer contours on digital histology under the training and supervision of J.A.G. and M.M., genitourinary pathologists with 13 and 22 years of experience, respectively) blinded to mpMRI. These contours and grades were edited if required and confirmed by a genitourinary pathologist (J.A.G. or M.M.) blinded to mpMRI. Regions on histological images were delineated as non-cancerous or with a Gleason score [25] of 3+3, 3+4, 4+3 or 4+4. The interpretation of these scores is as follows. A score of the form A+A (e.g. 3+3) denotes that only Gleason grade A (e.g. 3) was present. A score of the form A+B (e.g. 3+4) denotes that Gleason grades A and B (e.g. 3 and 4) were present and grade A (e.g. 3) tissue occupied more than 50% of the total area of the region. For scores of the form A+B, the exact proportions of Gleason A and Gleason B tissue within the region were not recorded due to practical considerations arising from the fine-scale admixing of tissues of the different grades. Correspondence establishment between *distinct* lesions identified on MRI and histological regions was facilitated by a submillimeter-accurate 3D reconstruction of the mid-gland histology sections (Chapter 4) [26]. This reconstruction makes explicit the plane-to-plane correspondence between each 2D histology section and the 3D context of an *ex vivo* MRI acquired prior to cutting the specimen and was intended to reduce the impact of observer variability during the subsequent interactive visual correspondence of the reconstructed histology with the MR images. Because histological regions were contoured and graded at a smaller scale than the MR *suspicious regions* and the relative proportions of tissues of different grades for histological regions with scores of the form A+B could not practically be recorded, the proportion of grade 3 and 4 within *distinct* lesions on MRI

containing both grades was estimated as an area-weighted average of the grades of histological regions, assigning 100% of the area when only one grade was present and assigning 75% of the area to the more prevalent grade and 25% of the area to the less prevalent grade when more than one grade was present. In this chapter, the terms *non-cancerous lesions*, *cancerous lesions*, and *Gleason 7 lesions* refer, respectively, to MRI-defined *distinct* lesions that did not corresponding to histological cancer, that corresponded to any histological cancer, and that corresponded to aggregated histological regions with Gleason score 7.



**Figure 6.2:** Illustrative (a) T2W, (b) ADC and (c) DCE mpMRI images for 3 subjects, showing *suspicious regions* from 4 observers (yellow, magenta, cyan and red contours). The Gleason scores (GS) and mean overall likelihood scores (OL) for each lesion are shown on the left side, and sequence-specific scores are shown with each image. Panel (d) shows post-prostatectomy histology corresponding to the *distinct* lesion in the top row with insets showing the fine-scale contouring of Gleason 3+3 (yellow), 3+4 (cyan), 4+3 (green) and 4+4 (black) regions.

#### 6.2.8.2 Anatomical location

Because of the differences in mpMRI appearance of tumors in the peripheral zone from those elsewhere in the prostate [10, 27, 28], the anatomical location of each *distinct* lesion was categorized as either peripheral zone or non-peripheral zone for the purpose of correlation analysis.

#### 6.2.8.3 PI-RADS scores

For the correlation of *interobserver contour agreement* (on each sequence for all three metrics) and *intersequence contour agreement* (on each pair of sequences for all three metrics) with PI-RADS scores, the scores were averaged over the observers.

#### 6.2.8.4 Tumor volume

Tumor volumes for the analyses in this study were estimated as the average of volumes of corresponding suspicious regions. For analyzing positive predictive values (PPVs) of PI-RADS likelihood scores as a function of tumor volume on MRI, volumes were averaged over all sequences for one observer. For the correlation of interobserver and intersequence contour agreements with tumor volumes, volumes were averaged over all observers and all sequences.

#### 6.2.9 Statistical analysis

Statistical analysis was performed using Matlab R2013a (The Mathworks Inc., Natick, USA). Reported overall PI-RADS likelihood scores included six non-integer scores lying between 2 and 3 (two lesions), 3 and 4 (one lesion), and 4 and 5 (three lesions); these were coded as scores of 2.5 (and excluded from analysis), 3.5 (and included) and 4.5 (and included), respectively. Similarly, reported sequence-specific PI-

RADS scores included five non-integer scores lying between 2 and 3 (one lesion) and 4 and 5 (four lesions), which were coded as scores of 2.5 and 4.5 (and included), respectively. Differences in PPVs for different PI-RADS likelihood scores were assessed using 95% confidence intervals on differences in proportions (combined Wilson score without continuity correction [29]). *Interobserver score agreements* were estimated as a mean and standard deviation of the score differences, and using Schouten's weighted kappa [24]. Spearman's rank correlation test was used to calculate correlations of *interobserver score agreements* and PI-RADS scores with histological grade (coded as ordinals: non-cancerous < Gleason 3+3 < Gleason 3+4 < Gleason 4+3), location (coded as 0 for non-peripheral zone and 1 for peripheral zone) and volume. *Interobserver* and *intersequence contour agreements* were estimated as the mean and standard deviation for each metric. For each metric, Wilcoxon signed rank tests were used to compare the *interobserver contour agreement* across sequences, corrected for multiple comparisons using Bonferroni-Holm correction [30]. Spearman's rank correlation test was used to calculate correlations of *interobserver* and *intersequence contour agreements* with histological grade (coded as above), location (coded as above), volume, mean PI-RADS likelihood score and mean PI-RADS sequence-specific scores (for *interobserver contour agreement*). We corrected for multiple comparisons using Bonferroni-Holm correction (separately for the *interobserver* and *intersequence contour agreements*).

## 6.3 Results

Overall, 49 *distinct* lesions were identified as having a PI-RADS likelihood score  $\geq 3$  by at least one of the observers, yielding 104 *score assignments* (each *distinct* lesion could be scored by 1 – 4 observers). 26 *distinct* lesions (and 68 *score assignments*)

corresponded to histological foci, of which 19 *distinct* lesions (and 48 *score assignments*) had mean contoured volumes  $\geq 0.2 \text{ cm}^3$  (the smallest clinically significant cancer according to [31]) and 9 *distinct* lesions (and 28 *score assignments*) had mean contoured volumes  $\geq 0.5 \text{ cm}^3$  (the smallest clinically significant cancer according to [32]). The PPVs of *score assignments* with PI-RADS likelihood scores of 3, 4 and 5 for any histological cancer and for Gleason 7 cancer, are given in Table 6.3. PPVs for PI-RADS likelihood scores of 5 were higher than those for PI-RADS likelihood scores of 3 or 4; however, we had insufficient power to statistically distinguish between PPVs for PI-RADS likelihood scores of 3 vs. 4.

**Table 6.3:** Positive predictive values (PPVs) for any histological cancer and for Gleason 7 cancer of score assignments with PI-RADS likelihood scores of 3, 4 and 5, individually, and scores above thresholds  $\geq 3$  and  $\geq 4$ , broken down by lesion volume on mpMRI.

<i>Lesion volume on mpMRI PI-RADS likelihood score</i>		<i>Positive predictive value</i>	
		<i>Any cancer</i>	<i>Gleason 7</i>
<b>Any volume</b>	<b>3</b>	59% (23/39)	46% (18/39)
	<b>4</b>	50% (17/34)	41% (14/34)
	<b>5</b>	89% (24/27)	85% (23/27)
	$\geq 3$	65% (68/104)	57% (59/104)
	$\geq 4$	69% (44/64)	63% (40/64)
<b>&gt; 0.2 cm<sup>3</sup></b>	<b>3</b>	52% (12/23)	43% (10/23)
	<b>4</b>	52% (12/23)	43% (10/23)
	<b>5</b>	87% (20/23)	83% (19/23)
	$\geq 3$	66% (48/73)	59% (43/73)
	$\geq 4$	71% (35/49)	65% (32/49)
<b>&gt; 0.5 cm<sup>3</sup></b>	<b>3</b>	100% (6/6)	83% (5/6)
	<b>4</b>	57% (8/14)	50% (7/14)
	<b>5</b>	93% (13/14)	93% (13/14)
	$\geq 3$	80% (28/35)	74% (26/35)
	$\geq 4$	76% (22/29)	72% (21/29)

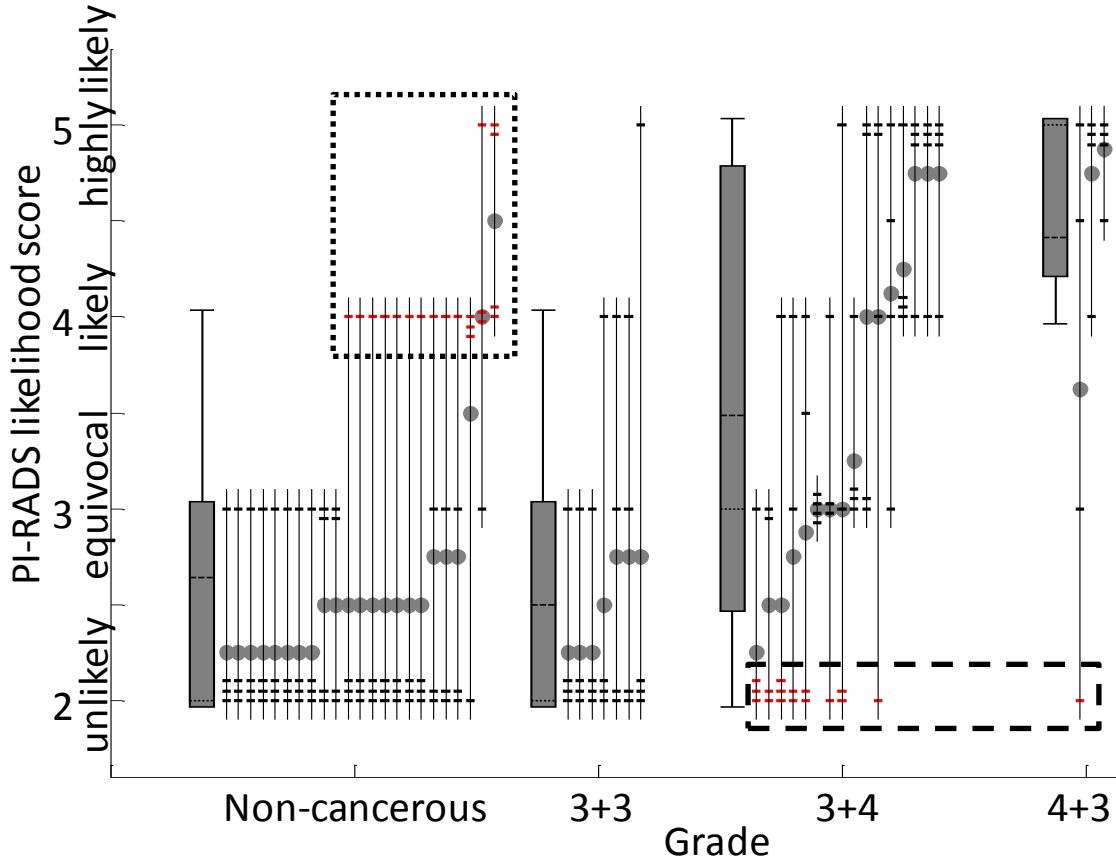
### 6.3.1 Image assessment

The key findings from this analysis were that (1) the mean PI-RADS likelihood score was correlated with histological grade and volume, (2) *interobserver score agreement* in PI-RADS likelihood scores yielded a kappa value of 0.30, and observers agreed more frequently on assigning likelihood scores  $< 4$  (unlikely cancerous or equivocal) vs.  $\geq 4$  (likely or highly likely cancerous) than on assigning  $< 3$  (unlikely cancerous) vs.  $\geq 3$  (equivocal, likely or highly likely cancerous), and (3) the mean interobserver differences in PI-RADS likelihood and sequence specific scores ranged from 0.6 to 1.2 points over a 5-point scale.

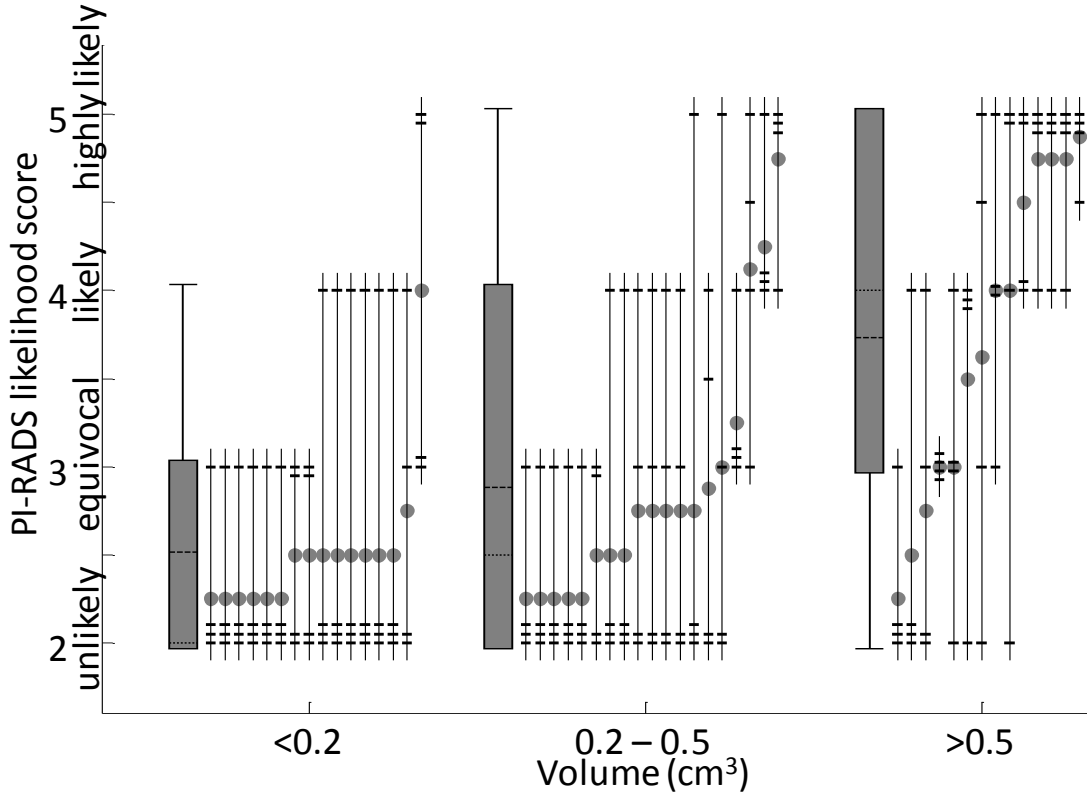
The PI-RADS likelihood scores were broken down by histological grade (Figure 6.3) and mean *suspicious region* volume (Figure 6.4). Analogous graphs for PI-RADS sequence-specific scores are included in Appendix C. As a basis for interpretation of these graphs, consider the task of robustly reporting histological foci with Gleason score 3+3 or higher as cancerous lesions. For this task, the ideal breakdown depicted in these graphs would show no non-cancerous foci recorded (i.e. as PI-RADS 3/“equivocal” or higher), and all lesions corresponding to cancerous foci recorded and given a PI-RADS likelihood score  $\geq 4$  by all 4 observers. In contrast, Figure 3 shows that 20 *score assignments* with PI-RADS likelihood scores of 4 or 5, denoting likely or highly likely cancerous lesions, were recorded for *non-cancerous lesions* (demarcated by the dotted box on Figure 3). Additionally, 16 of 26 *cancerous lesions* and 9 of 19 *Gleason 7 lesions* that were assessed and recorded by at least one observer, were denoted as a PI-RADS likelihood score of 2 for one or more other observers (recall that by *a priori* agreement, a default likelihood score of 2 was assigned for all observers that did not make a *score assignment* for a *distinct* lesion recorded by at least one observer). The Gleason 7 false negatives are demarcated by the dashed box on Figure 3. Figure 3 does, however, suggest a positive relationship between PI-RADS likelihood scores and histological grade. Similarly, Figure 4 shows a positive relationship between PI-RADS likelihood scores and volume as measured on mpMRI. We identified significant correlations of the mean PI-RADS likelihood score with volume ( $\rho=0.6$ ,  $p=0.0001$ ) and histological grade ( $\rho=0.6$ ,  $p=0.0001$ ). The correlation between histological grade and volume ( $\rho=0.3$ ,  $p=0.051$ ) was lower, but approached significance (without multiple comparison correction). We also identified trends of correlation ( $p<0.05$  before multiple comparison correction) relating



histological grade to the T2W, DCE and ADC PI-RADS scores ( $p=0.3$ ) and relating volume to the T2W PI-RADS scores ( $p=0.4$ ).



**Figure 6.3:** PI-RADS likelihood scores for lesions broken down by histological grade. Each vertical solid line corresponds to a *distinct* lesion, with horizontal dashes at the PI-RADS likelihood score for each observer. For each grade, Tukey box plots show the interquartile range (IQR), median (dotted line), mean (dashed line), and range of data within  $1.5 \times \text{IQR}$  of the IQR (whiskers). The dotted box denotes “false positives” where the PI-RADS likelihood scores denote likely or highly likely cancer for *non-cancerous lesions*. The dashed box denotes “higher grade false negatives” where the observer did not record a PI-RADS score for a *Gleason 7 lesion*. Note that by a *priori* agreement, a default likelihood score of 2 was assigned for all observers that did not make a *score assignment* for a *distinct* lesion recorded by at least one observer.



**Figure 6.4:** PI-RADS likelihood scores for lesions broken down by mean *suspicious region* volume. Each vertical solid line corresponds to a *distinct* lesion, with horizontal dashes at the PI-RADS likelihood score for each observer. For each volume range, Tukey box plots show the interquartile range (IQR), median (dotted line), mean (dashed line), and range of data within  $1.5 \times \text{IQR}$  of the IQR (whiskers). Note that by *a priori* agreement, a default likelihood score of 2 was assigned for all observers that did not make a *score assignment* for a *distinct* lesion recorded by at least one observer.

The *interobserver score agreement* for PI-RADS likelihood scores yielded a kappa value of 0.30 across all *distinct* lesions. On the subsets of *cancerous lesions* and *Gleason 7 lesions*, the weighted kappa values were 0.35 and 0.34, respectively.

*Interobserver score agreement* for assigning likelihood scores  $<3$  vs.  $\geq 3$  and  $<4$  vs.  $\geq 4$ , for all *distinct* lesions, for *cancerous lesions*, for *Gleason 7 lesions* and for *distinct* lesions with volumes  $\geq 0.5 \text{ cm}^3$  are shown in Table 6.4, suggesting higher agreement in distinguishing PI-RADS likelihood scores  $<4$  from  $\geq 4$  than in distinguishing scores  $<3$  from  $\geq 3$ , and slightly higher agreement for more clinically significant lesions (in terms of

grade and lesion volume as measured on MRI). The mean  $\pm$  SD interobserver difference in likelihood, T2W, DCE and ADC PI-RADS scores were  $0.8 \pm 0.4$ ,  $0.6 \pm 0.6$ ,  $1.2 \pm 0.9$  and  $0.6 \pm 0.5$  respectively. On *Gleason 7 lesions*, the differences were  $0.9 \pm 0.5$ ,  $0.7 \pm 0.6$ ,  $1.3 \pm 0.8$  and  $0.6 \pm 0.5$ , respectively. Our analysis failed to detect correlations relating *interobserver score agreement* to histological grade, anatomical location or volume.

**Table 6.4:** Interobserver agreement in assigning PI-RADS likelihood scores for two thresholds.

	<i>Threshold &lt;3 vs. <math>\geq 3</math></i>		<i>Threshold &lt;4 vs. <math>\geq 4</math></i>	
	<i>Weighted kappa</i>	<i>Consensus proportion</i>	<i>Weighted kappa</i>	<i>Consensus proportion</i>
<b>All scored lesions</b>	0.17	24% (12/49)	0.38	47% (23/49)
<b>Cancerous lesions</b>	0.26	38% (10/26)	0.44	46% (12/26)
<b>Gleason 7 lesions</b>	0.23	53% (10/19)	0.43	47% (9/19)
<b>Scored lesions with volumes <math>\geq 0.5 \text{ cm}^3</math></b>	0.24	50% (7/14)	0.45	50% (7/14)

The threshold  $<3$  vs.  $\geq 3$  separates unlikely cancerous from equivocal or likely cancerous. The threshold  $<4$  vs.  $\geq 4$  separates unlikely cancerous or equivocal from likely cancerous findings.

### 6.3.2 Interobserver and intersequence contour agreement

The key findings from this analysis were (1) that observers chose to delineate *suspicious regions* on T2W (103/104 *score assignments*) and ADC (96/104) images more frequently than on DCE (87/104) images; (2) ADC images had marginally superior *interobserver contour agreement* (not significant after multiple comparison correction) for all three measures compared to T2W images and DCE images; and (3) for all pairs of sequences, intersequence absolute differences in *suspicious region* volumes were correlated with PI-RADS likelihood scores and average *suspicious region* volumes.

The mean and standard deviation of the *interobserver* and *intersequence contour agreement* measurements are shown in Table 6.5. The mean Dice overlap ranged from 0.3 to 0.5. The mean MAD ranged from 1.5 to 2.4 mm; for comparison, the mean  $\pm$  standard deviation of the *distinct* lesions' longest diameters (averaged over all *suspicious regions* corresponding to each *distinct* lesion) was  $15.0 \pm 5.0$  mm. The mean differences in absolute volume ranged from 0.2 to  $0.5 \text{ cm}^3$ ; for comparison, the mean  $\pm$  standard deviation of the *distinct* lesions' volumes (averaged over all *suspicious regions* corresponding to each *distinct* lesion) was  $0.5 \pm 0.9 \text{ cm}^3$ . When including only *Gleason 7 lesions*, the *interobserver* and *intersequence contour agreement* were within 0.05 (Dice overlap), 0.2 mm (MAD) and  $0.1 \text{ cm}^3$  (absolute difference in volume) of those for all *distinct* lesions reported in Table 6.5. Our analysis failed to detect significant differences in *interobserver contour agreement* between the sequences; however, the *interobserver contour agreement* on ADC images was marginally superior for all 3 measures in our data, and the *intersequence contour agreement* between ADC and T2W was marginally superior than agreement between other pairs of sequences.

**Table 6.5:** Interobserver and intersequence contour agreement, and comparison of interobserver contour agreement across sequences.

		<i>Dice overlap mean <math>\pm</math> SD</i>	<i>MAD mean <math>\pm</math> SD (mm)</i>	<i>Absolute difference in volume mean <math>\pm</math> SD (cm<sup>3</sup>)</i>
<b>Interobserver contour agreement</b>	<b>T2W</b>	0.4 $\pm$ 0.2	2.4 $\pm$ 2.2	0.5 $\pm$ 0.6
	<b>DCE</b>	0.4 $\pm$ 0.3	2.2 $\pm$ 1.6	0.5 $\pm$ 0.7
	<b>ADC</b>	0.5 $\pm$ 0.2	1.8 $\pm$ 1.9	0.5 $\pm$ 0.5
<b>Intersequence contour agreement</b>	<b>T2W – DCE</b>	0.4 $\pm$ 0.2	1.6 $\pm$ 0.9	0.2 $\pm$ 0.3
	<b>T2W – ADC</b>	0.4 $\pm$ 0.1	1.5 $\pm$ 0.8	0.2 $\pm$ 0.2
	<b>ADC – DCE</b>	0.3 $\pm$ 0.2	2.0 $\pm$ 1.1	0.2 $\pm$ 0.5

SD denotes standard deviation, MAD denotes mean absolute distance between two contour boundaries,

Dice denotes the overlapping volume as a fraction of the average volume of two contours. See Section 6.2.6 for details.

Table 6.6 shows the correlation coefficients between *interobserver* and *intersequence contour agreement* and each of the interaction factors: the volume; the histological grade; the anatomical location; the PI-RADS likelihood score; and, for *interobserver contour agreement*, the corresponding PI-RADS sequence-specific scores. The intersequence absolute differences in contoured volumes were correlated with tumor volume ( $0.7 < p < 0.8$ ) and PI-RADS likelihood score ( $0.5 < p < 0.6$ ) on all sequences, and the Dice overlap was correlated with the volume for ADC–DCE images. The interobserver absolute differences in contoured volumes were correlated with volume as well ( $0.5 < p < 0.8$ ), and the interobserver Dice overlaps were correlated with the PI-RADS likelihood score on DCE. Our analysis failed to detect significant correlation of either the *interobserver* or *intersequence contour agreements* with histological grade, anatomical location or PI-RADS sequence specific scores.

**Table 6.6:** Spearman correlation coefficients\* of interobserver and intersequence contour agreement with interaction factors.

			<i>Volume</i>	<i>Histological</i>	<i>Anatomical</i>	<i>PI-RADS</i>	<i>PI-RADS</i>
				<i>grade</i>	<i>location</i>	<i>likelihood</i>	<i>sequence-specific</i>
						<i>score</i>	<i>score</i>
Interobserver	T2W	Dice	0.1	0.0	0.1	<b>0.5</b>	<b>0.5</b>
		MAD	0.3	0.1	-0.3	-0.1	-0.3
		$\Delta$ Volume	<b>0.8</b>	0.4	0.1	0.4	<b>0.4</b>
	DCE	Dice	0.1	0.3	-0.1	<b>0.7</b>	<b>0.5</b>
		MAD	0.2	-0.1	0.0	<b>-0.5</b>	<b>-0.5</b>
		$\Delta$ Volume	<b>0.5</b>	<b>0.5</b>	0.2	0.2	0.2
	ADC	Dice	0.0	-0.2	<b>0.5</b>	0.2	0.2
		MAD	0.4	0.2	<b>-0.5</b>	0.0	-0.1
		$\Delta$ Volume	<b>0.6</b>	0.3	-0.1	0.4	0.2
	T2W – DCE	Dice	0.3	0.2	0.0	0.2	N/a
		MAD	0.3	0.0	0.0	0.2	N/a
		$\Delta$ Volume	<b>0.7</b>	<b>0.3</b>	0.0	<b>0.6</b>	N/a
Intersequence	T2W – ADC	Dice	<b>0.4</b>	0.0	-0.1	<b>0.3</b>	N/a
		MAD	0.0	0.2	0.1	0.1	N/a
		$\Delta$ Volume	<b>0.7</b>	0.3	0.1	<b>0.5</b>	N/a
	ADC – DCE	Dice	<b>0.5</b>	0.2	-0.1	<b>0.3</b>	N/a
		MAD	-0.2	0.1	0.2	-0.1	N/a
		$\Delta$ Volume	<b>0.8</b>	0.3	0.1	<b>0.6</b>	N/a

\*Each value in the table represents the correlation coefficient between an interaction factor (by column; see

Section 6.2.8 for details) and a contour agreement metric (by row; see Section 6.2.6 for details). MAD

denotes mean absolute distance between two *suspicious region* boundaries, Dice denotes the overlapping

volume as a fraction of the average volume of two *suspicious regions*,  $\Delta$ Volume denotes absolute

difference in *suspicious region* volume. Bolded elements were significant ( $\alpha=0.05$ ) after multiple

comparison correction. Bold italicized coefficients were significant ( $\alpha=0.05$ ) before, but not after multiple

comparison correction.

## 6.4 Discussion

The accuracy and precision of localizing cancerous lesions is critical to the evaluation and clinical application of lesion-focused biopsies as well as lesion-focused therapies, such as cryotherapy, focal-laser ablation, high-intensity focused ultrasound, and lesion-targeted radiation boost.

This work investigated the variability in prostate cancer scoring and delineation on mpMRI to address the following four questions. (1) What is the interobserver agreement in scoring lesions as equivocal, likely cancerous or highly likely cancerous, following the PI-RADS guidelines? (2) What is the interobserver agreement in regions delineated on T2W, DCE and ADC images? (3) What is the intersequence agreement in regions delineated by an observer on these sequences? (4) What are the interactions between these aspects of agreement with the histological grade, anatomical location, and volume? In this study, prostate cancer PI-RADS scoring and contouring on T2W, ADC and DCE 3 T MRI show promise: likelihood scores correlated with Gleason grade and scores of 5 had an 85% PPV for Gleason 7 cancer, but we observed high interobserver score and contour variability.

The use of mpMRI for prostate cancer detection and staging has been widely reported [9-11, 17, 19-21, 33], and consensus guidelines have been proposed, including the PI-RADS structured reporting system used in this study [10]. In our data set, *score assignments* with PI-RADS overall likelihood scores of 5 had an 85% (23/27) PPV for Gleason 7 cancer, whereas those with likelihood scores of 3 and 4 had substantially lower PPVs of 46% (18/39) and 41% (14/34), respectively. A similar difference was seen in the PPVs for any cancer. In the context of clinical guidelines regarding the use of the PI-

RADS likelihood score to inform the decision to apply focal therapy to the lesion, setting a threshold criterion for the PI-RADS likelihood scores may be valuable; in our study, likelihood scores  $\geq 3$ ,  $\geq 4$  and  $=5$  had PPVs of 65% (68/104), 69% (44/64) and 89% (24/27). Although the small number of Gleason 4+3 tumors in our study precludes strong conclusions, we also observed that 75% (9/12) of the *score assignments* corresponding to Gleason 4+3 foci had PI-RADS likelihood scores  $>4$  (including 2 likelihood scores coded as 4.5). The prevalence of *score assignments* with PI-RADS likelihood scores denoting likely (17 *score assignments*) or highly likely (3 *score assignments*) cancer that did not correspond to histological foci suggests that further investigation of image features acting as confounders [34, 35] in PI-RADS scoring (particularly PI-RADS likelihood scores of 4) could be valuable. Such false positive *score assignments* were made by all four observers. There were 26 *cancerous lesions* identified by the observers, for which there was the potential for 104 *score assignments* (26 *cancerous lesions*  $\times$  4 observers = 104 potential sets of PI-RADS scores). Of these potential *score assignments*, there were 36 false negative instances where the observer did not assign a PI-RADS likelihood score of 3 or higher. For *Gleason 6 lesions*, the proportion of these false negatives was higher than for *Gleason 7 lesions*, which is consistent with previously described challenges in detecting lower grade cancer on mpMRI [36], and may have marginal clinical significance [37]. However, there were 17 false negative instances for *Gleason 7 lesions*, which potentially have higher clinical significance. All four observers made such false negative assessments. Note that these data may underestimate the false negative rate, as they exclude histological foci not recorded by any observer. The performance of mpMRI seen in this study may reflect, in part, the predominantly lower grade cancers in our study



population (with patient Gleason scores of G3+3 [n=5], G3+4 [n=11], and G4+3 [n=1], with no higher grade cancer), compared to previous studies [23]. Such a lower grade cohort may be more reflective of early stage cancer populations than a typical radical prostatectomy cohort.

We observed a correlation of  $\rho=0.6$  between histological grade and the PI-RADS likelihood score, and trends towards correlation for each of the sequence-specific scores. However, our analysis of *interobserver score agreement* yielded a weighted kappa value of 0.30 (consistent with the fair to moderate agreement reported in recent evaluations of PI-RADS assessments [38, 39]), and there was a 0.6 – 1.2 point average interobserver difference in the assignment of PI-RADS scores. This interobserver variability may represent a barrier that needs to be overcome to realize the potential of PI-RADS assessment. The high variability in PI-RADS DCE scores, in particular, may be related to the use of a low-temporal-resolution (90 seconds per phase) DCE protocol, or the lack of pharmacokinetic DCE maps; our investigation of this issue is currently ongoing.

mpMRI assessment has also been shown to support localization of cancer to prostate sextants [12-14]. This study complements existing studies of regional localization by directly evaluating the observer delineation of *suspicious regions* that could be used to define target volumes for lesion-focused diagnostics and therapies, and comparing delineations performed on different image sequences. The interobserver variability measured in this study suggests that there can be substantial disagreement regarding where the boundaries of such delineations should be. This variability injects uncertainty into clinical investigations of focal therapy, and may represent a barrier that needs to be overcome for the clinical implementation of focal therapy using mpMRI. The

intersequence variability observed in this study suggests that the choice of sequence for delineation may affect the resulting target volumes, and a marginally improved *interobserver contour agreement* for ADC images suggests further investigation of contouring on ADC images.

The conclusions of this study should be considered in the context of its limitations. Because each observer performed the scoring and delineation tasks independently of the other observers, interobserver lesion correspondence was defined *a posteriori* based on the overlap of *composite suspicious regions*, complicating accurate correspondence. For example, in theory, two partially overlapping *suspicious regions* intended by different observers to denote two *distinct* lesions could have been corresponded as a single *distinct* lesion. Alternately, a *distinct* lesion delineated as one lesion by one observer and as two separate lesions by another observer might have no well defined one-to-one correspondence between lesions. Such errors could result in overestimation of interobserver variability. This was mitigated by visual inspection of all corresponded regions, and manual correspondence was used in the two cases where it was deemed necessary. Observers' intersequence lesion correspondence was also complicated by patient motion between imaging sequences (1 – 13 mm in-plane and 2 – 10 mm out-of-plane), forcing the observers to mentally register the images during the *score assignment* on mpMRI. In calculating *intersequence contour agreement* metrics, however, rigid image registration with 1 mm mean target registration error was used to reduce metric error due to misalignment. The sample included in this study was small and contained only three tumors with Gleason score 4+3, only one subject whose overall Gleason score was 4+3, and no cancer with Gleason pattern 5. Although *interobserver*

*score and contour agreement* may be different on such a sample than on one including more advanced cancers, our sample may be more representative a group of patients suitable for focal therapy. Other limitations include a small number of readers and variability in experience of the readers. Examination of correlation of the lesion boundary as delineated by the radiologist against the histological boundary was out of the scope of this study – this analysis would be necessary to evaluate the accuracy of delineation strategies on mpMRI (e.g. whether a target should be delineated on a single sequence, or computed from the overlap or combination of delineations on multiple sequences).

