
5-2014

Methodological Development Of A Multi-Parametric Quantitative Imaging Biomarker Framework For Assessing Treatment Response With Mri

Ryan J. Bosca

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



Part of the [Medicine and Health Sciences Commons](#)

Recommended Citation

Bosca, Ryan J., "Methodological Development Of A Multi-Parametric Quantitative Imaging Biomarker Framework For Assessing Treatment Response With Mri" (2014). *Dissertations and Theses (Open Access)*. 443.

https://digitalcommons.library.tmc.edu/utgsbs_dissertations/443

This Dissertation (PhD) is brought to you for free and open access by the MD Anderson UTHealth Houston Graduate School at DigitalCommons@TMC. It has been accepted for inclusion in Dissertations and Theses (Open Access) by an authorized administrator of DigitalCommons@TMC. For more information, please contact digcommons@library.tmc.edu.

METHODOLOGICAL DEVELOPMENT OF A MULTI-PARAMETRIC
QUANTITATIVE IMAGING BIOMARKER FRAMEWORK FOR
ASSESSING TREATMENT RESPONSE WITH MRI

by
Ryan Joseph Bosca, M.S.

APPROVED:

Supervisory Professor
Edward F. Jackson, Ph.D.

Lei Dong, Ph.D.

Valen E. Johnson, Ph.D.

Anita Mahajan, M.D.

R. Jason Stafford, Ph.D.

APPROVED:

Dean, The University of Texas
Graduate School of Biomedical Sciences at Houston

METHODOLOGICAL DEVELOPMENT OF A MULTI-PARAMETRIC
QUANTITATIVE IMAGING BIOMARKER FRAMEWORK FOR
ASSESSING TREATMENT RESPONSE WITH MRI

A

DISSERTATION

Presented to the Faculty of
The University of Texas
Health Science Center at Houston
and
The University of Texas
MD Anderson Cancer Center
Graduate School of Biomedical Sciences
in Partial Fulfillment
of the Requirements
for the Degree of
DOCTOR OF PHILOSOPHY

by

Ryan Joseph Bosca, M.S.

Houston, Texas

May 2014

Dedications

This dissertation is dedicated to Joan BeVier whose compassionate encouragement to achieve excellence through hard, honest work continues to inspire her grandchildren.

Acknowledgements

The work performed herein was only possible through the knowledge, wisdom, and encouragement of innumerable mentors and colleagues. First, I would like to express my most sincere gratitude for my supervisory professor, Dr. Edward Jackson. Despite an incomprehensibly busy schedule, he always found time to entertain my questions and concerns, providing guidance with the utmost understanding and insight. I would also like to thank the members of my supervisory committee for their dedication and assistance. Dr. Dong provided invaluable insight with regard to radiation therapy and image registration. Dr. Johnson diligently, and with great patience, led me through the statistical concepts necessary for performing this work. Dr. Mahajan's clinical insights, especially the patient response contours that she carefully defined, were and are still greatly appreciated. Dr. Stafford's magnetic resonance expertise and many, lengthy, helpful discussions, scientific and otherwise, greatly facilitated the completion of this work.

Numerous staff members and fellow graduate students were also instrumental in the preparation of this work. Specifically, I would like to thank Dr. Ken Hwang and Dr. Adam Chandler for a number of helpful discussions and assistance in navigating the GE equipment used in this work, Dr. Paul D. Brown for his assistance in defining patient response ROIs, Hua Ai, Moiz Ahmad, Laura Rechner, and Daniel Robertson for many helpful discussions, Jan Harasim for being immensely accommodating during the final months of this work, and the support staff of the medical physics program, the Department of Imaging Physics, and the Graduate School of Biomedical Sciences. I would also like to express many thanks to those faculty members and peers who suffered patiently through and aided with the deployment of QUATTRO.

Additionally, I am and will always be grateful for the unwavering support and understanding of my parents, Donald and Tonya Bosca, whose generosity and encouragement throughout the years have been truly inspirational, allowing me the flexibility to follow my dreams. Moreover, the love of my life,

Sonja Harasim, selflessly supported and encouraged my efforts despite the often outlandish commitments required; her infectious compassion was an endless source of inspiration.

METHODOLOGICAL DEVELOPMENT OF A MULTI-PARAMETRIC
QUANTITATIVE IMAGING BIOMARKER FRAMEWORK FOR
ASSESSING TREATMENT RESPONSE WITH MRI

Ryan Joseph Bosca, M.S.

Supervisory Professor: Edward F. Jackson, Ph.D.

Abstract

Quantitative imaging biomarkers (QIBs) are increasingly being incorporated into early phase clinical trials as a means of non-invasively assessing the spatially heterogeneous treatment response to anticancer therapies, particularly as indicators for early response. MR QIBs are derived from the analysis of *in vivo* imaging data, such as that acquired via dynamic contrast enhanced (DCE), dynamic susceptibility enhanced (DSC), and diffusion tensor imaging (DTI). To date, preclinical and clinical applications of such QIBs have provided strong evidence for potential efficacy, but efforts to create meaningful estimates of localized treatment response using multiple QIBs have been stifled by the need for rigorous characterization of biases and variances inherent in MR equipment and analysis tools and a suitable means of associating QIB changes with treatment response. This research sought to develop such a framework, incorporating multiple MRI QIBs associated with the microvascular environment, *e.g.*, permeability, flow, and volume, and the cellular environment, *e.g.*, water diffusion, into a single classification model to generate maps of predicted locoregional response. To ensure treatment associated changes measured *in vivo* exceeded equipment related levels of bias and variance, two phantoms were developed. Weekly assessment of the MR imaging data from which the QIBs were derived resulted in coefficients of variation less than 15% for QIBs assessed, well below the expected treatment related changes (approximately 40%). Bias and variance associated with the software tools developed to facilitate longitudinal assessments of treatment response, QUATTRO, was also assessed using synthesized imaging data mimicking clinically relevant acquisitions schema, and found to introduce negligible levels of bias and variance. Finally, to develop an integrated approach to assessing response using multiple QIBs, two experienced radiation oncologists contoured regions of partial response (PR), stable disease (SD), and progressive disease (PD) on rigidly co-registered high grade brain tumor patient data sets, which included DCE, DSC, and DTI acquisitions. Response matched voxel-by-voxel QIBs were trained using an ordinal regression classifier. Using leave-one-out cross-validation, the prediction accuracies of the best model (single DTI QIB) were found to be, mean (standard error), 69.0 (11.1)% for SD, 35.2 (11.7)% for PD, and 52.3(9.7)% overall. In summary, this work resulted in the development of a comprehensive framework for predicting voxelwise radiological treatment response, including the development of phantoms and associated acquisitions for MR equipment quality control and establishment of system-related bias and variance, and a comprehensive software package for performing related image analyses and outcome prediction.

Table of Contents

Dedications	iii
Acknowledgements	iv
Abstract	vii
Table of Contents	viii
List of Figures.....	xi
List of Tables.....	xx
List of Abbreviations.....	xxiii
Chapter 1 Introduction.....	1
1.1 The Role of Quantitative Imaging Biomarkers in Treatment Assessment	1
1.2 Diffusion	3
1.2.1 Perfusion and Vascular Permeability.....	5
1.2.2 Challenges.....	8
1.2.3 Mapping Response using Multiple QIBs.....	9
1.3 Hypothesis and Specific Aims	10
Chapter 2 Background and Theory.....	12
2.1 Quantitative Imaging.....	12
2.1.1 Diffusion	12
2.1.2 T_1 Relaxometry	20
2.1.3 Dynamic-Contrast Enhanced-MRI	23
2.1.4 Dynamic-Susceptibility Contrast-MRI.....	34
2.2 Statistics.....	40
2.2.1 Measuring Agreement and Repeatability	40
2.2.2 Logistic and Ordinal Regression.....	41
2.2.3 Cross-Validation.....	44
Chapter 3 Assessing MR Equipment Bias and Variance	46

3.1	Introduction.....	46
3.2	Phantom Design	47
3.2.1	DCE Phantom.....	47
3.2.2	ADC Phantom	48
3.3	Methods	50
3.3.1	MR Acquisition Protocols	50
3.3.2	Equipment Variance Quantitation.....	52
3.4	Results	55
3.4.1	Equipment Variance	55
3.5	Discussion	75
Chapter 4	A Unified Framework for Assessing Treatment Response	79
4.1	Introduction.....	79
4.2	Software Implementation, Validation, and Optimization	80
4.2.1	Models and Model Validation	82
4.2.2	Image Registration and Registration Validation.....	91
4.2.3	Code Optimization.....	92
4.3	Results	93
4.3.1	Relaxometry Validation	93
4.3.2	Diffusion Validation	96
4.3.3	DCE Validation	99
4.3.4	DSC Validation	104
4.3.5	Image Registration.....	106
4.3.6	Code Optimization.....	107
4.4	Discussion	107
Chapter 5	Radiological Response Maps	110
5.1	Introduction.....	110
5.2	Patient Acquisitions.....	110

5.3	Methods and Materials	111
5.3.1	Image Registrations	112
5.3.2	Quantitative Image Analysis	114
5.3.3	Radiological Assessment	117
5.3.4	Statistical Classification and Response Maps	117
5.4	Results	120
5.4.1	Data Preparation	120
5.4.2	Qualitative Inspection	123
5.4.3	Statistical Classification and Response Maps	135
5.5	Discussion	145
Chapter 6	Conclusions and Future Directions.....	147
6.1	Evaluation of the Hypothesis.....	147
6.2	Future Directions	150
	References	153
	Vita	179

List of Figures

Figure 2-1. Two compartment kinetic model schematic. The pharmacokinetic model most frequently used in the analysis of DCE-MRI data consists of two compartments separated by the endothelium: an extracellular extravascular space (left) and a vascular space (right). K^{trans} describes the transfer rate constant of contrast agent from the vascular to EES space, v_e describes the fractional EES volume, and v_p describes the fractional plasma volume. The parameter k_{ep} is defined as K^{trans}/v_e . Arrows indicate blood flow through the vascular space.	26
Figure 2-2. Contrast agent recirculation example. This plot shows a VIF contrast uptake curve with the solid line representing a fitted gamma-variate and the black arrow specifying the characteristic recirculation peak.....	37
Figure 3-1. DCE and ADC phantom schematics. Axial cross-sectional schematic representation of the DCE (<i>left</i>) and ADC (<i>right</i>) phantom (units: cm). The DCE phantom inner ring (<i>red</i>) contains 8 solutions representative of VIF response (R_1 : 1.05-41.86 s ⁻¹) while the outer ring (<i>blue</i>) contains 8 solutions representative of tissue response (R_1 : 0.97-7.84 s ⁻¹). The ADC phantom contained vials with various sucrose concentrations (10 to 45 %w/w); the lowest sucrose concentration was placed at the center.	47
Figure 3-2. Comparison of VFA R_1 estimation with theory. Bland-Altman plots (<i>top</i>) of the longitudinally averaged VFA R_1 measurements vs. theoretical R_1 values at 1.5T (<i>left</i>) and 3.0T (<i>right</i>). The 8 VIF (<i>squares</i>) and tissue (<i>diamonds</i>) compartments were given different markers for clarity and the dashed lines are the limits of agreement. Correlation plots (bottom) are shown for comparison with the linear regression and corresponding coefficient of determination for 1.5T (<i>left</i>) displayed above the data and that for 3.0T (<i>right</i>) displayed below the data.	57
Figure 3-3. Comparison of VTI R_1 estimation with theory. Bland-Altman plots (<i>top</i>) of the longitudinally averaged VTI R_1 measurements vs. theoretical R_1 values at 1.5T (<i>left</i>) and 3.0T (<i>right</i>). The 8 VIF (<i>squares</i>) and tissue (<i>diamonds</i>) compartments were given different markers for clarity and the dashed	

lines are the limits of agreement. Correlation plots (bottom) are shown for comparison with the linear regression and corresponding coefficient of determination for 1.5T (*left*) displayed above the data and that for 3.0T (*right*) displayed below the data. 60

Figure 3-4. 1.5T VIF R_1 dependence as a function of distance from isocenter in the I-S direction. Plots of the VIF [NiCl₂] sequence R_1 estimates normalized to the corresponding measurements on the slice at isocenter vs. distance from isocenter in the I-S (negative-positive) direction averaged over all longitudinal acquisitions at 1.5T. Data from the two most superior and two most inferior slices were omitted to ensure the imaging volume was matched to the corresponding 3T data. For visualization purposes, a shape-preserving smooth interpolant was used. Error bars were omitted because limits were so small that the error markers were indistinguishable from the data markers. 61

Figure 3-5. 3.0T VIF R_1 dependence as a function of distance from isocenter in the I-S direction. Plots of the VIF [NiCl₂] sequence R_1 estimates normalized to the corresponding measurements on the slice at isocenter vs. distance from isocenter in the I-S (negative-positive) direction averaged over all longitudinal acquisitions at 3.0T. For visualization purposes, a shape-preserving smooth interpolant was used. Error bars were omitted because limits were so small that the error markers were indistinguishable from the data markers. 62

Figure 3-6. 1.5T tissue R_1 dependence as a function of distance from isocenter in the I-S direction. Plots of the tissue [NiCl₂] sequence R_1 estimates normalized to the corresponding measurements on the slice at isocenter vs. distance from isocenter in the I-S (negative-positive) direction averaged over all longitudinal acquisitions at 1.5T. Data from the two most superior and two most inferior slices were omitted to ensure the imaging volume was matched to the corresponding 3T data. For visualization purposes, a shape-preserving smooth interpolant was used. Error bars were omitted because limits were so small that the error markers were indistinguishable from the data markers. 63

Figure 3-7. 3.0T tissue R_1 dependence as a function of distance from isocenter in the I-S direction. Plots of the tissue [NiCl₂] sequence R_1 estimates normalized to the corresponding measurements on the slice

at isocenter vs. distance from isocenter in the I-S (negative-positive) direction averaged over all longitudinal acquisitions at 3.0T. For visualization purposes, a shape-preserving smooth interpolant was used. Error bars were omitted because limits were so small that the error markers were indistinguishable from the data markers. 64

Figure 3-8. Sucrose solution T_2 values. Measured T_2 values at 3.0T (*diamonds*) and 4.7T (*square*). The large difference in values between the two field strengths stems for the acquisition details, namely, only 4 echoes were used to acquire measurements at 3.0T compared to the 24 echoes used at 4.7T. The number of echoes was limited by the acquisition pulse sequence at 3.0T. Also, the T_2 value at 4.7T for the 45% vial was extrapolated since no data were acquired for that vial. 69

Figure 3-9. 1.5T DWI ADC dependence as a function of distance from isocenter in the I-S direction. Plots of ADC measured at 1.5T using DWI normalized to the corresponding measurements on the slice at isocenter vs. distance from isocenter in the I-S (negative-positive) direction averaged over all longitudinal acquisitions. Data volumes were matched to 3.0T acquisition coverage and a shape-preserving smooth interpolant was used for visualization purposes. Error bars were omitted because the error markers were indistinguishable from the data markers. 70

Figure 3-10. 3.0T DWI ADC dependence as a function of distance from isocenter in the I-S direction. Plots of ADC measured at 3.0T using DWI normalized to the corresponding measurements on the slice at isocenter vs. distance from isocenter in the I-S (negative-positive) direction averaged over all longitudinal acquisitions. Data volumes were matched to 3.0T acquisition coverage and a shape-preserving smooth interpolant was used for visualization purposes. Error bars were omitted because the error markers were indistinguishable from the data markers. 71