In conclusion, lesion assessments following the PI-RADS guidelines yielded cancer likelihood scores that correlated with Gleason scores assessed on post-prostatectomy histology. Scoring a lesion on mpMRI as having a PI-RADS likelihood score of 5 had a PPV of 85% for Gleason 7 cancer, with PPV of 93% for such lesions having volumes of  $> 0.5 \text{ cm}^3$  measured on mpMRI. However, interobserver agreement in the PI-RADS likelihood score yielded a weighted kappa value of 0.30 and average interobserver differences in the PI-RADS likelihood and sequence specific scores were 0.6 –1.2 points over a 5 point scale, suggesting variability that may impede interpretation of PI-RADS assessments. The prevalence of PI-RADS likelihood scores denoting likely or highly likely cancer for histologically non-cancerous lesions suggests a need for further investigation of image features that confound PI-RADS assessments. The delineation of suspicious regions was performed more frequently on T2W and ADC images than on DCE images. Suspicious regions delineated on ADC images showed a trend towards superior interobserver agreement compared to T2W and DCE images (with an average interobserver boundary distance of 1.8 mm for ADC images vs. 2.4 mm and

2.2 mm for T2W and DCE images, respectively). This suggests further investigation of contouring on ADC images instead of, or in addition to, contouring on the more commonly used T2W images and perhaps the inclusion of additional imaging information (e.g. pharmacokinetic maps computed from DCE).

## 6.5 References

1. S. Vourganti, A. Rastinehad, N. K. Yerram, J. Nix, D. Volkin, A. Hoang, B. Turkbey, G. N. Gupta, J. Kruecker, W. M. Linehan, P. L. Choyke, B. J. Wood and P. A. Pinto, "Multiparametric magnetic resonance imaging and ultrasound fusion biopsy detect prostate cancer in patients with prior negative transrectal ultrasound biopsies," *Journal of Urology* **188**, 2152–2157 (2012).
2. T. H. van der Kwast, M. B. Amin, A. Billis, J. I. Epstein, D. Griffiths, P. A. Humphrey, R. Montironi, T. M. Wheeler, J. R. Srigley, L. Egevad and B. Delahunt, "International Society of Urological Pathology (ISUP) Consensus Conference on Handling and Staging of Radical Prostatectomy Specimens. Working group 2: T2 substaging and prostate cancer volume," *Modern Pathology* **24**, 16–25 (2011).
3. E. Lecornet, C. Moore, H. U. Ahmed and M. Emberton, "Focal therapy for prostate cancer: fact or fiction?," *Urologic Oncology* **28**, 550–556 (2010).
4. G. Bozzini, P. Colin, P. Nevoux, A. Villers, S. Mordon and N. Betrouni, "Focal therapy of prostate cancer: energies and procedures," *Urologic Oncology* **31**, 155–167 (2013).
5. U. Lindner, R. A. Weersink, M. A. Haider, M. R. Gertner, S. R. Davidson, M. Atri, B. C. Wilson, A. Fenster and J. Trachtenberg, "Image guided photothermal focal therapy for localized prostate cancer: phase I trial," *Journal of Urology* **182**, 1371–1377 (2009).
6. H. U. Ahmed, R. G. Hindley, L. Dickinson, A. Freeman, A. P. Kirkham, M. Sahu, R. Scott, C. Allen, J. Van der Meulen and M. Emberton, "Focal therapy for localised unifocal and multifocal prostate cancer: a prospective development study," *Lancet Oncology* **13**, 622–632 (2012).
7. I. M. Lips, U. A. van der Heide, K. Haustermans, E. N. van Lin, F. Pos, S. P. Franken, A. N. Kotte, C. H. van Gils and M. van Vulpen, "Single blind randomized phase III trial to investigate the benefit of a focal lesion ablative microboost in prostate cancer (FLAME-trial): study protocol for a randomized controlled trial," *Trials* **12**, 255 (2011).

8. G. Bauman, M. Haider, U. A. Van der Heide and C. Menard, "Boosting imaging defined dominant prostatic tumors: a systematic review," *Radiotherapy and Oncology* **107**, 274–281 (2013).
9. A. Sciarra, J. Barentsz, A. Bjartell, J. Eastham, H. Hricak, V. Panebianco and J. A. Witjes, "Advances in magnetic resonance imaging: how they are changing the management of prostate cancer," *European Urology* **59**, 962–977 (2011).
10. J. O. Barentsz, J. Richenberg, R. Clements, P. Choyke, S. Verma, G. Villeirs, O. Rouviere, V. Logager and J. J. Futterer, "ESUR prostate MR guidelines 2012," *European Radiology* **22**, 746–757 (2012).
11. M. R. Engelbrecht, J. O. Barentsz, G. J. Jager, M. van der Graaf, A. Heerschap, J. P. Sedelaar, R. G. Aarnink and J. J. de la Rosette, "Prostate cancer staging using imaging," *BJU International* **86 Suppl 1**, 123–134 (2000).
12. S. Isebaert, L. Van den Bergh, K. Haustermans, S. Joniau, E. Lerut, L. De Wever, F. De Keyzer, T. Budiharto, P. Slagmolen, H. Van Poppel and R. Oyen, "Multiparametric MRI for prostate cancer localization in correlation to whole-mount histopathology," *Journal of Magnetic Resonance Imaging* **37**, 1392–1401 (2013).
13. M. A. Haider, T. H. van der Kwast, J. Tanguay, A. J. Evans, A. T. Hashmi, G. Lockwood and J. Trachtenberg, "Combined T2-weighted and diffusion-weighted MRI for localization of prostate cancer," *AJR. American Journal of Roentgenology* **189**, 323–328 (2007).
14. J. J. Futterer, T. W. Scheenen, S. W. Heijmink, H. J. Huisman, C. A. Hulsbergen-Van de Kaa, J. A. Witjes, A. Heerschap and J. O. Barentsz, "Standardized threshold approach using three-dimensional proton magnetic resonance spectroscopic imaging in prostate cancer localization of the entire prostate," *Investigative Radiology* **42**, 116–122 (2007).
15. L. Wang, Y. Mazaheri, J. Zhang, N. M. Ishill, K. Kuroiwa and H. Hricak, "Assessment of biologic aggressiveness of prostate cancer: correlation of MR signal intensity with Gleason grade after radical prostatectomy," *Radiology* **246**, 168–176 (2008).
16. C. K. Kim, B. K. Park, H. M. Lee and G. Y. Kwon, "Value of diffusion-weighted imaging for the prediction of prostate cancer location at 3T using a phased-array coil: preliminary results," *Investigative Radiology* **42**, 842–847 (2007).
17. I. Ocak, M. Bernardo, G. Metzger, T. Barrett, P. Pinto, P. S. Albert and P. L. Choyke, "Dynamic contrast-enhanced MRI of prostate cancer at 3 T: a study of pharmacokinetic parameters," *AJR. American Journal of Roentgenology* **189**, W192–W201 (2007).

18. J. Kurhanewicz, M. G. Swanson, S. J. Nelson and D. B. Vigneron, "Combined magnetic resonance imaging and spectroscopic imaging approach to molecular imaging of prostate cancer," *Journal of Magnetic Resonance Imaging* **16**, 451–463 (2002).
19. K. L. Zakian, K. Sircar, H. Hricak, H. N. Chen, A. Shukla-Dave, S. Eberhardt, M. Muruganandham, L. Ebor, M. W. Kattan, V. E. Reuter, P. T. Scardino and J. A. Koutcher, "Correlation of proton MR spectroscopic imaging with Gleason score based on step-section pathologic analysis after radical prostatectomy," *Radiology* **234**, 804–814 (2005).
20. B. Turkbey, V. P. Shah, Y. Pang, M. Bernardo, S. Xu, J. Kruecker, J. Locklin, A. A. Baccala, A. R. Rastinehad, M. J. Merino, J. H. Shih, B. J. Wood, P. A. Pinto and P. L. Choyke, "Is apparent diffusion coefficient associated with clinical risk scores for prostate cancers that are visible on 3-T MR images?," *Radiology* **258**, 488–495 (2011).
21. H. A. Vargas, O. Akin, T. Franiel, Y. Mazaheri, J. Zheng, C. Moskowitz, K. Udo, J. Eastham and H. Hricak, "Diffusion-weighted endorectal MR imaging at 3 T for prostate cancer: tumor detection and assessment of aggressiveness," *Radiology* **259**, 775–784 (2011).
22. H. C. Rischke, U. Nestle, T. Fechter, C. Doll, N. Volegova-Neher, K. Henne, J. Scholber, S. Knippen, S. Kirste, A. L. Grosu and C. A. Jilg, "3 Tesla multiparametric MRI for GTV-definition of dominant intraprostatic lesions in patients with prostate cancer--an interobserver variability study," *Radiation Oncology* **8**, 183 (2013).
23. M. Anwar, A. C. Westphalen, A. J. Jung, S. M. Noworolski, J. P. Simko, J. Kurhanewicz, M. Roach, 3rd, P. R. Carroll and F. V. Coakley, "Role of endorectal MR imaging and MR spectroscopic imaging in defining treatable intraprostatic tumor foci in prostate cancer: quantitative analysis of imaging contour compared to whole-mount histopathology," *Radiotherapy and Oncology* **110**, 303–308 (2014).
24. H. J. A. Schouten, "Measuring pairwise interobserver agreement when all subjects are judged by the same observers," *Statistica Neerlandica* **36**, 45–61 (1982).
25. J. I. Epstein, "An update of the Gleason grading system," *Journal of Urology* **183**, 433–440 (2010).
26. E. Gibson, C. Crukley, M. Gaed, J. A. Gómez, M. Moussa, J. L. Chin, G. S. Bauman, A. Fenster and A. D. Ward, "Registration of prostate histology images to ex vivo MR images via strand-shaped fiducials," *Journal of Magnetic Resonance Imaging* **36**, 1402–1412 (2012).
27. S. E. Viswanath, N. B. Bloch, J. C. Chappelow, R. Toth, N. M. Rofsky, E. M. Genega, R. E. Lenkinski and A. Madabhushi, "Central gland and peripheral zone

- prostate tumors have significantly different quantitative imaging signatures on 3 Tesla endorectal, in vivo T2-weighted MR imagery," *Journal of Magnetic Resonance Imaging* **36**, 213–224 (2012).
28. O. Akin, E. Sala, C. S. Moskowitz, K. Kuroiwa, N. M. Ishill, D. Pucar, P. T. Scardino and H. Hricak, "Transition Zone Prostate Cancers: Features, Detection, Localization, and Staging at Endorectal MR Imaging," *Radiology* **239**, 784–792 (2006).
  29. R. G. Newcombe, "Interval estimation for the difference between independent proportions: comparison of eleven methods," *Statistics in Medicine* **17**, 873–890 (1998).
  30. S. Holm, "A simple sequentially rejective multiple test procedure," *Scandinavian Journal of Statistics* **6**, 65–70 (1979).
  31. J. I. Epstein, P. C. Walsh, M. Carmichael and C. B. Brendler, "Pathologic and clinical findings to predict tumor extent of nonpalpable (stage T1c) prostate cancer," *Journal of the American Medical Association* **271**, 368–374 (1994).
  32. T. A. Stamey, F. S. Freiha, J. E. McNeal, E. A. Redwine, A. S. Whittemore and H. P. Schmid, "Localized prostate cancer: relationship of tumor volume to clinical significance for treatment of prostate cancer," *Cancer* **71**, 933–938 (1993).
  33. D. B. Vigneron, R. Males, S. Noworolski, S. J. Nelson, J. Scheidler, D. Sokolov, H. Hricak, P. Carroll and J. Kurhanewicz, "3D MRSI of prostate cancer: correlation with histologic grade," in *Proceedings of International Society for Magnetic Resonance in Medicine*, (Berkeley, California, 1998).
  34. A. B. Rosenkrantz and S. S. Taneja, "Radiologist, be aware: ten pitfalls that confound the interpretation of multiparametric prostate MRI," *AJR. American Journal of Roentgenology* **202**, 109–120 (2014).
  35. G. Litjens, H. J. Huisman, R. Elliott, N. Shih, M. D. Feldman, S. E. Viswanath, J. Futterer, J. Bomers and A. Madabhushi, "Distinguishing benign confounding treatment changes from residual prostate cancer on MRI following laser ablation," in *Proceedings of SPIE Medical Imaging*, (San Diego, 2014), pp. 90361D.
  36. J. Thompson, N. Lawrentschuk, M. Frydenberg, L. Thompson and P. Stricker, "The role of magnetic resonance imaging in the diagnosis and management of prostate cancer," *BJU International* **112 Suppl 2**, 6–20 (2013).
  37. H. B. Carter, A. W. Partin, P. C. Walsh, B. J. Trock, R. W. Veltri, W. G. Nelson, D. S. Coffey, E. A. Singer and J. I. Epstein, "Gleason score 6 adenocarcinoma: should it be labeled as cancer?," *Journal of Clinical Oncology* **30**, 4294–4296 (2012).

38. A. B. Rosenkrantz, S. Kim, R. P. Lim, N. Hindman, F. M. Deng, J. S. Babb and S. S. Taneja, "Prostate cancer localization using multiparametric MR imaging: comparison of Prostate Imaging Reporting and Data System (PI-RADS) and Likert scales," *Radiology* **269**, 482–492 (2013).
39. L. Reisaeter, M. Biermann, A. E., Y. Nygaard, C. Beisland, L. A. Akslen, O. Halvorsen, K. Gravdal, J. Futterer and J. Rorvik, "Reliability and diagnostic performance using PIRADS in preoperative multiparametric MRI of prostate cancer.," in *Proceedings of Radiological Society of North America*, (Chicago, USA, 2012).



## Chapter 7.

# **Toward prostate cancer contouring guidelines on MRI: dominant lesion gross and clinical target volume coverage via accurate histology fusion<sup>†</sup>**

### 7.1 Introduction

Delineation of prostate cancer target volumes on imaging has the potential to contribute to the treatment of men with low-intermediate-risk, localized prostate cancer. Therapies targeted at delineated lesions, including focal laser ablation [1], cryotherapy [2], and high-intensity focused ultrasound [3], may be suitable curative-intent treatments for many men with low-intermediate-risk cancers, and have lower morbidity than radical prostatectomy or radiation therapy [2, 4]. For men with more advanced cancers undergoing radiation therapy, focal boosting of the radiation dose within a delineated dominant lesion target volume may reduce the rate of recurrence at the dominant lesion and improve outcomes [5, 6]. Over-contouring the actual extent of cancer could result in increased damage to prostate tissue and surrounding sensitive

---

<sup>†</sup> A preliminary analysis of the data presented in this chapter have been accepted for presentation: E. Gibson, M. Gaed, J. A. Gómez, M. Moussa, C. Romagnoli, Z. Kassam, M. Bastian-Jordan, D. W. Cool, S. Pautler, J. L. Chin, C. Crukley, G. S. Bauman, A. Fenster, A. D. Ward. "Toward prostate cancer contouring guidelines on MRI: dominant lesion gross and clinical target volume coverage via accurate histology fusion." To be presented at the American Society for Radiation Oncology Annual Meeting; San Francisco, USA; September 2014. A version of this chapter is in preparation for submission: E. Gibson, C. Romagnoli, M. Bastian-Jordan, D. W. Cool, Z. Kassam, M. Gaed, M. Moussa, J. A. Gómez, S. Pautler, J. L. Chin, C. Crukley, G. S. Bauman, A. Fenster, A. D. Ward, "Toward prostate cancer contouring guidelines on MRI: dominant lesion gross and clinical target volume coverage via accurate histology fusion" Radiology (in preparation).

organs, while under-contouring could result in delivering a lower than intended treatment dose to cancerous tissue.

mpMRI has shown promise for detecting [7, 8] and staging [9] prostate cancer, as well as approximate localization of cancer to a sextant of the prostate [10-12]. The suitability and optimal use of mpMRI for accurately and precisely delineating intraprostatic cancer target volumes, however, has not been well characterized. A recent evaluation of delineations within single 2D MRI slices [13] suggests that delineations of the visible cancer (hereafter referred to as *gross target volumes* or *GTVs*) do not cover all histologically identified cancerous tissue for detected cancers, and that adding margins to GTVs to account for cancer not visible on imaging (creating *clinical target volumes* or *CTVs*) may be a suitable approach to achieve adequate histological coverage (defined as 95% in their analysis). However, a 2D analysis that uses 2D instead of 3D GTVs and that excludes cancer extending in the inferior-superior direction may underestimate the necessary margins. There is, therefore, a need to evaluate margins in a 3D context for CTVs defined in 3D (as would be done in practice).

Selection of appropriate margins is complex and depends on many factors including predictions of how far cancer typically extends beyond GTVs, the treatment-modality-specific effects of spatially over- and under-treating the cancer and the treatment-modality-specific feasibility of treating CTVs with specific shapes and sizes. For example, focal laser ablation increases temperatures within a defined volume of tissue surrounding the tip of an inserted optic laser fiber and the number of fiber applications in a procedure is limited by time and cost; thus, the volume of a CTV is important to assessing the feasibility of treatment with this modality.

Accurate registration of prostatectomy histology to mpMRI (and corresponding GTVs and CTVs) enables the retrospective estimation of the histological cancer covered by CTVs with specified margins, characterizing the typical extension of cancer beyond GTVs. Although the extent of histological cancer is not known at treatment, prediction intervals for histological coverage at specified margins could inform decisions about appropriate margins for unseen patients.

In this pilot study, we investigated the spatial relationships of GTVs contoured on different mpMRI sequences and CTVs with isotropic margin expansion from the GTVs to the prostate volume and to histologically identified cancer. Specifically, we addressed the following three questions:

- ***Question 1: What margins must be added to delineated GTVs to achieve adequate coverage of histological cancer?*** Specifically, for true positive GTVs detected following PI-RADS guidelines, what margin expansion will yield, with high likelihood, CTVs that leave 0%, 5% and 10% residual cancer from corresponding histologically identified cancer foci (Gleason score  $\geq 6$ ) and *high-grade* cancer foci (Gleason score  $\geq 7$ )?
- ***Question 2: What are the volumes of CTVs defined using such margins?*** The answer to this question illuminates the feasibility of delivering treatment to the CTVs using different focal therapy modalities.
- ***Question 3: What are the relative volumes of the CTVs defined using such margins with respect to the corresponding GTVs and the corresponding histological foci?*** The answer to this question provides information regarding the

necessary volumetric expansions of the GTVs, and amount of non-cancerous parenchyma that would be treated within the GTVs.

## 7.2 Materials and methods

This study was conducted with the approval of the Human Subjects Research Ethics Board of our institution and with the informed consent of all subjects.

### 7.2.1 Participants

Our study population comprised 25 men selected from patients being treated for prostate cancer by one of three collaborating urologists. The demographics and clinical characteristics of these subjects are given in **Table 7.1**. 107 patients were screened for this study between March 2010 and December 2011. Inclusion criteria included (1) age 18 or older, (2) clinical stage T1 or T2 prostate cancer confirmed on previous biopsy, and (3) suitability for and willingness to undergo radical prostatectomy. The exclusion criteria were (1) previous prostate cancer therapy (n=8); (2) use of 5-alpha reductase inhibitors in the 6 month period prior to enrollment in the study (n=6); (3) inability to undergo imaging prior to the surgery (n=1); (3) allergies to the contrast agents (n=0); (5) sickle cell or other anemia (n=0); (6) insufficient renal function (n=1); (7) sources of imaging artifacts in the pelvis (e.g. hip or penile prostheses) (n=1); (8) contraindications to MRI such as metal in the orbit, cerebral aneurysm clips, claustrophobia and morbid obesity (n=11); and (9) prostate sizes exceeding the dimensions of 2 inch by 3 inch whole-mount histology slides (n=15).

**Table 7.1:** Subject demographics and clinical characteristics.

<b>Patient age median, range (y)</b>	63, 45–72
<b>PSA level median, range (ng/mL)</b>	4.7, 1.3–18.8
<b>Prostate volume estimated from transrectal ultrasound median, range (cm<sup>3</sup>)</b>	35, 19–49
<b>Pathological stage (number of subjects)</b>	
<b>pT2c</b>	17
<b>pT3a</b>	7
<b>pT3b</b>	1
<b>Patient overall Gleason score (number of subjects)</b>	
<b>3+3</b>	6
<b>3+4</b>	17
<b>3+4+5</b>	1
<b>4+3</b>	1

### 7.2.2 Imaging

We acquired mpMRI of the prostate  $19 \pm 8$  weeks (mean  $\pm$  SD) weeks post-biopsy and  $2 \pm 1$  (mean  $\pm$  SD) weeks before surgery. mpMRI were acquired on a 3 T GE Discovery MR750 (GE Healthcare, Waukesha, WI, USA) with an endorectal coil (Prostate eCoil, Medrad, Inc., Warrendale, PA, USA) using T2-weighted (T2W), dynamic contrast-enhanced (DCE) and diffusion-weighted (DW) clinical sequences detailed in **Table 7.2**. An additional high-resolution T2W image was also acquired to support registration (also detailed in **Table 7.2**). DW images were processed on the MRI console to generate apparent diffusion coefficient (ADC) maps. Images from any sequence where 3 of the 4 radiologist noted that the images were non-diagnostic or suffered from artifacts or distortion were excluded from the analysis; four ADC images were excluded based on this criterion.

**Table 7.2:** mpMRI sequence details.

<i>Sequence</i>	<i>Clinical T2W</i>	<i>Clinical DCE</i>	<i>Clinical DWI</i>	<i>High resolution T2W</i>
<b>Repetition time (ms)</b>	4000–13000	5.6–5.9	4000	2000
<b>Echo time (ms)</b>	156–164	2.1–2.2	70–77	144–177
<b>Bandwidth (kHz)</b>	31.25	31.25	125.00	125
<b>Number of excitations</b>	2	1–2	3	0.5–2
<b>Field of view (cm)</b>	14	14	14	14
<b>Slice thickness (mm)</b>	2.2	2.8	3.3–3.6	1.4
<b>Slice spacing (mm)</b>	2.2	1.4	3.3–3.6	0.7
<b>Matrix</b>	320 × 192	256 × 192	128 × 256	320 × 192
<b>Number of slices</b>	40	42	20–34	84–144
<b>Flip angle (°)</b>	90	15	90	90
<b>b-value</b>	N/A	N/A	600–800	N/A
<b>Temporal spacing (s)</b>	N/A	90	N/A	N/A

After resection, the prostate gland was fixed for at least 48 hours in 10% buffered formalin. MRI- and histology-visible fiducials [14] were affixed to the specimen and *ex vivo* MR images of the gland were acquired to support histology fusion [15] (described in Section 7.2.5). The prostatic apex and base were removed for parasagittal histology, and the midgland was cut into 4.4 mm slices using a rotary cutter and was processed using our clinical laboratory’s whole-mount histology protocol. Only the midgland histology was used in the analysis in this study. Histology was stained with hematoxylin and eosin, and digitized using a ScanScope GL (Aperio Technologies, Vista, CA, USA) bright field slide scanner at 20x magnification (yielding a 0.5  $\mu\text{m}$  pixel size).

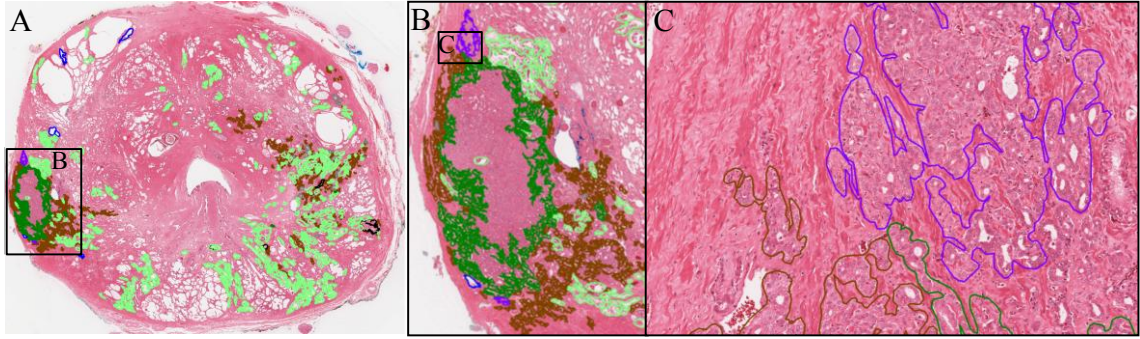
### 7.2.3 Lesion scoring and delineation

mpMRI images were independently read, blinded to histopathological results, by four observers, comprising one radiology resident (D.W.C.) and three radiologists (C.R.,

M.B.-J., and Z.K.) with 5, 6, 2.5 and 2.5 years of experience in prostate MRI assessment, respectively. Observers identified lesions using all mpMRI sequences together, and following the PI-RADS guidelines [8] assessed the overall likelihood score. For lesions judged to be equivocal (PI-RADS likelihood score of 3), likely cancer (4) or highly likely cancer (5), the observer attempted to delineate the lesions on images from each of the mpMRI sequences. Images were read using ClearCanvas 7.1 (ClearCanvas Inc., Toronto, Canada). Lesions were delineated on all image slices using a pen-enabled 21.5 inch widescreen display (Cintiq 22HD, Wacom Co., Ltd., Otone, Japan) with ITK-SNAP 2.4 software ([www.itksnap.org](http://www.itksnap.org)) in polygon or paintbrush mode.

#### 7.2.4 Histological contouring and grading

Histology images were delineated and graded, blinded to MR imaging, by a clinician (M.G.) with 3 years of experience in contouring prostate cancer on digital histology images under the supervision of 2 genitourinary pathologists (J.A.G. and M.M.) with 13 and 22 years of experience, respectively. Each region of homogenous or intermingled Gleason grade was graded and delineated, as illustrated in Figure 7.1. The resulting contours were edited, if required, and confirmed by one of the pathologists. Histology images were delineated and graded using a pen-enabled 21.5 inch widescreen display (Cintiq 22HD, Wacom Co., Ltd., Otone, Japan) with ImageScope 10 software (Aperio Technologies, Vista, CA, USA).



**Figure 7.1:** Illustrative example of a contoured and graded H&E-stained whole-mount histology image (A) with magnified insets (B) and (C), showing homogenous Gleason 3+3 (brown), intermingled Gleason 3+4 (dark green), intermingled Gleason 4+3 (purple), and prostatic intraepithelial neoplasia (light green) contours.

### 7.2.5 Histology-mpMRI registration

Histology images, and the corresponding contours, were registered to the mpMR images using a three stage registration. First, the histology images were reconstructed into the 3D context of *ex vivo* MR images using a method employing histology- and MRI-visible strand-shaped fiducial markers (Chapter 4) [15]. This method minimizes the fiducial registration error, under the space of affine transformations, between cross-sections of fiducial markers identified on histology and the corresponding curves of the strand-shaped fiducial markers on *ex vivo* MRI. Second, the *ex vivo* MR images (and corresponding histology and contours) were transformed into the space of the high resolution T2W *in vivo* image using a thin-plate-spline (TPS) deformation [16] computed from 33–91 control points defined interactively in 3D Slicer 4.2 (Surgical Planning Lab, Harvard Medical School, Boston, USA). Finally, the images from the three mpMRI sequences were registered to the high resolution 3D T2W image using a rigid transformation defined interactively in 3D Slicer. Where residual non-rigid spatial distortion was observed between *in vivo* images (which occurred in 6 ADC images), it was mitigated using an interactively defined TPS deformation.



The target registration error (TRE) for the histology reconstruction has been previously reported (Chapter 4) [15]. The TRE for the rigid mpMRI co-registration was estimated by identifying pairs of homologous landmarks and measuring the post-transformation 3D Euclidean distance between corresponding pairs. The TRE for the TPS deformations (*ex vivo*–*in vivo* registration and *in vivo*–*in vivo* distortion correction) was estimated using a leave-one-control-point-pair-out estimate,

$$TRE_i = \text{mean}_{1 \leq k_i \leq K_i} \left\| T_{\sim k_i}(\mathbf{c}_{k_i}^f) - \mathbf{c}_{k_i}^m \right\|, \quad 7.1$$

where  $\mathbf{c}_{k_i}^m$  and  $\mathbf{c}_{k_i}^f$  are the  $k_i$ -th control point pair (in the moving and fixed image, respectively) for the  $i$ -th image pair,  $T_{\sim k_i}(p)$  is the TPS deformation defined using all except the  $k_i$ -th control point pair, and  $K_i$  is the number of control point pairs in the  $i$ -th image pair.

### 7.2.6 Histology fusion

Because the histology was contoured and graded at a fine scale (see Figure 7.1), histology contours were aggregated into foci. Within each histology image, contours within 2 mm were considered to be part of the same aggregated histological focus. To aggregate contours between adjacent histology images, contours were reconstructed into 3D (in the space of the *ex vivo* image) and projected onto a plane midway between the reconstructed histology images; any contours that overlapped were considered to be part of the same aggregated histological focus. Aggregated foci smaller than  $0.1 \text{ cm}^3$  (estimated using planimetry as the in-plane area times the histological spacing) with a Gleason score of 3+3 were excluded from analysis. True positive GTVs were identified by the first author (not involved in mpMRI or histology contouring) using the co-

registered contoured histology and mpMRI images. Where two observer GTVs corresponded to a single aggregated focus, the union of the GTVs was considered as a single GTV.

### 7.2.7 CTV definition

Single-sequence CTVs were defined by expanding GTVs on images from each sequence (T2W, DCE and ADC) by 3D isotropic margins from 0–30 mm at 0.5 mm intervals, with the CTVs constrained to lie within the prostate. Binary masks representing the isotropic expansion of the GTVs (computed in the delineated image) were transformed to the high resolution T2W image space using the co-registration transformations described in Section 7.2.5. Binary masks of the prostate volume, computed by thresholding an *ex vivo* MRI of the resected specimen, were also transformed to the high resolution T2W image space, and used to constrain the CTVs to the prostate volumes.

In addition to the single-sequence CTVs, composite CTVs were defined by the union of single-sequence CTVs for each pair of mpMRI sequence (i.e. T2W+DCE, T2W+ADC and (DCE+ADC) and for the union of all three sequences (i.e. T2W+DCE+ADC). For true positive findings where an observer declined to contour on one of the sequences, composite CTVs including that sequence were not defined.

### 7.2.8 Histological coverage

The coverage of histologically identified cancer in the midgland was estimated for CTVs corresponding to each true positive GTV. Following a stereological approach [17], the volumetric proportion of residual cancer was estimated from the proportion by area

on the histology sections as the ratio between the total area of residual cancer and the total area of cancer across histology samples. CTVs were sampled along the best-fit oblique plane to the deformably co-registered histology surface, and histology contours were projected on this oblique plane. The histology contours and CTVs were discretized onto a 0.27 mm grid (matching the in-plane spacing of the highest resolution *in vivo* MR image). The proportional residual area of cancer for each CTV was computed as the proportion of discretized histology pixels containing cancer from the corresponding aggregated focus that were not contained within the CTV. The proportional residual area of high-grade cancer for each CTV was computed similarly, limiting the analysis to pixels containing Gleason grades 4 or 5, and excluding aggregated foci containing only Gleason grade 3. To the best of our knowledge, the prognostic impact of small volumes of residual cancer post-treatment are not known and may vary with treatment modality, challenging the definition of adequate histological coverage. To mitigate this uncertainty, we investigated 3 thresholds for residual cancer area: 0%, 5% (following [13]) and 10% residual area.

In order to identify the minimum margin with a high likelihood of achieving the specified residual area for a new patient, considering intertumor variability, we computed one-sided non-parametric prediction intervals (PIs) for the percentage of residual cancer area and for the percentage of residual high-grade cancer. PIs were computed for each observer for each type of CTV (3 single-sequence and 4 composite types) at each margin size. The confidence level and construction of the PIs are described in Section 7.2.10.

### 7.2.9 CTV volumetrics

Since individual sequences may consistently over- or under-estimate the extent of cancer and different margins could be specified for each CTV type to account for this, CTV volumetric analyses across CTV types were based on a fixed residual area criterion instead of fixed margin sizes (see discussion for more details). Specifically, for each observer and CTV type, the smallest margin that yielded a residual area prediction interval upper bound  $<5\%$  was defined as the *PI margin*. Different CTV types were compared at their respective PI margins. The 5% residual area criterion corresponds to an *adequate coverage* threshold described in previous literature [13].

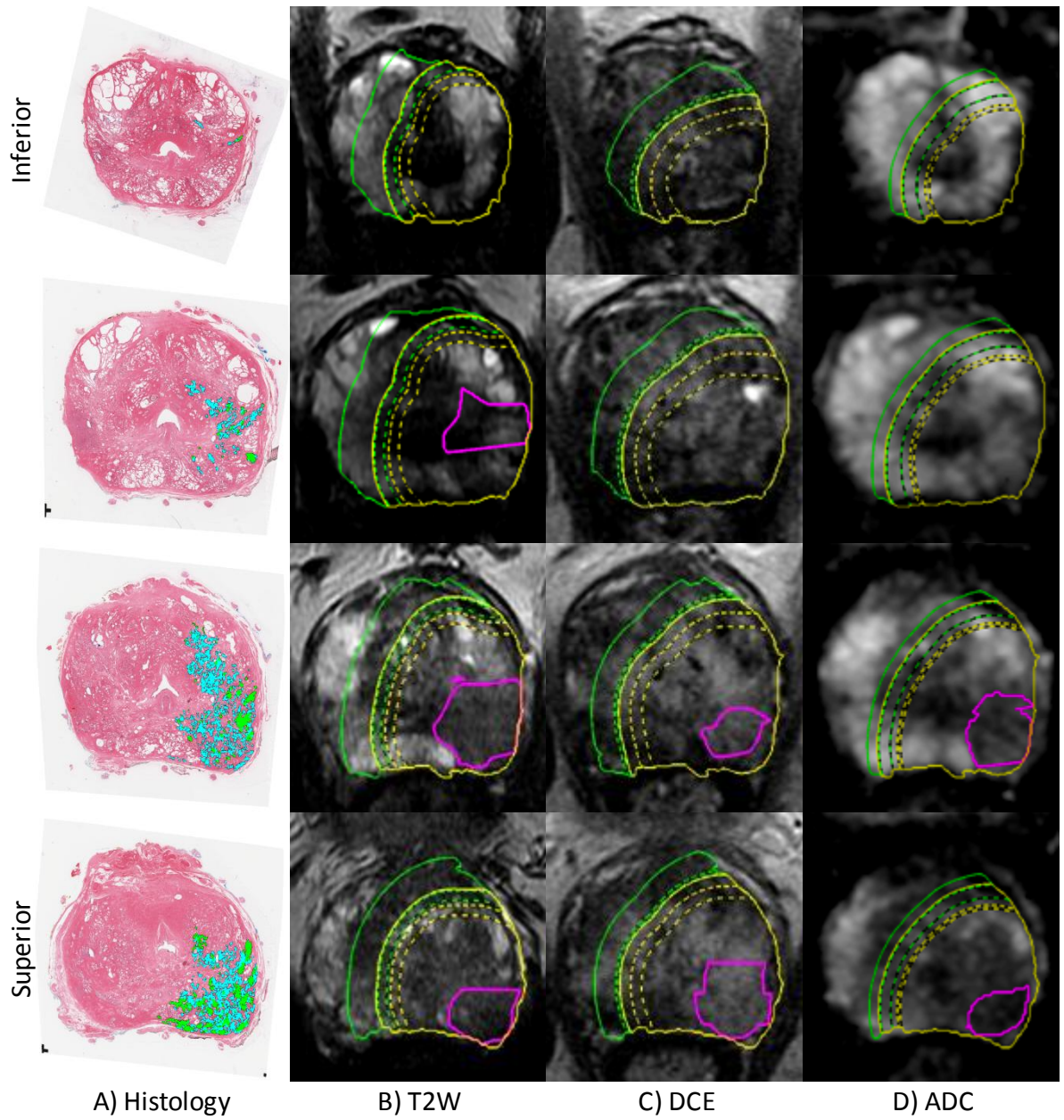
We computed the absolute volumes of the CTVs with PI margins, in the space of the high resolution T2W image. We compared these volumes to 2 thresholds:  $4.5 \text{ cm}^3$ , based on reported volumes for focal laser ablations [18]; and  $10 \text{ cm}^3$ , based on reported volumes for focal radiation boosting [6]. We also computed the relative sizes of the CTVs with PI margins with respect to the volumes of the GTVs, in the space of the high resolution T2W image. The volume ratios between the CTVs and GTVs were expressed as scaling factors (defined as the cube roots of the volume ratios) representing the equivalent expansions in each dimension. We computed the relative volumes of the CTVs with respect to the histologically identified cancers in a manner similar to the residual area calculation, using planimetrics in the space of the best-fit oblique planes to the deformably co-registered histology images. The volume ratios between the CTVs and cancer foci were expressed as scaling factors (defined as the square roots of the planimetric volume ratios) representing the equivalent expansion in each dimension.

### 7.2.10 Statistical analysis

Statistical analyses were performed in SPSS 21 (IBM Corp., Armonk, NY). The impact of tumor and observer variability (as random effects) and the incremental effect of contouring on each sequence (as fixed effects) on the minimum margin required to achieve 95% coverage were tested using a mixed-effects ANOVA (with an observer  $\times$  tumor interaction term and pairwise sequence inclusion interaction terms). Relative contributions of tumor and observer variability to margin variability were computed as variance components under the same mixed-effects model. Non-parametric, one-sided prediction intervals of residual cancer and residual high-grade cancer were calculated aggregated over the N true positive CTVs for each observer for each CTV type at each margin size. The confidence levels chosen for these intervals were  $(N-1)/(N+1)$ , such that the prediction interval upper bound corresponded to the penultimate order statistic of the sample [19]. This mitigated the sensitivity to outliers of the highest confidence level prediction interval (corresponding to the maximum on the sample).

## 7.3 Results

In our data set, 23/25 subjects had midland aggregated foci included in the analysis, yielding 59 aggregated foci. Of the foci larger than  $0.1 \text{ cm}^3$ , 76% (22/29) were contoured by at least one observer, whereas only 13% (4/30) of the smaller foci were contoured. The numbers of foci and the subset that were contoured are shown in Table 7.3, broken down by volume and Gleason score. Since each focus could be contoured by 1–4 observers, 72 sets of true positive GTVs were contoured and analyzed. Observers 1–4 contours 20, 16, 23 and 13 sets of true positive GTVs, respectively. An illustrative set of true positive GTVs and selected corresponding CTVs are shown in Figure 7.2.



**Figure 7.2:** Illustrative set of true positive GTVs for a Gleason 3+4 sparse cancer. Column A shows 4 sections of midland histology with regions of Gleason 7 cancer (green) and Gleason 6 cancer (cyan) highlighted. Columns B–D show corresponding oblique sections through the T2W, DCE and ADC images respectively, with overlaid boundaries for the corresponding GTVs (magenta), and CTVs with PI margins for 0% (solid line), 5% (dashed) and 10% (dotted) residual area for any cancer (green) and for high-grade cancer (yellow).

**Table 7.3:** For each volume range and Gleason score: number of aggregate foci (subset with at least on true positive GTV).

	<i>Volume (cm<sup>3</sup>)</i>			
	<i>&lt;0.1</i>	<i>0.1–0.2</i>	<i>0.2–0.5</i>	<i>&gt;0.5</i>
<b>Gleason score 6</b>	excluded	5 (3)	3 (1)	1 (1)
<b>Gleason score 7</b>	30 (4)	8 (7)	9 (7)	3 (3)

The key findings from our analysis were that (1) much of the variance in required margin size could be attributed to tumor variability, (2) the required margins to achieve 0%, 5% and 10% residual cancer with high probability was smallest (4.5–9.0 mm, 1.5–7.5 mm and 1.5–4.5, respectively, for residual high-grade cancer, and 6.0–11.0 mm, 5.5–10.0 mm, and 3.0–9.5 mm, respectively, for any residual cancer) for composite GTVs defined as the union of T2W, DCE and ADC GTVs; (3) these margins were lower than the margins for single-sequence GTVs but in many cases not lower than the margins for at least one 2-sequence GTV; and (4) the lower margins for 2- and 3-sequence composite GTVs did not result in clear reductions in volumes of the resulting CTVs in cases where the difference was <2 mm.

The residual areas as a function of margin size for each tumor individually are included in Appendix D. The per-tumor analysis of required margin to achieve adequate coverage of histological cancer was conducted for both all cancer and high-grade cancer, and at three thresholds for residual area (0%, 5% and 10%). The mean required margins (aggregated over all tumors, CTV types and observers) ranged from 3.5–6.8 mm for all cancer and from 2.4–5.0 mm for high-grade cancer. The maximum required margins in these analyses ranged from 15–22 mm for all cancer, and 16.5–21 mm for high-grade cancer. Details of these analyses are given in Table 7.4. For the 10% residual high-grade

cancer analysis, one set of GTVs required margins 6–8 standard deviations above the mean required margins for the remaining data. Table 7.4 shows the details of this analysis with the corresponding CTVs included and excluded. Excluding this analysis, all three random effects were significant factors in the required margins: tumor variability ( $p < 0.001$ ), tumor-observer interaction ( $p < 0.001$ ) and observer variability ( $p \leq 0.03$ ). In the 10% residual high-grade cancer analysis with all CTVs included, only the tumor-observer interaction was significant; however, with the aforementioned GTV excluded, all three random factors were significant. The factors representing additional delineation of images from each of the sequences were all significant, with an average reduction in required margin of 0.8–1.1 mm for contouring DCE GTVs, 0.3–0.7 mm for contouring T2W GTVs and 0.1–0.6 mm for contouring ADC GTVs. Consistently significant interaction effects between the addition of T2W and DCE GTVs suggest that their incremental values depend on the presence or absence of the other GTV. The variance component analysis suggests that tumor variability is the largest contributor to variability in required margins; however, tumor-observer interaction and observer effects were also significant.

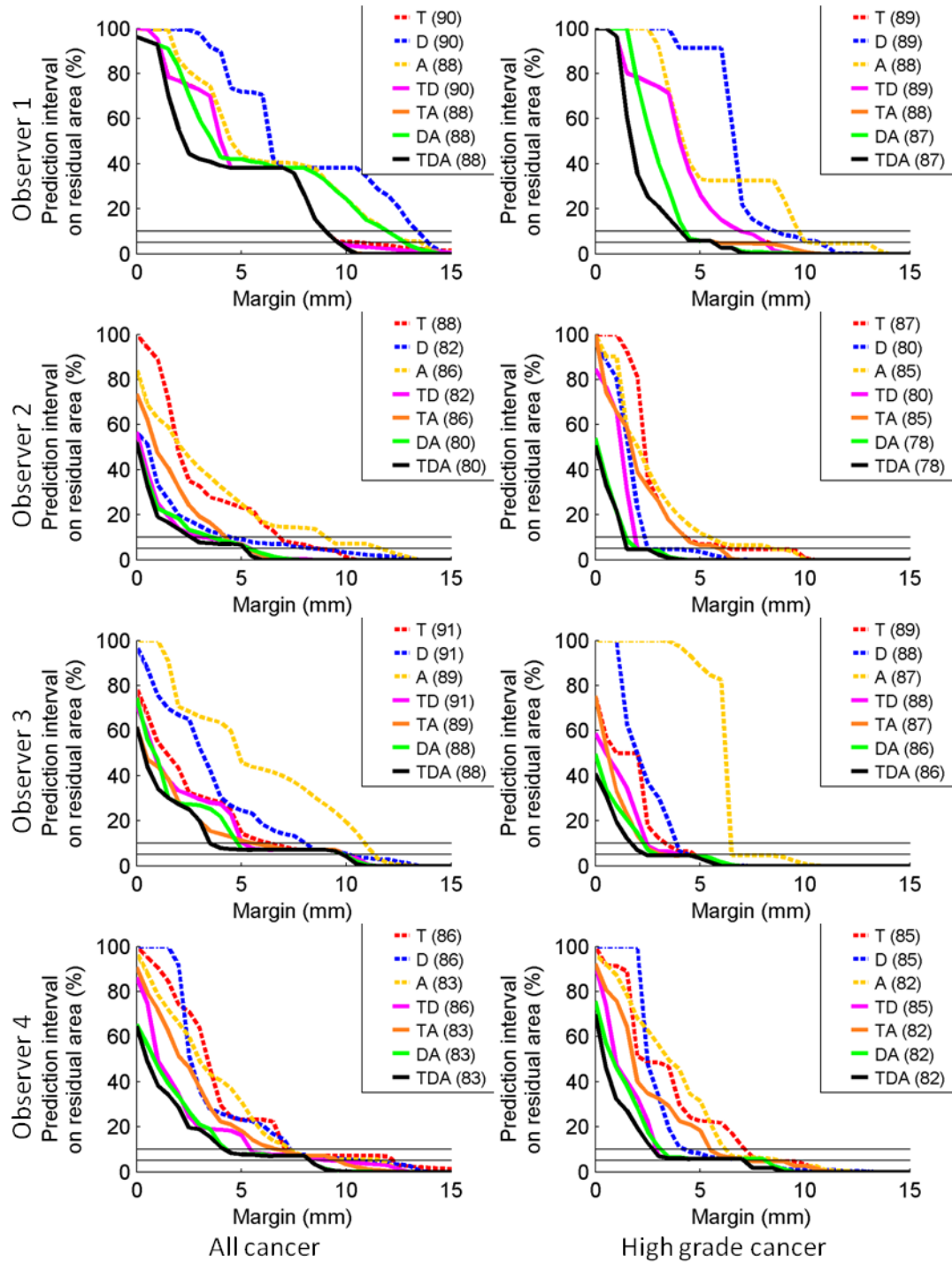


**Table 7.4:** Per-tumor statistical analysis of the minimum margins required to achieve specified residual areas. The variance components represent the relative weighting of different sources of random variability on the per-tumor margin variability. The final 3 rows show estimates of the incremental value of delineating images from an additional MR sequence, in terms of how much it decreased the margin, on average. The analysis in the final column (marked with \*) excludes one set of GTVs that required margins 6–8 standard deviations higher than the mean of the remaining data.

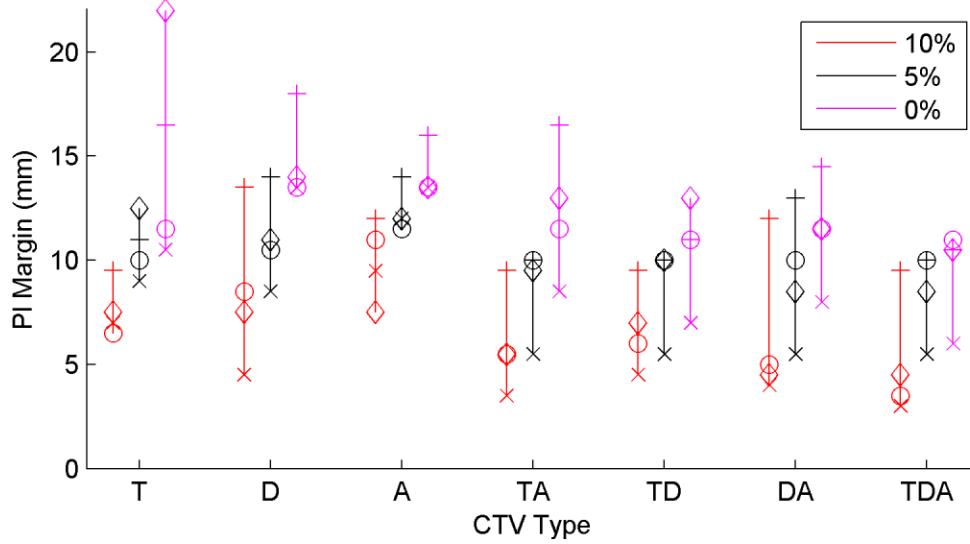
	<i>Residual cancer</i>			<i>Residual high-grade cancer</i>			
	<i>0%</i>	<i>5%</i>	<i>10%</i>	<i>0%</i>	<i>5%</i>	<i>10%</i>	<i>10%*</i>
Descriptive statistics on minimum margin (mm)							
Mean	6.8	4.7	3.5	5.0	3.3	2.4	2.2
Minimum	0	0	0	0	0	0	0
Maximum	22	18	15	21	18	16.5	11
Variance components of minimum margin (%)							
Var <sub>Tumor</sub>	71	49	43	48	44	2	23
Var <sub>Tumor-Observer</sub>	10	24	27	18	20	65	23
Var <sub>Observer</sub>	5	7	8	10	11	7	9
Var <sub>Unmodeled</sub>	13	20	22	24	25	26	45
Incremental decrease in minimum margin for contouring additional images (mm)							
T2W	0.7	0.5	0.4	0.3	0.3	0.4	0.4
DCE	1.0	0.9	0.8	1.1	0.9	0.8	0.8
ADC	0.6	0.5	0.4	0.4	0.3	0.2	0.1

The residual area PI upper bounds for all CTV types for each observer for all cancer and for high-grade cancer are shown in Figure 7.3, as a function of margin size. The PI margins (the smallest margin where the PI upper bound was <5% residual cancer) ranged widely from 5.5–14.0 mm for coverage of all cancer, and from 1.5–11.0 mm for high-grade cancer. The PI margins for all cancer and for high-grade cancer are summarized in Figure 7.4 and Figure 7.5, respectively. In general, for all observers, for all three thresholds on residual area for both all cancer and high-grade cancer, the PI margins for the three-sequence composite GTVs (defined as the union of T2W, DCE and ADC GTVs) were lower than the PI margins for single-sequence GTVs, with the

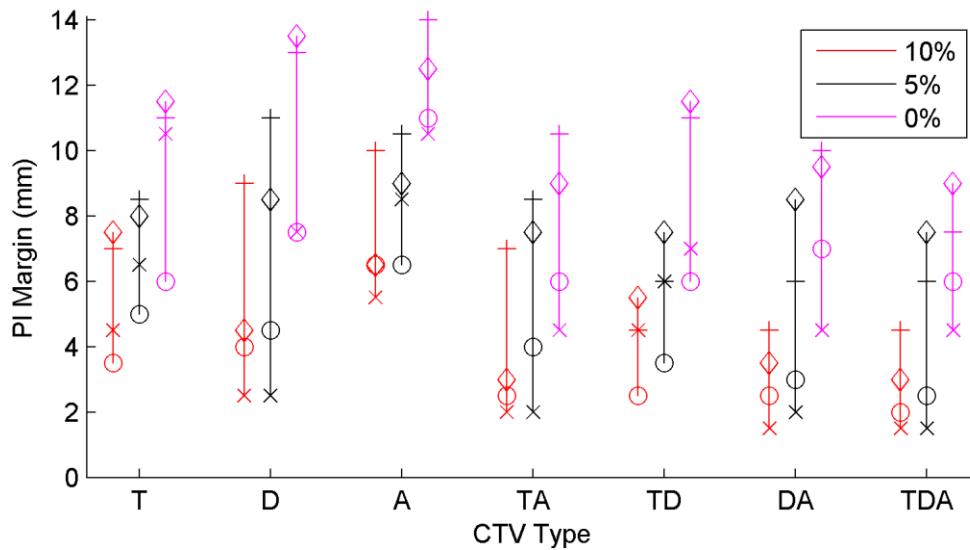
exception of three cases where the T2W PI margin was the same as the three-sequence margin for one observer. However, the PI margins for the 3-sequence composite GTVs were not consistently lower than those for the 2-sequence GTVs. The lowest required margins for high-grade cancer were 4.5 mm, 1.5 mm and 1.5 mm, for 0%, 5% and 10% residual area, respectively, for 3-sequence composite CTVs; however, the number of GTVs considered in this estimate was lower (due to fewer included true positive GTVs), resulting in a lower confidence level (78%) for these PIs. The 90% confidence PI margins for this observer/CTV type were 7.5 mm, 3 mm, and 2.5 mm for 0%, 5% and 10% residual area, respectively.



**Figure 7.3:** Upper bound on residual cancer (left) and high-grade cancer (right) prediction intervals for 4 observers (rows). Legend denotes CTV types by the first initials of the sequences (T=T2W, D=DCE, A=ADC), with the PI confidence level in parentheses. The 5% and 10% residual area levels are marked by the horizontal lines.



**Figure 7.4:** PI margins for coverage of low- and high-grade cancer for 4 observers (denoted by symbols) for 0%, 5% and 10% residual area for each CTV type. CTV types are denoted by the first initials of the sequences (T=T2W, D=DCE, A=ADC)



**Figure 7.5:** PI margins for coverage of high-grade cancer for 4 observers (denoted by symbols) for 0%, 5% and 10% residual area for each CTV type. CTV types are denoted by the first initials of the sequences (T=T2W, D=DCE, A=ADC)

The volumes of CTVs with PI margins for all cancer and for high-grade cancer are shown in Figure 7.6 and Figure 7.7, respectively. The median volumes of CTVs and the proportions of CTVs with PI margins smaller than  $10 \text{ cm}^3$  and smaller than  $4.5 \text{ cm}^3$

are given in Table 7.5. Of the 53 and 67 outliers shown in the 24 graphs in Figure 7.6 and Figure 7.7, respectively, 45/53 and 60/67 correspond to one histological focus, with a 21-mm longest diameter on whole-mount histology, substantial Gleason grade 4, and extraprostatic extension near the bladder neck. The two other foci that led to the remaining 8/53 and 7/67 outliers were large diffuse cancers, one Gleason score 3+3 with a largest diameter of 36 mm and extraprostatic extension, and one Gleason score 3+4 with tertiary grade 5, with a largest diameter of 24 mm and extraprostatic extension. Median CTV volumes for composite CTVs with PI margins were neither consistently lower nor consistently higher than those for single-sequence CTVs. However, for cases where the difference between the single-sequence and composite PI margins was  $< 2$  mm, the median single-sequence-CTV volumes were smaller by  $0\text{--}4\text{ cm}^3$ , while for cases where the difference was  $\geq 2$ , the median single-sequence-CTV volumes were larger by  $0\text{--}16\text{ cm}^3$ .

**Table 7.5:** Volumes of CTVs with PI margins.

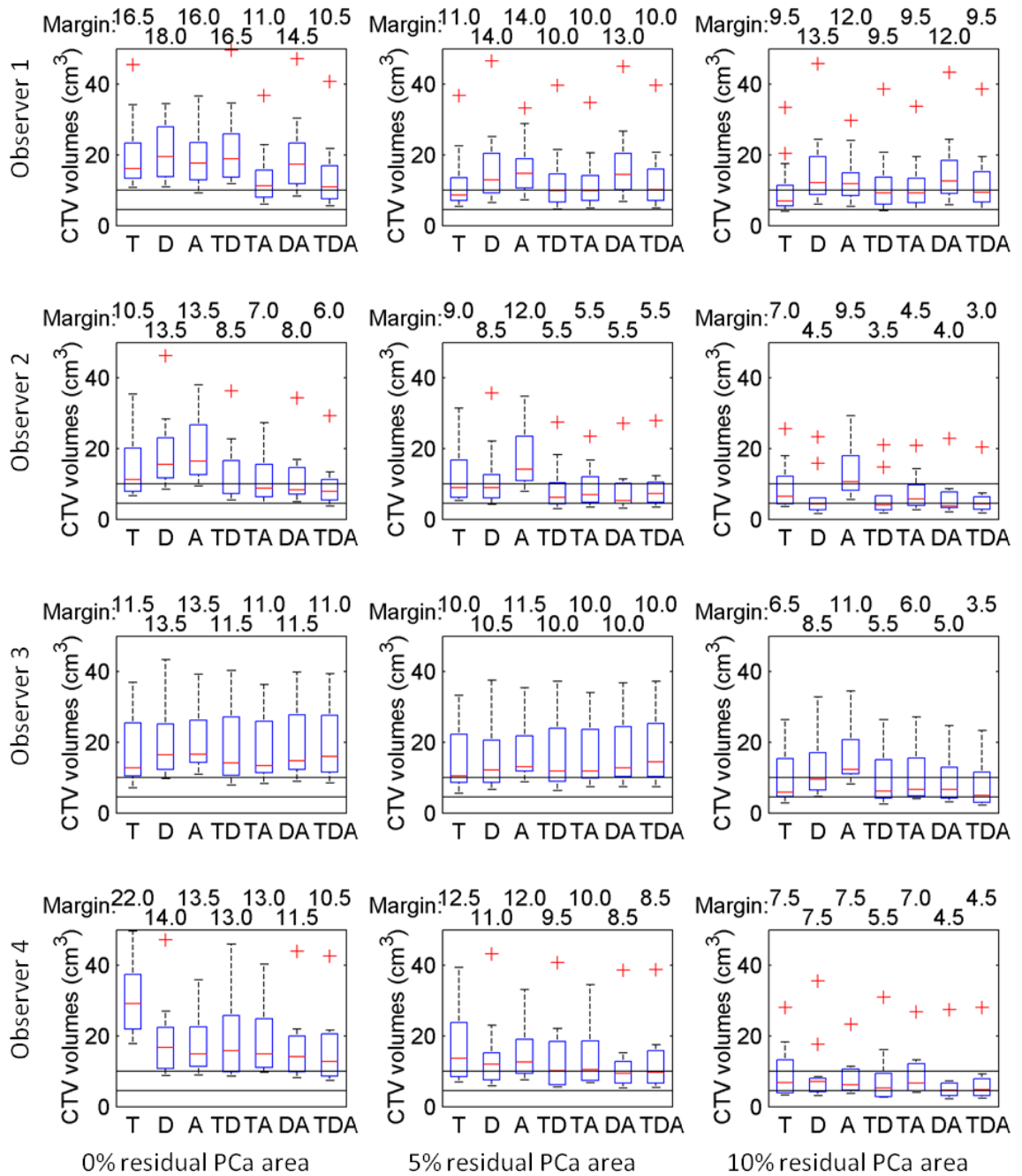
	<i>Residual cancer</i>			<i>Residual high-grade cancer</i>		
	<i>0%</i>	<i>5%</i>	<i>10%</i>	<i>0%</i>	<i>5%</i>	<i>10%</i>
Median volume T2W ( $\text{cm}^3$ )	16.3	9.8	6.7	9.2	5.8	4.2
Median volume DCE ( $\text{cm}^3$ )	16.8	12.0	8.1	10.4	7.0	4.3
Median volume ADC ( $\text{cm}^3$ )	16.5	13.5	11.3	13.4	8.9	6.7
Median volume T2W+DCE ( $\text{cm}^3$ )	15.8	10.1	6.7	8.4	6.1	3.7
Median volume T2W+ADC ( $\text{cm}^3$ )	12.7	10.5	6.8	10.3	6.3	4.7
Median volume DCE+ADC ( $\text{cm}^3$ )	14.4	11.7	7.0	9.1	5.4	3.2
Median volume T2W+DCE+ADC ( $\text{cm}^3$ )	12.4	10.7	6.0	8.3	5.1	3.2
Median volume (pooled) ( $\text{cm}^3$ )	15.4	11.4	7.7	10.3	6.4	4.4
Proportion with volume $\leq 4.5\text{ cm}^3$	0%	3%	22%	9%	32%	51%
Proportion with volume $4.5\text{--}10.0\text{ cm}^3$	20%	39%	41%	40%	35%	28%
Proportion with volume $>10.0\text{ cm}^3$	80%	58%	38%	51%	33%	21%

The expansion factors of the CTVs relative to their corresponding GTVs are shown in Figure 7.8 for each observer and level of residual PCa area. There are two key observations to be made from these plots. First, the CTVs with PI margin were substantially expanded from the delineated GTVs with median linear expansion factors of 1.3–4.9. Second, the median expansion factors for 3-sequence composite GTVs were consistently lower than those for single sequence GTVs. Note, however, that 3-sequence composite GTVs are by construction at least as large, and typically larger, than single-sequence GTVs.

The expansion factors of the CTVs relative to their corresponding histological foci are shown in Figure 7.9. The key observation to be made from these plots is that substantial non-cancerous parenchyma is covered by the CTVs with PI margins for 0%, 5% and 10% residual cancer area for all observers, with median linear expansion factors of 2.7–9.3. Note that these expansion factors incorporate volume due to CTV coverage of regions with no cancer as well as CTV coverage of stromal regions between diffuse cancerous gland (not included in histological focus volumes). Scaling factors for composite CTVs were not consistently larger or smaller than those for single-sequence CTVs.

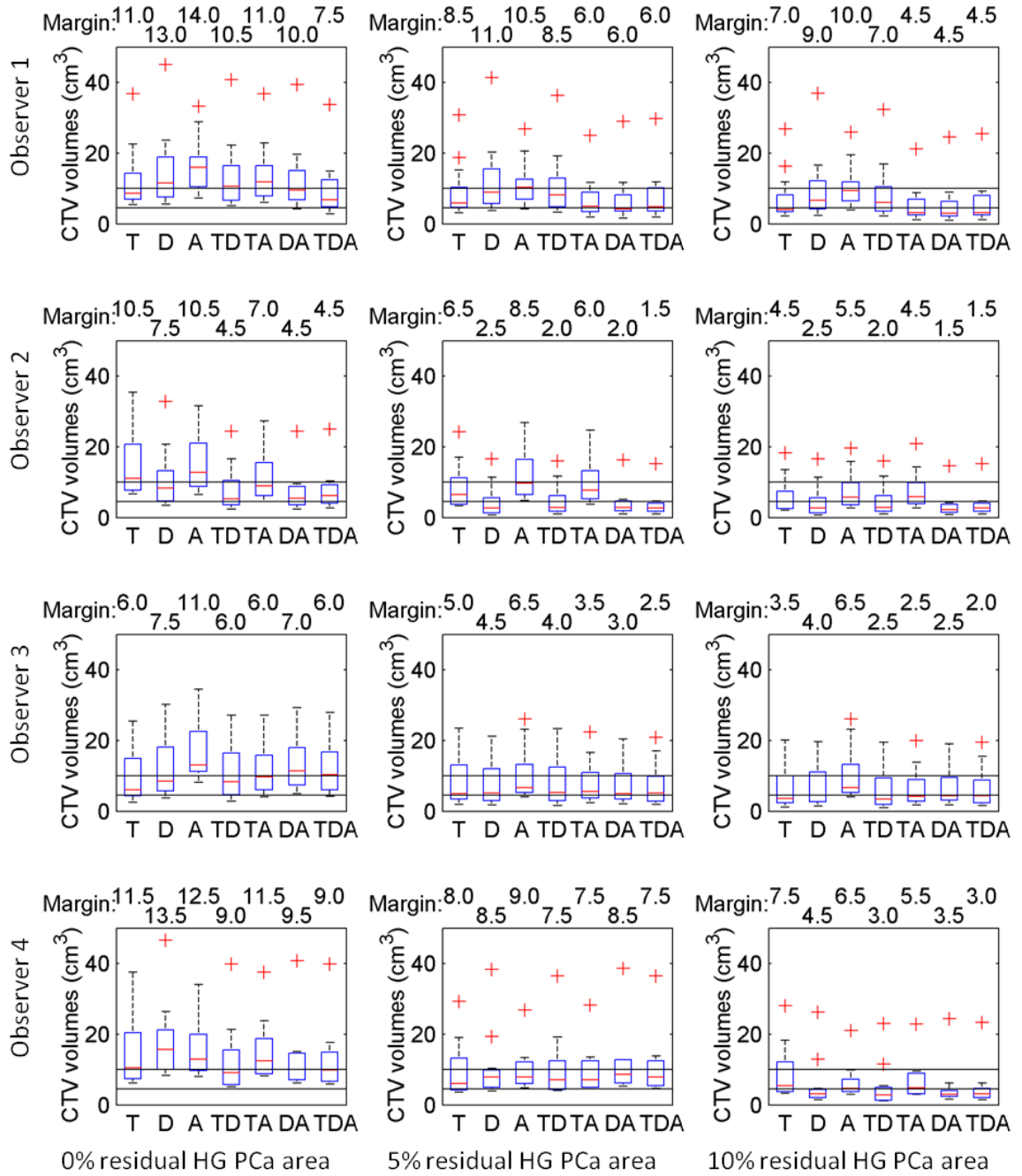
The mean $\pm$ SD TRE for the *ex vivo*–*in vivo* MR image TPS registrations was estimated to be 1.4 $\pm$ 0.2 mm aggregated across TRE estimates on 25 MR image pairs. Each TRE estimate was calculated using 33–90 control point pairs (1482 control point pairs total across all MR image pairs). The mean $\pm$ SD TREs for the rigid mpMRI co-registrations to the high resolution T2W images were estimated to be 0.7 $\pm$ 0.1 mm for the clinical T2W images, 1.0 $\pm$ 0.5 mm for the DCE images, and 1.0 $\pm$ 0.2 mm for the ADC

images, aggregated across TRE estimates on 3 MR image pairs. Each TRE estimate was calculated using 4–6 intrinsic landmark pairs (e.g. calcifications and atrophic cysts) per MR image pair (41 intrinsic landmark pairs total). For the distortion correction registering the 6 distorted ADC images to their respective high resolution T2W images, the mean $\pm$ SD TREs was estimated to be 1.4 $\pm$ 0.5 mm aggregated across TRE estimates on 6 MR image pairs. Each TRE estimate was calculated using 32–65 control point pairs (299 control point pairs total across all 6 MR image pairs). The mean $\pm$ SD TRE for the 3D histology reconstruction was previously reported to be 0.7 $\pm$ 0.4 mm [15] aggregated across 232 intrinsic landmark pairs (3–16 intrinsic landmark pairs per histology section for 37 histology sections).

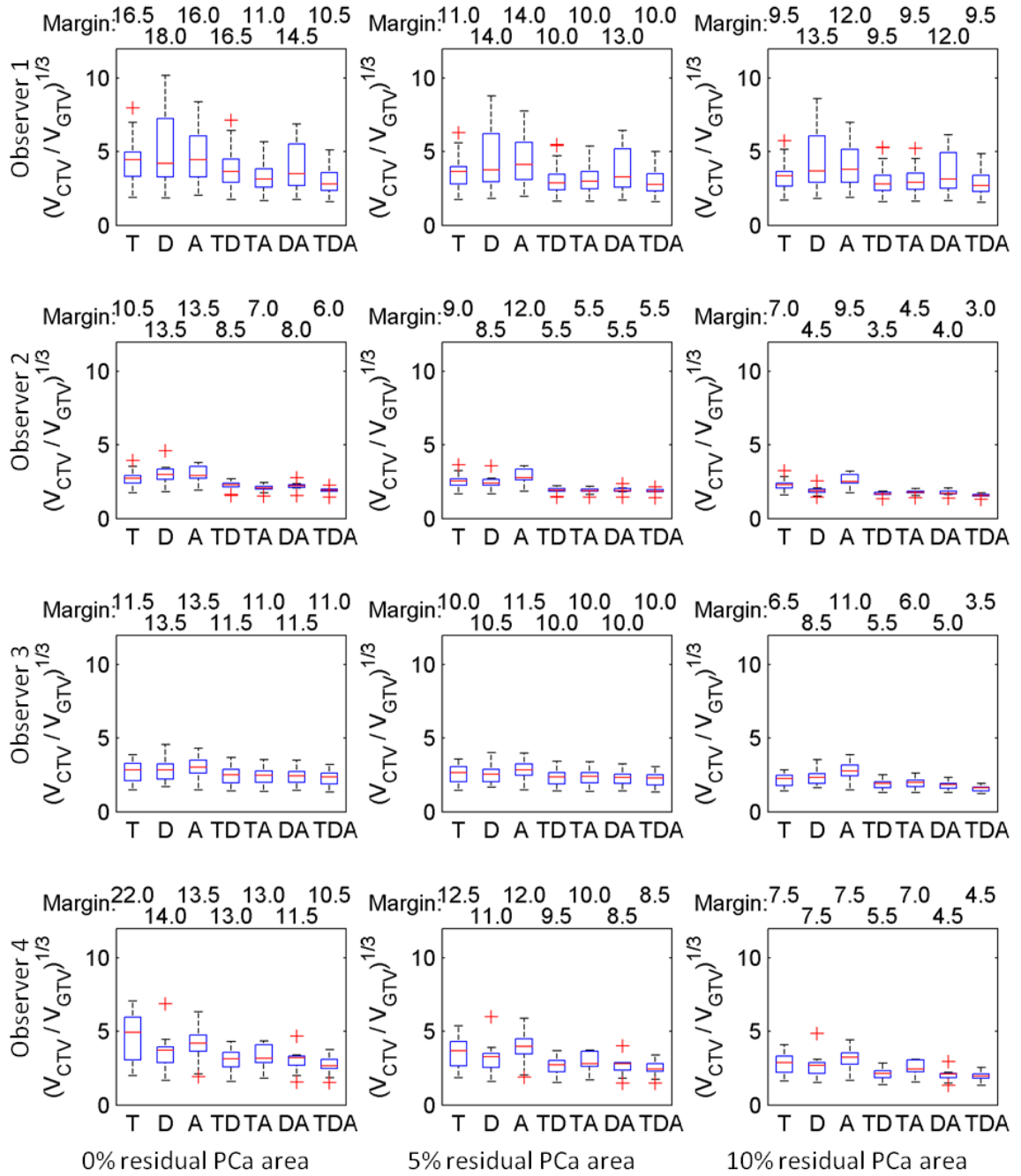


**Figure 7.6:** Volume encompassed by CTVs with PI margins for **any cancer** for 4 observers (rows) for 0%, 5% and 10% residual cancer area (columns). CTV types are denoted by the first initials of the sequences (T=T2W, D=DCE, A=ADC), with the PI margin shown above the graph. Box plots show the median, interquartile range (IQR), whiskers for the closest points within  $1.5 \times \text{IQR}$  from the IQR, and outliers beyond the whiskers. The horizontal lines mark  $4.5 \text{ cm}^3$  and  $10 \text{ cm}^3$ , reported treatment volumes for focal laser ablation and focal radiation boosting, respectively.