Figure 3-11. 1.5T DTI ADC dependence as a function of distance from isocenter in the I-S direction. Plots of ADC measured at 1.5T using DTI normalized to the corresponding measurements on the slice at isocenter vs. distance from isocenter in the I-S (negative-positive) direction averaged over all longitudinal acquisitions. Data volumes were matched to 3.0T acquisition coverage and a shape-

preserving smooth interpolant was used for visualization purposes. Error bars were omitted because the error markers were indistinguishable from the data markers	72
Figure 3-12. 3.0T DTI ADC dependence as a function of distance from isocenter in the I-S direction. Plots of ADC measured at 3.0T using DTI normalized to the corresponding measurements on the slice at isocenter vs. distance from isocenter in the I-S (negative-positive) direction averaged over all longitudinal acquisitions. Data volumes were matched to 3.0T acquisition coverage and a shape-preserving smooth interpolant was used for visualization purposes. Error bars were omitted because the error markers were indistinguishable from the data markers	73
Figure 3-13. 1.5T FA dependence as a function of distance from at isocenter in the I-S direction. Plots of FA measured at 1.5T normalized to the corresponding isocenter slice measurement as a function of distance from isocenter in the I-S (negative-positive) direction averaged over all longitudinal acquisitions. Data volumes were matched to 3.0T acquisition coverage and a shape-preserving smooth interpolant was used for visualization purposes. Error bars were omitted because limits were so small that the error markers were indistinguishable from the data markers.....	74
Figure 3-14. 3.0T FA dependence as a function of distance from at isocenter in the I-S direction. Plots of FA measured at 3.0T normalized to the corresponding isocenter slice measurement as a function of distance from isocenter in the I-S (negative-positive) direction averaged over all longitudinal acquisitions. Data volumes were matched to 3.0T acquisition coverage and a shape-preserving smooth interpolant was used for visualization purposes. Error bars were omitted because limits were so small that the error markers were indistinguishable from the data markers.....	75
Figure 4-1. QUATTRO GUI screenshot. This is a view of the QUATTRO GUI, showing the main features and primary use as an image analysis tool. The common exam workspace can be seen in the expanded dropdown box at the lower left.	82
Figure 4-2. Example variable inversion time digital reference images. An example image of the noiseless (left) and noisy (right) VTI DRO. The patches for most values of R1 can be easily seen with the top left	

square being the lowest value of R_1 and the bottom right corner the largest value of R_1 . To better illustrate the added noise, images simulated from two different TI values are displayed, hence the seemingly different test objects.....	87
Figure 4-3. Actual pharmacokinetic values of the DCE digital reference object. Actual values of K^{trans} (left), v_e (middle), and v_p (right) used to generate synthetic tissue uptake curves. The DRO is 250 pixels in height and 100 pixels in width.....	89
Figure 4-4. R_1 Bland-Altman plots for the VFA and VTI digital phantoms. Summary plots of the Bland-Altman analysis performed (averaged over all DRO voxels) for the VFA (left) and VTI (right) test objects. The red “x” symbols, which appear as a solid line because of the high density, denote the DRO averaged bias, where the black lines above and below these symbols represent the limits of agreement. For comparison visual comparisons, a line of zero bias was also plotted.....	94
Figure 4-5. R_1 percent error maps for the VFA and VTI digital phantoms. Percent error from the true value as a function of SNR and R_1 averaged over all voxels of each respective relaxation rate for VFA (left) and VTI (right) simulations. The color bar shows the percent difference from the actual parameter used to simulate the MR signal.	94
Figure 4-6. R_1 difference coefficient of variation maps for VFA and VTI digital phantoms. Maps showing the coefficient of variation of R_1 bias for VFA (left) and VTI (right) simulations averaged over all voxels for a given R_1 value. The color bar shows the COV in %. COV was chosen over other measures of spread because of the normalization, allowing quick comparison of different parameter values.	95
Figure 4-7. ADC Bland-Altman plot for the diffusion digital reference object. Summary plot of the Bland-Altman analysis performed (averaged over all DRO voxels) on the ADC test objects are shown. The red “x” symbols, which appear as a solid line because of the high density, denote the DRO averaged bias, where the black lines above and below these symbols represent the limits of agreement. For comparison visual comparisons, a line of zero bias was also plotted, although this line is not readily visible.....	96

Figure 4-8. ADC percentage bias and coefficient of variation maps for the diffusion DRO. Percentage error (<i>left</i>) and coefficient of variation (<i>right</i>) maps for ADC as a function of SNR. The color bars show the percent error (<i>left</i>) and the percent COV (<i>right</i>).	97
Figure 4-9. Bland-Altman plots of PK parameters as a function of SNR. Summary plots of the Bland-Altman analysis performed (averaged over all DRO voxels) on the DCE test objects are shown. The red “x” symbols denote the DRO averaged bias, where the black lines above and below these symbols represent the limits of agreement. For comparison visual comparisons, a solid black line of zero bias was also plotted.	98
Figure 4-10. DCE DRO percent bias maps for estimated K^{trans} at 1s temporal resolution. Color maps of the percent bias as a function of the actual value of v_p (0.05- <i>top left</i> , 0.1- <i>top right</i> , 0.15- <i>middle left</i> , 0.2- <i>middle right</i>), v_e on the y-axis, K^{trans} on the x-axis, and SNR on z-axis. The color bar represents the percentage difference between the estimated parameter and the actual parameter value.....	101
Figure 4-11. DCE DRO percent bias maps for estimated K^{trans} at 5s temporal resolution. Color maps of the percent bias as a function of increasing v_p (0.05- <i>top left</i> , 0.1- <i>top right</i> , 0.15- <i>middle left</i> , 0.2- <i>middle right</i>), v_e on the y-axis, K^{trans} on the x-axis, and SNR on z-axis. The color bar represents the percentage difference between the estimated parameter and the actual parameter value.	102
Figure 4-12. DCE DRO percent bias maps for estimated v_p at 5s temporal resolution. Color maps of the percent bias as a function of increasing v_p (0.51- <i>top left</i> , 0.1- <i>top right</i> , 0.15- <i>middle left</i> , 0.2- <i>middle right</i>), v_e on the y-axis, K^{trans} on the x-axis, and SNR on z-axis. The color bar represents the percentage difference between the estimated parameter and the actual parameter value.	103
Figure 4-13. DCE DRO percent bias maps for estimated v_e at 5s temporal resolution. Color maps of the percent bias as a function of increasing v_p (0.05- <i>top left</i> , 0.1- <i>top right</i> , 0.15- <i>middle left</i> , 0.2- <i>middle right</i>), v_e on the x-axis, K^{trans} on the y-axis, and SNR on z-axis. The color bar represents the percentage difference between the estimated parameter and the actual parameter value.	104

Figure 4-14. <i>rCBV</i> Bland-Altman plot and CCC plot as a function of SNR for the DSC DRO. Summary plots of the Bland-Altman analysis performed (averaged over all DRO voxels) on the DSC test objects are shown. The red “x” denotes the DRO averaged bias, where the black lines above and below these symbols represent the limits of agreement. For comparison visual comparisons, a solid black line of zero bias was also plotted.	105
Figure 4-15. <i>rCBV</i> percentage bias and COV maps for the DSC DRO. Percentage error (<i>left</i>) and coefficients of variation (<i>right</i>) maps for <i>rCBV</i> as a function of SNR. The color bars show the percent error (<i>left</i>) and the percent COV (<i>right</i>).	106
Figure 5-1. Example pre-contrast and recirculation cut-offs for DSC VIF. Two examples of a DSC VIF demonstrating the manually defined pre-enhancement (first dotted vertical line) frames and the bolus recirculation cut-off (second vertical line). The small squares are the measured data, converted from signal intensities to arbitrary units of [Gd], and the solid line is the fitted gamma-variate. Note that even though the pre-contrast frame is manually defined a value for the bolus arrival is actually one of the fitting parameters.....	116
Figure 5-2. DTI and DSC quantitative imaging biomarker difference histograms for patients receiving the placebo. Histograms displaying the distributions of QIB differences between baseline and the first mid-treatment study (<i>left</i>) and between the first and second mid-treatment studies (<i>right</i>). For a given parameter (label shown under each plot), the x and y limits are held constant, but vary between parameters.	124
Figure 5-3. DCE quantitative imaging biomarker difference histograms for patients receiving the placebo. Histograms displaying the distributions of QIB differences between baseline and the first mid-treatment study (<i>left</i>) and between the first and second mid-treatment studies (<i>right</i>). For a given parameter (label shown under each plot), the x and y limits are held constant, but vary between parameters.	127

Figure 5-4. Selected DCE and DSC quantitative imaging biomarker difference histograms for patients treated with bevacizumab. Histograms displaying the distributions of QIB differences between baseline and the first mid-treatment study (<i>left</i>) and between the first and second mid-treatment studies (<i>right</i>). For a given parameter (label shown under each plot), the x and y limits are held constant, but vary between parameters.	130
Figure 5-5. Probability histograms for patients receiving the placebo. Histograms displaying the probability of being a specific category, either SD or PD, are shown for DTI and DSC parameters (<i>left</i>) and DCE parameters (<i>right</i>). Each half of this figure (left/right) represents the data from baseline/first mid-treatment differences (<i>left</i>) and differences between the two mid-treatment data sets (<i>right</i>). ..	131
Figure 5-6. DTI and DSC QIB probability histograms for voxels from patients receiving bevacizumab. Histograms displaying the distributions of DTI and DSC derived QIB differences between baseline and the first mid-treatment study (<i>left</i>) and between the first and second mid-treatment studies (<i>right</i>) .	132
Figure 5-7. DCE QIB probability histograms for voxels from patients receiving bevacizumab. Histograms displaying the distributions of DTI and DSC derived QIB differences between baseline and the first mid-treatment study (<i>left</i>) and between the first and second mid-treatment studies (<i>right</i>)	133
Figure 5-8. Example ordinal regression response category scores. Response category scores for PR (<i>top left</i>), SD (<i>top right</i>), and PD (<i>bottom</i>) for the ordinal model incorporating <i>ADC</i> , <i>FA</i> , and <i>rCBV</i> . The left half of the region displayed in color was categorized as SD and the other half as PD.	140
Figure 5-9. Example logistic regression response category scores. Response category scores for PR/SD (<i>left</i>) and PD (<i>right</i>) for the logistic model incorporating <i>ADC</i> , <i>FA</i> , and <i>rCBV</i> . The left half of the region displayed in color was categorized as SD and the other half as PD.	141
Figure 5-10. Example ordinal regression response category scores. Response category scores for PR (<i>top left</i>), SD (<i>top right</i>), and PD (<i>bottom</i>) for the ordinal model incorporating <i>FA</i> and v_p . The left half of the region displayed in color was categorized as SD and the other half as PD.....	142

Figure 5-11. Example logistic regression response category scores. Response category scores for PR/SD (<i>left</i>) and PD (<i>right</i>) for the logistic model incorporating FA and v_p . The left half of the region displayed in color was categorized as SD and the other half as PD.	143
Figure 5-12. Example ordinal regression response category scores. Response category scores for PR/SD (<i>left</i>) and PD (<i>right</i>) for the logistic model incorporating ADC and v_p . The left half of the region displayed in color was categorized as SD and the other half as PD.	144
Figure 5-13. Example logistic regression response category scores. Response category scores for PR/SD (<i>left</i>) and PD (<i>right</i>) for the logistic model incorporating ADC and v_p . The left half of the region displayed in color was categorized as SD and the other half as PD.	145

List of Tables

Table 3-1. DCE phantom [NiCl ₂] and theoretical R_1 values. R_1 ranges and corresponding [NiCl ₂] theoretical calculations assuming $r_1 = 0.62 \text{ mM}^{-1}\text{s}^{-1}$ and $R_{10} = 0.3\text{s}^{-1}$. Note that neither r_1 nor R_{10} were quantified prior to phantom manufacture.....	49
Table 3-2. DCE and ADC phantom protocol acquisition parameters. Parenthetical values represent the acquisition parameters at 3.0T. For the DWI and DTI acquisitions, the number of diffusion encoding directions was 3 and 27, respectively. Abbreviations: Sl. = Slice, Th. = Thickness, DSE = dual spin echo. 51	
Table 3-3. Summary of the DCE phantom longitudinal R_1 measurements. Measured R_1 values in s^{-1} averaged over all repeatability measurements using a VFA technique (<i>left</i>) and VTI technique (<i>right</i>). Values for the centrifuge tubes containing the VIF sequence are in the top half of the table, while the representative tissue values are in the bottom half of the table.	56
Table 3-4. Analysis of PK parameters simulated from R_1 variance. Lower/upper limits of agreement and concordance correlation coefficient as calculated from the simulated PK parameter estimates for levels of ΔR_1 variance at 1.5T (<i>top</i>) and 3.0T (<i>bottom</i>).	65
Table 3-5. Summary of the diffusion phantom longitudinal ADC measurements. Measured ADC ($\times 10^{-3} \text{ mm}^2/\text{s}$) values averaged over all repeatability measurements using a DWI technique (<i>left</i>) and DTI technique (<i>right</i>) for sucrose concentrations ranging from 10% to 45% (top to bottom).....	67
Table 4-1. Description of QUATTRO classes. These classes, with the exception of <i>qt_options</i> were designed to perform as stand-alone command prompt software, meaning that users can utilize those classes even beyond the scope of the QUATTRO GUI (<i>e.g.</i> , in scripts or as an API).	81
Table 4-2. Digital reference objects construction parameters. Parameters used to construct the DROs used for software performance evaluation. Where possible, values were estimated (as described in the text) from phantom and/or patient data. The last column, “simulation parameters”, contains all other necessary model parameters used to perform the simulations. *denotes proton density estimated SNR	

(<i>not</i> image SNR). For example, the lowest SNR value used for VFA simulations corresponds to an FSPGR proton density image SNR of approximately 15.	85
Table 5-1. Patient DCE, DSC, and DTI phantom protocol acquisition parameters. The temporal resolution of the DCE scan was approximately 6.3s. *Denotes the number of diffusion encoding directions (one b_0 image was acquired). Abbreviations: Num. = Number, Sl. = Slice, Th. = Thickness	111
Table 5-2. QIB modeling thresholds. Voxels exceeding the thresholds defined in this table were excluded from the regression analyses, as these thresholds nominally represent physically realizable parameter quantities of biological tissue.....	118
Table 5-3. DCE and DSC VIF locations. Artery/vein used in defining the vascular input functions used for quantitation of the DCE and DSC studies.	121
Table 5-4. Voxel counts for the consensus ROI definitions of both radiation oncologists. This table summarizes the voxel counts of all ROI data defined in consensus between the oncologists gathered from all patients and separated by patient, response category, and treatment type. QIB thresholds were applied prior to calculating these voxel counts. *indicates that the true treatment was unknown.	122
Table 5-5. Voxel summary statistics grouped by response category and treatment type. This table shows summary statistics, mean (standard deviation), for differences in QIBs between at the baseline/first mid-treatment (top) and the second/first mid-treatment (bottom) time points for voxels from all patients separated by response category and treatment type. All mean values have physical units corresponding to the parameter of interest.	126
Table 5-6. Ordinal classifier leave-one-out accuracies for linear, first-order interaction, and squared terms. This table summarizes the prediction accuracies for the multi-parameter models exhibiting the highest categorical and overall accuracies. Table cells with a black background and white text represent the highest accuracy for the respective category, identifying both single and multi-parameter maximum accuracies.	135

Table 5-7. Logistic classifier leave-one-out accuracies and R^2 values for linear, first-order interaction, and squared terms. This table summarizes the prediction accuracies (standard error) for the multi-parameter models exhibiting the highest categorical and overall accuracies. Table cells with a black background and white text represent the highest accuracy for the respective category, identifying both single and multi-parameter maximum accuracies. 137

List of Abbreviations

ADC – Apparent Diffusion Coefficient

AIF – Arterial Input Function

API – Advanced Programming Interface

BAT – Bolus Arrival Time

BBB – Blood Brain Barrier

BW – Bandwidth

CA – Contrast Agent

CCC – Concordance Correlation Coefficient

CDF – Cumulative Density Function

CNS – Central Nervous System

COV – Coefficient of Variation

CT – Computed Tomography

DCE – Dynamic Contrast-Enhanced

DICOM – Digital Imaging and Communications in Medicine

DRO – Digital Reference Object

DSC – Dynamic Susceptibility Contrast

DSE – Dual Spin Echo

DTI – Diffusion Tensor Imaging

DWI – Diffusion Weighted Imaging

EES – Extracellular Extravascular Space

EPI – Echo Planar Imaging

ETL – Echo Train Length

FA – Fractional Anisotropy

FOV – Field of View

FSPGR – Fast Spoiled Gradient Echo

GUI – Graphical User Interface

ISMRM – International Society for Magnetic Resonance in Medicine

ITK – Insight Toolkit

IVIM – Intravoxel Incoherent Motion

LOA – Limits of Agreement

LOO – Leave-One-Out

MIPAV – Medical Image Processing, Analysis, and Visualization

MRI – Magnetic Resonance Imaging

NEX – Number of Excitations

OOP – Object Oriented Program

PGF – Pulsed Gradient Field

PET – Positron Emission Tomography

PK – Pharmacokinetic

QIB – Quantitative Imaging Biomarker

QIBA – Quantitative Imaging Biomarker Alliance

QUATTRO – **Q**uantitative **U**tility for **A**ssessment of **T**rea**T**ment **R**esp**O**nse

RANO – Revised Assessment in Neuro-Oncology

RECIST – Response Evaluation Criteria in Solid Tumors

SAR – Specific Absorption Rate

SE – Spin Echo

SNR – Signal-to-Noise Ratio

SR – Saturation Recovery

TE – Echo Time

TI – Inversion Time

TR – Repetition Time

US – Ultrasound

VEGF – Vascular Endothelial Growth Factor

VIF – Vascular Input Function

VFA – Variable Flip Angle

VTI – Variable Inversion Time

Chapter 1 Introduction

1.1 The Role of Quantitative Imaging Biomarkers in Treatment Assessment

Assessing treatment response in cancer patients offers a multitude of challenges. Traditional measures of response, such as the Response Evaluation Criteria in Solid Tumors (RECIST) [1] and the more recent Revised Assessment in Neuro-Oncology (RANO) [2], have found success in the simplicity of required measurements, namely evaluation of tumor size. However, structural assessment as defined by these evaluation criteria has many inherent limitations, such as image acquisition variability (especially for two-dimensional image slices), interpreter differences, involvement of lymph nodes, and tumor type/location to name a few, as Husband *et al.* [3] and Jaffe [4] point out. Resulting variability can deleteriously impact the applicability and efficacy of response assessment. Indeed, these inadequacies have inspired the development of specialized criteria such as the recently updated RANO [2] for assessing high-grade gliomas or the modified RECIST assessment for hepatocellular carcinoma [5]. However, assessment of structural information may be inadequate and/or delayed significantly, especially for novel targeted therapeutics. For example, therapies incorporating anti-angiogenic drugs may induce substantial changes in a lesion's vascularity, resulting in a lack or altogether absence of enhancement patterns in the tumor (a common image feature used in the RANO assessment [2]) regardless of the actual underlying tumor response. However, the use of functional information derived from imaging-based biological markers, or imaging biomarkers for short, such as vascular permeability and flow, is a promising new avenue for treatment assessment.

Biomarkers, especially those potentially suitable for use as surrogate endpoints, are playing an increasingly important role in clinical trials as a means of assessing response to novel and combined therapeutic interventions. As defined by the Biomarkers Definitions Working Group [6], these markers objectively characterize normal and pathological biological processes. Advancements in imaging

technology have resulted in the availability of many image acquisition techniques capable of quantifying, in a non- or minimally-invasive manner, various biological characteristics, a subset of markers referred to collectively as *quantitative imaging biomarkers* (QIBs). Quantifiable functional information, such as cellular proliferation, perfusion, vascular flow, volume, and permeability, cellularity, and metabolism, comprise a subset of the class of QIBs. Techniques to quantify this information have been developed for modalities as diverse as magnetic resonance imaging (MRI), computed tomography (CT), positron emission tomography (PET), nuclear medicine, and ultrasound [7]. However, the myriad image contrast mechanisms, use of non-ionizing radiation, and continual improvements in MR hardware have led to strengthening interest in applications of functional MR QIBs. Aside from providing functional information, some biomarkers have been shown to exhibit measureable changes predictive of tissue response within days following the initial treatment [7-8], substantially sooner than would be detectable by monitoring morphological changes alone. This need for rapid assessment of patient response to novel or combined cancer therapies has motivated the incorporation of many of these MR imaging techniques into early phase clinical trials.

The remainder of this chapter will focus on the introduction and discussion of the functional MRI biomarkers specific to this work (*i.e.*, diffusion, perfusion, and vascular permeability), their biological significance and use as a means of assessing cancer therapies, challenges associated with their incorporation into the clinical setting, and the motivation for developing multi-parametric response maps as a means of assessing treatment response. Thorough development of the theory, acquisition, and analysis of these biomarkers is covered in Chapter 2. Extensive discussions of QIB applications are available from a number of authors. For example, oncologic applications of QIBs derived from contrast enhanced MR acquisitions are discussed in the text by Jackson [10], or a more generic discussion of diffusion-weighted imaging derived QIB applications can be found in the text by Koh and Thoeny [11]. For brain applications, the text by Tofts [12] serves as an excellent reference for commonly used MR

imaging biomarkers in the investigation of a number of pathological processes. The reader is referred to these texts, and other reviews [13]–[18], for a broader, more detailed perspective regarding specific applications of QIBs.

1.2 Diffusion

Perhaps one of the most commonly investigated physical processes in the assessment of treatment response is the molecular diffusion of water, a process of random thermal molecular motion. In tissue, this stochastic motion of water molecules is affected by a number of biophysical mechanisms, such as intracellular-extracellular water exchange, active transport, flow/perfusion, bulk movements, and tortuosity (a property describing the cellular density, volume, and extracellular space organization) [19]. Often the property of most interest in investigating observed differences in diffusion processes is the interaction of water molecules with structures that inhibit water mobility (*e.g.*, hydrophilic cellular membranes and organelles within cells). In fact, differences in the microenvironments of the intracellular and extracellular spaces give rise to unique diffusion coefficients. However, in biological tissues, diffusion is often characterized by a single parameter describing the general mobility of water molecules, incorporating the several aforementioned physical processes, hence the name *apparent diffusion coefficient* (ADC). Diffusion-weighted imaging (DWI) is the MRI method used to quantify this scalar quantity.

Most tissues in the human body are characterized by water diffusion rates that are equal in all directions (*isotropic* diffusion), although the existence of, for example, cell membranes restricts the ability of water molecules to freely diffuse, giving rise to the process called *restricted diffusion*. The degree of restricted water diffusion is generally thought to correspond to the cellular density of tissue, which has been confirmed by numerous researchers. A majority of studies comparing histology to ADC measurements have shown a statistically significant (negative) correlation between the histologically

derived cellularity and ADC values in tumors [20]–[31]. However, some investigators, for example Sadeghi *et al.* [32], have found a non-significant relationship between cellular density and the ADC.

Consensus has yet to be reached regarding the exact mechanisms responsible for affecting change in longitudinal ADC measurements during treatment regimens [19]. This incomplete mechanistic understanding likely stems from the use of over-simplified mathematical models, confounding factors from data acquisition, and lack of sensitivity/specificity in the acquisition techniques to specific biophysical processes. Regardless, numerous processes corresponding to known pathological processes have been proposed for the observed changes in ADC values during and following treatment. As Patterson *et al.* [18] summarize, events such as cellular swelling, reduction in blood flow, and tissue compaction/fibrosis are proposed to result in lower ADC values, while processes such as apoptotic death and necrosis might result in increased ADC values as the result of decreased cellular density. Strong evidence exists to support increases in the ADC as an indicator of partial response to treatment [18], [33]–[36]. Intuitively, these observations suggest cell death/necrosis resulting from successful treatment will yield increased mobility of extracellular water (*i.e.*, increased ADC). Cellular proliferation resulting from progressing disease processes is thought to cause reductions in observed ADC values. Higano *et al.* [37] observed a strong negative correlation between ADC and cellular proliferation, confirming that hypothesis in patients with glioblastomas and astrocytomas. Researchers have utilized this rationale and observed decreased ADC values [33], [37]–[39] in patients with progressive disease, and although some groups observed no change in the ADC values with disease progression [40], [41] this discrepancy is likely the result of small patient cohorts in these initial studies.

Although the ADC has been extensively investigated in numerous tissues, under certain circumstances (*e.g.*, in muscle or myelinated nerve fibers) diffusion becomes preferentially restricted in one direction, resulting in diffusion *anisotropy* and necessitating additional measures to provide a more complete description of water diffusion. Quantities such as the fractional anisotropy (FA) are often used

to summarize the strength of ADC's dependence on direction, with values falling in the interval between 0 (representing purely isotropic diffusion) and 1 (representing highly anisotropic diffusion) [42].

Quantification of anisotropy is performed using diffusion tensor imaging (DTI), which encodes this additional directional information in the acquired images. The investigation of such quantities outside of the central nervous system (CNS) is severely limited because most tissues lack the necessary microstructural properties capable of generating anisotropic diffusion. However, in the CNS changes in measured FA values might be indicative of the disease state. Comparison of histology and measured FA values exhibited a negative correlation with cellularity in brain tumors [43]. Furthermore, investigations of peritumoral FA values in brain tumors have revealed patterns of tumor infiltration [43]–[45]. Mechanistically, these observed differences in FA measures likely correspond to the increased cellular disorganization and/or extracellular water content of tumors [45]–[47].

Despite the lack of general consensus regarding the mechanisms of ADC and FA change, substantial evidence exists that suggests these QIBs are potentially important markers of treatment response or proliferative disease when observed longitudinally during the course of therapeutic intervention [18], [33]–[39]. ADC quantitation has been shown to track with changes in cellular density and cellular proliferation in tumors, and associated changes in the ADC can be indicative of response and proliferative disease. Moreover, some researchers have noted that diffusion QIBs provide early indications of response [33]–[39], which could afford the opportunity to adapt treatments to non-responsive disease. These current findings suggest that quantitative diffusion imaging is a promising technique capable of assessing therapeutic response, even in shorter time frames than traditional measures of response.

1.2.1 Perfusion and Vascular Permeability

Perfusion is the biophysical process by which blood is delivered to tissue via the capillary bed, and is often summarized by a number of QIBs such as the relative cerebral blood flow (*rCBF*), relative

cerebral blood volume ($rCBV$), and K^{trans} , a rate constant describing an amalgamation of vascular permeability and blood flow. These biophysical properties are investigated most commonly through the use of low molecular weight extracellular exogenous contrast agents (CA). An intravenous bolus injection and subsequent rapid imaging of the CA provides a means of assessing the distribution of the contrast agent within the vasculature and, following extravasation, extracellular extravascular space (EES). Pharmacokinetic models applied to these bolus tracking techniques provide a means of deriving the aforementioned QIBs in addition to estimates of the relative EES v_e . T_2^* -weighted imaging methods are most commonly used to derive $rCBF$ and $rCBV$ (or more generally rBF and rBV in extracranial applications) and are referred to as dynamic susceptibility contrast MRI or DSC-MRI techniques, whereas K^{trans} and v_e are derived from T_1 -weighted methods collectively called dynamic contrast-enhanced (DCE) MRI. As with diffusion imaging, the pathophysiology of cancer provides a unique opportunity for the investigation of tumor perfusion and vascular permeability characteristics.

Nutrients are delivered to avascular tumors by passive diffusion, which limits the size of the tumor to the approximate distance oxygen is capable of passively diffusing (1-2 millimeters). Further viable growth is accomplished only through the recruitment of neovasculature (*i.e.*, angiogenesis), providing the necessary infrastructure to allow nutrients to reach the most interior regions of the tumor to support the uninhibited growth of new cancerous cells [48]. Factors of the tumor microenvironment such as pH, glucose concentration, and even certain oncogenes (*e.g.*, p53) can result in the up-regulation of vascular endothelial growth factor (VEGF), which, as the name suggests, promotes the growth of endothelial cells eventually culminating in a neovascular network [10]. Histologically, this neovascularization has been shown to correlate with increased microvascular density (MVD), which has been shown to be an important biomarker of tumor stage and prognosis [49], [50]. In addition to rapid growth of new blood vessels, vasculature recruited in this manner tends to be highly heterogeneous and often contains numerous morphological defects, such as coarse capillaries with irregularly shaped

vessels and tortuous paths [51]–[53]. These gross malformations, as Jackson *et al.* summarized [10], can result in increased spacing between inter-endothelial junctions and a basement membrane that is discontinuous or altogether absent, resulting in vascular hyperpermeability [54].

Both DCE- and DSC-MRI have shown promise in assessing these microvascular properties. In well perfused tumors, histological measurements of MVD have been found to correlate with K^{trans} in a number of extracranial lesion types [55]–[57], although similar results in brain tumors remain unestablished [58], [59]. However, Tynnenen *et al.* [60] studied human gliomas and found a positive correlation between MRI enhancement and cellular proliferation/MVD, but quantitative biomarkers were uninvestigated. Since investigation of these parameters requires the extravasation of contrast agent into the EES, varying levels of blood brain barrier (BBB) disruption in pathological processes, especially in tumor borders, of the CNS may result in a general lack of concordance between T_1 -weighted quantities and histological markers. However, infiltration of brain tumor cells and vasogenic edema can cause disruptions in the BBB, allowing passage of contrast agents into the extravascular extracellular space [61]. The reduced permeability caused by an intact or partially disrupted BBB provides an ideal scenario for DSC imaging. Specifically, several authors, using T_2^* -weighted imaging, have shown correlations between MVD and $rCBV$ in brain tumors [32], [56], [62]–[64].

These techniques have been employed in a number of studies to assess response to radiation and/or chemotherapy treatments in cancers of the bladder [65], breast [66], [67], cervix [57], [68], rectum [69], and brain [70], where, similar to changes in the ADC , reductions in vascular permeability (K^{trans}) and relative blood volume generally corresponding to responding disease and increase tend to be associated with disease progression. Despite these promising results, the assessment of antiangiogenic agents is of substantially more interest in the application of QIB describing perfusion and vascular permeability. Jackson *et al.* [71] and O'Connor *et al.* [72] have provided excellent reviews cataloging the most up-to-date efforts employing perfusion/permeability imaging to assess these novel antivascular

drugs. Initial results from a number of studies have shown K^{trans} to be a predictor of both partial response and progression. However, substantial heterogeneity exists in the results, and is likely the result of small patient cohorts and a large variety in the treatment regimens and tumors sites.

1.2.2 Challenges

Standardization of data acquisition and analysis strategies is, perhaps, the most detrimental factor preventing the widespread incorporation of quantitative imaging techniques into the clinical setting [9], [19]. This lack of standardization has resulted in a number of issues, making the comparison of published results difficult, preventing large multi-center trials, and slowing the validation of the many proposed QIBs. In fact, much of the variation in the cited results of the previous sections can likely be attributed to a combination of the lack of standard acquisitions/analyses and small patient cohorts. Recent consensus documents have been published in an effort to address, specifically, the use of DCE-MRI [73] and DW-MRI [19], and more generally, quantitative imaging recommendations [74]–[77].