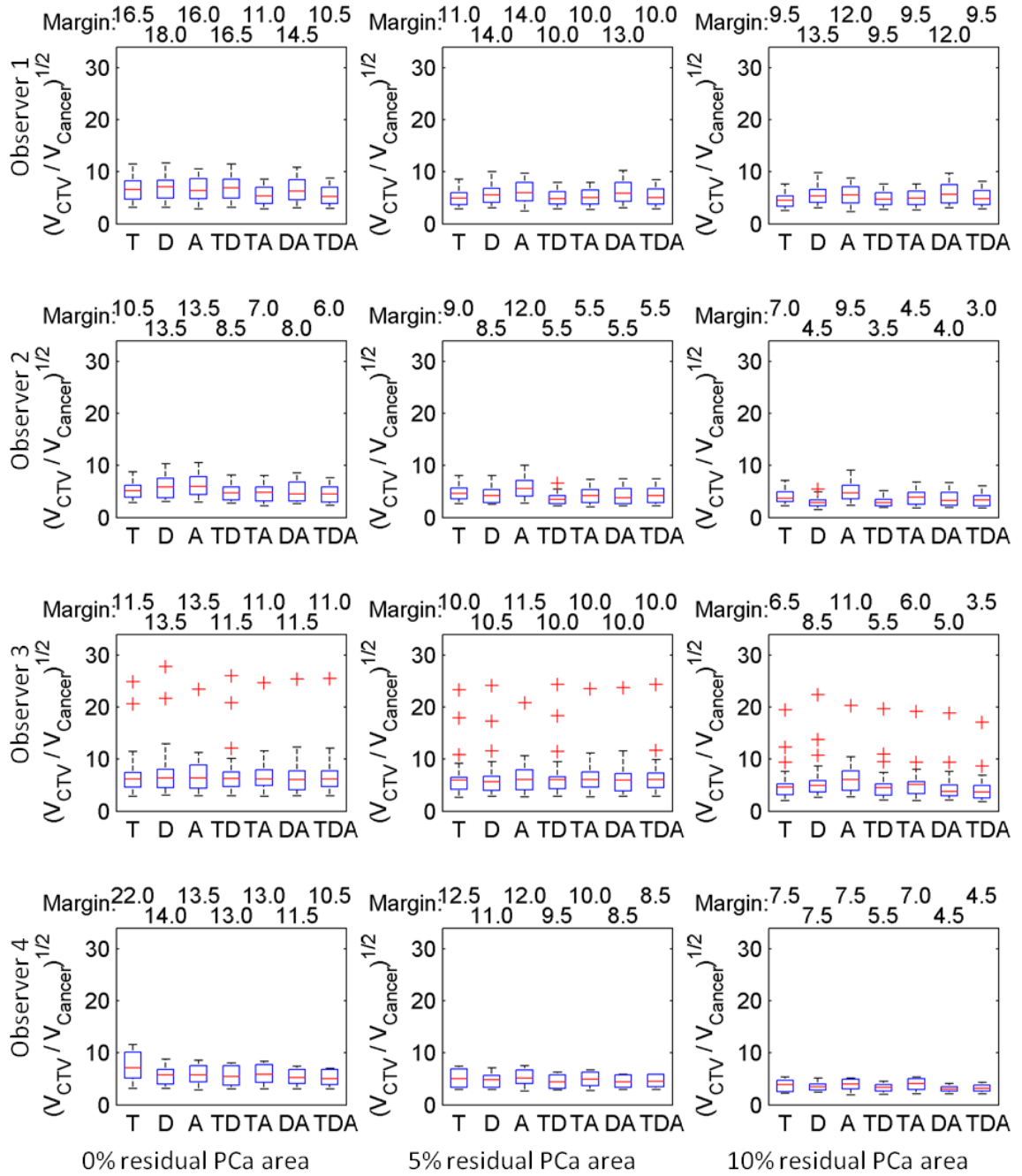




**Figure 7.7:** Volume encompassed by CTVs with PI margins for **high-grade cancer** for 4 observers (rows) for 0%, 5% and 10% residual high-grade cancer area (columns). CTV types are denoted by the first initials of the sequences (T=T2W, D=DCE, A=ADC), with the PI margin shown above the graph. Box plots show the median, interquartile range (IQR), whiskers for the closest points within  $1.5 \times \text{IQR}$  from the IQR, and outliers beyond the whiskers. The horizontal lines mark  $4.5 \text{ cm}^3$  and  $10 \text{ cm}^3$ , reported treatment volumes for focal laser ablation and focal radiation boosting, respectively.



**Figure 7.8:** Scaling factors of the CTVs with PI margins **relative to their corresponding GTVs** for 4 observers (rows) for 0%, 5% and 10% residual cancer area (columns). Y-axis shows the cube root of the volume ratio (equivalent scaling factor in each dimension). CTV types are denoted by the first initials of the sequences (T=T2W, D=DCE, A=ADC), with the PI margin shown above the graph. Box plot shows the median, interquartile range (IQR), whiskers for the closest points within 1.5×IQR from the IQR, and outliers beyond the whiskers. A similar set of graphs for high-grade cancer only is shown in Appendix D.



**Figure 7.9:** Scaling factors of the CTVs **relative to their corresponding aggregated foci** for 4 observers (rows) for 0%, 5% and 10% residual area (columns). Y-axis shows the square root of the planimetric volume ratio (equivalent scaling factor in each dimension). CTV types are denoted by the sequences initials (T=T2W, D=DCE, A=ADC), with the PI margin shown above the graph. Box plot shows the median, interquartile range (IQR), whiskers for the closest points within  $1.5 \times IQR$  from the IQR, and outliers beyond the whiskers. A similar set of graphs for high-grade cancer only is shown in Appendix D.

## 7.4 Discussion

Delineation of prostate cancer on imaging has the potential to support multiple stages of the clinical workflow. Although mpMRI has shown promise for prostate cancer detection, staging and localization, its suitability for delineation is not yet well-characterized. In this work, we addressed 3 questions related to the suitability of mpMRI for delineating target volumes: (1) What margins must be added to delineated GTVs to achieve adequate coverage of histological cancer? (2) What are the volumes of CTVs defined using such margins? (3) What are the relative volumes of the CTVs defined using such margins with respect to the corresponding GTVs and the corresponding histological foci?

The mean margins required for adequate coverage in our data set were 6.8 mm, 4.7 mm and 3.5 mm for 0%, 5% and 10% residual cancer area, respectively. However, the margins required for adequate coverage of specific tumors in our data set ranged as high as 22 mm for 0% residual area, 18 mm for 5% residual area and 15 mm for 10% residual area. A substantial component of this variability was attributable to tumor variability. This suggests a strong need to characterize individual lesions during therapy planning to identify lesions that require larger margins. Imaging characteristics used to detect cancer, such as the "erased charcoal sign" or "lenticular shape" described in the PI-RADS reporting guidelines, could be investigated to assess their correlations with tumors that need larger margins.

Additional variability in required margin sizes was attributed to observer variability and tumor-observer interaction. Furthermore, there was considerable interobserver variability in the identified PI margins for each type of CTV. For CTVs

derived from GTVs on 1 or 2 sequences, there was interobserver variability not only in the sizes of the margins, but also in which modality yielded the smallest margins. This variability might be addressed through training targeted at delineation. Specific training in target volume contouring on mpMRI with histological feedback, as has been recommended for reading and reporting mpMRI [20], may play a role in reducing the observed variability and reducing required margins. Consensus guidelines for how to delineate target volumes for focal therapy may also reduce variability. If large interobserver variability remains despite training and guidelines, recommendation of generally applicable margin guidelines may be challenging. A possible alternative would be to develop benchmark databases of co-registered mpMRI and histology with tools to allow clinicians to characterize their own performance and develop individualized delineation guidelines in terms of optimal sequences and appropriate margins.

The observed variability suggests that precise identification of the margin for individual tumors may be challenging and supports the use of conservative guidelines that take into account the variability in required margins, such as prediction interval upper bounds.

The PI margins identified in this study were defined such that an observer using these margins on a new patient could have a high confidence (78%–91%) of leaving 0%, 5%, or 10% residual cancer (or high-grade cancer) area beyond the CTV boundary. The CTVs defined based on the composite of the T2W+DCE+ADC GTVs required the smallest margins. These three-sequence PI margins for high-grade cancer ranged from 4.5–9.0 mm, 1.5–7.5 mm and 1.5–4.5 mm for 0%, 5%, and 10% residual high-grade cancer area. For covering low and high-grade cancer, the PI margins were higher, ranging

from 6.0–11.0 mm, 5.5–10.0 mm, and 3.0–9.5 mm. These margins are higher than recently reported margins for adequate tumor coverage [13]. The reported margins are not directly comparable, as the previous report aggregated histology-MRI boundary distances across all tumors and observers. However, there are two additional key differences. First, our analysis was performed in 3D whereas the previous report looked only at a single slice of histology (the slice with the maximum cross-sectional area). Second, the patients included in the margin estimation in the previous study had higher patient Gleason scores (2 G3+3, 1 G3+4, 3 G4+3, 2 G4+4 and 2 G4+5) than patients included in our margin estimation (2 G3+3, 18 G3+4, 1 G4+3). Less aggressive cancers are harder to detect [21], and may also be harder to delineate reliably.

Three-sequence CTVs required the smallest margins, and, with few exceptions, CTVs defined based on the composite GTVs from two sequences required smaller margins than those defined based on a single sequence. This suggests there may be a trade-off between the size of the required margins and the amount of target volume contouring required. Alternatively, there may be potential for tools that facilitate contouring on fused, co-registered mpMRI to reduce required margins.

Interpreting the need for different margins for different CTV types is complex. As a simplified illustration, suppose that GTVs on sequence A were consistently 2 mm isotropically outside GTVs on sequence B. A 6-mm-margin CTV on sequence A would be identical to an 8-mm-margin CTV on sequence B, and contouring on sequence A with a 2 mm smaller margin would confer no advantage. The true situation is more complex, as GTVs on different sequences are not necessarily concentric. Notably, however, the composite GTVs (defined as the union of single-sequence GTVs) may be larger, in

general, than the single-sequence GTVs, which may account, in part, for the smaller required margins. For the volumetric analysis, this interaction is mitigated by using PI margins. Although composite CTVs with PI margins have smaller treatment volumes and cover a smaller proportion of the prostate for some observers, this is not consistent across observers. Thus, we are unable to conclude that contouring on multiple sequences and applying smaller margins reduces the CTV volumes required to get adequate histological coverage with high probability. In particular, where the margins differed by  $< 2$  mm, the single-sequence CTVs were smaller, but where the margins differed by  $\geq 2$  mm, the three-sequence CTVs were smaller. There may, however, be other aspects of treatment planning where smaller required margins are advantageous. An investigation of the types of tissue (prostate parenchyma vs. organ-at-risk) covered by these CTVs may clarify whether there is a clinical need to contour GTVs on all three sequences.

The median volumes of CTVs defined using PI margins ranged from 4–30 cm<sup>3</sup>, for CTVs with PI margins for all cancer, and from 2–16 cm<sup>3</sup> for CTVs with PI margins for high-grade cancer, depending on the residual cancer area, the observer and CTV type. Assessment of whether these CTVs could be feasibly treated depends on many factors including the treatment modality, and the locations of the CTVs relative to organs at risks. However, focal laser ablations up to 4.5 cm<sup>3</sup> using 2 fibers have been reported [18] and focal radiation boosting typically targets volumes  $<10$  cm<sup>3</sup> [6]. Based on volume considerations alone, 50–80% of the CTVs with PI margins for 0–10% residual high-grade cancer area could be targeted with a modality capable of treating targets  $<10$  cm<sup>3</sup>, and 9–51% could be targeted with a modality capable of treating targets  $<4.5$  cm<sup>3</sup>. For the larger CTVs with PI margins for all cancer, 20–63% could be treated with a modality

capable of treating targets  $<10\text{ cm}^3$ , but only 0–22% could be targeted with a modality capable of treating targets  $<4.5\text{ cm}^3$ . Although there were outliers with CTV volumes substantially larger than these treatment volumes, they corresponded to large foci with extraprostatic extension that would likely not be candidates for tissue sparing focal therapies such as focal laser ablation.

Considering the trade-offs between time for delineation, required margin and treatment constraints, one could envision multiple workflows for target volume delineation in focal therapy planning incorporating PI margins. In one workflow, one could initially attempt to delineate GTVs on all sequences. If all sequences could be delineated, smaller margins for three-sequence composite GTVs could be used; otherwise, the larger margins required for two-sequence composite or single-sequence GTVs could be used. In an alternative workflow, one could initially delineate a single sequence and apply larger margins. If the resulting CTV was feasible to deliver and met guidelines for sparing organs at risk, it could be delivered; otherwise, additional sequences could be delineated and smaller margins used, potentially resulting in target volumes that could more feasibly be delivered or that better spared organs at risk.

The volumes of CTVs with PI margins were large compared to the underlying histological cancer (median linear expansion factor: 2.7–9.3). Some of this expansion can be attributed to the CTV covering stromal tissue between cancerous glands, particularly in diffuse cancers. However, the expansion also includes non-cancerous regions within the isotropic expansions used in this study. This suggests potential value in evaluating non-isotropic margin expansions, such as different radial and tangential expansions or expansion constrained to remain within the prostatic zone containing the GTV.



Some of the unmodeled variability in required margins may be due to registration error. Misalignment of a focus on the histological reference standard with respect to the GTVs could result in an increase or in a decrease in required margins for specified coverage, depending on their spatial configuration, with an overall effect of increasing variability and introducing a positive bias (smaller, in general, than the quantified registration error). Since high confidence prediction intervals are sensitive to extreme values, the additional variability would introduce a positive bias to the PI upper bound as well.

The conclusions of this study should be considered in the context of its strengths and limitations. Estimates of residual and total cancer area could only be made within the histology sections, representing 3–5 thin surfaces within the mid-gland. If the CTV coverage characteristics in the apex and base differ from those in the mid-gland, our conclusions may not generalize to the entire prostate. Expansion of this analysis to include the apex and base histology is ongoing. The sparsity of the sampling, however, is a limitation of the clinical workflow, and may contribute to the observed variability. With 4 observers and high interobserver variability, we were unable to draw conclusions about the expected performance of an unseen observer. Reporting the observer characteristics separately, however, did elucidate this variability and led to identifying variability as a key challenge. The sample size in each analysis, limited to each observer's GTVs separately, combined with the skewed distribution of residual areas, resulted in prediction interval boundaries that were sensitive to outliers. This was mitigated in our analysis by using lower confidence levels. Due to the fine scale of the cancer delineations on histology in our study, the planimetric volume estimates do not include stromal tissue that

would be included in more typical coarse scale delineations; thus, our volume estimations may be equivalent to larger histological tumor volumes reported in the literature.

In conclusion, the delineation of GTVs on mpMRI has high interobserver variability in the optimal sequence(s) to use for delineation and in the expansion margins needed to achieve a 0–10% residual cancer area with high probability. These margins were smallest for targets defined by the union of delineations on T2W, DCE and ADC images: 4.5–9.0 mm, 1.5–7.5 mm and 1.5–4.5 mm for CTVs intended to leave 0%, 5% and 10% residual high-grade (Gleason score  $\geq 7$ ) cancer area, respectively, and 6.0–11.0 mm, 5.5–10.0 mm, and 3.0–9.5 mm for CTVs intended to leave 0%, 5% and 10% residual cancer of any grade. However, the smaller margins for the composite three-sequence GTVs did not result in consistently smaller CTVs, compared with the CTVs resulting from the larger margins around single-sequence GTVs. The CTVs resulting from these margins had volumes that may preclude focal treatment with some volume-constrained modalities for many patients. In order to facilitate margin recommendations that yield CTVs that can be feasibly targeted in most patients, reduction of interobserver and intraobserver variability in delineation should be a priority.

## 7.5 References

1. U. Lindner, R. A. Weersink, M. A. Haider, M. R. Gertner, S. R. Davidson, M. Atri, B. C. Wilson, A. Fenster and J. Trachtenberg, "Image guided photothermal focal therapy for localized prostate cancer: phase I trial," *Journal of Urology* **182**, 1371–1377 (2009).
2. E. Lecornet, C. Moore, H. U. Ahmed and M. Emberton, "Focal therapy for prostate cancer: fact or fiction?," *Urologic Oncology* **28**, 550–556 (2010).
3. H. U. Ahmed, R. G. Hindley, L. Dickinson, A. Freeman, A. P. Kirkham, M. Sahu, R. Scott, C. Allen, J. Van der Meulen and M. Emberton, "Focal therapy for localised

- unifocal and multifocal prostate cancer: a prospective development study," *Lancet Oncology* **13**, 622–632 (2012).
4. G. Bozzini, P. Colin, P. Nevoux, A. Villers, S. Mordon and N. Betrouni, "Focal therapy of prostate cancer: energies and procedures," *Urologic Oncology* **31**, 155–167 (2013).
  5. I. M. Lips, U. A. van der Heide, K. Haustermans, E. N. van Lin, F. Pos, S. P. Franken, A. N. Kotte, C. H. van Gils and M. van Vulpen, "Single blind randomized phase III trial to investigate the benefit of a focal lesion ablative microboost in prostate cancer (FLAME-trial): study protocol for a randomized controlled trial," *Trials* **12**, 255 (2011).
  6. G. Bauman, M. Haider, U. A. Van der Heide and C. Menard, "Boosting imaging defined dominant prostatic tumors: a systematic review," *Radiotherapy and Oncology* **107**, 274–281 (2013).
  7. A. Sciarra, J. Barentsz, A. Bjartell, J. Eastham, H. Hricak, V. Panebianco and J. A. Witjes, "Advances in magnetic resonance imaging: how they are changing the management of prostate cancer," *European Urology* **59**, 962–977 (2011).
  8. J. O. Barentsz, J. Richenberg, R. Clements, P. Choyke, S. Verma, G. Villeirs, O. Rouviere, V. Logager and J. J. Futterer, "ESUR prostate MR guidelines 2012," *European Radiology* **22**, 746–757 (2012).
  9. M. R. Engelbrecht, J. O. Barentsz, G. J. Jager, M. van der Graaf, A. Heerschap, J. P. Sedelaar, R. G. Aarnink and J. J. de la Rosette, "Prostate cancer staging using imaging," *BJU International* **86 Suppl 1**, 123–134 (2000).
  10. S. Isebaert, L. Van den Bergh, K. Haustermans, S. Joniau, E. Lerut, L. De Wever, F. De Keyzer, T. Budiharto, P. Slagmolen, H. Van Poppel and R. Oyen, "Multiparametric MRI for prostate cancer localization in correlation to whole-mount histopathology," *Journal of Magnetic Resonance Imaging* **37**, 1392–1401 (2013).
  11. M. A. Haider, T. H. van der Kwast, J. Tanguay, A. J. Evans, A. T. Hashmi, G. Lockwood and J. Trachtenberg, "Combined T2-weighted and diffusion-weighted MRI for localization of prostate cancer," *AJR. American Journal of Roentgenology* **189**, 323–328 (2007).
  12. J. J. Futterer, T. W. Scheenen, S. W. Heijmink, H. J. Huisman, C. A. Hulsbergen-Van de Kaa, J. A. Witjes, A. Heerschap and J. O. Barentsz, "Standardized threshold approach using three-dimensional proton magnetic resonance spectroscopic imaging in prostate cancer localization of the entire prostate," *Investigative Radiology* **42**, 116–122 (2007).

13. M. Anwar, A. C. Westphalen, A. J. Jung, S. M. Noworolski, J. P. Simko, J. Kurhanewicz, M. Roach, 3rd, P. R. Carroll and F. V. Coakley, "Role of endorectal MR imaging and MR spectroscopic imaging in defining treatable intraprostatic tumor foci in prostate cancer: quantitative analysis of imaging contour compared to whole-mount histopathology," *Radiotherapy and Oncology* **110**, 303–308 (2014).
14. A. D. Ward, C. Crukley, C. A. McKenzie, J. Montreuil, E. Gibson, C. Romagnoli, J. A. Gomez, M. Moussa, J. Chin and G. Bauman, "Prostate: registration of digital histopathologic images to in vivo MR images acquired by using endorectal receive coil," *Radiology* **263**, 856–864 (2012).
15. E. Gibson, C. Crukley, M. Gaed, J. A. Gómez, M. Moussa, J. L. Chin, G. S. Bauman, A. Fenster and A. D. Ward, "Registration of prostate histology images to ex vivo MR images via strand-shaped fiducials," *Journal of Magnetic Resonance Imaging* **36**, 1402–1412 (2012).
16. F. L. Bookstein, "Principal warps: thin-plate splines and the decomposition of deformations," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11**, 567–585 (1989).
17. T. M. Mayhew and L.-M. Cruz Orive, "Caveat on the use of the Delesse principle of areal analysis for estimating component volume densities," *Journal of Microscopy* **102**, 195–207 (1974).
18. U. Lindner, N. Lawrentschuk, R. A. Weersink, S. R. H. Davidson, O. Raz, E. Hlasny, D. L. Langer, M. R. Gertner, T. Van der Kwast and M. A. Haider, "Focal laser ablation for prostate cancer followed by radical prostatectomy: validation of focal therapy and imaging accuracy," *European Urology* **57**, 1111–1114 (2010).
19. P. Hall and A. Rieck, "Improving coverage accuracy of nonparametric prediction intervals," *Journal of the Royal Statistical Society Series B-Statistical Methodology* **63**, 717–725 (2001).
20. B. G. Muller, J. J. Fütterer, R. T. Gupta, A. Katz, A. Kirkham, J. Kurhanewicz, J. W. Moul, P. A. Pinto, A. R. Rastinehad, C. Robertson, J. de la Rosette, R. Sanchez-Salas, J. S. Jones, O. Ukimura, S. Verma, H. Wijkstra and M. Marberger, "The role of magnetic resonance imaging (MRI) in focal therapy for prostate cancer: recommendations from a consensus panel," *BJU International* **113**, 218–227 (2014).
21. J. Thompson, N. Lawrentschuk, M. Frydenberg, L. Thompson and P. Stricker, "The role of magnetic resonance imaging in the diagnosis and management of prostate cancer," *BJU International* **112 Suppl 2**, 6–20 (2013).

## Chapter 8.

# **Contributions of the thesis, applications and suggestions for future work**

### 8.1 Contributions of the thesis

This thesis contributes advances in methods, concepts and knowledge in several areas related to the four research questions posed in the introduction:

- How does registration error impact the statistical power of imaging validation studies?
- What is the performance (in terms of reconstruction error and robustness) of an extrinsic fiducial-based histology reconstruction method, and how does this compare to alternative approaches based on image-guided slicing and intensity-based image registration?
- What is the variability of lesion scoring (using consensus-panel-recommended PI-RADS guidelines [1]) and contouring on mpMRI?
- What are appropriate margins around observers' target volumes to achieve histological coverage of detected cancers?

The following sections address contributions related to each question in turn.

### 8.1.1 Research question 1: How does registration error impact the statistical power of imaging validation studies?

**Chapter 2 introduced a new approach to evaluating registration error, by considering registration in the context of imaging validation studies seeking to make an inference, and evaluating registration error by its impact on the statistical power of a study.** Registration techniques can be evaluated based on their registration error; however, criteria for whether a registration is sufficiently accurate are application-dependent [2] and depend on the success criteria for the application. In the context of studies that aim to make an inference, one success criterion is whether the study outcome is a true positive, correctly rejecting the null hypothesis. Statistical power is the probability of rejecting the null hypothesis when it is false. By evaluating registration error with respect to the statistical power of a study incorporating the registration, we can express a success criterion for registration in terms of an established success criterion for the application. *Thus, this thesis has advanced knowledge in registration evaluation by contributing a statistically justifiable approach for evaluating registration accuracy in studies that culminate in a statistical test.*

**Chapter 2 also derived and evaluated new power calculation formulae that relate sample size, the minimum detectable difference and registration error in one class of imaging validation study.** In addition to proposing an approach for evaluating registration accuracy, we reduced the concept to practice, by deriving statistical power calculation formulae for studies aiming to detect a significant signal difference between foreground and background regions defined using a registered reference standard, and applying the formulae to a case study. The derived formulae enable study designers to

answer three key questions related to study design. (1) Given a specified sample size and a specified effect size to detect, what is the maximum acceptable mean error for a registration method in this study? (2) Given a quantified registration error and a specified effect size to detect, what is the minimum sample size to accrue? (3) Given a specified sample size and a quantified registration error, how small an effect can be detected? *Thus, this thesis advanced knowledge in registration evaluation by contributing a statistically rigorous way to answer key questions in the design of one class of imaging validation studies.*

8.1.2 Research question 2: What is the performance (in terms of reconstruction error and robustness) of an extrinsic fiducial-based histology reconstruction method, and how does this compare to alternative approaches based on image-guided slicing and intensity-based image registration?

**Chapter 3 characterized spatial relationships of histology sections relative to the tissue slices from which they were cut, and identified that an isotropic scaling deformation accounts for the majority of the in-plane deformation, and that the *front face assumption*— that histology sections correspond to the front face of tissue sections from which they were cut — introduces seemingly small but potentially impactful reconstruction errors.** The process of acquiring histology sections involves chemically processing tissue slices, embedding them in paraffin, mounting them onto a microtome, cutting through the tissue slices until full cross-sections are exposed, cutting 4- $\mu$ m-thick sections of tissue, floating the sections on a water bath and mounting them to slides. These processes could potentially result in substantial removal of tissue from the

tissue slice and potentially induce substantial deformation. By comparing intrinsic landmarks on histology and MR images of the tissue slices, we identified where the histology sections were cut from within the tissue slices and quantified the deformation. Since previous histology reconstruction methods have used a wide range of transformations (rigid [3], similarity [4, 5], affine [6, 7], or thin-plate-spline [8, 9]) to account for deformation during processing, we assessed how well different transformations compensated for the in-plane deformation. Since many previous histology methods [4, 5, 7, 9-12] have made the front face assumption, we assessed its strength by measuring the depth and orientation of the section within the tissue slice, and estimated the impact on reconstruction error of making the assumption. The incremental error due to different in-plane deformations and the front face assumption was seemingly small (1.4 mm); however, analyzing the resulting reconstruction errors in the context of a hypothetical imaging validation study of  $0.2 \text{ cm}^3$  tumors (using the approach presented in Chapter 2), suggested that the assumption could result in a 1.5 fold increase in sample size. *Thus, the thesis advanced knowledge in histology reconstruction by quantifying the strength and impact of assumptions that are commonly made in histology reconstruction.*

**Chapter 4 presented a fiducial-based method for 3D reconstruction of histology images that was not disruptive to the clinical diagnosis, could be implemented with little specialized equipment (beyond *ex vivo* MR imaging), and compensates for variability described in Chapter 3.** As described in Section 1.2.3.3, there have been three main approaches to histology reconstruction. Image-guided-slicing-based methods that constrain the orientation of gross slicing have the potential to disrupt standard clinical slicing rules and typically assume histology corresponds to the front face



of tissue slices. Methods based on additional imaging of the specimen allow more flexible slicing, but also typically assume histology corresponds to the front face of tissue slices. Image-registration-based methods have the potential to minimize disruption to the clinical workflow and mitigate the variability introduced by histology sectioning; however, they rely on image features that may be disrupted by the presence of cancer or other abnormalities. The approach presented in this chapter leverages extrinsic fiducials to reconstruct histology to the context of an *ex vivo* MR image without making assumptions about the position of histology relative to their corresponding tissue slices, thereby (1) allowing flexibility in the gross slicing to suit pathologists' requirements, and (2) mitigating the variability introduced by histology sectioning. The use of extrinsic fiducials instead of intrinsic image information allows the method to be robust to imaging characteristics of the prostate tissue. *Thus, this thesis advanced knowledge in histology reconstruction by presenting a new method for reconstruction that is less disruptive and more robust than existing approaches.*

**Chapter 4 also presented an approach to identify pairs of homologous point fiducials on imaging to support the measurement of the TRE of histology reconstruction.** Identification of homologous point landmarks on histology and MR imaging has been identified as a challenging problem, and a previous histology reconstruction method used this as a justification to avoid measuring TRE for their method [9]. Many other reconstruction methods also reported surrogates for registration error instead of TRE (overlap of the prostate gland [7, 8], prostate boundary distances [13, 14], qualitative assessment of prostate cancer tumor alignment [11, 15], post-optimization similarity metric values [15], or image intensity properties [16]). In this

chapter, we depicted and described illustrative homologous point landmarks that could be identified on histology and *ex vivo* MR, and validated the reconstruction using 184 such landmark pairs across 34 histology sections. *Thus, this thesis advanced knowledge in histology reconstruction by describing homologous point landmarks that could be used to measure TRE between histology and ex vivo MRI.*

**Chapter 4 also presented a direct comparison of the fiducial-based method to an alternative approach based on image-guided slicing, showing that the fiducial-based algorithm had a lower mean error.** Many existing approaches for histology reconstruction guided the slicing of the prostate specimen in a controlled way to facilitate reconstruction. In this chapter, we evaluated this approach by directly comparing it to the fiducial-based method presented in this chapter in terms of the TRE. To the best of my knowledge, this represented the first direct comparison of two substantially different approaches to histology reconstruction. We demonstrated that the fiducial-based method had a lower TRE than the image-guided-slicing-based method. This was consistent with the observations in Chapter 3 that assuming histology sections correspond to the front faces of tissue blocks, as was assumed in the image-guided-slicing-based approach, introduces a small incremental error. *Thus, this thesis advanced knowledge in histology reconstruction by demonstrating the higher accuracy of the presented fiducial-based approach over the image-guided-slicing-based approach.*

**Chapter 5 presented a direct comparison of the accuracy and robustness of the fiducial-based method described in Chapter 4 to an alternative image-registration-based approach based on maximizing normalized mutual information, and identified that the image-registration-based approach was not a robust solution**

**to histology reconstruction.** Decades of research in image registration have produced flexible tools, such as the normalized-mutual-information similarity metric [17], that have been widely applied to the registration of images from different modalities. One could imagine that a flexible image registration approach maximizing normalized mutual information between histology and *ex vivo* MR images could trivially yield an accurate and robust reconstruction without the need for fiducials. In this chapter, we evaluated this approach by directly comparing it to the fiducial-based method. We compared the robustness to initializations perturbed away from the reference position defined by manually identified landmarks, and the accuracy after a practically achievable initialization. While the fiducial-based approach had negligible sensitivity to initialization, the image-registration-based approach was sensitive to initialization and further analysis demonstrated that the similarity metric was less optimal at the reference point than at convergence points away from the reference point (suggesting that the similarity metric is a contributor to the lack of robustness). *Thus, the thesis advanced knowledge in histology reconstruction by demonstrating that a non-customized image-registration-based approach using the normalized mutual information did not yield a reconstruction technique robust to initialization error.* This does not, however, preclude the possibility that image-registration-based methods customized to the task could result in robust histology reconstructions.

8.1.3 Research question 3: What is the variability of lesion scoring (using consensus-panel-recommended PI-RADS guidelines [1]) and contouring on mpMRI?

**Chapter 6 investigated lesion scoring according to the established PI-RADS structured reporting guidelines for mpMRI. We found a correlation between PI-RADS overall likelihood score and Gleason grade, and corroborated recent reports of fair to moderate agreement in PI-RADS scores and a high (85% in our analysis) PPV for a PI-RADS overall likelihood score of 5.** Consensus guidelines for prostate cancer detection and reporting on mpMRI have been published recently, and characterizing the performance of observers applying these guidelines may contribute to the evaluation and optimal use of the guidelines. In this chapter, we looked at the level of agreement between observers, and the relationship between lesion scoring and histological cancer. Our analysis of interobserver score agreement yielded a kappa of 0.3 in the PI-RADS overall likelihood score, consistent with recent evaluations of the guidelines [18, 19] noting fair to moderate agreement. The high positive predictive value of a PI-RADS overall likelihood score of 5 was also consistent with a recent report [20], although their observed PPVs were somewhat higher than ours for overall likelihood scores of 3–5. Although these are not new findings, the contribution of findings from multiple centers is important to the evaluation of consensus guidelines. The correlation between the PI-RADS overall likelihood score and prostatectomy Gleason grade on whole-mount histopathology has not, to the best of my knowledge, been previously identified, and corroborates a previous finding of correlation between the sum of PI-RADS sequence-specific scores and biopsy Gleason grade [21]. *Thus, the thesis*

*advanced the knowledge of a lesion scoring under a consensus-recommended structured reporting system by presenting an unreported correlation between the PI-RADS overall likelihood scores and Gleason grade, and by accumulating evidence regarding the interobserver variability and predictive value of PI-RADS overall likelihood scores.*

**Chapter 6 also investigated lesion delineation on multi-parametric MRI and characterized the variability in delineation between observers and between different mpMRI sequences.** Delineations of cancer target volumes may support delivery of lesion focused therapies; however, the accuracy and variability of such delineations are not well characterized. In this study, four observers delineated lesions on different mpMRI sequences, enabling the measurement of variability in lesion delineation between observers, and between different mpMRI sequences. Although no clear advantage of any one sequence emerged, our descriptive measurements of contour agreement quantified the high variability between observers and between mpMRI sequences, a finding which had not been previously reported. *Thus, this thesis advanced knowledge in lesion delineation on mpMRI by characterizing the interobserver and intersequence variability in contouring.*

8.1.4 Research question 4: What are appropriate margins around observers' target volumes to achieve histological coverage of detected cancers?