Another common theme among these reports is the necessity for the development and implementation of quality assurance methods capable of assessing sources of bias and variance in the QIB acquisition and analysis pipelines [74]–[77]. As Leach *et al.* summarized [73], a number of international committees have been formed, such as the *ad hoc* committee on standards for quantitative MRI within the International Society for Magnetic Resonance in Medicine (ISMRM) and the perfusion, diffusion, and flow subcommittee of the Quantitative Imaging Biomarker Alliance (QIBA). Efforts by the perfusion, diffusion, and flow subcommittee have manifested in the development [78] and ongoing accrual of multi-center data to assess equipment related bias and variance of DCE-MRI in body applications, in addition to other independent efforts to quantify bias and variance in DWI and DTI applications [79], [80]. These efforts are important not only for assessing bias and variance, but also for the detection, evaluation, and mitigation of changes in scanner performance as the result of equipment failure or software/hardware upgrades.

As the development and implementation of standardized methodology and quality assurance programs continues, incorporation of these techniques into clinical trials will expand, particularly those encompassing multiple QIB acquisitions. Moreover, standardization will also facilitate the development of robust software tools for quantitative imaging analysis, another set of tools requiring thorough validation and characterization. Thorough understanding of the equipment and analysis related bias and variance will provide a means of determining detectable limits of change in, and ultimately, in conjunction with large patient cohorts, validation of quantitative imaging biomarkers.

1.2.3 Mapping Response using Multiple QIBs

Treatment-induced changes in normal tissue and tumors result in a multi-faceted, temporally varying, spatially heterogeneous, biological response, especially in highly conformal and combination therapies. Overwhelming motivation for the employment of quantitative imaging techniques for assessing treatment response has been discussed in the preceding sections. Even a handful of studies have simultaneously investigated multiple MRI QIBs to differentiate benign and malignant tumors [81], assess treatment response [82], and develop mathematical tumor models [83]. Extracting useful quantitative information from these techniques requires special consideration, and, moreover, instilling utility into such extracted QIBs remains an immense challenge.

Research performed in much of the literature discussed previously is based on summary statistics calculated over some region or volume of interest. Potential for incorporating multiple tissues or missing nuance in tissue changes using this method is substantial. Appreciation for the spatial heterogeneity of disease related changes has led many researchers to consider voxel-by-voxel analysis methods such as histogram analyses [7], [82], [84], [85]. Although quantitative by nature, histogram analyses destroy important spatial information contained in QIB parameter maps. Detecting spatial heterogeneous information using such maps can be highly advantageous, but quantifying such heterogeneity is extremely challenging. Perhaps the most important attempt at quantifying such

information to date is the development of the functional diffusion map [34] and, more recently, the parametric response map [86]. These techniques utilize parameter map (*ADC* and *rCBV*, respectively) differences, defining thresholds for positive and negative change from which voxel volume fractions are calculated. By setting appropriate thresholds on the volume fractions of positive and negative change, the functional diffusion map analysis provided a means of differentiating partial responders from those patients with stable or progressive disease [34]. Similar findings were reported for the parametric response map [86]. Although these analyses represented significant contributions to the detection of disease states using parametric maps, spatial information relating QIBs directly to response was still disregarded. As increasingly complex therapies are performed, establishing local relationships between tissue QIBs and treatment response provides incredible promise to elucidate treatment associated changes, potentially guiding clinical decisions, especially if multiple complimentary quantitative imaging biomarkers are incorporated into such an analysis. To date, no reports of techniques capable of assessing locoregional response in this way have been published.

1.3 Hypothesis and Specific Aims

Quantitative imaging biomarkers have shown promise in assessing response to a variety of treatments in a number of anatomic sites. However, incorporating the complimentary information contained in the QIBs of diffusion, perfusion, and vascular permeability into locoregional assessment of treatment response remains as of unexplored. The objective of this research is to develop a comprehensive and self-contained methodology for incorporating multi-parametric quantitative imaging biomarkers into a statistical model capable of predicting locoregional radiological response. **We hypothesize that the incorporation of multiple quantitative imaging biomarkers into ordinal regression models will better predict locoregional radiological response in post-surgical glioblastoma**

patients treated with intensity-modulated radiation therapy and temozolomide with or without bevacizumab compared with the response predicted by a single quantitative imaging biomarker.

Specific Aim 1: Assess MRI equipment bias and variance. Phantoms will be developed to assess equipment variance related to quantitative acquisitions, specifically dynamic contrast-enhanced and diffusion-weighted (and diffusion tensor) imaging techniques. Weekly data acquisitions will be performed to investigate longitudinal equipment level bias and variance. Additionally, these measured variance levels will be used to simulate associated variances pharmacokinetic models that are not able to be assessed through direct measurement of phantom properties.

Specific Aim 2: Implement and characterize software for assessing treatment response. Comprehensive software tools for integrating quantitative modeling, image registration, and statistical analyses will be developed. All quantitative models will be characterized to assess the algorithm performance using digital reference objects.

Specific Aim 3: Predict locoregional response in a cohort of glioblastoma patients. Using the software tools developed in specific aim 2, locoregional response will be defined by an experienced radiation oncologist and used to develop ordinal regression models incorporating individual QIBs and combined QIBs. These single parameter and multi-parametric statistical models will be used to predict locoregional response.

Chapter 2 Background and Theory

2.1 Quantitative Imaging

2.1.1 Diffusion

Robert Brown first described, in 1828, the random motion of pollen particulates in suspension. Later, Alfred Fick derived an expression, commonly known as Fick's First Law of diffusion [87], relating the molecular motion of a solvent to an existing concentration gradient, written succinctly as

$$\vec{J} = -D\vec{\nabla}C, \quad 2.1$$

where \vec{J} is the particle flux, $\vec{\nabla}C$ is the concentration gradient, and D is the diffusion coefficient. Albert Einstein showed in 1905 using kinetic theory that the random molecular motion first described by Brown (*i.e.*, Brownian motion) was, in fact, a diffusive process dependent on temperature and molecular mobility [88]. In that seminal paper, Einstein showed, using Fick's second law (in one dimension),

$$\frac{\partial C}{\partial t} = D \frac{\partial^2 C}{\partial x^2}, \quad 2.2$$

that the general solution to equation 2.2 for an n particle system is,

$$C(x, t) = \frac{n}{\sqrt{4\pi Dt}} e^{-x^2/4Dt}, \quad 2.3$$

and that the diffusion coefficient can be expressed in terms of the mean square displacement $\langle R^2 \rangle$ and observation time t , given by

$$D = \frac{\langle R^2 \rangle}{2t}, \quad 2.4$$

From this zero mean Gaussian (equation 2.3), the famous Einstein result is easily extracted

$$\sqrt{\langle R^2 \rangle} = \sqrt{2Dt}, \quad 2.5$$

yielding the simple interpretation that the root-mean-square displacement is proportional to the square root of the time that has elapsed since the initial state of the system. In other words, the longer a system of particles is allowed to undergo random molecular motion, the greater the root-mean-square displacement of those particles. This result is easily extensible to three dimensions, noting that the integral used to generate equation 2.5 becomes a volume integral over all space, resulting in

$$\sqrt{\langle R^2 \rangle} = \sqrt{6Dt}. \quad 2.6$$

Even in the absence of chemical gradients these general results still hold, in which case, D is referred to as the self-diffusion coefficient. In the context of imaging measurements, D is generally the value of interest and the term “apparent” is appended to draw a distinction between the measurement of self-diffusion and the measurement of molecular motion in a complex biological system, yielding the term apparent diffusion coefficient or *ADC*. Biological systems, as will be discussed later in this section, can encompass numerous biophysical phenomena other than self-diffusion alone. Despite this distinction, *ADC* and D are used interchangeably in the following discussions, especially when mathematical expression can be simplified.

When considering the measurement of diffusion, especially in biological tissues, directional dependence of diffusion may be encountered. Muscle and myelinated nerve fibers are examples of biological structures in which water diffusion becomes directionally restricted or, in other words, diffusion of molecules is oriented preferentially along the length of the fiber. In such cases, a single proportionality constant describing the ability of particles to disperse as a function of time becomes insufficient. For a homogenous medium, this process can be represented by a second-rank tensor, D_{ij} for i and $j=x, y, z$ or

$$\mathbf{D} = \begin{pmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{pmatrix}, \quad 2.7$$

where D_{ij} represents the i^{th} component of diffusion along the j^{th} axis. In practice, this can be interpreted as the i^{th} component of diffusion process measured through observations along the j^{th} direction. Fick's second law can be extended to incorporate the directional dependence (*i.e.*, the diffusion tensor). When considering self-diffusion in the absence of other forces influencing the process (*e.g.*, active transport), the diffusion tensor becomes symmetric (*i.e.*, $D_{ij}=D_{ji}$). This additional symmetry results in the diffusion tensor being uniquely determined by six terms, the terms along the diagonal and either the three elements above or below the diagonal. If the coordinate systems of the physical process and the observations were perfectly aligned, the off-diagonal components of the diffusion tensor would be identically zero, with the values along the diagonal known as the principle diffusion coefficients. In practice, the orientation of the diffusion processes is unknown and the unique components of the tensor must be observed. Diagonalization can then be performed to recover the principle diffusion components, which are invariant regardless of the orientation of the measurement or laboratory coordinate system. This reduces equation 2.7 to

$$\mathbf{D}' = \begin{pmatrix} \lambda_{x'} & 0 & 0 \\ 0 & \lambda_{y'} & 0 \\ 0 & 0 & \lambda_{z'} \end{pmatrix} \begin{pmatrix} \varepsilon_{x'} \\ \varepsilon_{y'} \\ \varepsilon_{z'} \end{pmatrix}, \quad 2.8$$

where the three principle diffusion components or eigenvalues $\lambda_{i'}$ represent the diffusion coefficient in the new “prime” coordinate system specified by the basis of eigenvectors $\varepsilon_{i'}$ or diffusion directions, where the result of equation 2.5 can be shown to hold for each of the three eigenvalues (*i.e.*, $\sqrt{2\lambda_{i'}t}$ for $i'=x',y',z'$). Note that in an isotropic medium, this formalism reduces to that considered by Einstein, where the scalar

$$D = \begin{cases} D_{ij} & \text{for } i = j \\ 0 & \text{otherwise} \end{cases} \quad 2.9$$

which is equivalent to the initial formulation by Einstein. Tensor formalism gives a complete description of the physical system, but generates an enormous amount of data, three eigenvalues and eigenvectors to be exact, although knowledge of certain systems might suggest that there are only two unique eigenvalues and eigenvectors. Especially when these values are computed on a voxel-by-voxel basis, rotationally invariant summary parameters (*i.e.*, those that depend only on the underlying physical process, not the frame of observation) become increasingly important. Two of the most commonly employed parameters are the *mean diffusivity (MD)*

$$MD = \frac{\text{Trace}(\mathbf{D}')}{3} \quad 2.10$$

and *fractional anisotropy (FA)*

$$FA = \sqrt{\frac{3}{2} \frac{\sum_{i=1}^3 (\lambda_i - MD)^2}{\sum_{i=1}^3 \lambda_i^2}} \quad 2.11$$

both of which have been investigated frequently as QIBs [18], [19]. The latter, FA, was proposed by Basser and Pierpaoli [42], and achieves values ranging from 0, which represents a purely isotropic medium (equation 2.9), to 1, indicative of a cylindrical anisotropic medium (*i.e.*, $\lambda_x \gg \lambda_y = \lambda_z$, or any other permutation of the directional indices).

Stejskal and Tanner [89] were the first researchers to describe a method for sensitizing the MRI signal to the diffusion properties of the object or patient being imaged. The proposed pulse sequence incorporated two rectangular pulsed gradient fields (PGF) of equal magnitude G and duration δ separated by a time interval of Δ , and followed by a standard spin-echo (SE) imaging sequence. During the interval Δ , a 180° RF pulse is applied, which, in combination with the second gradient pulse, causes a

reversal of the phase accumulated during the first gradient pulse. Stationary spins experience the same local magnetic susceptibilities throughout this pulse sequence and the resulting signal intensity is unaffected. However, diffusing spins will experience changing magnetic susceptibilities while traversing the molecular environment, resulting in a net signal loss following this pulse sequence. More specifically, solving the Bloch-Torrey equation [90] for the PGF sequence gives

$$S(\vec{G}, \delta, \Delta; \mathbf{D}, TE) = S_0 e^{-\gamma^2 \delta^2 \left(\Delta - \frac{1}{3}\delta\right) \vec{G} \cdot \vec{G} \cdot \mathbf{D}} e^{-TE \cdot R_2^*}, \quad 2.12$$

where γ is the gyromagnetic ratio, \vec{G} is the applied gradient vector, S_0 is the thermal equilibrium signal intensity¹, TE is the echo time, and R_2^* is the transverse relaxation rate; all other parameters are as defined previously. The direct product or dyadic of the magnetic gradient vector is defined as

$$\vec{G} \vec{G} = |\vec{G}|^2 \begin{pmatrix} \alpha_x^2 & \alpha_x \alpha_y & \alpha_x \alpha_z \\ \alpha_y \alpha_x & \alpha_y^2 & \alpha_y \alpha_z \\ \alpha_z \alpha_x & \alpha_z \alpha_y & \alpha_z^2 \end{pmatrix}, \quad 2.13$$

where α_x , α_y , and α_z are the directional cosines of the applied diffusion encoding gradient. In most applications, a single user-adjustable imaging parameter b (or the b -value) is used *in lieu* of the timing/amplitude values, and is defined as

$$b(\vec{G}, \delta, \Delta) = \gamma^2 \delta^2 \left(\Delta - \frac{1}{3}\delta\right) \vec{G} \vec{G}, \quad 2.14$$

which allows recasting of equation 2.12 into the canonical form

$$S(\mathbf{b}; \mathbf{D}, TE) = S_0 e^{-\mathbf{b} \cdot \mathbf{D}} e^{-TE \cdot R_2^*}. \quad 2.15$$

¹ S_0 , more specifically, is a combination of the proton density, receiver gains, and, potentially, image reconstruction techniques.

Although this development has assumed an ideally rectangular pulsed gradient, analytical expressions can be derived for many other physically realizable gradient waveforms [91], including effects of the imaging gradients on the b -values [92].

The theoretical developments by Stejskal and Tanner made no assumptions regarding the underlying microstructural tissue properties or other physiological properties affecting DW signal intensities. Le Bihan *et al.* [93] posited that compartmentalization of water in perfused capillaries can generate diffusion-like (or pseudodiffusion) reductions in signal intensity that actually result from perfusion effects. A subsequent publication [94] by that same group resulted in the development of a model to describe pseudodiffusion, yielding

$$S(\mathbf{b}; \mathbf{D}, TE) = S_0 e^{-\mathbf{b}:\mathbf{D}} [(1 - f) + f \cdot e^{-\mathbf{b}:\mathbf{D}^*}] e^{-TE \cdot R_2^*} \quad 2.16$$

where \mathbf{D}^* is the pseudodiffusion coefficient describing water movement in capillaries that appears as a diffusion process (i.e. exponentially decaying), f is the perfusion fraction, and all other symbols are as defined previously. Interestingly, accounting for the perfusion fraction (essentially the fractional volume of plasma space) in this way allows comparisons with T_1 -weighted pharmacokinetic models that explicitly account for fractional plasma volume, which recent results suggest agree well in prostate cancer and the corresponding contralateral normal tissue [95]. Many other models describing, for example, the non-Gaussian characteristics of restricted diffusion and compartmental models have been proposed and investigated [96], but the models described by equations 2.15 and 2.16 are, by far, the most commonly employed.