**Chapter 7 evaluated the histological coverage of clinical target volumes defined on multi-parametric MR images and (1) identified a range of expansion margins that yielded 90%, 95% and 100% histological coverage with high probability in study observers; (2) identified that composing delineations on multiple modalities reduced the required margins, but did not consistently reduce**

**target volumes; and (3) identified intraobserver and interobserver variability as a priority for achieving histological coverage with high probability with target volumes small enough to feasibly treat.** In targeted treatments, a common way to address uncertainty in target volumes is to add margins to the GTVs, but the appropriate margins for cancer GTVs defined on mpMRI are not yet established. A recent study [22] reported that a 5 mm margin around GTVs delineated on mpMRI was sufficient to cover 95% of histologically defined cancer, and 8 mm was sufficient to cover 100%. However, this analysis aggregated histology-MRI contour distances across all observer and tumors, and was performed in 2D, ignoring the extent of cancer in the inferior-superior direction and possibly underestimating the required margins. In this chapter, we evaluated in 3D the histological coverage of GTVs defined on mpMRI at a range of margins, separating our analysis by sequence(s) used to define the GTVs and by observer. The key findings, described above, do not result in a clear recommendation on margins that should be used for focal therapy trials; however, they provide guidance on ranges of margins that could be considered for evaluation and elucidated the presence of substantial variability that may challenge the evaluation of lesion-focused therapies using target volume delineation on mpMRI. *Thus, this chapter advanced knowledge in lesion delineation by reporting margin sizes to achieve 90–100% histological coverage with high confidence for four observers and elucidating the interobserver variability in margin sizes and variability in which sequences are optimal for contouring.*

## 8.2 Applications and future directions

The methods developed in this thesis support research applications in several directions. In the following sections, a number of these potential applications will be discussed, and remaining gaps in knowledge will be described.

### 8.2.1 Evaluation and refinement of existing and novel imaging

mpMRI has shown potential to support many aspects of the clinical workflow; however, detection of central gland tumors, assessing cancer aggressiveness and planning treatment remain challenging. Thus, prostate cancer imaging remains an active research direction. New imaging technologies are being developed, such as endogenous sodium imaging [23, 24] and hyperpolarized carbon imaging [25, 26]. New derived images computed from existing imaging are being developed, such as high b-value DW images [27], diffusion compartment model imaging [28] and alternative DCE parametric maps [29]. Computer-aided detection and delineation tools are also being developed [30, 31]. All of these methods need to be evaluated to assess their clinical utility. Because of the slow natural history of prostate cancer, evaluating novel imaging technology with respect to outcome may introduce unacceptable delays. Initial evaluation of imaging technology with respect to a surrogate endpoint, such as agreement with a co-registered histological reference standard, has the potential to identify promising imaging technology early for more thorough validation in clinical trials.

This research direction can be pursued already using techniques presented in my thesis, with adaptations to register the images of interest to the *in vivo* or *ex vivo* MRI. Preliminary investigations of <sup>18</sup>F-fluorocholine PET/CT and PET/MR [32], endogenous sodium imaging [33], radiofrequency time series ultrasound [34] and derived images

computed from DCE MRI [29] using techniques presented in my thesis are currently ongoing, and a multi-centre evaluation of additional imaging technologies incorporating these techniques is scheduled to begin in the near future. However, this research direction could be facilitated substantially by the automation of an accurate registration from *ex vivo* to *in vivo* MRI with appropriately quantified registration error. Nir et al. [35] presented an approach that uses magnetic resonance elastography; however, this technology is not widely available and could require substantial alteration of clinical imaging to implement. Surface-based techniques for ultrasound to MRI registration in fused TRUS-MRI biopsy systems [36, 37] may be adaptable to this problem as well.

Histology-based imaging evaluation has the potential to rapidly identify imaging technologies with the potential for accurate diagnosis, patient stratification and treatment delivery without invasive procedures and hasten their clinical evaluation and adoption.

### 8.2.2 Use in clinical training

Accuracy and variability in prostate cancer image interpretation is related to reader experience. Reading and reporting MRI images with pathological feedback has been identified as an important component of developing and maintaining expertise in prostate cancer image interpretation [38, 39]. Educational curricula for delineation (of whole prostates) have improved contouring variability, and curricula for interpretation involving (unregistered) pathological feedback have improved accuracy and variability in prostate cancer detection [40]. The development of an interactive training platform that presents imaging alongside a co-registered histological reference standard may support clinicians in developing and maintaining expertise in image interpretation and delineation.



The techniques presented in my thesis enable the accurate spatial alignment of histology with mpMRI. Presenting these data, which are typically not available to radiologists, in an interactive training platform may support training. However, effective presentation of this material to maximize learning may pose challenges, and substantial research and development may be needed to achieve the educational potential of these data. Educational cognitive psychology may provide direction in addressing some of these challenges. For example, high cognitive load due to complex interactions between information from multiple sources may interfere with learning and may need to be addressed within the training tools [41]. Embedding training tools within the context that the knowledge will be applied (e.g. in a radiologist's clinical picture archiving and communication system [PACS]) may improve learning [42]. Integrating such principles into a training tool may facilitate improvements in clinician performance. Furthermore, although ongoing training with pathological feedback has been suggested in consensus findings [38, 39] and unregistered pathology feedback has improved detection rates [40], it has not yet been demonstrated that training with co-registered pathological feedback will result in better patient outcomes; this hypothesis needs to be tested to demonstrate the clinical value of such an educational platform.

A fused imaging-histology training platform that supports clinicians in developing and maintaining expertise in prostate cancer image interpretation could improve the accuracy and reduce the intra- and interobserver variability in imaging interpretation and delineation. This could enable more consistent clinical use of imaging, and more precise characterization of the diagnostic, prognostic and therapeutic value of imaging. It could support imaging applications, such as focal therapy planning, where variability remains a

challenge. Such a tool could also enable the wide-spread adoption of prostate cancer imaging beyond academic hospitals by enabling clinicians with smaller prostate imaging case loads to develop and maintain expertise.

### 8.2.3 Development and evaluation of consensus guidelines for contouring

Consensus guidelines have been established for many stages of the prostate cancer clinical workflow, including prostate cancer detection [1], focal therapy [39] and pathology interpretation [43]. Although mpMRI has been recommended for focal therapy planning [39], there are currently no clinical guidelines for how to contour prostate cancer lesions on mpMRI for lesion-focused therapies.

This thesis could support the development of consensus guidelines directly by providing evidence regarding lesion delineation variability and predicted histological coverage for target volume margins, and indirectly by supporting evaluation of lesion delineation in future studies. One could also imagine a guideline consensus panel process that involved (1) the interpretation and delineation of mpMRI by panel members, (2) comparison to a co-registered histological reference, and (3) consensus discussion about the results of this comparison. However, consensus guidelines are typically developed based on multiple sources of information and considering a substantial body of clinical experience and literature. Further studies, and in particular large multi-centre studies, of lesion delineation are required to establish the evidence to support consensus.

Consensus guidelines for lesion delineation on prostate cancer imaging could improve consistency in lesion delineation, support the meaningful comparisons of outcomes across studies of targeted biopsy and lesion-focused therapies, and facilitate the translation of such procedures to widespread clinical use.

## 8.3 Remaining gaps in histology-imaging fusion methods

This thesis has made contributions towards addressing the 4 criteria for histology-imaging fusion methods described in Section 1.2.3.3 (non-disruptive, spatially accurate, robust, and widely implementable). There remains room for further improvement in histology-imaging fusion methods. In particular, there is no appropriately validated existing method that can register histology collected following standard clinical pathology protocols to mpMRI images collected using standard radiology protocols, without collection of intermediate data. Such a method could potentially enable large studies leveraging retrospective imaging and pathology data, enable studies of older data with known patient outcome, and enable the incorporation of histology-imaging fusion into routine clinical reporting. Appropriate prospective evaluation of the accuracy and variability of such a method is critical because in retrospective or ongoing use, there may be insufficient information to confirm the accuracy of a fusion. Ideally, the evaluation would include the estimation of the distribution of target registration errors measured using homologous point landmarks across a data set large enough to allow the construction of high-confidence prediction intervals for unseen cases. The evaluation should also include an assessment of robustness including the variability in accuracy, and the rate and causes of failed registrations. Although these criteria may be challenging to meet, further progress towards these ideals would be valuable and meeting them could potentially transform the use of histology-imaging fusion in prostate cancer imaging.

## 8.4 References

1. J. O. Barentsz, J. Richenberg, R. Clements, P. Choyke, S. Verma, G. Villeirs, O. Rouviere, V. Logager and J. J. Futterer, "ESUR prostate MR guidelines 2012," *European Radiology* **22**, 746–757 (2012).

2. M. H. Loew and C. E. Rodriguez-Carranza, "Technical issues in multimodality medical image registration," in *Proceedings of IEEE Computer-Based Medical Systems*, (1998), pp. 2–7.
3. C. Hughes, O. Rouvière, F. Mege-Lechevallier, R. Souchon and R. Prost, "Robust alignment of prostate histology slices with quantified accuracy," *IEEE Transactions on Biomedical Engineering* **60**, 281–291 (2012).
4. G. Groenendaal, M. R. Moman, J. G. Korporaal, P. J. van Diest, M. van Vulpen, M. E. Philippens and U. A. van der Heide, "Validation of functional imaging with pathology for tumor delineation in the prostate," *Radiotherapy and Oncology* **94**, 145–150 (2010).
5. B. Castaneda, K. Hoyt, M. Zhang, D. Pasternack, L. Baxter, P. Nigwekar, A. di Sant'Agnese, J. Joseph, J. Strang, D. J. Rubens and K. J. Parker, "P1C-9 prostate cancer detection based on three dimensional sonoelastography," in *Proceedings of IEEE Ultrasonics Symposium*, (New York, USA, 2007), pp. 1353–1356.
6. E. Gibson, C. Crukley, M. Gaed, J. A. Gómez, M. Moussa, J. L. Chin, G. S. Bauman, A. Fenster and A. D. Ward, "Registration of prostate histology images to ex vivo MR images via strand-shaped fiducials," *Journal of Magnetic Resonance Imaging* **36**, 1402–1412 (2012).
7. L. S. Taylor, B. C. Porter, G. Nadasdy, P. A. di Sant'Agnese, D. Pasternack, Z. Wu, R. B. Baggs, D. J. Rubens and K. J. Parker, "Three-dimensional registration of prostate images from histology and ultrasound," *Ultrasound in Medicine and Biology* **30**, 161–168 (2004).
8. Y. Zhan, Y. Ou, M. Feldman, J. Tomaszewski, C. Davatzikos and D. Shen, "Registering histologic and MR images of prostate for image-based cancer detection," *Academic Radiology* **14**, 1367–1381 (2007).
9. H. Park, M. R. Piert, A. Khan, R. Shah, H. Hussain, J. Siddiqui, T. L. Chenevert and C. R. Meyer, "Registration methodology for histological sections and in vivo imaging of human prostate," *Academic Radiology* **15**, 1027–1039 (2008).
10. A. D. Ward, C. Crukley, C. A. McKenzie, J. Montreuil, E. Gibson, C. Romagnoli, J. A. Gómez, M. Moussa, J. Chin, G. Bauman and A. Fenster, "Prostate: registration of digital histopathologic images to in vivo MR images acquired by using endorectal receive coil," *Radiology* **263**, 856–864 (2012).
11. V. Shah, T. Pohida, B. Turkbey, H. Mani, M. Merino, P. A. Pinto, P. Choyke and M. Bernardo, "A method for correlating in vivo prostate magnetic resonance imaging and histopathology using individualized magnetic resonance-based molds," *Review of Scientific Instruments* **80**, 104301–104306 (2009).

12. A. S. Jackson, S. A. Reinsberg, S. A. Sohaib, E. M. Charles-Edwards, S. Jhavar, T. J. Christmas, A. C. Thompson, M. J. Bailey, C. M. Corbishley, C. Fisher, M. O. Leach and D. P. Dearnaley, "Dynamic contrast-enhanced MRI for prostate cancer localization," *British Journal of Radiology* **82**, 148–156 (2009).
13. S. Bart, P. Mozer, P. Hemar, G. Lenaour, E. Comperat, R. Renaud-Penna, E. Chartier-Kastler and J. Troccaz, "MRI-histology registration in prostate cancer," in *Proceedings of Surgetica*, (2005).
14. P. Patel, J. Chappelow, J. Tomaszewski, M. D. Feldman, M. Rosen, N. Shih and A. Madabhushi, "Spatially weighted mutual information (SWMI) for registration of digitally reconstructed ex vivo whole mount histology and in vivo prostate MRI," in *Proceedings of IEEE Engineering in Medicine & Biology Society*, (2011), pp. 6269–6272.
15. J. Chappelow, B. N. Bloch, N. Rofsky, E. Genega, R. Lenkinski, W. DeWolf, S. Viswanath and A. Madabhushi, "COLLINARUS: collection of image-derived non-linear attributes for registration using splines," in *Proceedings of SPIE Medical Imaging*, (2009), pp. 72592N.
16. J. Chappelow, A. Madabhushi, M. Rosen, J. Tomaszewski and M. Feldman, "Multimodal image registration of ex vivo 4 Tesla MRI with whole mount histology for prostate cancer detection," in *Proceedings of SPIE Medical Imaging, Vol. 6512* (2007), pp. 65121S
17. F. Maes, A. Collignon, D. Vandermeulen, G. Marchal and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Transactions on Medical Imaging* **16**, 187–198 (1997).
18. A. B. Rosenkrantz, S. Kim, R. P. Lim, N. Hindman, F. M. Deng, J. S. Babb and S. S. Taneja, "Prostate cancer localization using multiparametric MR imaging: comparison of Prostate Imaging Reporting and Data System (PI-RADS) and Likert scales," *Radiology* **269**, 482–492 (2013).
19. L. Reisaeter, M. Biermann, A. E., Y. Nygaard, C. Beisland, L. A. Akslen, O. Halvorsen, K. Gravdal, J. Futterer and J. Rorvik, "Reliability and diagnostic performance using PIRADS in preoperative multiparametric MRI of prostate cancer," in *Proceedings of Radiological Society of North America*, (Chicago, USA, 2012).
20. J. E. Thompson, D. Moses, R. Shnier, P. Brenner, W. Delprado, L. Ponsky, M. Pulbrook, M. Böhm, A.-M. Haynes, A. Hayen and P. D. Stricker, "Multiparametric Magnetic Resonance Imaging Guided Diagnostic Biopsy Detects Significant Prostate Cancer and could Reduce Unnecessary Biopsies and Over Detection: A Prospective Study," *Journal of Urology* **192**, 67–74 (2014).

21. D. Junker, G. Schäfer, M. Edlinger, C. Kremser, J. Bektic, W. Horninger, W. Jaschke and F. Aigner, "Evaluation of the PI-RADS Scoring System for Classifying mpMRI Findings in Men with Suspicion of Prostate Cancer," *BioMed Research International* **2013**, 252939 (2013).
22. M. Anwar, A. C. Westphalen, A. J. Jung, S. M. Noworolski, J. P. Simko, J. Kurhanewicz, M. Roach, 3rd, P. R. Carroll and F. V. Coakley, "Role of endorectal MR imaging and MR spectroscopic imaging in defining treatable intraprostatic tumor foci in prostate cancer: quantitative analysis of imaging contour compared to whole-mount histopathology," *Radiotherapy and Oncology* **110**, 303–308 (2014).
23. D. Hausmann, S. Konstandin, F. Zöllner, S. Haneder, F. Wetterling, A. Nagel, D. Dinter, S. Schönberg and L. Schad, "Sodium Imaging of the Prostate at 3T," in *Proceedings of International Society for Magnetic Resonance in Medicine*, (Melbourne, Australia, 2012).
24. J. Near and R. Bartha, "Quantitative sodium MRI of the mouse prostate," *Magnetic Resonance in Medicine* **63**, 822–827 (2010).
25. M. J. Albers, R. Bok, A. P. Chen, C. H. Cunningham, M. L. Zierhut, V. Y. Zhang, S. J. Kohler, J. Tropp, R. E. Hurd, Y.-F. Yen, S. J. Nelson, D. B. Vigneron and J. Kurhanewicz, "Hyperpolarized <sup>13</sup>C Lactate, Pyruvate, and Alanine: Noninvasive Biomarkers for Prostate Cancer Detection and Grading," *Cancer Research* **68**, 8607–8615 (2008).
26. S. J. Nelson, J. Kurhanewicz, D. B. Vigneron, P. E. Z. Larson, A. L. Harzstark, M. Ferrone, M. van Criekinge, J. W. Chang, R. Bok, I. Park, G. Reed, L. Carvajal, E. J. Small, P. Munster, V. K. Weinberg, J. H. Ardenkjaer-Larsen, A. P. Chen, R. E. Hurd, L.-I. Odegardstuen, F. J. Robb, J. Tropp and J. A. Murray, "Metabolic Imaging of Patients with Prostate Cancer Using Hyperpolarized [1-<sup>13</sup>C]Pyruvate," *Science Translational Medicine* **5**, 198ra108 (2013).
27. A. B. Rosenkrantz, H. Chandarana, N. Hindman, F.-M. Deng, J. S. Babb, S. S. Taneja and C. Geppert, "Computed diffusion-weighted imaging of the prostate at 3 T: impact on image quality and tumour detection," *European Radiology* **23**, 3170–3177 (2013).
28. E. Panagiotaki, R. W. Chan, N. Dikaio, H. Ahmed, D. Atkinson, S. Punwani and D. C. Alexander, "Microstructural characterisation of normal and malignant human prostate tissue with VERDICT-MRI," in *Proceedings of International Society of Magnetic Resonance in Medicine*, (Milan, Italy, 2014).
29. W. T. Hrinivich, E. Gibson, M. Gaed, J. A. Gómez, M. Moussa, C. A. McKenzie, G. S. Bauman, A. D. Ward, A. Fenster and E. Wong, "A dimensionless dynamic contrast enhanced MRI parameter for intra-prostatic tumour target volume delineation: initial

- comparison with histology," in *Proceedings of SPIE Medical Imaging*, Vol. 9036 (San Diego, USA, 2014), pp. 90362I–90367.
30. V. Shah, B. Turkbey, H. Mani, Y. Pang, T. Pohida, M. J. Merino, P. A. Pinto, P. L. Choyke and M. Bernardo, "Decision support system for localizing prostate cancer based on multiparametric magnetic resonance imaging," *Medical Physics* **39**, 4093–4103 (2012).
  31. G. Litjens, O. Debats, J. Barentsz, N. Karssemeijer and H. J. Huisman, "Computer-aided detection of prostate cancer in MRI," *IEEE Transactions on Medical Imaging* **33**, 1083–1092 (2014).
  32. G. Bauman, J. Mandel, C. Romagnoli, I. Rachinsky, A. D. Ward, E. Gibson, M. Gaed, M. Moussa, J. A. Gómez, J. L. Chin, S. Pautler, J. Butler, E. Wong and T.-Y. Lee, "18-fluorocholine imaging of localized prostate cancer: first experience with hybrid PET/MRI," in *Proceedings of Canadian Association of Radiation Oncology/Canadian Organization Medical Physicists Joint Meeting*, (Montreal, 2013).
  33. A. Farag, J. Peterson, T. Wade, T. Szekeres, A. Ouriadove, E. Gibson, A. D. Ward, J. Chin, S. Pautler, G. Bauman, C. Romagnoli, R. Bartha and T. Scholl, "Combined Transmit-Only Asymmetric Birdcage and Receive-Only Endorectal Surface RF Coils," in *Proceedings of International Society for Magnetic Resonance in Medicine*, (Milan, Italy, 2014).
  34. F. Imani, P. Abolmaesumi, E. Gibson, A. Khojaste, M. Gaed, M. Moussa, J. A. Gómez, C. Romagnoli, D. R. Siemens, M. Leviridge, S. Chang, A. Fenster, A. D. Ward and P. Mousavi, "Ultrasound-Based Characterization of Prostate Cancer: An in vivo Clinical Feasibility Study," in *Proceedings of Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013*, Vol. 8150 (Springer Berlin Heidelberg, Nagoya, Japan, 2013), pp. 279–286.
  35. G. Nir, R. S. Sahebjavaher, P. Kozlowski, S. D. Chang and R. Sinkus, "Model-based registration of ex vivo and in vivo MRI of the prostate using elastography," *IEEE Transactions on Medical Imaging* **32**, 1068–1080 (2013).
  36. Y. Sun, W. Qiu, C. Romagnoli and A. Fenster, "3D non-rigid surface-based MR-TRUS registration for image-guided prostate biopsy," in *Proceedings of SPIE Medical Imaging*, Vol. 9036 (2014), pp. 90362J.
  37. W. J. M. van de Ven, Y. Hu, J. O. Barentsz, N. Karssemeijer, D. Barratt and H. J. Huisman, "Surface-based prostate registration with biomechanical regularization," in *Proceedings of SPIE Medical Imaging*, Vol. 8671 (San Diego, USA, 2013), pp. 86711R.

38. A. P. S. Kirkham, P. Haslam, J. Y. Keanie, I. McCafferty, A. R. Padhani, S. Punwani, J. Richenberg, G. Rottenberg, A. Sohaib, P. Thompson, L. W. Turnbull, L. Kurban, A. Sahdev, R. Clements, B. M. Carey and C. Allen, "Prostate MRI: Who, when, and how? Report from a UK consensus meeting," *Clinical Radiology* **68**, 1016–1023 (2013).
39. B. G. Muller, J. J. Fütterer, R. T. Gupta, A. Katz, A. Kirkham, J. Kurhanewicz, J. W. Moul, P. A. Pinto, A. R. Rastinehad, C. Robertson, J. de la Rosette, R. Sanchez-Salas, J. S. Jones, O. Ukimura, S. Verma, H. Wijkstra and M. Marberger, "The role of magnetic resonance imaging (MRI) in focal therapy for prostate cancer: recommendations from a consensus panel," *BJU International* **113**, 218–227 (2014).
40. O. Akin, C. Riedl, N. Ishill, C. Moskowitz, J. Zhang and H. Hricak, "Interactive dedicated training curriculum improves accuracy in the interpretation of MR imaging of prostate cancer," *European Radiology* **20**, 995–1002 (2010).
41. P. Devolder, B. Pynoo, T. Voet, L. Adang, J. Vercruysse and P. Duyck, "Optimizing Physicians' Instruction of PACS Through E-Learning: Cognitive Load Theory Applied," *Journal of Digital Imaging* **22**, 25–33 (2009).
42. A. E. Flanders, "What Is the Future of Electronic Learning in Radiology?," *Radiographics* **27**, 559–561 (2007).
43. T. H. van der Kwast, M. B. Amin, A. Billis, J. I. Epstein, D. Griffiths, P. A. Humphrey, R. Montironi, T. M. Wheeler, J. R. Srigley, L. Egevad and B. Delahunt, "International Society of Urological Pathology (ISUP) Consensus Conference on Handling and Staging of Radical Prostatectomy Specimens. Working group 2: T2 substaging and prostate cancer volume," *Modern Pathology* **24**, 16–25 (2011).



## Appendix A. Derivation of the variance of $\tilde{R}$

The variance of this distribution is the mean of the covariance matrix elements

$Cov_{l,m} = E[(\tilde{R}_l - \mu_{\tilde{R}_l})(\tilde{R}_m - \mu_{\tilde{R}_m})]$ , where  $\tilde{R}_m$  is the random variable representing the intensity of the  $m$ -th sample, and  $\mu_{\tilde{R}_m}$  is the mean of  $\tilde{R}_m$ . The distribution of the  $m$ -th sample depends on whether the sample is from a region  $R_i$  (denoted  $\tilde{R}_m \sim R$ ) or  $B'_i$  (denoted  $\tilde{R}_m \sim B$ ). For any pair of samples  $l, m$ , there are four possible outcomes:  $\tilde{R}_l \sim R$  and  $\tilde{R}_m \sim R$ ;  $\tilde{R}_l \sim R$  and  $\tilde{R}_m \sim B$ ;  $\tilde{R}_l \sim B$  and  $\tilde{R}_m \sim R$ ; or  $\tilde{R}_l \sim B$  and  $\tilde{R}_m \sim B$ . We denote the probabilities of these outcomes as  $p_{RR}$ ,  $p_{RB}$ ,  $p_{BR}$ , and  $p_{BB}$ , and the expected values of these outcomes as  $E_{RR}$ ,  $E_{RB}$ ,  $E_{BR}$  and  $E_{BB}$ . We can express the covariance as a weighted sum of the four possible outcomes:

$$Cov_{l,m} = E_{RR}p_{RR} + E_{RB}p_{RB} + E_{BR}p_{BR} + E_{BB}p_{BB} \quad A.1$$

We can express these expected values in terms of population parameters by noting that  $R_l - \mu_{\tilde{R}_l} = R_l - \mu_R + (1 - \mu_F)\mu_d$ , and  $B_l - \mu_{\tilde{R}_l} = B_l - \mu_B - \mu_F\mu_d$ . Expanding the polynomial in the expected value, and noting that  $E[R_l - \mu_R] = 0$ , and  $E[(R_l - \mu_R)(R_m - \mu_R)]$  is the covariance of  $R_l$  and  $R_m$ , yields the expressions

$$E_{RR}(\delta) = \sigma_{DR}^2 + \sigma_{IR}^2\delta + (1 - \mu_F)^2\mu_d^2, \quad A.2$$

$$E_{RB} = -(1 - \mu_F)\mu_F\mu_d^2, \quad A.3$$

$$E_{BR} = -(1 - \mu_F)\mu_F\mu_d^2, \text{ and} \quad A.4$$

$$E_{BB}(\delta) = \sigma_{DB}^2 + \sigma_{IB}^2\delta + \mu_F^2\mu_d^2, \quad A.5$$

if the  $l$ -th and  $m$ -th samples are from the same sampling region, and 0 otherwise, where

$\delta = 1$  if  $l = m$ , and 0 otherwise. Note that  $E_{RR}(1) = E_{RR}(0) + \sigma_{IR}^2$ , and  $E_{BB}(1) =$

$E_{BB}(0) + \sigma_{IB}^2$ , and that  $E_{RR}(0)$  and  $E_{BB}(0)$  are independent of  $l$  and  $m$ .

Because samples from different regions are independent (assumption 1 in Section 2.2.1) and regions have a constant size (assumption 6 in Section 2.2.1), the covariance matrix is blockwise diagonal, where each block represents one sampling region. By isolating the terms of the expected value that are independent of  $l$  and  $m$ , we can express the sum of the covariance within each block as

$$\begin{aligned}
& E_{RR}(0) \sum_l \sum_m p_{RR} + E_{RB} \sum_l \sum_m p_{RB} + E_{BR} \sum_l \sum_m p_{BR} + E_{BB}(0) \sum_l \sum_m p_{BB} \\
& + \sigma_{IR}^2 \sum_l p(\tilde{R}_l \sim R) + \sigma_{IB}^2 \sum_l p(\tilde{R}_l \sim B).
\end{aligned} \tag{A.6}$$

If we express the probabilities as integrations of conditional probabilities, conditional on the value of the fractional overlap, and then reorder the summation terms, we show that

$$\sum_l \sum_m p_{RR} = \int p(F_i = f_i) \sum_l \sum_m p(\tilde{R}_l \sim R, \tilde{R}_m \sim R | F_i = f_i) df_i, \tag{A.7}$$

and we can show that similar expressions hold for  $p_{RB}$ ,  $p_{BR}$ ,  $p_{BB}$ ,  $p(\tilde{R}_l \sim R)$  and  $p(\tilde{R}_l \sim B)$ . The conditional probabilities can be expressed as polynomials of  $f_i$ , which can be expressed in terms of population parameters of  $F$ :

$$\sum_l \sum_m p_{RR} = v^2 \int p(F_i = f_i) f_i^2 df_i = v^2(\mu_F^2 + \sigma_F^2), \tag{A.8}$$

$$\sum_l \sum_m p_{RB} = v^2 \int p(F_i = f_i) f_i(1 - f_i) df_i = v^2(\mu_f - \mu_F^2 - \sigma_F^2), \tag{A.9}$$

$$\sum_l \sum_m p_{BR} = v^2 \int p(F_i = f_i) f_i(1 - f_i) df_i = v^2(\mu_f - \mu_F^2 - \sigma_F^2), \tag{A.10}$$

$$\sum_l \sum_m p_{BB} = v^2 \int p(F_i = f_i) (1 - f_i)^2 df_i = v^2(1 - 2\mu_F + \mu_F^2 + \sigma_F^2), \tag{A.11}$$

$$\sum_l p(\tilde{R}_l \sim R) = v^2 \int p(F_i = f_i) f_i df_i = v\mu_F, \text{ and} \quad \text{A.12}$$

$$\sum_l p(\tilde{R}_l \sim B) = v^2 \int p(F_i = f_i) (1 - f_i) df_i = v(1 - \mu_F). \quad \text{A.13}$$

By substituting these equations into Expression A.6, we can calculate the mean over all covariance matrix elements to be

$$\sigma_{\tilde{R},nv}^2 = \frac{\mu_d^2 \sigma_F^2 + (\sigma_{DR}^2 + \sigma_{DB}^2)(\mu_F^2 + \sigma_F^2) + \frac{(\sigma_{IR}^2 - \sigma_{IB}^2 - 2v\sigma_{DB}^2)\mu_F}{v} + \sigma_{DB}^2 + \frac{\sigma_{IB}^2}{v}}{n}. \quad \text{A.14}$$

For conciseness, we introduce two new terms:  $\sigma_D^2 = \sigma_{DR}^2 + \sigma_{DB}^2$  and  $\sigma_{B,k}^2 = \sigma_{IB}^2/v +$

$k\sigma_{DB}^2$ . With this notation,  $\sigma_{\tilde{R},nv}^2$  can be rewritten as

$$\sigma_{\tilde{R},nv}^2 = \frac{\mu_d^2 \sigma_F^2 + \sigma_D^2(\mu_F^2 + \sigma_F^2) + \left(\frac{\sigma_{IR}^2}{v} - \sigma_{B,2}^2\right)\mu_F + \sigma_{B,1}^2}{n}. \quad \text{A.15}$$

## Appendix B. Large specimen pathology processing schedule

London Health Sciences Centre's standard clinical pathology laboratory protocol for large specimens consists of processing tissue slices through baths of graded alcohols, xylene, and paraplast on a Tissue-Tek vacuum infiltration tissue processor (Sakura Finetek USA, Inc., Torrance, USA) following the schedule shown in Table B.1.

**Table B.1:** Our hospital's standard clinical pathology laboratory protocol for large specimens.

<i>Solution</i>	<i>Duration (hours)</i>	<i>Temperature (°C)</i>	<i>Vacuum (inches)</i>	<i>Pressure (PSI)</i>
<b>80% ethanol, 20% formalin</b>	1	40	15	7
<b>95% ethanol, 5% formalin</b>	2	40	15	7
<b>100% ethanol</b>	1	40	15	7
<b>100% ethanol</b>	2	40	15	7
<b>100% ethanol</b>	2	40	15	7
<b>100% xylene</b>	1	40	15	7
<b>100% xylene</b>	1	40	15	7
<b>100% xylene</b>	2	40	15	7
<b>100% paraplast</b>	1	60	15	7
<b>100% paraplast</b>	1	60	15	7
<b>100% paraplast</b>	1	60	15	7
<b>100% paraplast</b>	1	60	15	7

## Appendix C. Supplementary material to Chapter 6

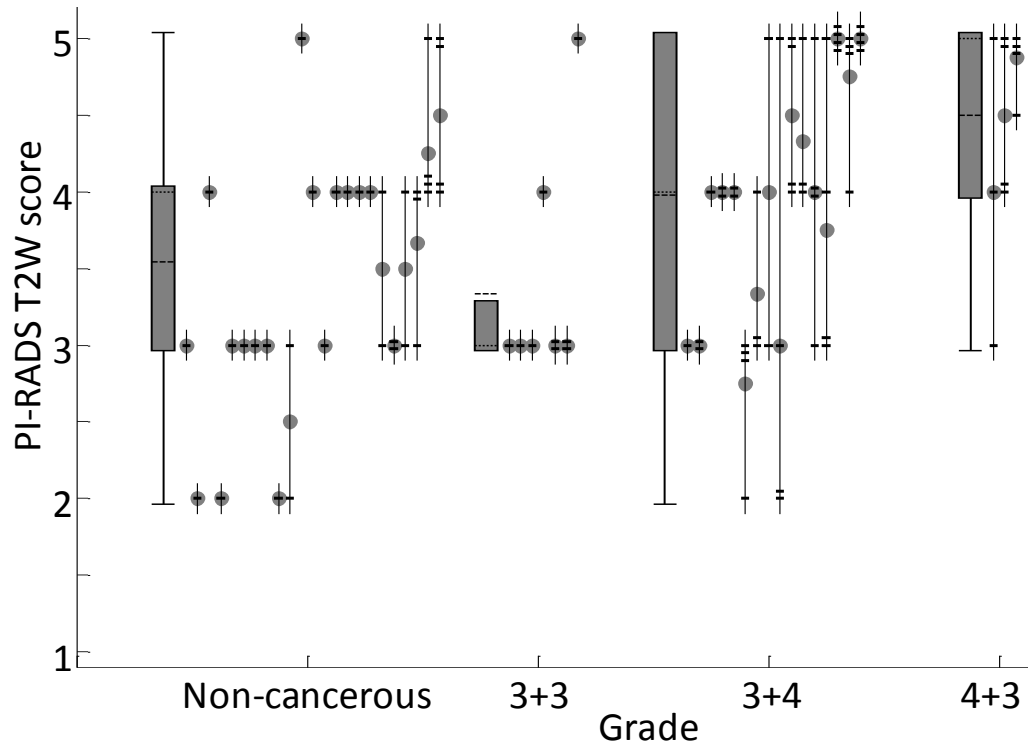
### C.1 Manual correspondence for two *distinct* lesions

During the interobserver correspondence of *suspicious regions*, two cases had spatial relationships where overlap of *composite suspicious regions* did not yield one-to-one lesion correspondences. In the first case, a *composite suspicious region*  $A_1$  from one observer overlapped in different amounts with two *composite suspicious regions*  $B_1$  (Dice metric of 0.7) and  $C_1$  (Dice metric of 0.09) defined by another observer; in this case,  $A_1$  was taken to correspond with  $B_1$ , and not with  $C_1$ . In a second case, three observers each made a *score assignment* and delineated large mutually overlapping *composite suspicious regions*  $A_2$ ,  $B_2$  and  $C_2$  and the fourth observer made two *score assignments* and delineated two smaller *composite suspicious regions*  $D_2$  and  $E_2$ . It was clear by inspection of the contours that the fourth observer contoured two subregions of the *distinct* lesion contoured by the other observers; in this case, the fourth observer's two lesions were taken to be a single *scored* lesion, the PI-RADS scores in the corresponding *score assignments* were averaged and the union of the corresponding *suspicious regions* were taken to correspond with  $A_2$ ,  $B_2$  and  $C_2$ .

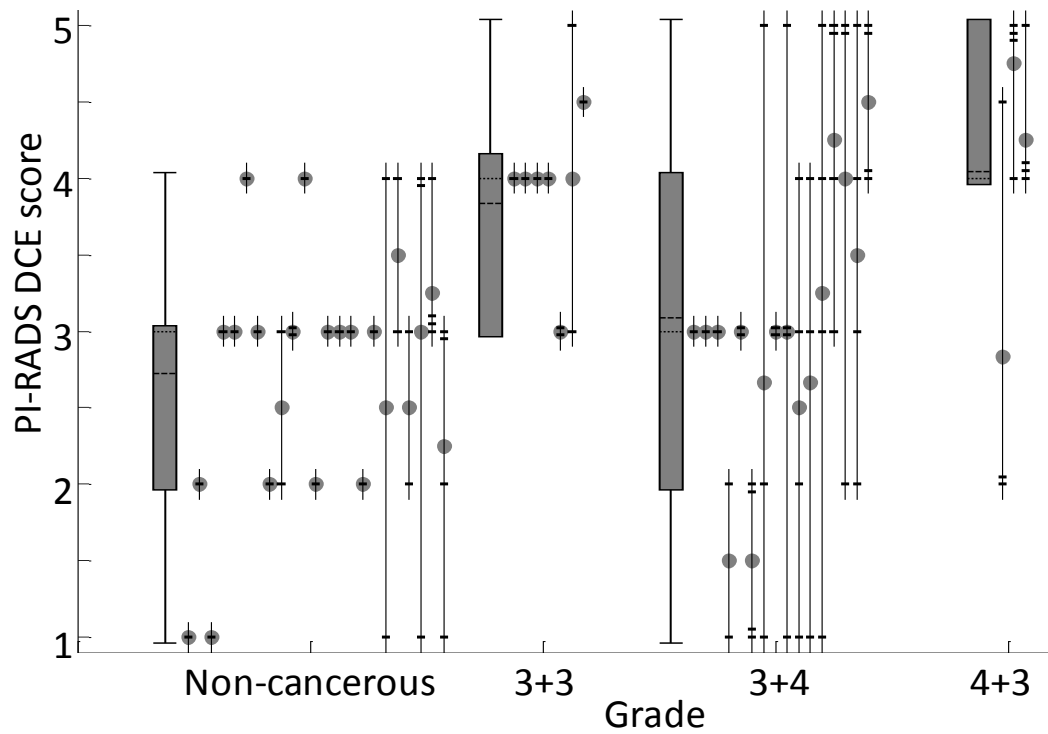
### C.2 PI-RADS sequence-specific scores broken down by grade and volume

In each of the following figures, the PI-RADS sequence-specific scores (i.e. T2W, DCE and ADC scores) from *score assignments* are broken down by histological grade or mean *suspicious region* volume. Each vertical solid line corresponds to a *distinct* lesion, with horizontal dashes at the PI-RADS sequence-specific score for each observer. For

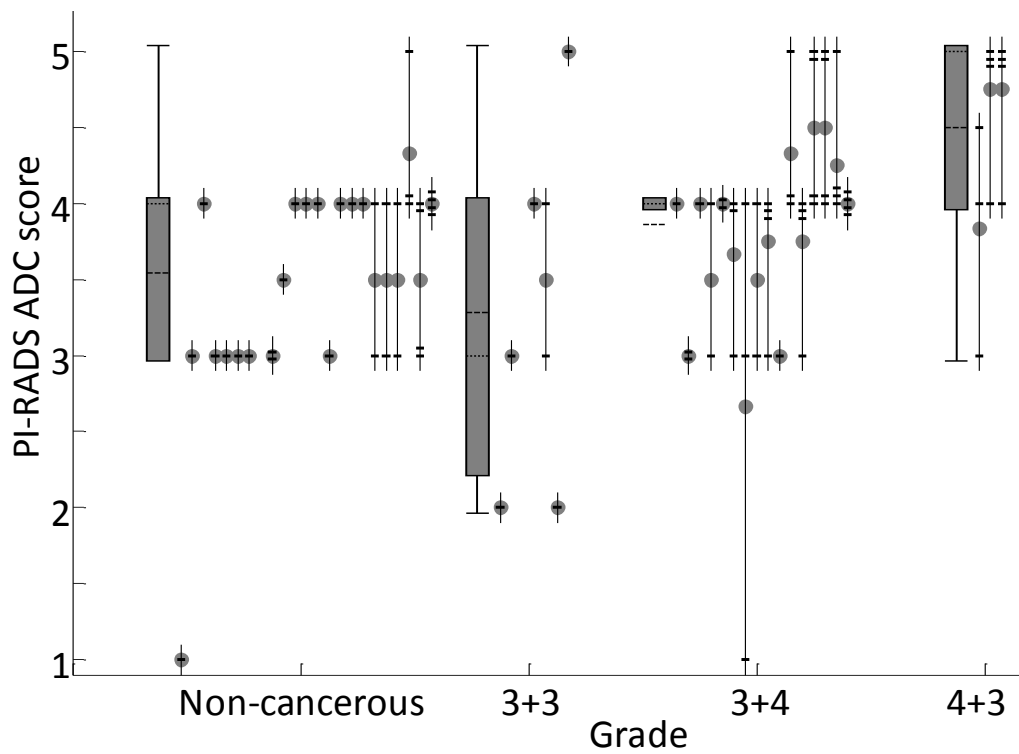
each grade/volume range, Tukey box plots show the interquartile range (IQR) box, median (dotted line), mean (dashed line), and range of data within  $1.5 \times \text{IQR}$  of the IQR (whiskers). Note that sequence-specific scores, unlike likelihood scores, were not assigned by default when no *score assignment* was made, so not all observers are represented for each *distinct* lesion.



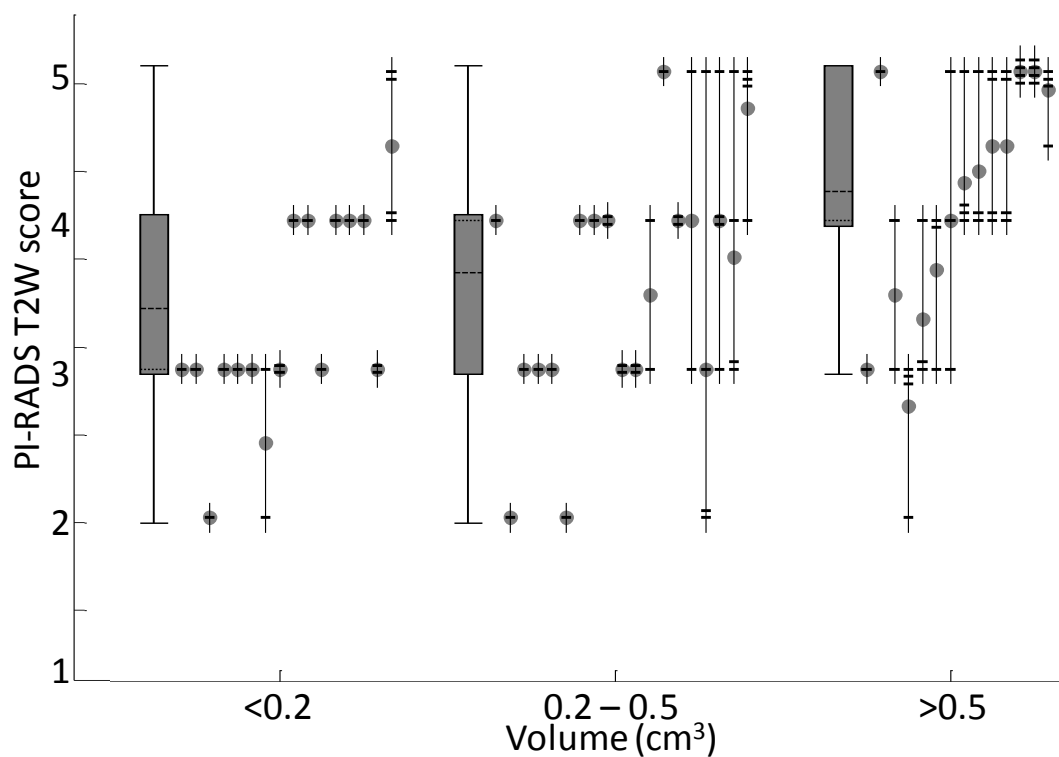
**Figure C.1:** PI-RADS T2W scores broken down by histological grade.



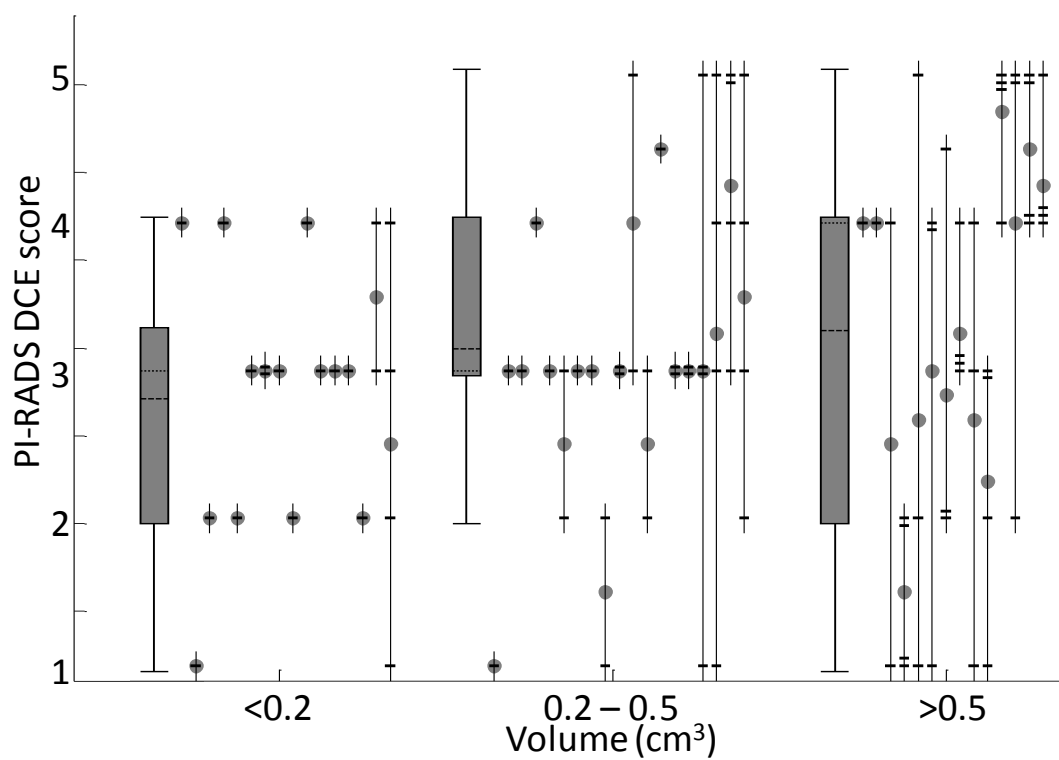
**Figure C.2:** PI-RADS DCE scores broken down by histological grade.



**Figure C.3:** PI-RADS ADC scores broken down by histological grade.

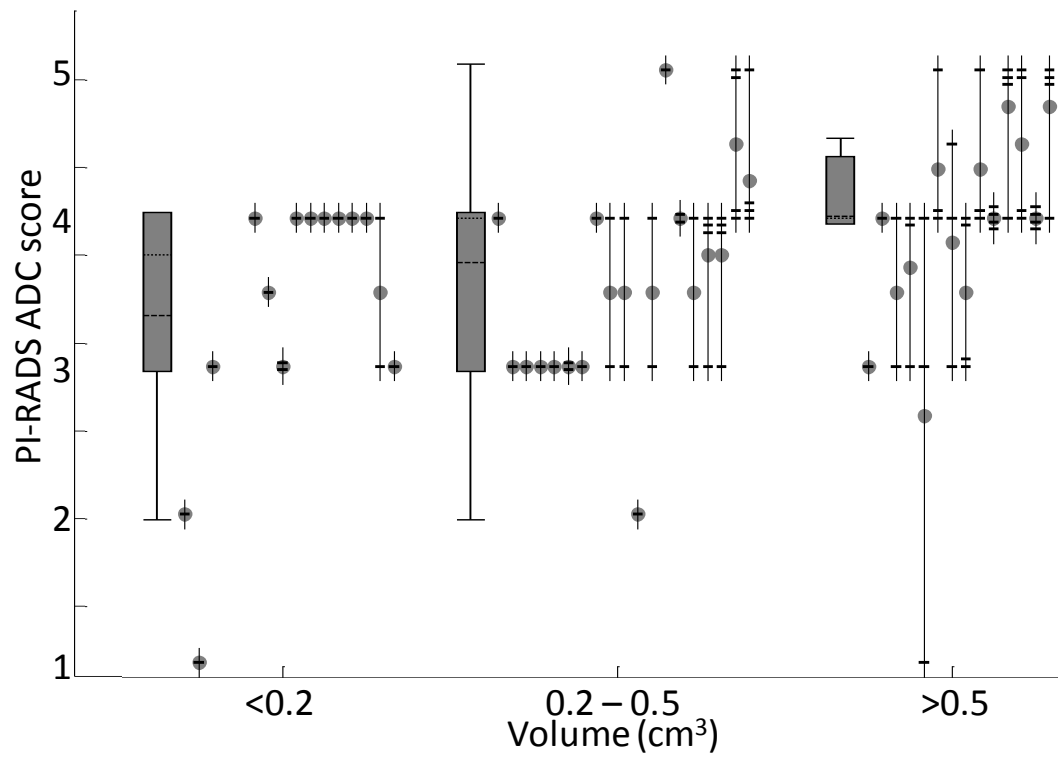


**Figure C.4:** PI-RADS T2W scores broken down by mean *suspicious region* volume.



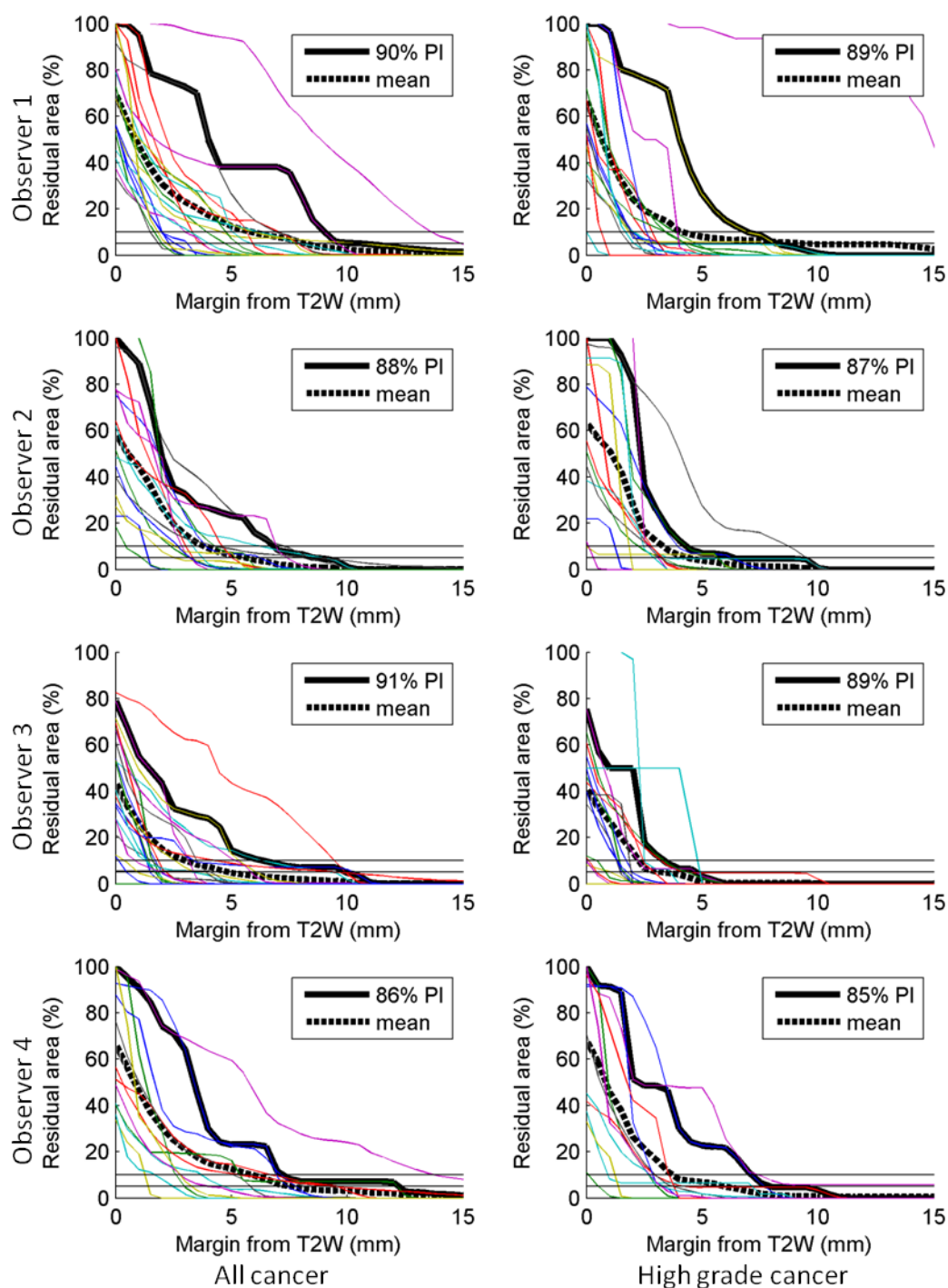
**Figure C.5:** PI-RADS DCE scores broken down by mean *suspicious region* volume.



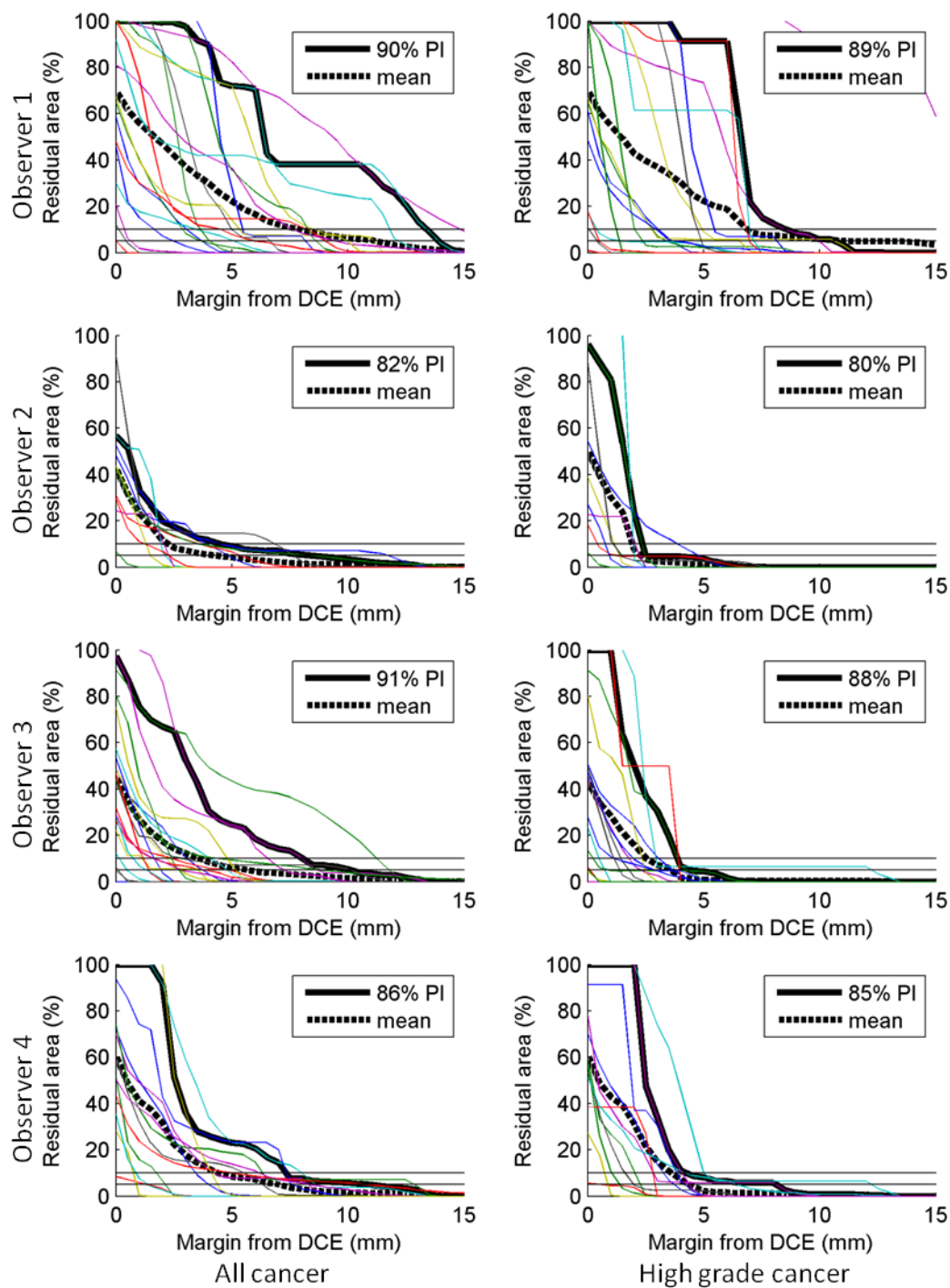


**Figure C.6:** PI-RADS ADC scores broken down by mean *suspicious region* volume.

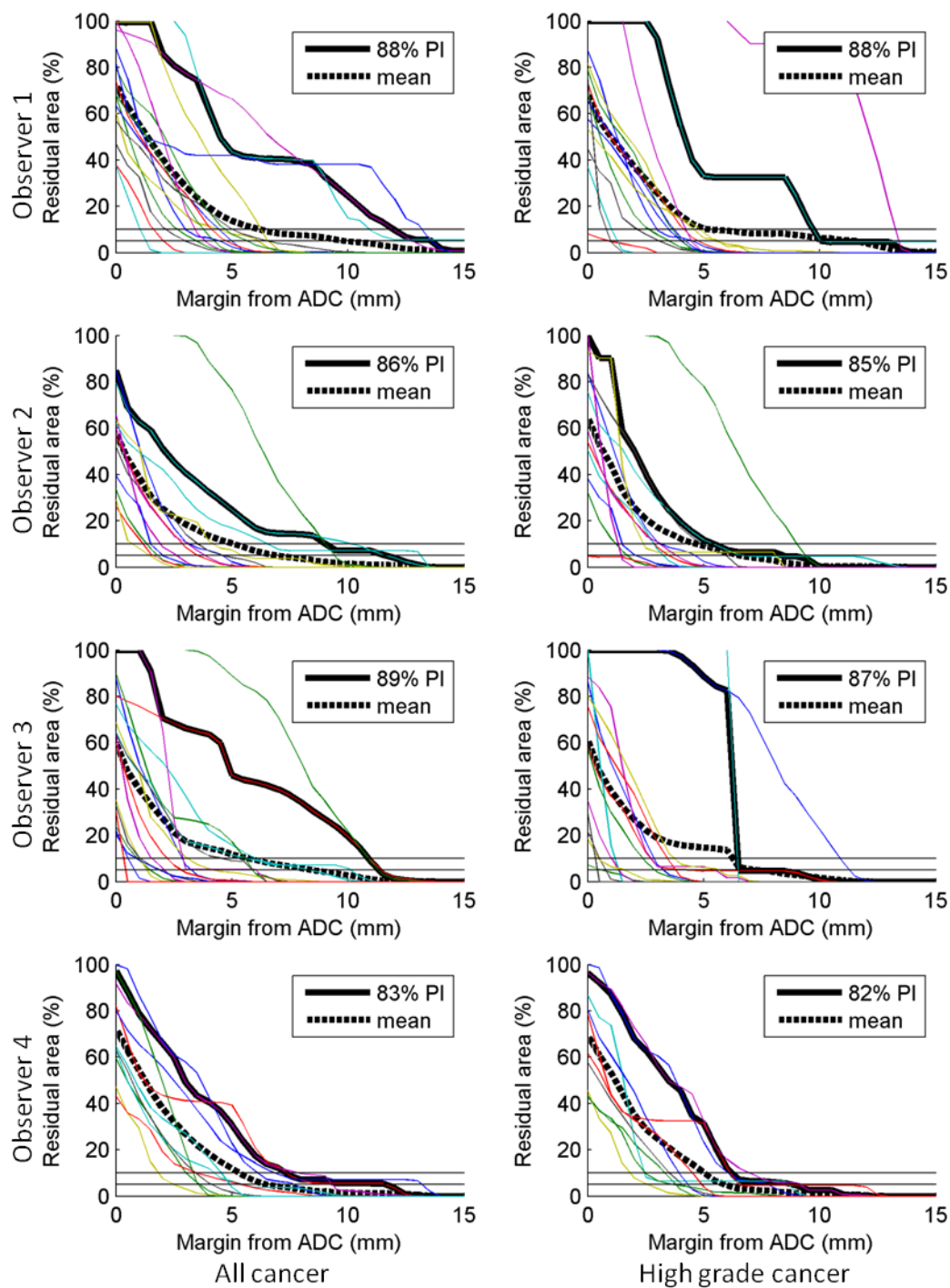
## Appendix D. Supplementary material to Chapter 7



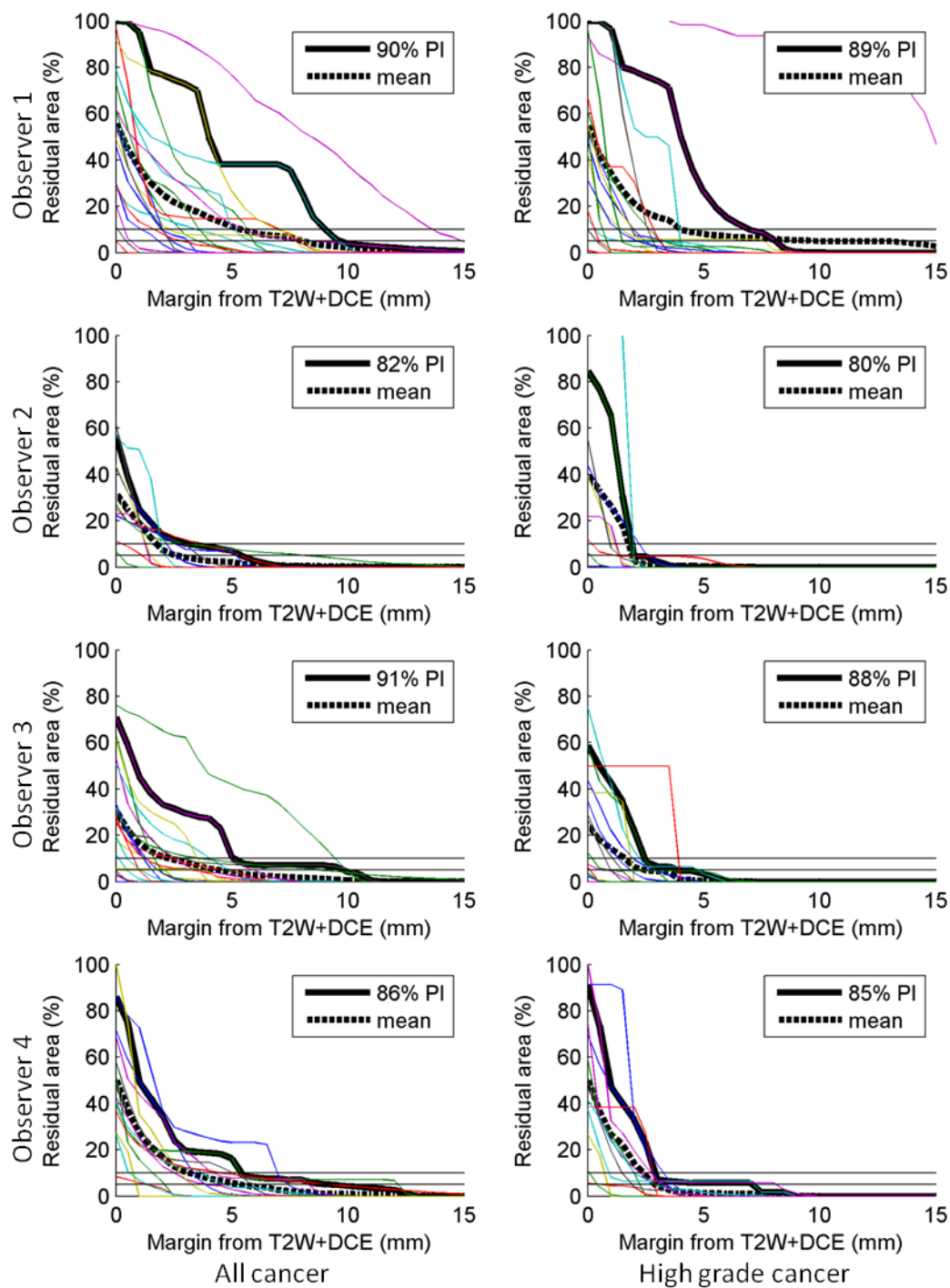
**Figure D.1:** Residual cancer (left) and high-grade cancer (right) for CTVs expanded by various margins from individual GTVs delineated on T2W images for 4 observers (rows). Legend shows prediction interval (PI) confidence levels. The 5% and 10% residual area levels are marked by the horizontal lines.



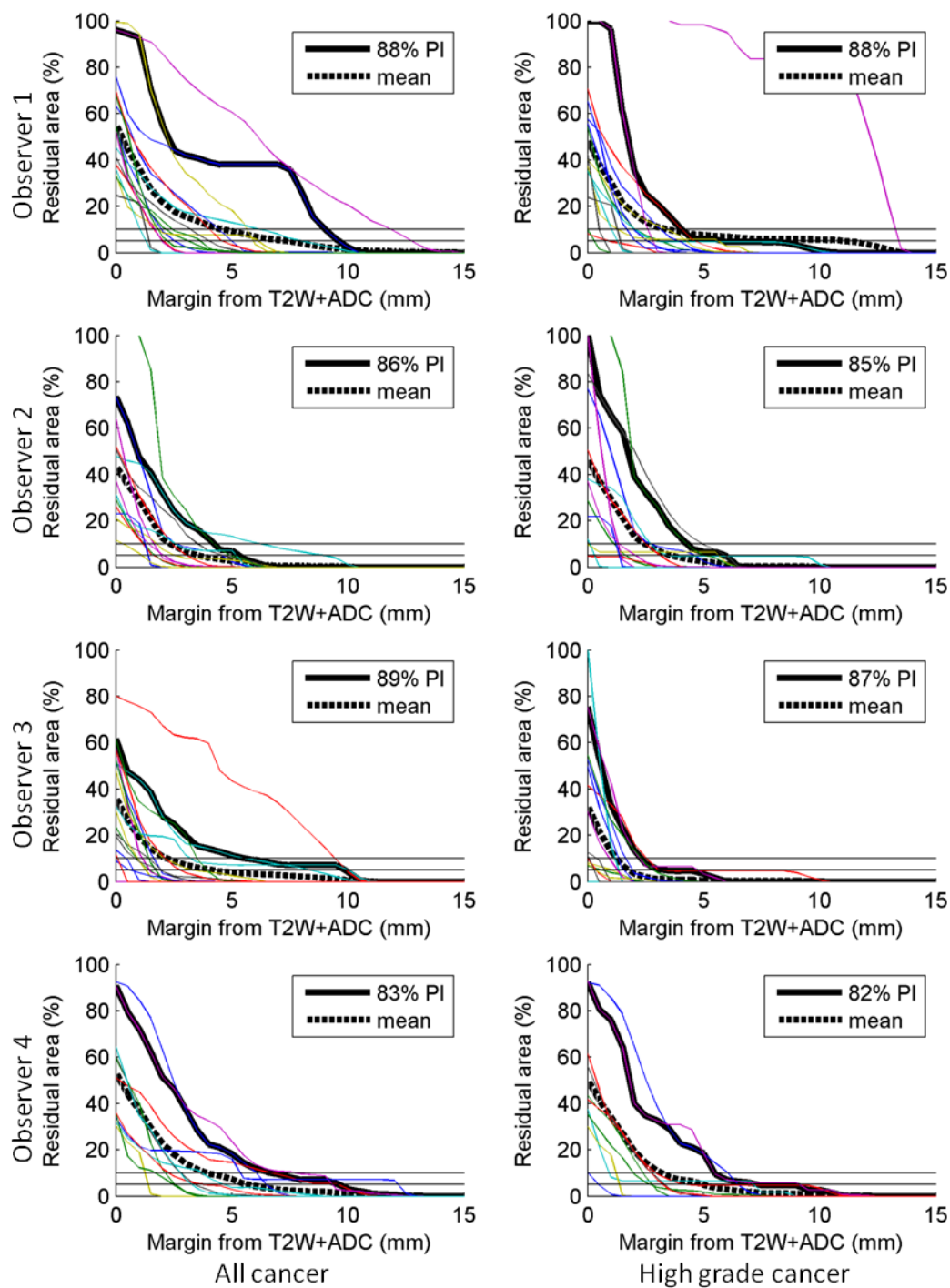
**Figure D.2:** Residual cancer (left) and high-grade cancer (right) for CTVs expanded by various margins from individual GTVs delineated on DCE images for 4 observers (rows). Legend shows prediction interval (PI) confidence levels. The 5% and 10% residual area levels are marked by the horizontal lines.



**Figure D.3:** Residual cancer (left) and high-grade cancer (right) for CTVs expanded by various margins from individual GTVs delineated on ADC images for 4 observers (rows). Legend shows prediction interval (PI) confidence levels. The 5% and 10% residual area levels are marked by the horizontal lines.