In vivo quantitation of diffusion, especially for the purposes of assessing treatment response, is performed primarily under two paradigms, one for tissues in which isotropic diffusion is expected and the other for the case of anisotropic diffusion. In this discussion of diffusion modeling the effects of diffusion on the MRI signal, no assumptions regarding the diffusion isotropy have been made despite

continued appearance of the diffusion tensor in equations 2.12 through 2.16. Rather, a more general approach was taken, which results in greatly simplified models under circumstances when anisotropy can be or is purposely neglected, and will be the focus of section 2.1.1.1.

2.1.1.1 Quantitative Diffusion – Acquisition Strategies

As described earlier in section 2.1.1, most biological tissues exhibit restricted isotropic diffusion of water molecules and a single summary parameter, ADC , suffices to describe the system. Under these circumstances, the diffusion tensor reduces to a constant or zeroth-rank tensor, and by ignoring the directional component of \mathbf{b} reduces the previous linear algebra operators of equation 2.15 are reduced to the scalar products of b and D . By acquiring images using at least two unique b -values, keeping the TE and timing parameters of b constant, regression analysis (linear or non-linear) can be performed to derive the diffusion coefficient. Historically, two acquisitions were used to derive the ADC , and consisted of one image without diffusion weighting (*i.e.*, $b \approx 0$) and a second image with a non-zero b -value. However, when modeling is performed with equation 2.15, mounting evidence, as reviewed in a recent consensus report of DWI by Padhani and Liu [19], suggests that multiple b -values (*i.e.*, more than two) should be used and that both small and large b -values ($<100 \text{ s/mm}^2$ and $>1000 \text{ s/mm}^2$) should be carefully considered in the design of acquisition protocols to avoid perfusion effects and the bi-exponential behavior exhibited by the DW signal at high b -values, respectively. More advanced modeling, such as the IVIM model, necessitates the acquisition of more than two b -values to uniquely determine the model parameters.

In addition to varying the b -values, acquisitions are commonly performed by utilizing the same set of b -values acquired in multiple directions, most commonly in three mutually orthogonal directions yielding ADC_x , ADC_y , and ADC_z . A number of useful acquisition schemes have been proposed for the acquisition of these three ADC values. Perhaps the simplest and most commonly used schema acquires images with diffusion encoding parallel to the physical gradient axes (G_x , G_y , and G_z). So called

tetrahedron encoding can provide improvements in signal-to-noise ratio (SNR) (recall the TE dependence of equation 2.15) over this simple scheme. By combining pairs of gradients (G_x/G_y , G_x/G_z , and G_y/G_z), which effectively increases the maximum gradient strength by a factor of 1.4 and reduces the pulse duration Δ , the echo time can be reduced, resulting in less T_2^* dephasing and increased SNR. For isotropic tissues, the resulting ADC values should be equal, disregarding differences in gradient performance, but in tissues with anisotropic diffusion, these values are rotationally variant and can depend on external factors such as patient orientation. Despite the lack of rotational invariance, DWI acquisitions have found clinical utility in the neurological evaluation of strokes, infections, head trauma, tumors, and hemorrhages [97]. Furthermore, multi-directional DWI acquisitions can provide a rapid means of qualitatively assessing anisotropy (or confirming assumptions about isotropy).

Anisotropic tissues are described only partially by DWI quantification methods, requiring computation of the diffusion tensor for a more adequate description. Recall that a minimum of six elements are required to uniquely determine the diffusion tensor. By solving equation 2.15 for the argument of the exponential term containing the tensor and absorbing the T_2^* decay term into S_0 , an expression relating the b -values and diffusion tensor elements is given by

$$\ln \left(\frac{S(\mathbf{b}; TE)}{S(0; TE)} \right) = -\mathbf{b} : \mathbf{D}. \quad 2.17$$

Expansion of the Frobenius inner product of the \mathbf{b} -matrix and diffusion tensor gives rise to a six term linear equation for each applied direction of the diffusion gradient, assuming that $D_{ij}=D_{ji}$. If six non-zero, non-collinear, non-coplanar, \mathbf{b} -values are acquired, in addition to an image without diffusion weighting, equation 2.17 represents a linear system of six equations that uniquely determine all elements of \mathbf{D} . Therefore, a minimum of six diffusion weighted images and one non-diffusion weighted image, for a total of seven images, are required to determine the diffusion tensor. Basser *et al.* [98] first described a method of multiple linear regression for determining the elements of \mathbf{D} when using this minimal

acquisition strategy. As the number of diffusion weighted images (N) increases beyond six, the method of singular value decomposition is usually employed to find a solution to the N linear equations of 2.17. Acquisitions on commercially available clinical systems usually present the user with a choice only for the number of diffusion directions to acquire and a pre-computed set of gradient strengths and directions are used to perform serial imaging.

Diffusion tensor imaging is wrought with trade-offs. In clinical settings, limited scanner time and issues such as patient motion (DTI and DWI are, after all, serial image acquisitions) place constraints on the number of quality images that are able to be acquired. Two primary acquisition parameters determine the image quality and acquisition time: the number of diffusion encoding directions and the number of signal averages to perform on each acquired image. A theoretical study of noise by Bastin *et al.* [99] showed that, below an SNR value of 20, components of the diffusion tensor rapidly diverged from ground truth. SNR can be controlled in the acquisition by averaging the signal, essentially acquiring the same data multiple times and averaging, or by increasing the number of diffusion encoding directions. For diffusion gradient schemes acquiring the absolute minimum number (or near that) of directions, signal averaging must be used to improve image SNR, especially as the magnitude of the b -value increases. However, as the number of acquired encoding directions increases, the amount of time required to perform signal averaging becomes too expensive, but the acquisition of these additional images with, diffusion encoded in different directions, also improves image SNR. Additionally, Armitage and Bastin [100] and Jones *et al.* [101] have shown that too few encoding directions unnecessarily biases tensor estimates, while Poonawalla and Zhou [102] showed that the SNR improvements fall off drastically after the number of encoding directions exceeds approximately 25.

2.1.2 T_1 Relaxometry

Quantitation of longitudinal relaxation rates, R_1 (usually quoted in units of s^{-1}), generally requires modeling of the signal intensity, averaged over an ROI or taken from an individual voxel, as a function of

one variable acquisition parameter. Two common techniques are based on the extension of common imaging techniques (*e.g.*, FSPGR and IR-SE) to include multiple imaging parameter settings such as variable flip angles (VFA) and variable inversion times (VTI). Numerous other techniques, such as the Look-Locker sequence, saturation recovery, and stimulated echo acquisitions [103], have been developed and employed frequently for research purposes. However, these newer techniques remain unavailable on commercially-available clinical equipment, and since the goal of this work was to address quantitative imaging in a clinical setting, the following discussions are limited to those acquisition strategies currently available on clinical systems.

Spoiled gradient echo sequences provide a time efficient method of estimating relaxation rates for 2- or 3-D image sets. This is accomplished by acquiring images using two or more unique RF excitation flip angles, hence the name variable flip angle, while keeping all other acquisition parameters constant. By solving the Bloch equations, assuming that steady state has been reached and that perfect RF spoiling is occurring, the signal intensity as a function of the RF excitation angle (α) can be written as

$$S(\alpha; R_1, R_2^*, S_0, TE, TR) = S_0 \cdot \sin(\alpha) \frac{1 - e^{-TR \cdot R_1}}{1 - \cos(\alpha) e^{-TR \cdot R_1}} e^{-TE \cdot R_2^*} \quad 2.18$$

where S_0 is the thermal equilibrium signal intensity, TR is the repetition time, TE is the echo time, R_1 ($=1/T_1$) is the longitudinal relaxation rate, and R_2^* is the transverse relaxation rate. Equation 2.18 can be simplified by noting that TE is generally held constant as a function of flip angle in these types of acquisitions, and, moreover, that the scan parameter TE can, in many cases, be assumed to satisfy the relation $T_2^* (=1/R_2^*) \ll TE$, yielding the product $TE \cdot R_2^*$ is approximately zero and an exponential term of nearly unity. This results in the following simplified expression for the MRI signal intensity:

$$S(\alpha; R_1, S_0, TR) = S_0 \cdot \sin(\alpha) \frac{1 - e^{-TR \cdot R_1}}{1 - \cos(\alpha) e^{-TR \cdot R_1}} \quad 2.19$$

In certain instances, particularly when the specific absorption rate (SAR) limit is being approached, manufacturers will change the RF pulse structure causing slight changes in TE , but this assumption still holds to a first approximation, and these limits are rarely reached in brain applications. Generally speaking, all parameters (except for TR) of equation 2.19 are functions of spatial position. Although the acquisition and processing of serial images implicitly accounts for the spatial distribution of signal and relaxation rates, a more stringent assumption is often placed on the flip angle, namely that of spatial invariance. Therefore, any deviations in the flip angle, which can arise from transmit miscalibrations or spatial heterogeneity of the B_1 field, can and are known to cause errors in R_1 estimates [104].

Inversion recovery based relaxometry methods offer enhanced robustness with respect to systematic errors that plague other strategies (*e.g.*, RF heterogeneity of VFA acquisitions), and superb signal-to-noise ratio (SNR) efficiency [103]. Specifically, a spin-echo inversion sequence consists of a single 180° RF pulse followed by a programmable delay and a standard spin-echo imaging sequence. Solving the Bloch equations for arbitrary inversion times, inversion flip angle, and relaxation rates provides the following result,

$$S(TI; R_1, R_2, S_0, TE, TI, TR, \theta) = S_0 \left(1 + 2e^{-(TR-TE/2) \cdot R_1} - e^{-TR \cdot R_1} - (1 - \cos\theta)e^{-TI \cdot R_1} \right) e^{-TE \cdot R_2} \quad 2.20$$

where TI is the inversion time, θ is the inversion pulse flip angle, and all other quantities are as defined previously. Assuming that $TE \ll T_2$, a reasonable assumption for most acquisitions, the R_2 dependency is removed, as discussed for the FSPGR acquisition, resulting in

$$S(TI; R_1, S_0, TI, TR, \theta) = S_0 \left(1 + 2e^{-(TR-TE/2) \cdot R_1} - e^{-TR \cdot R_1} - (1 - \cos\theta)e^{-TI \cdot R_1} \right) \quad 2.21$$

Further simplifications can be made if the echo time is assumed to be much less than TR , which is generally the case in T_1 relaxometry techniques since TR is frequently chosen specifically to satisfy $TR > 5 \cdot T_1$, which allows full or nearly full recovery of the longitudinal magnetization to thermal

equilibrium. In combination, these assumptions provide the following simplified form of the Bloch equations solution:

$$S(TI; R_1, S_0, \theta) = S_0(1 - (1 - \cos \theta)e^{-TI \cdot R_1}) \quad 2.22$$

Rarely in practice is the nominal inversion flip angle of exactly 180° achieved, which is corrected here by the model parameter, θ . To appropriately satisfy all of the simplifying assumptions of this model, imaging times for this sequences can become prohibitively long (*e.g.*, a single slice, 256x128 matrix requires a 10 minute acquisition with $TR=5s$), requiring more temporally efficient imaging sequences such as fast spin-echoes, which can introduce additional errors in parameter estimation. When processing VTI data, another potential concern arises from the fact that images are generally computed and stored as magnitude data; the acquired complex raw data are not stored. Computation of the magnitude images changes the structure of noise in the image from a zero mean Gaussian distribution to a Rician distribution [105], particularly in the low SNR regions surrounding the null point – $TI_{null}(=\ln(2)/R_1)$ [12]. This can create bias in the parameter fits, which is mitigated by using the raw images or, to a lesser extent, by inverting the signal intensity of voxels in the magnitude images satisfying the condition $TI < TI_{null}$.

2.1.3 Dynamic-Contrast Enhanced-MRI

DCE-MRI refers to a collection of MR imaging techniques that utilize exogenous contrast agents, most frequently injected intravenously, to increase (*i.e.*, enhance) the observed MRI signal intensity. Gadolinium(III)-based, or Gd for short, contrast agents exhibit favorable properties for preferentially increasing longitudinal relaxation rates, *i.e.*, relatively large T_1 relaxivity, while minimally disrupting magnetic susceptibility, *i.e.*, relative small T_2 relaxivity, and, although these properties are not unique to Gd (*e.g.*, manganese exhibits similar paramagnetic properties), FDA and other regulatory organization approval and wide clinical availability of the low molecular weight (<1 kDa), extracellular agents, such as

gadopentetate dimeglumine (Gd-DTPA) delivered intravenously have provided the scaffold for innumerable quantitative clinical studies [106], [107]. Within several seconds following contrast agent (CA) injection, these agents pass through the vasculature and begin to distribute in the extracellular extravascular space (EES) of tissues exhibiting vascular permeability to such agents, which include most extracranial tissues and regions of the CNS with blood brain barrier (BBB) disruptions. This distribution of CA within the vascular system and, subsequently, in the EES of tissues forms the basis for the application of pharmacokinetic models. Although other approaches to quantification exist, such as CA time course shape analysis and semi-quantitative (or model-free) descriptors of tracer kinetics, the following development focuses on pharmacokinetic (PK) models with physiologically meaningful parameters [10].

2.1.3.1 Tracer Pharmacokinetic Models

Kety first described the kinetics of a freely diffusible tracer in which the rate of equilibration between the arterial and extravascular tracer concentrations was modeled [108]. In this early model, two compartments were considered: a vascular compartment and a tissue compartment. If a diffusible contrast agent is well-mixed and distributed uniformly in these two compartments with direct access to the water population, then a simple rate equation can describe the system. Tofts and Kermode [109] showed that the model developed by Kety [108] could be recast in terms of an extracellular CA concentration in the EES, $C_e(t)$, fractional volume of the EES, v_e , and the CA concentration in the plasma space, $C_p(t)$, giving

$$v_e \frac{dC_e(t)}{dt} = K^{trans} (C_p(t) - C_e(t)), \quad 2.23$$

where K^{trans} is the initial CA transfer rate from the plasma to the EES using the proposed consensus notation of Tofts *et al.* [110]. This model and these parameters are represented schematically in **Figure 2-1**. The quantity $C_p(t)$ is known as the arterial input function, or more generally as the vascular input

function (VIF), and $C_e(t)$ represents the CA concentration. Calculation of an integration factor turns this simple ordinary differential equation (2.23) into a separable differential equation, the solution of which is

$$v_e C_e(t) = K^{trans} \int_0^t C_p(\tau) \cdot e^{-\frac{K^{trans}}{v_e}(t-\tau)} d\tau \quad 2.24$$

or more compactly

$$v_e C_e(t) = C_p(t) \otimes H(t), \quad 2.25$$

where \otimes is the convolution operator and $H(t)$ is the impulse response function given by

$$H(t) = K^{trans} e^{-\frac{K^{trans}}{v_e}t}. \quad 2.26$$

Using this formulation, equation 2.25, in conjunction with 2.26, can be described intuitively as follows: extravasation of plasma tracer into the EES proceeds initially at a rate given by K^{trans} and results in the immediate exponential decay of EES tracer concentration at a rate of k_{ep} ($\equiv K^{trans}/v_e$).

In quantitative imaging studies, $C_e(t)$ is measured from either a voxel or region of interest, which is relatively large when compared with the size of the EES and/or plasma space. If contributions from the latter can be ignored, then the tissue concentration is given by

$$C_t(t) = v_e C_e(t). \quad 2.27$$

In normal brain, even when EES concentrations are estimated over relatively large volumes that might include plasma contributions, this relationship holds to a first order approximation as the vascular volume is on the order of 2-4% [109]. However, in highly vascular pathologies, such as glioblastomas, non-negligible vascular contributions may unduly bias the estimated tissue concentrations, resulting in overestimates of K^{trans} (known as pseudopermeability) [111]. A simple model accounting for the additional plasma volume is given by assuming tissue concentrations of the form

$$C_t(t) = v_p C_p(t) + v_e C_e(t), \quad 2.28$$

where v_p is the fractional plasma volume, and noting that, in general, $v_p + v_e \neq 1$. Substitution of equation

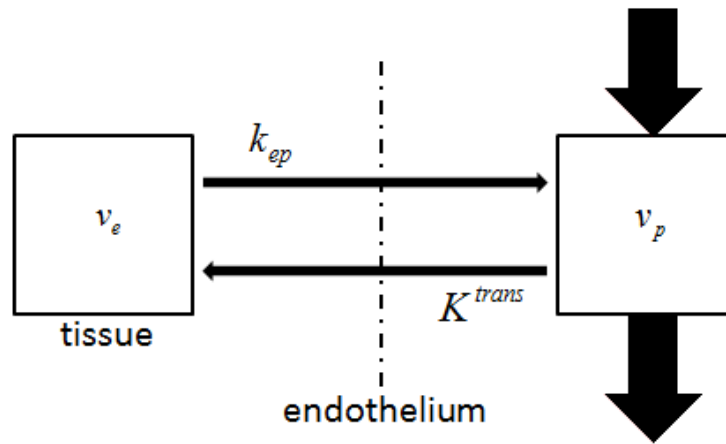


Figure 2-1. Two compartment kinetic model schematic. The pharmacokinetic model most frequently used in the analysis of DCE-MRI data consists of two compartments separated by the endothelium: an extracellular extravascular space (left) and a vascular space (right). K^{trans} describes the transfer rate constant of contrast agent from the vascular to EES space, v_e describes the fractional EES volume, and v_p describes the fractional plasma volume. The parameter k_{ep} is defined as K^{trans}/v_e . Arrows indicate blood flow through the vascular space.

2.28 into 2.25 yields the so called modified Tofts-Kermode or generalized kinetic model (GKM) [110], given by

$$C_t(t) = v_p C_p(t) + C_p(t) \otimes H(t). \quad 2.29$$

At this point, the model parameters describing the fractional intravascular and EES volumes have been described, but K^{trans} remains a simple initial rate constant without any other physical interpretation. Returning to equation 2.23 and considering the initial state of the system provides a means of interpreting K^{trans} . Prior to and during the initial arrival of contrast in the tissue, the rate of change in the EES CA concentration will be proportional to the blood flow. Moreover, for some time following the injection of CA, the EES CA concentration is negligible and, and equation 2.23 can be written as

$$v_e \frac{dC_e(t)}{dt} = F(1 - Hct)\rho C_p(t), \quad 2.30$$

where F is the flow of whole blood per gram of tissue, ρ is the tissue density, and the parenthetical term accounts for the fact that the CA in this discussion is extracellular agents. Equating the right hand sides of equations 2.23 and 2.30, after some algebraic manipulation, gives the following results,

$$K^{trans} = F(1 - Hct)\rho \frac{(C_p(t) - C_e(t))}{C_p(t)} = F(1 - Hct)\rho E, \quad 2.31$$

where the fraction of tracer concentration initially extracted into the EES is known as the extraction fraction E [110]. Renkin, in a study of postassium-42 tracer, modeled capillaries as a permeable cylinder with permeability P and surface area S [112], and Tofts *et al.* [110] reformulated this result using the proposed standard notation, giving

$$E = \frac{(c_p(t) - c_e(t))}{c_p(t)} = 1 - e^{-PS/F(1-Hct)}. \quad 2.32$$

The general form of K^{trans} can thus be written as

$$K^{trans} = F(1 - Hct)\rho \left(1 - e^{-PS/F(1-Hct)} \right), \quad 2.33$$

where K^{trans} is a combination of the permeability surface area PS and capillary flow. Two mathematically limiting cases of this model are often discussed: flow limited ($F \ll PS$) and permeability limited ($PS \ll F$) [10], [12], [110]. The former results in the original model proposed by Kety [108] (*i.e.*, K^{trans} is proportional to F) and the latter, with the help of a Taylor expansion, yields

$$K^{trans} \approx F(1 - Hct)\rho \left(1 - \left(1 - \frac{PS}{F(1-Hct)} \right) \right) = PS\rho. \quad 2.34$$

While these limiting cases provide simple interpretations of K^{trans} , in general, the underlying physiology frequently renders both limiting cases unreasonable approximations and K^{trans} must be interpreted as an extraction-flow product dependent upon capillary flow and permeability surface area.

Not surprisingly, a plethora of other pharmacokinetic models have been proposed. Li and Jackson [113] proposed a simplified method for extracting GKM parameters from only the first pass of tracer. St. Lawrence and Lee [114] proposed the adiabatic approximation to tissue homogeneity model, which exploits the early flow-limited (*i.e.*, K^{trans} is proportional to F) phases of contrast enhancement followed by a general mixture of flow and permeability (equation 2.33) to allow the computation of blood flow. Models of increasing complexity introduce additional challenges to acquisition and analysis strategies, which will be discussed in detail in the following sections, and have slowed the wide-spread adoption of more sophisticated models, leaving the GKM the most widely employed pharmacokinetic model in the analysis of low molecular weight Gd-based CA.

2.1.3.2 Quantitative DCE-MRI – Acquisition and Analysis Strategies

Gadolinium based CAs, as discussed at the beginning of section 2.1.3, can be used to affect the longitudinal relaxation rates of biological systems, and in conjunction with a T_1 -weighted imaging strategy and PK model allow investigation of vascular permeability/flow effects. More specifically, the interaction of protons with this paramagnetic complex result in an increase in the native longitudinal relaxation rate R_{10} , which can be described by a simple linear relationship

$$R_1 = R_{10} + r_1[Gd], \quad 2.35$$

where R_1 is the resulting longitudinal relaxation rate of the sample, r_1 is the T_1 relaxivity of the Gd-based CA in units of $\text{mM}^{-1}\text{s}^{-1}$, and $[Gd]$ is the concentration of the CA in mM. Dipole-dipole interactions (the dominant source of longitudinal relaxation mechanisms in liquids and soft tissues) are inherently short range, requiring direct access to a water population, and therefore require a substantive distribution of Gd throughout the observed compartment to affect changes in the MRI signal. Quantitative DCE-MRI modeling utilizes imaging protocols that minimize the effect of this R_{20} shortening (namely, short TE acquisitions) and exploit the changes in R_{10} , manifested as changes in MRI signal intensity, to indirectly observe the distribution of Gd throughout the intravascular and EES compartments by rapidly acquiring T_1 -weighted images prior to and following injection of the CA.

T_1 -weighted FSPGR sequences (discussed in some additional detail in section 2.1.2) have been used extensively for DCE-MRI acquisitions as these techniques are capable of acquiring relatively large volumes with reasonable spatial (sub-millimeter in plane) and temporal ($\sim 5\text{s}$) resolution, allowing coverage of the anatomy being investigated while providing adequate temporal estimation of CA concentration used in the PK model. Prior to CA injection, a cinematic sequence of acquisitions is initialized capturing the pre-contrast signal intensities. After a number of pre-contrast frames have been

acquired, the CA is injected intravenously followed by several more minutes of image acquisition. Before PK modeling can occur, the signal intensity must ideally be converted into [Gd].

Assuming that the [Gd] remains low enough such that $T_2^* \ll TE$ and that the presence of CA has an inconsequential impact on the proton density (*i.e.*, S_0 is constant), then the signal intensity as a function of time is given by combining equations 2.19 and 2.35

$$S([Gd](t); \alpha, r_1, R_{10}, S_0, TR) = S_0 \cdot \sin(\alpha) \frac{1 - e^{-TR(R_{10} + r_1[Gd](t))}}{1 - \cos(\alpha) e^{-TR(R_{10} + r_1[Gd](t))}}, \quad 2.36$$

where α is the flip angle. For convenience, define the constant quantity

$$\Gamma = \frac{1 - e^{-TR \cdot R_{10}}}{1 - \cos(\alpha) e^{-TR \cdot R_{10}}} \quad 2.37$$

and consider the ratio of pre-contrast signal intensities, say S_{pre} , to the post-contrast time-varying signal intensity of equation 2.36. Through careful algebraic manipulation, [Gd] as a function of observed pre- and post-contrast signal intensities can be shown to be

$$[Gd](t; r_1, R_{10}, \alpha, TR) = \frac{1}{r_1} \left\{ \frac{1}{TR} \ln \left(\frac{\Gamma \cdot \cos(\alpha) - S_{pre}/S(t)}{\Gamma - S_{pre}/S(t)} \right) - R_{10} \right\}. \quad 2.38$$

At this point, the relationship between [Gd] and MRI signal intensity can be seen to depend only on the known acquisition parameters α and TR , the Gd relaxivity r_1 , and the native longitudinal relaxation rate R_{10} . Generally, a single value is assumed for the relaxivity, which will depend on temperature, magnetic field strength, and the physiochemical environment (*e.g.*, tissue type) [115], [116]. However, there are principally two approaches to providing a value for R_{10} . In one approach, R_{10} is estimated using, for example, the FSPGR method of section 2.1.2. The resultant relaxation rate maps can then be used in conjunction with the dynamic series images to calculate [Gd] on a voxel-by-voxel basis, but additional uncertainties in the estimation of R_{10} will propagate through to the calculation of [Gd] and ultimately to the PK parameters. Another common approach is to assume that R_{10} varies minimally across various

tissues and pathologies (see, for example, the reviews of Bottomley *et al.* [117], [118]) and assume a single relaxation rate for tissue (usually 1000ms) and one for the intravascular compartment (usually 1440ms). Haacke *et al.* [85] showed that assuming a constant value for the intrinsic longitudinal relaxation rate resulted in more reproducible results. Although biases may be present, there is a paucity of evidence suggesting that these biases would necessarily vary significantly over relatively short durations of chemotherapeutic or radiation therapies and would, there, have little impact on PK parameters changes [119].

A second, more parameter independent, approach is also used commonly in literature. Again, assuming that S_0 and T_{20}^* are unaffected by the presence of [Gd], and that the $TR \ll 1/R_1$, a Taylor series expansion of the exponential terms in equation 2.36 results in

$$S([Gd](t); \alpha, r_1, R_{10}, S_0, TR) = S_0 \cdot \sin(\alpha) \frac{1 - (1 - TR \cdot R_1 + \dots)}{1 - \cos(\alpha) \cdot (1 - TR \cdot R_1 + \dots)}. \quad 2.39$$

Furthermore, in the limit as the flip angle α approaches 90° this expression reduces to

$$S([Gd](t); r_1, R_{10}, S_0, TR) \propto S_0 \cdot TR \cdot R_1. \quad 2.40$$

Applying these same approximations to the pre-contrast signal intensity and taking the difference in post- and pre-contrast signal intensities yields a simple result,

$$\Delta SI = S - S_{pre} \approx S_0 \cdot TR \cdot r_1 [Gd], \quad 2.41$$

Therefore, for large flip angles, the change in signal intensity following administration of the CA is proportional to the [Gd] within a constant scaling factor ($S_0 \cdot TR \cdot r_1$). However, as the [Gd] increases, the assumption that $TR \ll 1/R_1$ rapidly degrades the validity of this approximation, and the resulting concentrations are underestimated. Regardless, these approximations provide a method of quantifying PK parameters without *a priori* knowledge of the inherent relaxation rates.

The last crucial component in the quantitative analysis of DCE-MRI data is the selection of an appropriate VIF. Early researchers, Tofts and Kermode [109], Weinmann *et al.* [120], and Brix *et al.* [121], for example, used empirically derived population averaged models to describe the VIF. However, the VIF time course is intimately dependent on exam specific factors such as CA injection rate and dose, and, more importantly, on patient specific factors such cardiac output, pharmacokinetic distribution of the CA, and kidney glomerular filtration rates, all of which can exhibit considerable variance. As hardware performance has improved, fast imaging techniques capable of capturing the rapid changes in signal intensity resulting from a CA bolus have become commonplace. Although some disagreement regarding the reproducibility of the respective methods exists [122], [123], measurement of individual patient VIFs is generally accepted to be more physiologically accurate [10], [12].

2.1.3.3 Quantitative DCE-MRI – Additional Considerations

Throughout the preceding discussions of DCE-MRI modeling, the distribution of CA in the intravascular and tissue compartments was described appropriately as a continuous physical process. However, in practice, the image sampling rate is limited as a result of hardware limitations and application specific acquisition parameter choices. To date, several researchers have investigated the effects of temporal resolution on PK estimates. Early investigations by Karmonik and Jackson [124] found a temporal resolution of less than 4s was required to ensure that the 95% confidence interval of K^{trans} estimates was within 10% of the true value. Henderson *et al.* [125] later corroborated these findings, showing that sampling intervals of once every 4s are required to ensure that K^{trans} and v_e are within 10% of the true values. For the adiabatic approximation to tissue homogeneity model, Kershaw and Cheng [126] showed unavoidable bias and variance were introduced unless images were sampled at least once every 3s. Notably, these sampling rates allow adequate measurement of the rapid VIF kinetics, with tissue kinetics requiring substantially lower temporal resolution [125]. To circumvent these stringent temporal resolution requirements, Kovar *et al.* [127] proposed a reference region that utilizes

a nearby enhancing region, such as muscle, *in lieu* of measuring a VIF, a method that was further refined by Yankeelov *et al.* [128]. Heisen *et al.* [129], using such tissue reference methods, showed significant reduction in the temporal resolution requirements, *i.e.*, data acquisition only once every 20s, still provided accurate quantification. However, in brain applications, the lack of enhancing structures in the imaging volume is problematic for reference region models. Moreover, in applications assessing treatment response, identification of a suitable region that remains unaffected by the treatment regimen becomes increasingly problematic.

Aside from the strict temporal requirements of DCE-MRI, the discussions of section 2.1.3.2 assumed that the tissue and intravascular compartments contained a homogenous water population that interacted directly with the Gd complex. In biological tissues, water is further compartmentalized within tissues, with a majority contained inside the constituent cells. Inter-compartmental transit of these water molecules can render the simple linearity described by equation 2.35 invalid. If little difference exists between the relaxation rates of the compartments, then the observed compartments will exhibit a nearly single valued relaxation rate (an average of the compartments' rates) and the linear relationship between [Gd] and R_1 holds. Under these circumstances, the compartments are said to be in the fast exchange limit or fast exchange regime [115]. Conversely, the slow exchange regime results from large differences in compartment relaxation rates. Donahue *et al.* [115] showed that both regimes can be present in a given DCE-MRI study, with intravascular-EES compartments exhibiting slow exchange and intracellular-EES compartments giving rise to fast exchange. Implications of the former effect were studied by Larsson *et al.* [130], and concluded to have minimal impact on the quantification of PK parameters, specifically K^{trans} , at clinical doses, a result corroborated by Buckley [131]. Evidence strongly suggests that the fast exchange limit reasonably approximates intracellular-EES water exchange for clinically relevant doses, usually 0.1 mmol/kg [115], [132]. Despite this evidence, researchers such as Landis *et al.* [133] and Yankeelov *et al.* [134] argued that this linearity holds for only a limited range of

tissue concentrations, which prompted the investigation of transcytolemmal water exchange and the development of the “shutter-speed” model to address the effect of compartmental water exchange rates. However, Buckley *et al.* [135] showed that the clinically relevant acquisitions were insensitive to this exchange, warranting the use of simpler models that assume the fast exchange limit, especially when coupled with short TR values and large flip angle FSPGR acquisitions [136].