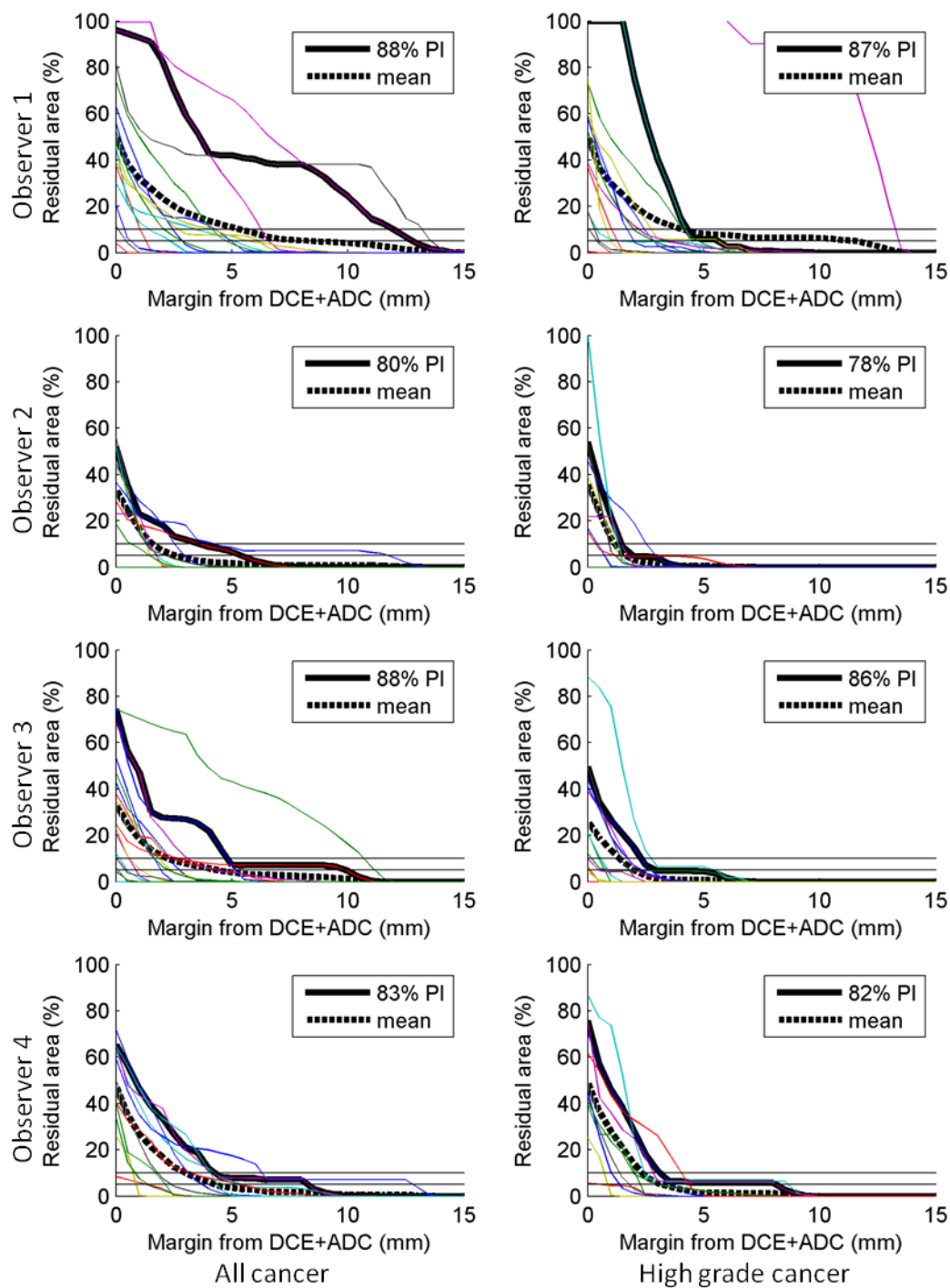


**Figure D.4:** Residual cancer (left) and high-grade cancer (right) for CTVs expanded by various margins from individual composite T2W+DCE GTVs for 4 observers (rows). Legend shows prediction interval (PI) confidence levels. The 5% and 10% residual area levels are marked by the horizontal lines.

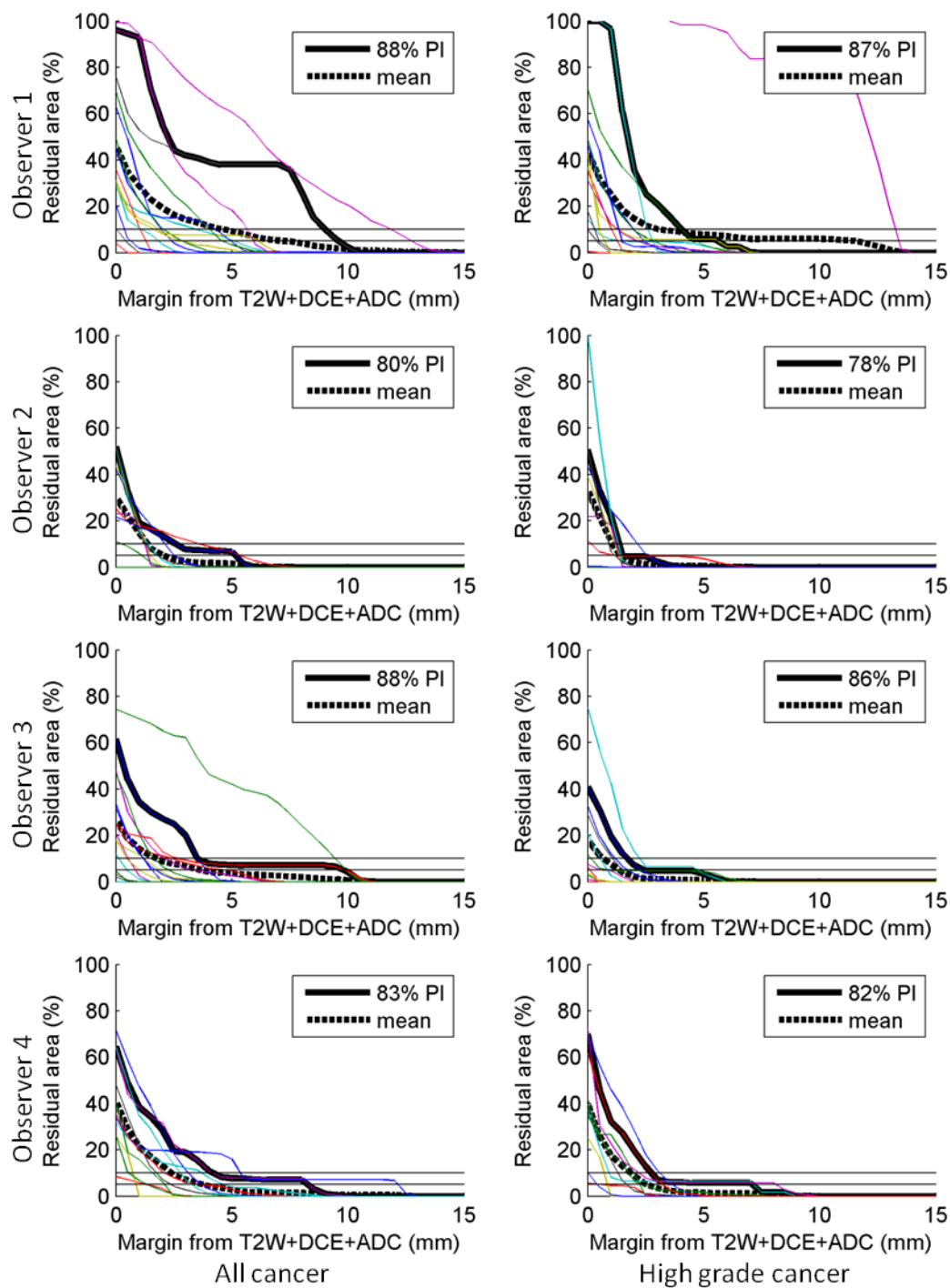


**Figure D.5:** Residual cancer (left) and high-grade cancer (right) for CTVs expanded by various margins from individual composite T2W+ADC GTVs for 4 observers (rows). Legend shows prediction interval (PI) confidence levels. The 5% and 10% residual area levels are marked by the horizontal lines.



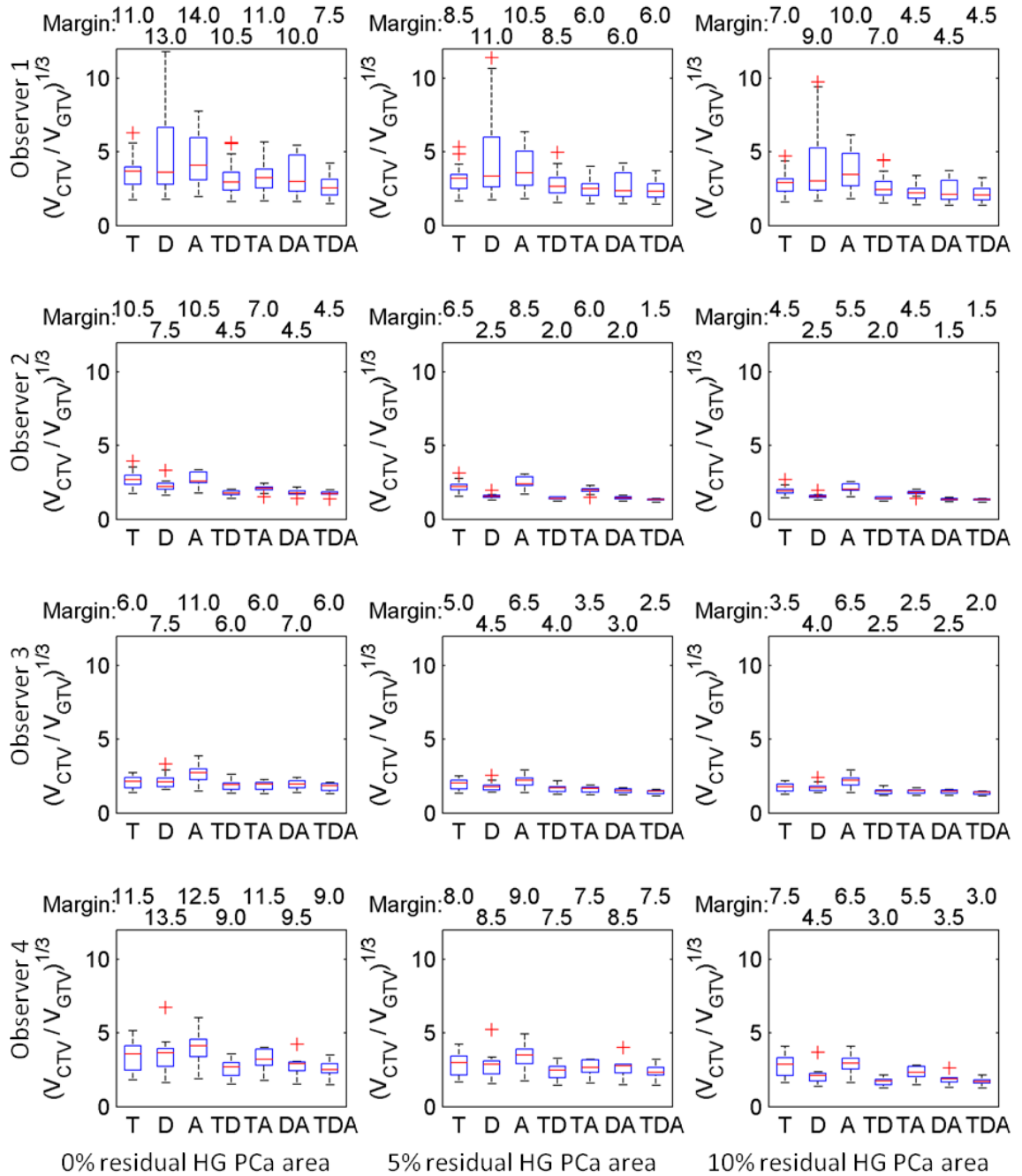


**Figure D.6:** Residual cancer (left) and high-grade cancer (right) for CTVs expanded by various margins from individual composite DCE+ADC GTVs for 4 observers (rows). Legend shows prediction interval (PI) confidence levels. The 5% and 10% residual area levels are marked by the horizontal lines.

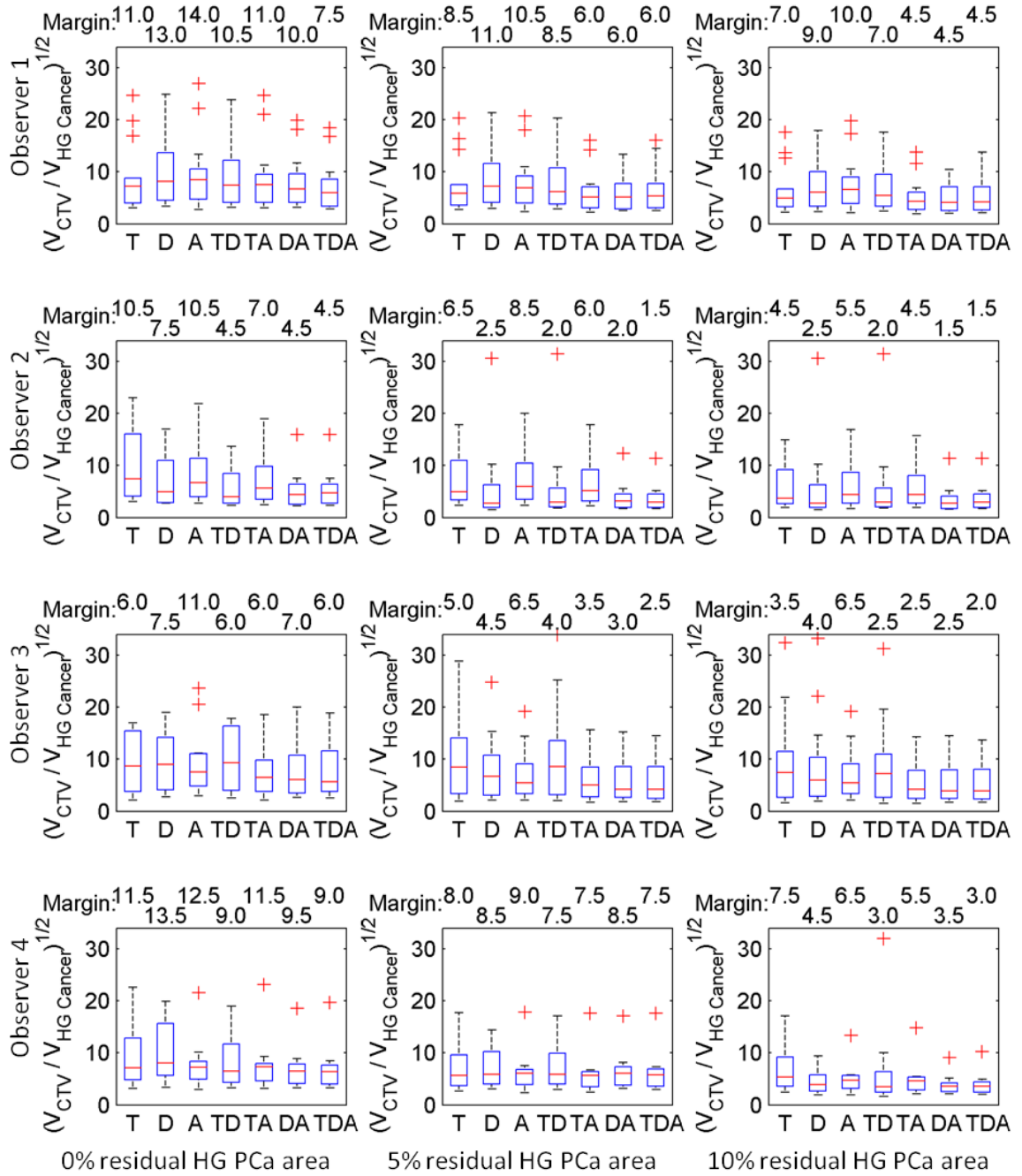


**Figure D.7:** Residual cancer (left) and high-grade cancer (right) for CTVs expanded by various margins from individual composite T2W+DCE+ADC GTVs for 4 observers (rows). Legend shows prediction interval (PI) confidence levels. The 5% and 10% residual area levels are marked by the horizontal lines.





**Figure D.8:** Scaling factors of the CTVs with PI margins for high-grade cancer **relative to their corresponding GTVs** for 4 observers (rows) for 0%, 5% and 10% residual high-grade cancer area (columns). Y-axis shows the cube root of the volume ratio (the equivalent scaling factor in each dimension). CTV types are denoted by the first initials of the sequences (T=T2W, D=DCE, A=ADC), with the PI margins for high-grade cancer shown above the graph. Box plots show the median, interquartile range (IQR), whiskers for the closest points within  $1.5 \times IQR$  from the IQR, and outliers beyond the whiskers.



**Figure D.9:** Scaling factors of the CTVs **relative to their corresponding aggregated high-grade foci** for 4 observers (rows) for 0%, 5% and 10% residual area (columns). Y-axis shows the square root of the planimetric volume ratio (the equivalent scaling factor in each dimension). CTV types are denoted by the first initials of the sequences (T=T2W, D=DCE, A=ADC), with the PI margins for high-grade cancer shown above the graph. Box plots show the median, interquartile range (IQR), whiskers for the closest points within 1.5×IQR from the IQR, and outliers beyond the whiskers.

# Appendix E. Permissions

## E.1 Permission to reproduce previously published material in Chapter 1

### WOLTERS KLUWER HEALTH LICENSE TERMS AND CONDITIONS

Jun 23, 2014

This is a License Agreement between Eli Gibson ("You") and Wolters Kluwer Health ("Wolters Kluwer Health") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Wolters Kluwer Health, and the payment terms and conditions.

**All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.**

License Number	3415010517252
License date	Jun 23, 2014
Licensed content publisher	Wolters Kluwer Health
Licensed content publication	American Journal of Surgical Pathology, The
Licensed content title	The 2005 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma
Licensed content author	Jonathan Epstein, William Allsbrook, Mahul Amin, et al
Licensed content date	Jan 1, 2005
Volume Number	29
Issue Number	9
Type of Use	Dissertation/Thesis
Requestor type	Individual
Portion	Figures/table/illustration
Number of figures/tables/illustrations	1
Figures/tables/illustrations used	12
Author of this Wolters Kluwer article	No
Title of your thesis / dissertation	3D FUSION OF HISTOLOGY TO MULTI-PARAMETRIC MRI FOR PROSTATE CANCER IMAGING EVALUATION AND LESION-TARGETED TREATMENT PLANNING
Expected completion date	Aug 2014
Estimated size(pages)	320
Billing Type	Invoice
Billing address	

Canada  
Total 0.00 USD  
Terms and Conditions

### Terms and Conditions

1. A credit line will be prominently placed and include: for books - the author(s), title of book, editor, copyright holder, year of publication; For journals - the author(s), title of article, title of journal, volume number, issue number and inclusive pages.
2. The requestor warrants that the material shall not be used in any manner which may be considered derogatory to the title, content, or authors of the material, or to Wolters Kluwer.
3. Permission is granted for a one time use only within 12 months from the date of this invoice. Rights herein do not apply to future reproductions, editions, revisions, or other derivative works. Once the 12-month term has expired, permission to renew must be submitted in writing.
4. Permission granted is non-exclusive, and is valid throughout the world in the English language and the languages specified in your original request.
5. Wolters Kluwer cannot supply the requestor with the original artwork or a "clean copy."
6. The requestor agrees to secure written permission from the author (for book material only).
7. Permission is valid if the borrowed material is original to a Wolters Kluwer imprint (Lippincott-Raven Publishers, Williams & Wilkins, Lea & Febiger, Harwal, Igaku-Shoin, Rapid Science, Little Brown & Company, Harper & Row Medical, American Journal of Nursing Co, and Urban & Schwarzenberg - English Language).
8. If you opt not to use the material requested above, please notify Rightslink within 90 days of the original invoice date.
9. Please note that articles in the ahead-of-print stage of publication can be cited and the content may be re-used by including the date of access and the unique DOI number. Any final changes in manuscripts will be made at the time of print publication and will be reflected in the final electronic version of the issue. Disclaimer: Articles appearing in the Published Ahead-of-Print section have been peer-reviewed and accepted for publication in the relevant journal and posted online before print publication. Articles appearing as publish ahead-of-print may contain statements, opinions, and information that have errors in facts, figures, or interpretation. Accordingly, Lippincott Williams & Wilkins, the editors and authors and their respective employees are not responsible or liable for the use of any such inaccurate or misleading data, opinion or information contained in the articles in this section.
10. This permission does not apply to images that are credited to publications other than Wolters Kluwer journals. For images credited to non-Wolters Kluwer journal publications, you will need to obtain permission from the journal referenced in the figure or table legend or credit line before making any use of the image(s) or table(s).
11. In case of **Disease Colon Rectum, Plastic Reconstructive Surgery, The Green Journal, Critical Care Medicine, Pediatric Critical Care Medicine, the American Heart Publications, the American Academy of Neurology** the following guideline applies: no drug brand/trade name or logo can be included in the same page as the material re-used
12. When requesting a permission to translate a full text article, Wolters Kluwer/Lippincott Williams & Wilkins requests to receive the pdf of the translated document
13. "Adaptations of single figures do not require Wolters Kluwer **further** approval if the permission has been granted previously. However, the adaptation should be credited as follows: "Adapted with permission from Lippincott Williams and Wilkins/Wolters Kluwer Health: [JOURNAL NAME] (reference citation), copyright (year of publication)" **Please note that modification of text within figures or full-text articles is strictly forbidden.**
14. The following statement needs to be added when reprinting the material in Open Access journals only: "promotional and commercial use of the material in print, digital or mobile device format is prohibited without the permission from the publisher Lippincott Williams & Wilkins. Please contact further information".
15. Other Terms and Conditions:

v1.8

**If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 48 hours of the license date. Payment should be in the form of a check or money order referencing your account number and this invoice number 501334905. Once you receive your invoice for this order, you may pay your invoice by credit card. Please follow instructions provided at that time.**

**Make Payment To:**

**For suggestions or comments regarding this order, contact RightsLink Customer Support:  
or (toll free in the US) or**

**Gratis licenses (referencing \$0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.**

---

## E.2 Permission to reproduce previously published material in Chapter 2

### ELSEVIER LICENSE TERMS AND CONDITIONS

Jun 30, 2014

---

This is a License Agreement between Eli Gibson ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

**All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.**

Supplier	Elsevier Limited
Registered Company Number	1982084
Customer name	Eli Gibson
Customer address	
License number	3418801267600
License date	Jun 30, 2014
Licensed content publisher	Elsevier
Licensed content publication	Medical Image Analysis
Licensed content title	The impact of registration accuracy on imaging validation study design: A novel statistical power calculation
Licensed content author	Eli Gibson,Aaron Fenster,Aaron D. Ward
Licensed content date	October 2013
Licensed content volume number	17
Licensed content issue number	7
Number of pages	11
Start Page	805
End Page	815
Type of Use	reuse in a thesis/dissertation
Intended publisher of new work	other
Portion	full article
Format	both print and electronic
Are you the author of this Elsevier article?	Yes

Will you be translating?	No
Title of your thesis/dissertation	3D FUSION OF HISTOLOGY TO MULTI-PARAMETRIC MRI FOR PROSTATE CANCER IMAGING EVALUATION AND LESION-TARGETED TREATMENT PLANNING
Expected completion date	Aug 2014
Estimated size (number of pages)	320
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 USD
VAT/Local Sales Tax	0.00 USD / 0.00 GBP
Total	0.00 USD
Terms and Conditions	

## INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at <http://myaccount.copyright.com>).

## GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.

3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:

“Reprinted from Publication title, Vol./edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER].” Also Lancet special credit - “Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier.”

4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.

5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier at )

6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.

7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.

9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.

10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

11. No Transfer of License: This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.

12. No Amendment Except in Writing: This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).

13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

#### **LIMITED LICENSE**



The following terms and conditions apply only to specific license types:

15. **Translation:** This permission is granted for non-exclusive world **English** rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article. If this license is to re-use 1 or 2 figures then permission is granted for non-exclusive world rights in all languages.

16. **Posting licensed content on any Website:** The following terms and conditions apply as follows: Licensing material from an Elsevier journal: All content posted to the web site must maintain the copyright information line on the bottom of each image; A hyper-text must be included to the Homepage of the journal from which you are licensing at <http://www.sciencedirect.com/science/journal/xxxxx> or the Elsevier homepage for books at <http://www.elsevier.com>; Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

Licensing material from an Elsevier book: A hyper-text link must be included to the Elsevier homepage at <http://www.elsevier.com>. All content posted to the web site must maintain the copyright information line on the bottom of each image.

**Posting licensed content on Electronic reserve:** In addition to the above the following clauses are applicable: The web site must be password-protected and made available only to bona fide students registered on a relevant course. This permission is granted for 1 year only. You may obtain a new license for future website posting.

**For journal authors:** the following clauses are applicable in addition to the above: Permission granted is limited to the author accepted manuscript version\* of your paper.

**\*Accepted Author Manuscript (AAM) Definition:** An accepted author manuscript (AAM) is the author's version of the manuscript of an article that has been accepted for publication and which may include any author-incorporated changes suggested through the processes of submission processing, peer review, and editor-author communications. AAMs do not include other publisher value-added contributions such as copy-editing, formatting, technical enhancements and (if relevant) pagination.

You are not allowed to download and post the published journal article (whether PDF or HTML, proof or final version), nor may you scan the printed edition to create an electronic version. A hyper-text must be included to the Homepage of the journal from which you are licensing at <http://www.sciencedirect.com/science/journal/xxxxx>. As part of our normal production process, you will receive an e-mail notice when your article appears on Elsevier's online service ScienceDirect ([www.sciencedirect.com](http://www.sciencedirect.com)). That e-mail will include the article's Digital Object Identifier (DOI). This number provides the electronic link to the published article and should be included in the posting of your personal version. We ask that you wait until you receive this e-mail and have the DOI to do any posting.

**Posting to a repository:** Authors may post their AAM immediately to their employer's

institutional repository for internal use only and may make their manuscript publically available after the journal-specific embargo period has ended.

Please also refer to [Elsevier's Article Posting Policy](#) for further information.

18. **For book authors** the following clauses are applicable in addition to the above: Authors are permitted to place a brief summary of their work online only.. You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version. **Posting to a repository:** Authors are permitted to post a summary of their chapter only in their institution's repository.

20. **Thesis/Dissertation:** If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission.

### **Elsevier Open Access Terms and Conditions**

Elsevier publishes Open Access articles in both its Open Access journals and via its Open Access articles option in subscription journals.

Authors publishing in an Open Access journal or who choose to make their article Open Access in an Elsevier subscription journal select one of the following Creative Commons user licenses, which define how a reader may reuse their work: Creative Commons Attribution License (CC BY), Creative Commons Attribution – Non Commercial - ShareAlike (CC BY NC SA) and Creative Commons Attribution – Non Commercial – No Derivatives (CC BY NC ND)

#### **Terms & Conditions applicable to all Elsevier Open Access articles:**

Any reuse of the article must not represent the author as endorsing the adaptation of the article nor should the article be modified in such a way as to damage the author's honour or reputation.

The author(s) must be appropriately credited.

If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source it is the responsibility of the user to ensure their reuse complies with the terms and conditions determined by the rights holder.

#### **Additional Terms & Conditions applicable to each Creative Commons user license:**

**CC BY:** You may distribute and copy the article, create extracts, abstracts, and other revised versions, adaptations or derivative works of or from an article (such as a translation), to include in a collective work (such as an anthology), to text or data mine the article, including for commercial purposes without permission from Elsevier

**CC BY NC SA:** For non-commercial purposes you may distribute and copy the article, create extracts, abstracts and other revised versions, adaptations or derivative works of or from an article (such as a translation), to include in a collective work (such as an anthology), to text and data mine the article and license new adaptations or creations under identical terms without permission from Elsevier

**CC BY NC ND:** For non-commercial purposes you may distribute and copy the article and include it in a collective work (such as an anthology), provided you do not alter or modify the article, without permission from Elsevier

Any commercial reuse of Open Access articles published with a CC BY NC SA or CC BY NC ND license requires permission from Elsevier and will be subject to a fee.

Commercial reuse includes:

- Promotional purposes (advertising or marketing)
- Commercial exploitation ( e.g. a product for sale or loan)
- Systematic distribution (for a fee or free of charge)

Please refer to [Elsevier's Open Access Policy](#) for further information.

## 21. Other Conditions:

v1.7

**If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 48 hours of the license date. Payment should be in the form of a check or money order referencing your account number and this invoice number 501339969. Once you receive your invoice for this order, you may pay your invoice by credit card. Please follow instructions provided at that time.**

**Make Payment To:**

**For suggestions or comments regarding this order, contact RightsLink Customer Support:  
or (toll free in the US) or**

**Gratis licenses (referencing \$0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.**

---

## E.3 Permission to reproduce previously published material in Chapter 4

### JOHN WILEY AND SONS LICENSE TERMS AND CONDITIONS

Jun 30, 2014



This is a License Agreement between Eli Gibson ("You") and John Wiley and Sons ("John Wiley and Sons") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by John Wiley and Sons, and the payment terms and conditions.

**All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.**

License Number	3418800692399
License date	Jun 30, 2014
Licensed content publisher	John Wiley and Sons
Licensed content publication	Journal of Magnetic Resonance Imaging
Licensed content title	Registration of prostate histology images to ex vivo MR images via strand-shaped fiducials
Licensed copyright line	Copyright © 2012 Wiley Periodicals, Inc.
Licensed content author	Eli Gibson, Cathie Crukley, Mena Gaed, José A. Gómez, Madeleine Moussa, Joseph L. Chin, Glenn S. Bauman, Aaron Fenster, Aaron D. Ward
Licensed content date	Jul 31, 2012
Start page	1402
End page	1412
Type of use	Dissertation/Thesis
Requestor type	Author of this Wiley article
Format	Print and electronic
Portion	Full article
Will you be translating?	No
Title of your thesis / dissertation	3D FUSION OF HISTOLOGY TO MULTI-PARAMETRIC MRI FOR PROSTATE CANCER IMAGING EVALUATION AND LESION-TARGETED TREATMENT PLANNING
Expected completion date	Aug 2014
Expected size (number of pages)	320
Total	0.00 USD
Terms and Conditions	

### TERMS AND CONDITIONS

This copyrighted material is owned by or exclusively licensed to John Wiley & Sons, Inc. or one of

its group companies (each a "Wiley Company") or handled on behalf of a society with which a Wiley Company has exclusive publishing rights in relation to a particular work (collectively "WILEY"). By clicking accept in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the billing and payment terms and conditions established by the Copyright Clearance Center Inc., ("CCC's Billing and Payment terms and conditions"), at the time that you opened your Rightslink account (these are available at any time at <http://myaccount.copyright.com>).

### **Terms and Conditions**

- The materials you have requested permission to reproduce or reuse (the "Wiley Materials") are protected by copyright.
- You are hereby granted a personal, non-exclusive, non-sub licensable (on a stand-alone basis), non-transferable, worldwide, limited license to reproduce the Wiley Materials for the purpose specified in the licensing process. This license is for a one-time use only and limited to any maximum distribution number specified in the license. The first instance of republication or reuse granted by this licence must be completed within two years of the date of the grant of this licence (although copies prepared before the end date may be distributed thereafter). The Wiley Materials shall not be used in any other manner or for any other purpose, beyond what is granted in the license. Permission is granted subject to an appropriate acknowledgement given to the author, title of the material/book/journal and the publisher. You shall also duplicate the copyright notice that appears in the Wiley publication in your use of the Wiley Material. Permission is also granted on the understanding that nowhere in the text is a previously published source acknowledged for all or part of this Wiley Material. Any third party content is expressly excluded from this permission.
- With respect to the Wiley Materials, all rights are reserved. Except as expressly granted by the terms of the license, no part of the Wiley Materials may be copied, modified, adapted (except for minor reformatting required by the new Publication), translated, reproduced, transferred or distributed, in any form or by any means, and no derivative works may be made based on the Wiley Materials without the prior permission of the respective copyright owner. You may not alter, remove or suppress in any manner any copyright, trademark or other notices displayed by the Wiley Materials. You may not license, rent, sell, loan, lease, pledge, offer as security, transfer or assign the Wiley Materials on a stand-alone basis, or any of the rights granted to you hereunder to any other person.
- The Wiley Materials and all of the intellectual property rights therein shall at all times remain the exclusive property of John Wiley & Sons Inc, the Wiley Companies, or their respective licensors, and your interest therein is only that of having possession of and the right to reproduce the Wiley Materials pursuant to Section 2 herein during the continuance of this Agreement. You agree that you own no right, title or interest in or to the Wiley Materials or any of the intellectual property rights therein. You shall have no rights hereunder other than the license as provided for above in Section 2. No right, license or interest to any trademark, trade name, service mark or other branding ("Marks") of WILEY or its licensors is granted

hereunder, and you agree that you shall not assert any such right, license or interest with respect thereto.

- NEITHER WILEY NOR ITS LICENSORS MAKES ANY WARRANTY OR REPRESENTATION OF ANY KIND TO YOU OR ANY THIRD PARTY, EXPRESS, IMPLIED OR STATUTORY, WITH RESPECT TO THE MATERIALS OR THE ACCURACY OF ANY INFORMATION CONTAINED IN THE MATERIALS, INCLUDING, WITHOUT LIMITATION, ANY IMPLIED WARRANTY OF MERCHANTABILITY, ACCURACY, SATISFACTORY QUALITY, FITNESS FOR A PARTICULAR PURPOSE, USABILITY, INTEGRATION OR NON-INFRINGEMENT AND ALL SUCH WARRANTIES ARE HEREBY EXCLUDED BY WILEY AND ITS LICENSORS AND WAIVED BY YOU
- WILEY shall have the right to terminate this Agreement immediately upon breach of this Agreement by you.
- You shall indemnify, defend and hold harmless WILEY, its Licensors and their respective directors, officers, agents and employees, from and against any actual or threatened claims, demands, causes of action or proceedings arising from any breach of this Agreement by you.
- IN NO EVENT SHALL WILEY OR ITS LICENSORS BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR ENTITY FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL, INDIRECT, EXEMPLARY OR PUNITIVE DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, PROVISIONING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.
- Should any provision of this Agreement be held by a court of competent jurisdiction to be illegal, invalid, or unenforceable, that provision shall be deemed amended to achieve as nearly as possible the same economic effect as the original provision, and the legality, validity and enforceability of the remaining provisions of this Agreement shall not be affected or impaired thereby.
- The failure of either party to enforce any term or condition of this Agreement shall not constitute a waiver of either party's right to enforce each and every term and condition of this Agreement. No breach under this agreement shall be deemed waived or excused by either party unless such waiver or consent is in writing signed by the party granting such waiver or consent. The waiver by or consent of a party to a breach of any provision of this Agreement shall not operate or be construed as a waiver of or consent to any other or subsequent

breach by such other party.