Other notable sources of bias and, potentially, variance exist in the analysis of DCE-MRI data, such as differences in artery and capillary hematocrits, choice of the VIF, spatial R_{10} estimation, and environment dependent r_1 values. For these factors, some remain uninvestigated in the literature, and, for others, consensus regarding appropriate analysis methods has yet to be reached. By maintaining consistency in the choice of analysis and acquisition techniques, the model associated variances will be minimized despite potential biases. Moreover, in the context of treatment assessment, reproducibly biased methods reign supreme over those exhibiting lower bias and higher variance, as ultimately the goal is to detect changes, not absolutely quantitate physiological parameters.

2.1.4 Dynamic-Susceptibility Contrast-MRI

In the scientific literature, the terms DCE and DSC are occasionally used interchangeably, but for the following discussions DSC-MRI refers specifically to the collection of MR imaging techniques that exploit the magnetic susceptibility of paramagnetic contrast agents, in a manner analogous to DCE-MRI, to modify the observed MRI signal intensity. Unlike DCE, T_2 - or T_2^* -weighted sequences are used to observe the associated changes in the bulk magnetic susceptibility, affecting a decrease in signal intensities, rather than the enhancement seen in the T_1 -weighted counterparts. When these extracellular contrast agents are confined to relatively small regions, such as the intravascular space, for example, large local magnetic fields resulting from the paramagnetic properties of Gd induce increases in the transverse relaxation rate. However, the compartmentalized CA is unable to directly interact with the large extravascular water population, minimizing the associated T_1 changes, and resulting in a

disproportionately large change in R_2 or R_2^* . Analogously to T_1 -weighted techniques, rapid T_2 - or T_2^* -weighted imaging sequences provide a means of temporally estimating the CA distribution through the vasculature, allowing quantitation of physiological parameters and, if desired, semi-quantitative descriptors.

2.1.4.1 Tracer Pharmacokinetic Models

Early work by Meier and Zierler [137] culminated in the basic mathematical description used most frequently to describe the pharmacokinetics of non-diffusible tracers. These methods were later extended beyond initial applications utilizing injected dyes, ultimately providing a means of using extracellular contrast agents to imaging modalities such as MRI [138] to quantify cerebral blood flow and volume. While early researchers such as Østergaard *et al.* [138] and Sorensen *et al.* [139] quantified blood flow in brain tumors, Aronen and Perkiö [140] noted that the use of blood flow in evaluating high-grade brain tumors (the subject of this work) may be less appropriate than the cerebral blood volume owing to the poorly formed highly heterogeneous (spatially and temporally) vasculature. With this consideration, the following discussions will focus primarily on the quantification of the relative cerebral blood volume ($rCBV$).

Consider a closed vascular system (i.e., the CA remains intravascular) in which a CA is delivered through a single inlet, subsequently exiting the system through a single outlet. If CA recirculation is prevented, the amount of CA that passes through some volume of interest, $C(t)$, normalized by the total amount contrast injected into the system measured from a AIF, $C_a(t)$, can be represented as [10], [12]

$$rCBV = \frac{H_f}{\rho} \frac{\int_{t_0}^{\infty} C(t) dt}{\int_{t_0}^{\infty} C_a(t) dt}, \quad 2.42$$

where H_f is the hematocrit factor, ρ is the tissue density, and t_0 is the time at which the CA concentration becomes different from zero. The hematocrit factor, H_f , accounts for the difference in the hematocrits of small and large blood vessels, and is given by

$$H_f = (1 - Hct_{art}) / (1 - Hct_{cap}). \quad 2.43$$

Clearly, in clinical practice, preventing CA recirculation is impossible, resulting in concentration time curve that include contributions from the first-pass kinetics and subsequent CA recirculation. An example [Gd] recirculation as seen in a VIF is given in **Figure 2-2**. To avoid overestimating the $rCBV$ by including recirculation effects, the upper limit of the integrals can be modified to include only first-pass CA contributions, or

$$rCBV = \frac{H_f \int_{t_0}^{t_{fp}} C(t) dt}{\rho \int_{t_0}^{t_{fp}} C_a(t) dt}, \quad 2.44$$

where t_{fp} is the time at which reperfusion of the CA begins. However, performing the integration of equation 2.44 will systematically underestimate the $rCBV$, as the result of neglecting the first-pass concentration that is superimposed with CA recirculation shortly following t_{fp} . Thompson *et al.* [141] showed that a gamma-variate,

$$C(t) = \kappa(t - t_0)^\alpha e^{-(t-t_0)/\beta}, \quad 2.45$$

provided an excellent empirical description of the first-pass kinetics, where α , β , and κ are free model parameters, and t_0 is the bolus arrival time, which can be assumed or, more often, fitted as another model parameter. This approach is illustrated in **Figure 2-2**, in which a gamma-variate was fitted to the first-pass peak of a VIF. Integrating the resultant fitted gamma-variate in this way provides a means of extracting first-pass kinetics without unnecessarily biasing the resultant measurement by purposely neglecting or including concentration contributions that violate the assumptions of the system model.

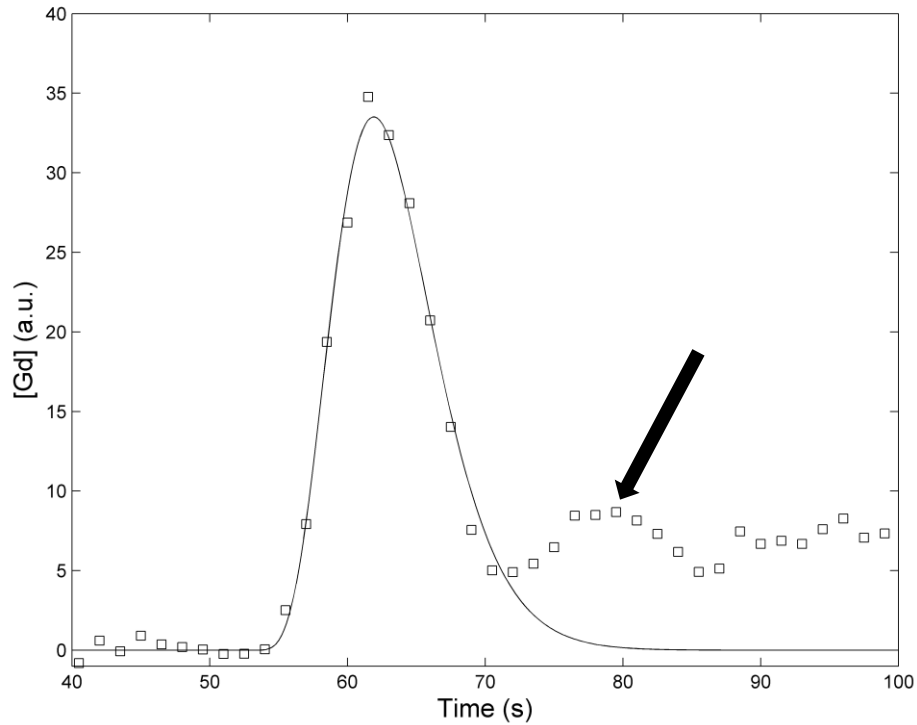


Figure 2-2. Contrast agent recirculation example. This plot shows a VIF contrast uptake curve with the solid line representing a fitted gamma-variate and the black arrow specifying the characteristic recirculation peak.

While this approach is used frequently to calculate the $rCBV$ independent of $rCBF$, the requirement of a non-linear fitting algorithm, especially with four free parameters, requires substantial computational resources, and still requires the determination of a reperfusion cut-off, neglecting all data following the cut-off when performing fits to equation 2.45.

2.1.4.2 Quantitative DSC-MRI – Acquisition and Analysis Strategies

When confined to intravascular spaces, the gadolinium based CAs, as discussed at the beginning of section 2.1.4, can be used to affect the transverse relaxation rates of biological systems, and in conjunction with a T_2 - or T_2^* -weighted imaging strategy and PK model allow investigation of the relative vascular volume. As with the longitudinal relaxation rate, the presence of Gd affects the local

transverse relaxation rate based on the T_2 relaxivity, r_2 , of the CA, and can be, to a first order approximation described by

$$R_2 = R_{20} + r_2[Gd], \quad 2.46$$

where R_2 is the resulting transverse relaxation rate of the sample and $[Gd]$ is the CA concentration. In addition to the increase in R_2 , Gd also alters the local bulk magnetic susceptibility resulting in relatively large inhomogeneous local magnetic fields that interact with intra- and extravascular water protons, inducing further dephasing beyond that of the R_2 mechanisms. When using pulse sequences that fail to refocus dephasing arising from local field inhomogeneities, the transverse relaxation rate is described more completely by a superposition of relaxation mechanisms accounting for the thermal signal decay (R_2) and inhomogeneous field induced decay (R_2'), represented mathematically as

$$R_2^* = R_2 + R_2', \quad 2.47$$

from which an expression analogous to equation 2.46 can be written to describe the resultant relaxation rate, namely

$$R_2^* = R_{20}^* + r_2^*[Gd] . \quad 2.48$$

Intravascular accumulation of the CA modifies both the R_2 and R_2^* relaxation rates, but, for compactness and because the acquisitions of this work were T_2^* -weighted, R_2^* will be used exclusively in the discussion to follow.

In the same manner as DCE acquisitions (section 2.1.3.2), a fast imaging sequence capable of acquiring a volume of images every 1-2s is required to adequately estimate the temporal and spatial distribution of the CA. Generally, SE or gradient recalled echo (GRE) echo planar imaging (EPI) sequence is used to rapidly acquire T_2 - or T_2^* -weighted images, respectively, although other sequences have been used [12]. Several pre-contrast phases are acquired to establish the steady-state baseline signal

intensity, followed by another few minutes of cinematic image acquisitions. The resulting serial images can then be used to estimate $[Gd]$, which is, in turn, used to estimate $rCBV$.

Assuming that the introduction of CA negligibly impacts the proton density (*i.e.*, S_0 is constant) and longitudinal relaxation rates (*i.e.*, R_{10} is constant), then for simple GRE (or SE by removing the * superscript) imaging sequence, the solution to Bloch's equations can be written compactly as

$$S([Gd](t); r_2^*, R_{20}^*, S_0) = S_0 \cdot \Omega(TE, TR, \alpha, \dots) e^{-TE \cdot (R_{20}^* + r_2^*[Gd])} \quad 2.49$$

where Ω is a constant, pulse sequence specific function incorporating constant imaging parameters such as TE , TR , and the flip angle α . A similar expression can be easily written for the pre-contrast signal intensity, S_{pre} , resulting in an expression for $[Gd]$ after performing some algebraic manipulations on the ratio of pre- and post-contrast signal intensities and is given by

$$[Gd](t) = \frac{1}{TE \cdot r_2^*} \ln \left(\frac{S_{pre}}{S(t)} \right) \quad 2.50$$

Unlike the DCE acquisition strategies discussed previously (section 2.1.3.2), access to the CA concentration is independent of the pre-contrast transverse relaxation rate, greatly simplifying quantitative analysis. However, quantitation still requires the determination of a VIF or, using the notation of section 2.1.4.1, $C_a(t)$. Common measurement locations for the VIF include the cerebral arteries (or some peripheral branch) and the internal carotid arteries [12].

2.1.4.3 Quantitative DSC-MRI – Additional Considerations

Perhaps the most important consideration in modeling DSC-MRI as described in this section is the validity of assuming that the CA remains compartmentalized in the intravascular space. In fact, use of the DCE-MRI pharmacokinetic models described in the previous section (2.1.3) relies heavily on the assumption that the CA will extravasate, an assumption that is likely invalid in high-grade brain tumors. Two unique issues face PK modeling under these conditions: R_1 increases resulting from extravascular

CA distribution will, at least partially, mask the associated R_2^* effects [142] and the extravasation of CA will reduce intravascular $[Gd]$ resulting in an underestimations of the $rCBV$. Kassner *et al.* [142] showed that pre-loading the system with Gd effectively minimized the R_1 shine-through, a viable solution to the former issue. However, compensating for the $rCBV$ underestimation requires the use of more sophisticated PK models such as the first-pass pharmacokinetic model proposed by Johnson *et al.* [143].

2.2 Statistics

2.2.1 Measuring Agreement and Repeatability

One method used frequently to detect agreement between known values and measurements is Bland-Altman analysis, which provides the limits of agreement (LOA), calculated by

$$\langle d \rangle \pm 1.96\sigma \left(1 + \frac{1}{n} \right) \quad 2.51$$

where $\langle d \rangle$ is the average of differences between, in this case, the estimated and actual PK parameters, σ is the standard deviation of differences, and n is the number of samples [144]. This formulation essentially provides the 95% confidence interval for normally distributed differences, where agreement occurs when an application specific acceptable difference encompasses these limits. Another frequently used method is the concordance correlation coefficient [145], or CCC for short, which estimates the expected deviation from a perfect correlation line (*i.e.*, slope of one and intercept of zero) and is given by

$$\rho_c = \frac{2\rho\sigma_A\sigma_E}{\sigma_A^2 + \sigma_E^2 + (\mu_A - \mu_E)^2} \quad 2.52$$

where σ_A and σ_E are the variance of the actual and estimated values, respectively, μ_A and μ_E are the means of the actual and estimated values, respectively, and ρ is the Pearson correlation coefficient.

2.2.2 Logistic and Ordinal Regression

Logistic regression is a type of statistical classification technique that provides a means of modeling a binary response, and can be considered a special case of ordered logistic regression or, stated more succinctly, ordinal regression. These flexible models provide a means of modeling categorical (ordinal) data, such as radiological response (partial response, stable disease, progressive disease, etc.). In particular, ordinal regression allows multiple categories to be modeled simultaneously through the assignment of categorical probabilities. Discussions herein were adapted from the text by Johnson and Albert [146].

Let p_i denote the probability of success for the i^{th} observation, X_{ij} denote the vector of explanatory variables for the i^{th} observation, and β_j is the regression coefficient for the j^{th} explanatory variable. In the context of modeling ordinal data, failure and success can be restated as classification into categories, say, C_1 and C_2 , respectively. Using this nomenclature, the probability that the i^{th} observation falls in category C_2 (success) can be written as

$$p_i = F(\beta_0 + \sum_{j=1}^n \beta_j X_{ij}), \quad 2.53$$

or more compactly, by indexing j such that j assumes all integer values between 0 and n , and X_{i0} is one for all values of i ,

$$p_i = F(\boldsymbol{\beta}'\mathbf{X}_i) = F(Z_i), \quad 2.54$$

where F is known as the link function and Z is referred to as latent variable or latent trait. Link functions, usually chosen from the class of cumulative density functions (CDF), map the latent variable's range, described in the argument of equation 2.54, to the interval $(0, 1)$, providing a means of probabilistic

interpretation. Common link functions include CDFs of the normal, logistic, and log-logistic distributions, and although the choice is somewhat arbitrary, the logistic distribution will provide the foundation for this work. The CDF for the logistic distribution is given by

$$F(Z) = \frac{1}{1+e^{-Z}} . \quad 2.55$$

After a link function has been selected and covariates are chosen, values of p_i can be estimated from the response of a training data set, and maximum likelihood estimation can be used to calculate the regression coefficients. Predictions can then be performed on new data sets using the newly estimated logistic model.