- This Agreement may not be assigned (including by operation of law or otherwise) by you without WILEY's prior written consent.
- Any fee required for this permission shall be non-refundable after thirty (30) days from receipt by the CCC.
- These terms and conditions together with CCC's Billing and Payment terms and conditions (which are incorporated herein) form the entire agreement between you and WILEY concerning this licensing transaction and (in the absence of fraud) supersedes all prior agreements and representations of the parties, oral or written. This Agreement may not be amended except in writing signed by both parties. This Agreement shall be binding upon and inure to the benefit of the parties' successors, legal representatives, and authorized assigns.
- In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall prevail.
- WILEY expressly reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.
- This Agreement will be void if the Type of Use, Format, Circulation, or Requestor Type was misrepresented during the licensing process.
- This Agreement shall be governed by and construed in accordance with the laws of the State of New York, USA, without regards to such state's conflict of law rules. Any legal action, suit or proceeding arising out of or relating to these Terms and Conditions or the breach thereof shall be instituted in a court of competent jurisdiction in New York County in the State of New York in the United States of America and each party hereby consents and submits to the personal jurisdiction of such court, waives any objection to venue in such court and consents to service of process by registered or certified mail, return receipt requested, at the last known address of such party.

## WILEY OPEN ACCESS TERMS AND CONDITIONS

Wiley Publishes Open Access Articles in fully Open Access Journals and in Subscription journals offering Online Open. Although most of the fully Open Access journals publish open access articles under the terms of the Creative Commons Attribution (CC BY) License only, the subscription journals and a few of the Open Access Journals offer a choice of Creative Commons Licenses: Creative Commons Attribution (CC-BY) license [Creative Commons Attribution Non-Commercial \(CC-BY-NC\) license](#) and [Creative Commons Attribution Non-Commercial-NoDerivs \(CC-BY-NC-ND\) License](#). The license type is clearly identified on the article.

Copyright in any research article in a journal published as Open Access under a Creative Commons License is retained by the author(s). Authors grant Wiley a license to publish the article and identify itself as the original publisher. Authors also grant any third party the right to use the article freely as long as its integrity is maintained and its original authors, citation details and publisher are identified as follows: [Title of Article/Author/Journal Title and Volume/Issue. Copyright (c) [year] [copyright owner as specified in the Journal]]. Links to the final article on Wiley's website are encouraged where applicable.

### **The Creative Commons Attribution License**

The [Creative Commons Attribution License \(CC-BY\)](#) allows users to copy, distribute and transmit an article, adapt the article and make commercial use of the article. The CC-BY license permits commercial and non-commercial re-use of an open access article, as long as the author is properly attributed.

The Creative Commons Attribution License does not affect the moral rights of authors, including without limitation the right not to have their work subjected to derogatory treatment. It also does not affect any other rights held by authors or third parties in the article, including without limitation the rights of privacy and publicity. Use of the article must not assert or imply, whether implicitly or explicitly, any connection with, endorsement or sponsorship of such use by the author, publisher or any other party associated with the article.

For any reuse or distribution, users must include the copyright notice and make clear to others that the article is made available under a Creative Commons Attribution license, linking to the relevant Creative Commons web page.

To the fullest extent permitted by applicable law, the article is made available as is and without representation or warranties of any kind whether express, implied, statutory or otherwise and including, without limitation, warranties of title, merchantability, fitness for a particular purpose, non-infringement, absence of defects, accuracy, or the presence or absence of errors.

### **Creative Commons Attribution Non-Commercial License**

The [Creative Commons Attribution Non-Commercial \(CC-BY-NC\) License](#) permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.(see below)

### **Creative Commons Attribution-Non-Commercial-NoDerivs License**

The [Creative Commons Attribution Non-Commercial-NoDerivs License](#) (CC-BY-NC-ND) permits use, distribution and reproduction in any medium, provided the original work is properly cited, is not used for commercial purposes and no modifications or adaptations are made. (see below)

### **Use by non-commercial users**

For non-commercial and non-promotional purposes, individual users may access, download, copy, display and redistribute to colleagues Wiley Open Access articles, as well as adapt, translate, text-



and data-mine the content subject to the following conditions:

- The authors' moral rights are not compromised. These rights include the right of "paternity" (also known as "attribution" - the right for the author to be identified as such) and "integrity" (the right for the author not to have the work altered in such a way that the author's reputation or integrity may be impugned).
- Where content in the article is identified as belonging to a third party, it is the obligation of the user to ensure that any reuse complies with the copyright policies of the owner of that content.
- If article content is copied, downloaded or otherwise reused for non-commercial research and education purposes, a link to the appropriate bibliographic citation (authors, journal, article title, volume, issue, page numbers, DOI and the link to the definitive published version on **Wiley Online Library**) should be maintained. Copyright notices and disclaimers must not be deleted.
- Any translations, for which a prior translation agreement with Wiley has not been agreed, must prominently display the statement: "This is an unofficial translation of an article that appeared in a Wiley publication. The publisher has not endorsed this translation."

#### **Use by commercial "for-profit" organisations**

Use of Wiley Open Access articles for commercial, promotional, or marketing purposes requires further explicit permission from Wiley and will be subject to a fee. Commercial purposes include:

- Copying or downloading of articles, or linking to such articles for further redistribution, sale or licensing;
- Copying, downloading or posting by a site or service that incorporates advertising with such content;
- The inclusion or incorporation of article content in other works or services (other than normal quotations with an appropriate citation) that is then available for sale or licensing, for a fee (for example, a compilation produced for marketing purposes, inclusion in a sales pack)
- Use of article content (other than normal quotations with appropriate citation) by for-profit organisations for promotional purposes
- Linking to article content in e-mails redistributed for promotional, marketing or educational purposes;
- Use for the purposes of monetary reward by means of sale, resale, licence, loan, transfer or other form of commercial exploitation such as marketing products

- Print reprints of Wiley Open Access articles can be purchased from:

Further details can be found on Wiley Online Library  
<http://olabout.wiley.com/WileyCDA/Section/id-410895.html>

Other Terms and Conditions:

v1.9

**If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 48 hours of the license date. Payment should be in the form of a check or money order referencing your account number and this invoice number 501339952. Once you receive your invoice for this order, you may pay your invoice by credit card. Please follow instructions provided at that time.**

**Make Payment To:**

**For suggestions or comments regarding this order, contact RightsLink Customer Support:  
or (toll free in the US) or**

**Gratis licenses (referencing \$0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.**

---

## E.4 Permission to reproduce previously published material in Chapter 5



**American Association of Physicists in Medicine**

<http://www.aapm.org>

*Office of the Executive Director*  
Angela R. Keyser

**DATE OF REQUEST:** June 30, 2014

**FROM:**

Eli Gibson

**1. Permission is granted to:**

Eli Gibson

**2. Permission is requested to use the following material:**

E. Gibson, M. Gaed, J. A. Gómez, M. Moussa, C. Romagnoli, S. Pautler, J. L. Chin, C. Crukley, G. S. Bauman, A. Fenster and A. D. Ward, "3D prostate histology reconstruction: an evaluation of image-based and fiducial-based algorithms," *Medical Physics* 40, 093501 pp. 1-15 (2013)

**3. For what purpose:**

As a chapter in my PhD thesis

Authors seeking permission must also notify the first author of the article from which permission is being sought.

Permission is hereby granted:

 Signature

\_\_\_\_\_  
Date

The Association's Journals are *Medical Physics* and *Journal of Applied Clinical Medical Physics*  
Member Society of the American Institute of Physics and the International Organization of Medical Physics

# Curriculum Vitae

## University Education

- 2010– Ph.D. Biomedical Engineering (in progress)  
*3D fusion of histology to multi-parametric MRI for prostate cancer imaging evaluation and lesion-targeted treatment planning*  
 Supervisors: Aaron Fenster, Ph.D., FCCPM; Aaron D. Ward, Ph.D.  
 Roberts Research Institute, The University of Western Ontario, London, Canada
- 2007–2009 M.A.Sc. Engineering  
*On the analysis of cortical surface biomarkers*  
 Supervisor: Mirza Faisal Beg, Ph.D.  
 Medical Image Analysis Laboratory, Simon Fraser University, Vancouver, Canada
- 2001–2006 B.A.Sc. Computer Engineering; Minor in Computing Science  
 Simon Fraser University, Vancouver, Canada

## Undergraduate Honours, Scholarships and Awards

- 2014 Imaging Network Ontario Symposium — second place poster award
- 2014 CIHR Postdoctoral fellowship (\$100,000 award over 2 years, 170 awarded in Canada)
- 2013 Prostate Cancer Canada travel award
- 2013 London Imaging Discovery Day — oral presentation honourable mention
- 2013 Ontario Graduate Scholarship with Distinction (\$16,500 award, 1 of 20 awarded at UWO)
- 2013 Imaging Network Ontario Symposium — first place poster award
- 2012 OICR/CIHR Prostate Cancer Imaging Workshop — poster award
- 2012 MICCAI — Young Scientist Award (248 8-page submissions accepted with 68% rejection rate; 5 Young Scientist Awards given out)
- 2012 London Imaging Discovery Day — first place poster award
- 2012 Queen Elizabeth II Graduate Scholarship in Science and Technology (\$15,000 award)
- 2012 Imaging Network Ontario Symposium — first place poster award
- 2011 Pathology Informatics — oral presentation honorable mention
- 2011 London Imaging Discovery Day — third place poster award
- 2011 Oncology Research and Education Day — poster award
- 2011 Imaging Network Ontario Symposium — second place poster award
- 2010– NSERC CREATE CAMI Training Program (\$30,000 stipend over 2 years)
- 2010–2011 NSERC Postgraduate Scholarship — Doctoral Level (\$42,000 award over 2 years)
- 2007–2008 NSERC Canadian Graduate Scholarship — Master's Level (\$35,000 award over 2 years)
- 2007 Dean's Convocation Medal for Undergraduate Studies — Applied Science
- 2007 Kaiser Foundation
- 2007 APEGBC Gold Medal (most outstanding graduating student)
- 2006 NSERC Undergraduate Research Award
- 2006 MDSI – Peter Kam Scholarship (innovative engineering project work)
- 2006 Western Engineering Competition — first in Innovative Design
- 2006 Canadian Engineering Competition — third in Innovative Design
- 2001–2006 Dean's Honour Roll
- 2001–2006 SFU Open Scholarship
- 2005 NSERC Undergraduate Research Award
- 2003 Vancouver Island Engineering Institute Scholarship
- 2002 Top GPA in first year class
- 2002 Top 5 Canadian candidates for Lucent Global Science Scholarship
- 2001 SFU Summit Entrance Scholarship

## Publications, Presentations and Abstracts

### **Articles in refereed journals**

1. M. Shahedi, D. W. Cool, C. Romagnoli, G. Bauman, M. Bastian-Jordan, **E. Gibson**, G. B. Rodrigues, B. Ahmad, M. Lock, A. Fenster, A. D. Ward. Spatially Varying Accuracy and Repeatability of Prostate Segmentation in Magnetic Resonance Images using Manual and Semi-Automated Methods. *Medical Physics*, 2014. (accepted pending revisions)
2. Y. Xu, J. G. Pickering, Z. Nong, **E. Gibson**, A. D. Ward. A method for evaluating normal and regenerated microvasculature using 3D histopathology. *PLOS ONE*, 2014. (submitted)
3. W. T. Hrinivich, **E. Gibson**, T. Thang, C. A. McKenzie, G. S. Bauman, A. D. Ward, A. Fenster, E. Wong. A comparison of pharmacokinetic parameters in the prostate determined using linear reference region models and measured arterial input function from dynamic contrast enhanced MRI. *Magnetic Resonance in Medicine*, 2014. (submitted)
4. **E. Gibson**, M. Gaed, J. A. Gómez, M. Moussa, S. Pautler, J. L. Chin, C. Crukley, G. S. Bauman, A. Fenster, A. D. Ward. 3D prostate histology image reconstruction: quantifying the impact of tissue deformation and histology section location. *Journal of Pathology Informatics*, 4(1), 31, 2013.
5. **E. Gibson**, M. Gaed, J. A. Gómez, M. Moussa, C. Romagnoli, S. Pautler, J. L. Chin, C. Crukley, G. S. Bauman, A. Fenster, A. D. Ward. 3D prostate histology reconstruction: an evaluation of image-based and fiducial-based algorithms. *Medical Physics*, 40(9), 093501, 2013.
6. **E. Gibson**, A. Fenster, A. D. Ward. The impact of registration accuracy on imaging validation study design: a novel statistical power calculation. *Medical Image Analysis*, 17(7), pp. 805–15, 2013.
7. **E. Gibson**, C. Crukley, M. Gaed, J. A. Gómez-Lemus, M. Moussa, J. Chin, G. Bauman, A. Fenster, A. D. Ward. Registration of prostate histology images to ex vivo MR images via strand-shape fiducials. *Journal of Magnetic Resonance Imaging*, 36(6), pp. 1402–12, 2012.
8. A. D. Ward, C. Crukley, C. McKenzie, J. Montreuil, **E. Gibson**, C. Romagnoli, J. A. Gómez-Lemus, M. Moussa, J. Chin, G. Bauman, A. Fenster. Registration of prostate digital histopathology images to in vivo MRI acquired using an endorectal receive coil. *Radiology*, 263(3), pp. 856–64, 2012.
9. H. L. More, J. Chen, **E. Gibson**, J. M. Donelan, M. F. Beg. A semi-automated method for identifying and measuring myelinated nerve fibers in scanning electron microscope images. *Journal of Neuroscience Methods*, 201(1), pp. 149–158, 2011.
10. M. Sarunic, A. Yazdanpanah, **E. Gibson**, J. Xu, Y. Bai, S. Lee, H. U. Saragovi, M. F. Beg. Longitudinal Study of Retinal Degeneration in Rats using Spectral Domain Optical Coherence Tomography. *Optics Express*, 18(22), pp. 23435–23441, 2010.
11. **E. Gibson**, M. Young, M. Sarunic, M. F. Beg. Optical Nerve Head Registration via Hemispherical Surface and Volume Registration. *IEEE Transactions on Biomedical Engineering Letters*, 57(10), pp. 2593–2595, 2010.

### **Book chapters**

1. **E. Gibson**, C. Crukley, M. Gaed, J. A. Gómez-Lemus, M. Moussa, C. Romagnoli, S. Pautler, J. Chin, G. Bauman, A. Fenster, A. D. Ward. 3D registration of whole-mount prostate histology images to ex vivo magnetic resonance images using strand-shaped fiducials. In *Abdomen and Thoracic Imaging: An Engineering & Clinical Perspective*, A. S. El-Baz, L. Saba, J. Suri (Ed.). (

### **Refereed conference proceedings**

1. **E. Gibson**, M. Gaed, T. Hrinivich, J. A. Gómez, M. Moussa, C. Romagnoli, J. Mandel, M. Bastian-Jordan, D. Cool, S. Ghoul, S. Pautler, J. L. Chin, C. Crukley, G. S. Bauman, A. Fenster, A. D. Ward. Multi-parametric MR imaging of prostate cancer foci: assessing the detectability and localizability of Gleason 7 peripheral zone cancers based on image contrasts. In *SPIE Medical Imaging*, San Diego, USA, Feb. 2014.
2. T. Hrinivich, **E. Gibson**, M. Gaed, J. A. Gómez, M. Moussa, C. McKenzie, G. Bauman, A. D. Ward, A. Fenster, E. Wong. A dimensionless dynamic contrast enhanced MRI parameter for intra-prostatic

- tumour target volume delineation: initial comparison with histology. In *SPIE Medical Imaging*, San Diego, USA, Feb. 2014.
3. M. Salarian, **E. Gibson**, M. Shahedi, M. Gaed, J. A. Gómez, M. Moussa, C. Romagnoli, D. W. Cool, M. Bastian-Jordan, J. L. Chin, S. Pautler, G. S. Bauman, A. D. Ward. Accuracy and variability of tumour burden measurement on multi-parametric MRI. In *SPIE Medical Imaging*, San Diego, USA, Feb. 2014. (**podium** presentation by M. Salarian)
  4. Y. Xu, J. Pickering, Z. Nong, **E. Gibson**, A. D. Ward. 3D reconstruction of digitized histological sections for vasculature quantification in the mouse hind limb. In *SPIE Medical Imaging*, San Diego, USA, Feb. 2014. (**podium** presentation by Y. Xu)
  5. F. Imani, P. Abolmaesumi, **E. Gibson**, A. Khojaste, M. Gaed, M. Moussa, J. A. Gómez, C. Romagnoli, D. R. Siemens, M. Leviridge, S. Chang, Fenster, A. D. Ward, P. Mousavi. Ultrasound-based characterization of prostate cancer: an in vivo clinical feasibility study. In *Medical Image Computing and Computer-Assisted Intervention*, Lecture Notes in Computer Science 8150, pp. 279–286, Nagoya, Japan, Sep. 2013. (peer-reviewed manuscript; 67% rejection rate)
  6. **E. Gibson**, M. Gaed, J. A. Gómez-Lemus, M. Moussa, C. Romagnoli, J. L. Chin, C. Crukley, G. S. Bauman, A. Fenster, A. D. Ward. 3D prostate histology reconstruction: an evaluation of image-based and fiducial-based algorithms. In *SPIE Medical Imaging: Digital Pathology*, 8676, pp. 86760A, Orlando, USA, Mar. 2013. (**podium** presentation by **E. Gibson**)
  7. M. Salarian, M. Shahedi, **E. Gibson**, M. Gaed, J. A. Gómez-Lemus, M. Moussa, G. S. Bauman, A. D. Ward. Toward quantitative digital histopathology for prostate cancer: comparison of inter-slide interpolation methods for tumour measurement. In *SPIE Medical Imaging: Digital Pathology*, 8676, pp. 86760F, Orlando, USA, Mar. 2013. (**podium** presentation by M. Salarian)
  8. **E. Gibson**, A. Fenster, A. D. Ward. Registration accuracy: how good is good enough? A statistical power calculation incorporating image registration uncertainty. In *Medical Image Computing and Computer-Assisted Intervention*, Lecture Notes in Computer Science 7511, pp. 643–650, Nice, France, Oct. 2012. (**podium** presentation by **E. Gibson**, **Young Scientist Award**: 248 8-page submissions accepted with 68% rejection rate; 5 Young Scientist Awards given out)
  9. **E. Gibson**, C. Crukley, J. A. Gómez-Lemus, M. Moussa, G. Bauman, A. Fenster, A. D. Ward. 3D reconstruction of prostate histology based on quantified tissue cutting and deformation parameters. In *SPIE Medical Imaging*, 8317, pp. 83170N, San Diego, USA, Mar. 2012. (**podium** presentation by **E. Gibson**)
  10. **E. Gibson**, C. Crukley, V. Karnik, M. Gaed, J. A. Gómez-Lemus, M. Moussa, J. L. Chin, G. Bauman, A. Fenster, A. D. Ward. Co-registration framework for histology-registration-based validation of fused multimodality prostate cancer imaging. In *IEEE International Conference on Intelligent Computation and Bio-Medical Instrumentation*, pp. 93–6, Wuhan, China, Dec. 2011. (**podium** presentation by **E. Gibson**)
  11. **E. Gibson**, C. Crukley, M. Gaed, J. A. Gómez-Lemus, M. Moussa, J. L. Chin, G. Bauman, A. Fenster, A. D. Ward. Validation of direct registration of whole-mount prostate digital histopathology to ex vivo MR images. In *Medical Image Computing and Computer-Assisted Intervention Prostate Cancer Imaging Workshop*, Lecture Notes in Computer Science 6963, pp. 134–145, Toronto, Canada, Sep. 2011. (**podium** presentation by **E. Gibson** and J. L. Chin)
  12. **E. Gibson**, C. Crukley, J. A. Gómez-Lemus, M. Moussa, G. Bauman, A. Fenster, A. D. Ward. Tissue block MRI for slice-orientation-independent registration of digital histology images to ex vivo MRI of the prostate. In *IEEE International Symposium on Biomedical Imaging*, pp. 566–569, Chicago, USA, Mar. 2011.
  13. **E. Gibson**, A. Fenster, C. Crukley, C. McKenzie, J. A. Gómez-Lemus, M. Moussa, G. Bauman, A. D. Ward. Image-guided prostate sectioning supporting registration of graded cancerous foci from digital histopathology images to in vivo MRI: An interactive 3D visualization tool. In *SPIE Medical Imaging*, 7965, pp. 79650D, Orlando, USA, Feb. 2011. (**podium** presentation by **E. Gibson**)
  14. A. D. Ward, C. Crukley, C. McKenzie, J. Montreuil, **E. Gibson**, J. A. Gómez-Lemus, M. Moussa, G. Bauman, A. Fenster. Registration of in vivo prostate magnetic resonance images to digital histopathology images. In *Medical Image Computing and Computer-Assisted Intervention Prostate Imaging Workshop*, Lecture Notes in Computer Science 6367, pp. 66–76, Beijing, China, Sep. 2010.

15. **E. Gibson**, A. Khan, M. F. Beg. A combined surface and volumetric registration (SAVOR) framework to study cortical biomarkers and functional neuroimaging data. In *Medical Image Computing and Computer Assisted Intervention*, Lecture Notes in Computer Science 5761, pp. 713–720, London, UK, Sep. 2009. (peer-reviewed manuscript; 68% rejection rate)
16. M. Young, S. Lee, **E. Gibson**, K. Hsu, M. F. Beg, P. J. Mackenzie, M. V. Sarunic. Morphometric analysis of the optic nerve head with optical coherence tomography. In *SPIE Optical Coherence Tomography and Coherence Domain Optical Methods in Biomedicine*, 7554, pp. 75542L, San Francisco, USA, Jan. 2010
17. J. A. Hoffer, **E. Gibson**, B. Mädler, M. F. Beg. Estimation of nerve dimensions from MRI of the human thigh. 13th Ann. Conf. Int'l. Functional Electrical Stimulation Society, Freiburg, Germany, Sep. 2008.

### Refereed conference abstracts

1. D. W. Cool, E. Gibson, M. Gaed, J. Gómez, M. Moussa, A. D. Ward, S. Pautler, J. L. Chin, J. I. Izawa, A. Fenster, G. S. Bauman, C. Romagnoli. Prostate Imaging Reporting and Data System (PIRADS) for Multiparametric Prostate MRI (MP-MRI) Interpretation: Radiologic-Pathologic Correlation. Radiological Society of North America, Chicago, USA, Dec 2014.
2. E. Gibson, M. Gaed, J. A. Gómez, M. Moussa, C. Romagnoli, M. Bastian-Jordan, Z. Kassam, D. W. Cool, S. Pautler, J. L. Chin, C. Crukley, G. S. Bauman, A. Fenster, A. D. Ward. Toward prostate cancer contouring guidelines on MRI: dominant lesion gross and clinical target volume coverage via accurate histology fusion. American Society for Radiation Oncology Annual Meeting, San Francisco, USA, Sep. 2014. (accepted)
3. A. Farag, J. Peterson, T. Wade, T. Szekeres, A. Ouriadove, **E. Gibson**, A. D. Ward, J. L. Chin, S. Pautler, G. Bauman, C. Romagnoli, R. Bartha, T. Scholl. Combined Transmit-Only Asymmetric Birdcage and Receive-Only Endorectal Surface RF Coils. International Society of Magnetic Resonance in Medicine, Milan, Italy, May 2014.
4. **E. Gibson**, M. Gaed, J. A. Gómez, M. Moussa, C. Romagnoli, J. Mandel, M. Bastian-Jordan, S. Ghoul, D. Cool, S. Pautler, J. L. Chin, C. Crukley, G. S. Bauman, A. Fenster, A. D. Ward. Toward contouring guidelines for prostate cancer focal therapy planning on MRI: characterisation of tumour boundary contrast via accurate pathology fusion. Ann. Meeting Radiological Society of North America, Chicago, USA, Dec. 2013. (**podium** presentation by **E. Gibson**)
5. G. Bauman, J. Mandel, C. Romagnoli, I. Rachinsky, A. D. Ward, **E. Gibson**, M. Gaed, M. Moussa, J. Gómez-Lemus, J. Chin, S. Pautler, J. Butler, E. Wong, T.-Y. Lee. 18-fluorocholine imaging of localized prostate cancer: first experience with hybrid PET/MRI. Ann. Meeting Canadian Organization of Medical Physicists-Canadian Association of Radiation Oncology, Montreal, Canada, Sep. 2013.
6. M. Salarian, M. Shahedi, **E. Gibson**, M. Gaed, J. A. Gómez, M. Moussa, D. W. Cool, C. Romagnoli, G. S. Bauman, A. D. Ward. Imaging validation and quantitative pathology in prostate cancer: shape interpolation methods for tumour measurement. Ann. Meeting Canadian Organization of Medical Physicists-Canadian Association of Radiation Oncology, Montreal, Canada, Sep. 2013. (**podium** presentation by M. Salarian)
7. **E. Gibson**, M. Gaed, J. A. Gómez, M. Moussa, C. Romagnoli, J. Mandel, M. Bastian-Jordan, S. Ghoul, D. Cool, S. Pautler, J. L. Chin, C. Crukley, G. S. Bauman, A. Fenster, A. D. Ward. Toward contouring guidelines for prostate cancer focal therapy planning on MRI: characterisation of tumour boundary contrast via accurate pathology fusion. Australian-Canadian Prostate Cancer Research Alliance Symposium, Port Douglas, Australia, Aug. 2013. (**Prostate Cancer Canada travel award**; one of seven posters selected for **podium** presentation)
8. **E. Gibson**, L. J. Friesen-Waldner, A. M. Hamilton, E. J. Ribot, T. P. Wade, C. N. Wiens, K. Thind, J. K. Harris, N. M. Borradaile, C. A. McKenzie, A. D. Ward. 3D co-registration of MRI and histology in a mouse model of obesity and non-alcoholic fatty liver disease. Ann. Meeting International Society for Magnetic Resonance in Medicine, Salt Lake City, USA, Apr. 2013.
9. **E. Gibson**, I. Rachinsky, M. Gaed, C. Romagnoli, T.-Y. Lee, C. Crukley, J. A. Gómez, M. Moussa, J. L. Chin, G. S. Bauman, A. Fenster, A. D. Ward. Ontario's first 18F choline prostate PET images:

- histology and multi-parametric MRI correlation. 67th Ann. Meeting Canadian Urological Association, Banff, Canada, Jun. 2012.
10. M. Alturkustani, K. Derry, **E. Gibson**, C. Crukley, A. Fenster, J. D. Spence, J. Brennan, R. deKemp, R. Beanlands, Y. Yarofeyeva, M. Yaffe, R. Hammond. 3D Radiology/Pathology Correlates in Carotid Atheroma (Canadian Atherosclerosis Imaging Network). Pathology Visions, San Diego, USA, Oct. 2011.
  11. **E. Gibson**, C. Crukley, V. Karnik, M. Gaed, J. A. Gómez, M. Moussa, J. L. Chin, G. Bauman, A. Fenster, A. D. Ward. 3D reconstruction of prostate histopathology for co-registration with in vivo imaging. Pathology Visions, San Diego, USA, Oct. 2011.
  12. **E. Gibson**, C. Crukley, J. A. Gómez-Lemus, M. Moussa, G. Bauman, A. Fenster, A. D. Ward. 3D prostate histology reconstruction informed by quantified tissue cutting and deformation parameters. Pathology Informatics, Pittsburgh, USA, Oct. 2011. (**podium presentation by E. Gibson, Honorable Mention**)
  13. M. Alturkustani, K. Derry, M. Tsoi, **E. Gibson**, C. Crukley, A. Fenster, J. D. Spence, G. Youssef, R. deKemp, R. Beanlands, Y. Yarofeyeva, M. Yaffe, R. Hammond. The role of 3D digital quantitative histopathology coregistration to ultrasound, PET-CT and MRI (Canadian Atherosclerosis Imaging Network). 51th Ann. Meeting Canadian Association of Neuropathologist, Vancouver, Canada, Sep. 2011. (**podium presentation by M Alturkustani, Morrison H. Finlayson award**)
  14. **E. Gibson**, C. Crukley, J. A. Gómez-Lemus, M. Moussa, J. L. Chin, G. Bauman, A. Fenster, A. D. Ward. Fiducial-based registration of digital histopathology to ex vivo prostate MRI. Ann. Meeting American Association of Physicists in Medicine and Canadian Organization of Medical Physicists, Vancouver, Canada, Jul. 2011. (**podium presentation by E. Gibson**)
  15. A. D. Ward, C. Crukley, C. McKenzie, J. Montreuil, **E. Gibson**, C. Romagnoli, J. A. Gomez, M. Moussa, J. Chin, G. Bauman, A. Fenster. Registration of digital histology images to in vivo MRI for prostate cancer focal therapy research. Ann. Meeting American Urological Association, Washington, DC, May 2011. (**podium presentation by J. Chin**)
  16. E. Lebed, **E. Gibson**, M. F. Beg. Surface-smoothing based gyrification index and its use in localizing changes in cortical folding. 16th Ann. Meeting Org. Human Brain Mapping, Barcelona, Spain, Jun. 2010.
  17. M. Young, S. Lee, **E. Gibson**, M. Rateb, M. F. Beg, P. J. Mackenzie, M. V. Sarunic. Comparison of reference planes in volumetric SD OCT for optic nerve head morphometry. Ann. Meeting Assoc. Research in Vision and Ophthalmology, Fort Lauderdale, USA, May 2010.
  18. **E. Gibson**, L. Wang, M. F. Beg. Cortical thickness measurement using Eulerian PDEs and surface-based global topological information. 15th Ann. Meeting Org. Human Brain Mapping, San Francisco, USA, Jun. 2009.
  19. J. Xu, A. Yazdanpanah, M. Young, S. Lee, **E. Gibson**, Y. Bai, Z. Shi, M. F. Beg, H. U. Saragovi, M. V. Sarunic. Non-invasive quantitative measurement of axotomy-induced retinal degeneration in rats using Fourier domain optical coherence tomography. Ann. Meeting Assoc. Research in Vision and Ophthalmology, Fort Lauderdale, USA, May 2009.

### Invited talks

1. **E. Gibson**, Imaging-histology fusion-based validation of prostate MRI for cancer therapy planning. Centre for Medical Image Computing Seminar Programme. London, UK. Jun. 4, 2014. Invited by Dean Barratt.
2. **E. Gibson**, Imaging-histology fusion-based validation of prostate MRI for cancer therapy planning. NIH Molecular Imaging Program Guest Lecture. Bethesda, USA. Apr. 24, 2014. London, UK. Invited by Peter Choyke.
3. **E. Gibson**, Prostate cancer MRI: a new evolving paradigm - what can we learn from MRI-histology fusion. Citywide Cancer Imaging Seminar. London, Canada. Jan. 22, 2014. Invited by David Palma and Aaron Ward.



4. **E. Gibson**, Imaging-pathology fusion supporting development of contouring guidelines for prostate cancer focal therapy planning on multi-parametric MRI. Techna Institute seminar series. Toronto, Canada. Nov. 18, 2013. Invited by Vladimir Pekar.
5. **E. Gibson**, Update on CIHR Team Project: Imaging-pathology fusion supporting development of contouring guidelines for prostate cancer focal therapy planning on multi-parametric MRI. CIHR Team Grant & OICR Smarter Imaging Program Prostate Workshop. London, Canada. Nov. 15, 2013. Invited by Glenn Bauman.
6. **E. Gibson**, The statistics of registration error. Robarts MedIA lecture series 2013. London, Canada. Sep. 10, 2013. Invited by Jing Yuan.
7. **E. Gibson**, Histology-based validation of imaging. Wuhan, China. December 17, 2011. Invited by Mingyue Ding.
8. A. D. Ward, **E. Gibson**, Update on CIHR Team Project: Imaging-Pathology Registration for Image-Guided Diagnosis and Therapy of Prostate Cancer, CIHR Team Grant & OICR Imaging Pipeline Platform One Millimetre Cancer Challenge Workshop. London, Canada. November 18, 2011. Invited by Glenn Bauman. (presentation by all authors)
9. **E. Gibson**, Using digital histology images to validate patient MRI of prostate cancer. University Hospital Histotechnologist Lunch and Learn. London, Canada. May 20, 2011. Invited by Debbie Gaskin.
10. J. A. Gómez, A. D. Ward, **E. Gibson**, Update on CIHR Team Project: Imaging-pathology registration for prostate cancer diagnosis CIHR workshop, CIHR Team Grant & OICR Imaging Pipeline Platform One Millimetre Cancer Challenge Workshop. London, Canada. November 19, 2010. Invited by Glenn Bauman. (presentation by all authors)

### Teaching Experience

2014	The University of Western Ontario BME 9519B/ECE 9202B/ECE9022/MBP 9519B/ECE 4438B – Advanced Image Processing and Analysis 2 hour invited lecture
2013	The University of Western Ontario BME 9511B/ECE 9511B/ECE9511B – Image-guided Interventions 2 hour invited lecture
2012	The University of Western Ontario BME 9599 – Graduate Seminar – Teaching Assistant
2012	The University of Western Ontario MBP 4475/MBP 9515/BME 9513 – Introduction to Medical Imaging – Teaching Assistant
2012	The University of Western Ontario BME 9519B/ECE 9202B/ECE9022/MBP 9519B/ECE 4438B – Advanced Image Processing and Analysis 2 hour invited lecture
2011	The University of Western Ontario ENGSCI 1036A – Programming Fundamentals for Engineers – Teaching Assistant
2010	The University of Western Ontario ENGSCI 1036A – Programming Fundamentals for Engineers – Teaching Assistant

### Teaching Training

2014	The University of Western Ontario MBP 9573 – Advanced Studies in Biophysics: Design of Medical imaging Labs
2013	The University of Western Ontario Advanced Teaching Program – 24 hour seminar series