Latent variables, as introduced in equation 2.54, can be used to model unobservable traits, and, moreover, provide a framework for intuitively extending the categorical response modeled by logistic regression to more than two categories (*i.e.*, ordinal regression). These traits provide a means of determining the categorical response. Given some observation, the latent trait for that observation can be modeled as

$$Z_i = \boldsymbol{\beta}'\mathbf{X}_i + \epsilon_i , \quad 2.56$$

where the error terms, ϵ_i , are assumed to be identically distributed according to the distribution of the link function and are also assumed to be independent. Then for two categories, say C_1 and C_2 (failure or success, respectively, as discussed previously for logistic models), the response can be written as

$$y_i = \begin{cases} C_1 & \text{if } Z_i \leq 0 \\ C_2 & \text{if } Z_i > 0 \end{cases} . \quad 2.57$$

This notion can be extended easily to accommodate k ordered categories by introducing category cut-offs, γ_c , resulting in the following response definition,

$$y_i = \begin{cases} C_1 & \text{if } \gamma_0 < Z_i \leq \gamma_1 \\ C_1 & \text{if } \gamma_1 < Z_i \leq \gamma_2 \\ \vdots & \\ C_k & \text{if } \gamma_{k-1} < Z_i \end{cases}, \quad 2.58$$

where γ_0 is defined to be negative infinity, a mere notational convenience. Defining the response in this manner allows for the development of ordinal regression models.

To formulate ordinal regression in an analogous manner to logistic regression (equation 2.54), the individual category probabilities must be computed. Since the response probabilities are parameterized by the CDF F and the latent trait Z , the probability of attaining a particular category, say category c , is simply the integral of the probability density function (the first derivative of the CDF) between the cut-offs γ_c and γ_{c-1} , or, using the first fundamental theorem of calculus,

$$p_{ic} = F(\gamma_c - \boldsymbol{\beta}'\mathbf{X}_i) - F(\gamma_{c-1} - \boldsymbol{\beta}'\mathbf{X}_i). \quad 2.59$$

By defining the cumulative probability of the c^{th} category as

$$\theta_{ic} = p_{i1} + p_{i2} + \cdots + p_{ic}, \quad 2.60$$

one can easily show, using equations 2.59 and 2.60, that a system of $k-1$ equations following the operational form of the logistic regression described previously (*i.e.*, equation 2.54) is given by

$$\theta_{ic} = F(\gamma_c - \boldsymbol{\beta}'\mathbf{X}_i). \quad 2.61$$

To avoid overparameterization of the model, the cut-off γ_1 is fixed to zero, and all remaining cut-offs and regression parameters are calculated using maximum likelihood estimation. If the observations are assumed to be independent, then responses, y_i , will follow a multinomial distribution, a generalization of the binomial distribution (the distribution describing the response of logistic models).

In standard linear regression, the coefficient of determination or R^2 is of considerable utility in assessing the explanatory nature of the independent variable when defined as the fraction of explained

model variance, and is often used a goodness-of-fit indicator. Nagelkerke [147], recognizing the utility of R^2 , generalized the notion to encompass models such as ordinal regression. Let $L(0)$ define the likelihood of the null model (*i.e.*, only model constants such as the category cut-offs and β_0 are determined) and $L(\beta)$ be the likelihood of the full regression model (*i.e.*, all cut-offs and regression coefficients are calculated), then the coefficient of determination proposed by Nagelkerke [147] is given by

$$\bar{R}^2 = \frac{1 - (L(0)/L(\beta))^{2/n}}{1 - L(0)^{2/n}}, \quad 2.62$$

where n is the number of observations, and can be interpreted analogously to the classical R^2 . In fact, when applied to linear regression and the maximum likelihood estimate approach is equivalent to ordinary least squares, the classical definition and that proposed by Nagelkerke give the same results.

Statistical classification techniques utilize a simple construction for performing predictions. Let \mathbf{p}_i be the vector of probabilities for a k category ordinal regression model for which the category cut-offs and regression coefficients have been determined, written mathematically as

$$\mathbf{p}_i = (p_{i1}, p_{i2}, \dots, p_{ik}), \quad 2.63$$

then the predicted category given by

$$\hat{y}_i = \arg \max_c \mathbf{p}_i. \quad 2.64$$

This succinct notation simply states that the predicted response for some new observation yields a vector of probabilities \mathbf{p}_i for which the predicted classification is given by the category with the maximum probability.

2.2.3 Cross-Validation

Cross-validation is a technique for assessing the generalizability of classification models, a thorough discussion of which can be found in the text by Hastie *et al.* [148], for example. Given some set

of training data, $\{X_{ij}, Y_i\}$, the goal of cross-validation is to determine the misclassification (or classification error) rate on some larger set of data. If predictions are performed on some set of observations different from the training set, the error rate for logistic/ordinal models discussed previously is given by the ratio of misclassified predictions to the total number of observations, or written succinctly as

$$Err = \frac{1}{N} \sum_i I(y_i \neq \hat{y}_i), \quad 2.65$$

where $I(\dots)$ is the indicator variable for prediction misclassification. When a specific observation is correctly classified, the indicator assumes a value of zero and one otherwise. Prediction accuracy can be defined in a similar manner, or simply computed simply as $1-Err$. In the context of this work, the goal of cross-validation is two-fold: identify an appropriate model and assess that model's performance.

K-fold cross-validation attempts to estimate the model prediction error rate by partitioning the data into K subsets. By training the classifier on K-1 sets and performing predictions on the K^{th} test set that was removed from the training set, an estimate of the classification error rate can be obtained. If K is chosen to be the number of observations in the training data, then the so called leave-one-out (LOO) cross-validation technique is attained, the focus of this work.

Chapter 3 Assessing MR Equipment Bias and Variance

3.1 Introduction

Longitudinal assessment of quantitative imaging biomarkers for the purpose of treatment response requires scrupulous assessment of measurement bias and variance attributable to the equipment to ensure that the measured biological changes of interest are distinguishable from such errors. The necessity of characterizing sources of equipment bias and variance, particularly those related to quantitation, is of significant interest and importance as evidenced by numerous associated initiatives from major organizations such as the American Association of Physicists in Medicine, Radiological Society of North America, International Society of Magnetic Resonance in Medicine, and the National Cancer Institute [73]–[76]. Within each of these organizations, several committees focusing on different modalities and different aspects of image quantitation have been formed and are beginning to amass data that consistently elucidate the need for longitudinal system assessment [78]–[80], [149].

To assess relaxometry, an important property intrinsic to biological tissues that is exploited in contrast enhanced exams, and diffusion acquisition properties using physical phantoms, researchers have proposed numerous approaches of varying complexity [150]–[153], including solutions and gels, in addition to more complicated phantoms that attempt to mimic the actual physiological properties being measured, such as the dynamic lesion phantom developed by Freed *et al.* [154] to assess DCE acquisitions or the fiber phantom developed by Laun *et al.* [155] to assess DTI quantitation. To achieve an optimal assessment of equipment bias and variance, sources of error, such as those introduced by the manufacturing materials and manufacture processes, solution stabilities, and experimental conditions (e.g., temperature), must also be mitigated where possible. With this goal in mind, the primary design considerations for the phantoms proposed for these studies included: easily reproducible manufacturing procedures, non-volatile chemicals, measurement values representative of

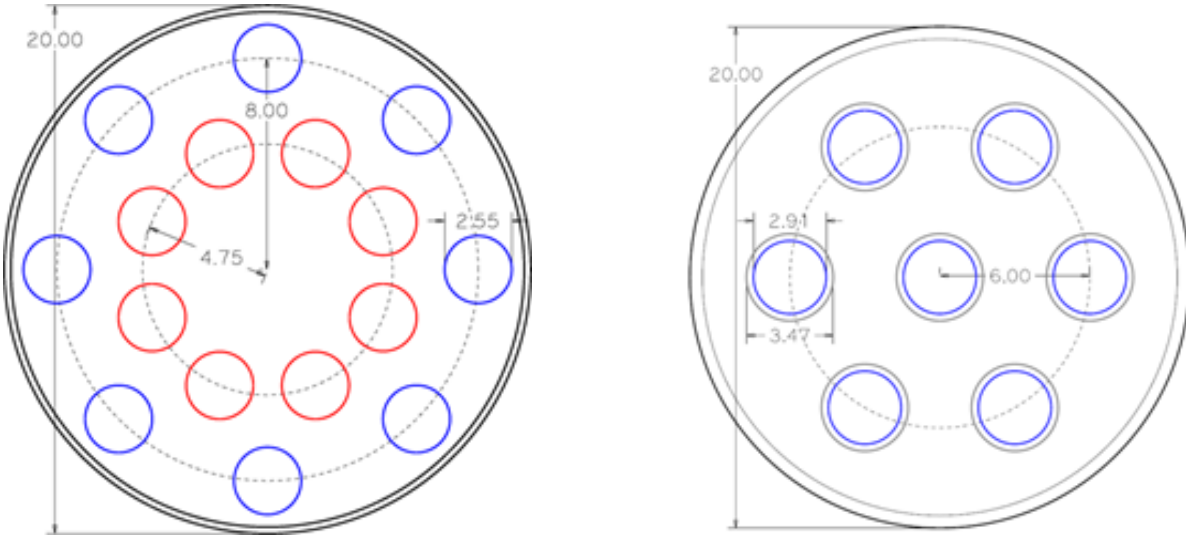


Figure 3-1. DCE and ADC phantom schematics. Axial cross-sectional schematic representation of the DCE (*left*) and ADC (*right*) phantom (units: cm). The DCE phantom inner ring (*red*) contains 8 solutions representative of VIF response (R_1 : 1.05-41.86 s^{-1}) while the outer ring (*blue*) contains 8 solutions representative of tissue response (R_1 : 0.97-7.84 s^{-1}). The ADC phantom contained vials with various sucrose concentrations (10 to 45 %w/w); the lowest sucrose concentration was placed at the center.

those found *in vivo*, long-term solution stability, well characterized temperature dependences, and reproducible acquisition strategies. Phantoms matching these criteria were designed and manufactured with the primary purpose of assessing equipment related variances using clinically relevant diffusion and contrast-enhanced acquisitions.

3.2 Phantom Design

3.2.1 DCE Phantom

DCE acquisitions are designed to capture dynamic processes and, through modeling, estimate physiological parameters. Currently, no viable dynamic phantom capable of mimicking these processes

is available despite a recent proof-of-concept dynamic phantom [154]. Therefore, the approach taken in this work focused on quantifying variances in the measurement of R_1 relaxation rates and propagating those measured variances to the estimation of PK parameters through simulated data using a static phantom consisting of a large compartment filled with a homogeneous fill solution and several smaller compartments with solutions having a range of R_1 values. To adjust the R_1 values, NiCl_2 was used as a paramagnetic doping material. As an R_1 altering dopant, NiCl_2 is approximately independent of magnetic field strength for common clinical magnets (*i.e.*, 0.2-4.0T) and temperature, ideal properties for solutions used to estimate equipment bias and variance.

A schematic of the DCE phantom is shown **Figure 3-1**. The cylindrical phantom was constructed from a polycarbonate cylinder measuring 20 cm in diameter with a depth of 12 cm and featured a removable lid for installing various vials and a threaded plug providing access to the fill solution for temperature measurements. Using an NiCl_2 relaxivity of $0.62 \text{ mM}^{-1}\text{s}^{-1}$ [156], [157], theoretical $[\text{NiCl}_2]$ values were calculated for a range of relaxivities representative of expected *in vivo* gadolinium concentrations occurring during the course of a DCE study in a tissue compartment (R_1 : $0.67\text{-}7.54 \text{ s}^{-1}$) and a vascular compartment (R_1 : $0.75\text{-}41.53 \text{ s}^{-1}$). Two batch solutions, one for each of the highest tissue and vascular R_1 values, were prepared using degassed, deionized water. Serial dilutions of these batch solutions were performed to achieve the theoretical $[\text{NiCl}_2]$ listed in **Table 3-1**. These solutions were poured into 30 mL sealing, polypropylene copolymer, centrifuge tubes (Sigma-Aldrich, USA), placing the tissue $[\text{NiCl}_2]$ solutions at the outer radius with R_1 values increasing in the clockwise direction and the representative vascular vials at the inner radius with R_1 values decreasing in the clockwise direction. The remaining volume of the phantom was filled with a 30 mM $[\text{NaCl}]$ solution to simulate coil loading by a patient and to minimize static, magnetic field inhomogeneities [158].

3.2.2 ADC Phantom

The ADC phantom is represented schematically in **Figure 3-1**. Similar to the DCE phantom, this phantom was constructed from polycarbonate with an exterior diameter of 20 cm and depth of 12 cm, featuring the same removable lid and threaded plug design. Seven sucrose solutions, 10, 20, 25, 30, 35, 40 and 45 %(w/w), were used to achieve ADC values ranging from approximately 0.5 to $1.7 \times 10^{-3} \text{ mm}^2/\text{s}$, which encompass reported ADCs in normal and tumor tissues [159]–[161]. These values were determined by estimating a linear fit to data reported by Laubach *et al.* [153]. Solutions were prepared by stirring the appropriate weight of sucrose ($\geq 99.5\%$, Sigma-Aldrich, USA) with degased, deionized water, adding sodium azide (3 mg/mL of H_2O) to all vials to prevent alteration of the desired ADC values by microbial growth. These solutions were then poured into 50 mL polypropylene centrifuge tubes (Corning®, USA), and the caps were glued shut to minimize exchange of sucrose between the fill solution and centrifuge tubes. The 10 %(w/w) vial was placed in the center with all other vials distributed in increasing concentration in a counter-clockwise arrangement. In addition to the 30 mM NaCl fill solution, a low concentration (165 mg/L) of MnCl_2 was added to reduce the T_2 of the fill solution to approximately 30 ms, bringing the measured signal intensity close to that of background and thus minimizing image artifacts resulting from the fill solution signal.

Vial Number	Tissue Compartment		Vascular Compartment	
	$R_1 \text{ (s}^{-1}\text{)}$	$[\text{NiCl}_2] \text{ (mg/L)}$	$R_1 \text{ (s}^{-1}\text{)}$	$[\text{NiCl}_2] \text{ (mg/L)}$
1	0.97	139.4	1.05	156.8
2	1.24	197.1	2.93	548.7
3	1.63	278.7	6.86	1371.8
4	2.19	394.2	11.86	2416.9
5	2.97	557.4	17.86	3671.1
6	4.07	788.3	24.86	5134.3
7	5.63	1114.8	32.86	6806.6
8	7.84	1576.6	41.86	8687.8

Table 3-1. DCE phantom $[\text{NiCl}_2]$ and theoretical R_1 values. R_1 ranges and corresponding $[\text{NiCl}_2]$

theoretical calculations assuming $r_1 = 0.62 \text{ mM}^{-1}\text{s}^{-1}$ and $R_{10} = 0.3\text{s}^{-1}$. Note that neither r_1 nor R_{10} were quantified prior to phantom manufacture.

3.3 Methods

3.3.1 MR Acquisition Protocols

Assessing the equipment longitudinal variance was accomplished by performing a single set of same day repeatability measurements in combination with ongoing weekly quantitative acquisitions using the ADC and DCE phantoms on a 1.5T GE HDxt scanner and 3.0T GE MR750 scanner. Where possible, acquisition parameters for these phantom studies were matched to corresponding clinical protocols. To minimize variances due to variations in temperature, the phantoms were stored in the scanner room overnight, allowing equilibration between the fill solution and scanner room temperature. Additionally, the temperature of the phantom fill solution was measured prior to each imaging session. Following the temperature measurements, the DCE phantom was scanned first, followed by the ADC phantom.

The imaging protocol for the DCE phantom included, on a weekly basis, VFA R_1 measurements and a full DCE acquisition. The VTI R_1 measurements were performed once a month. Each phantom was set up on edge in the RF coil, aligning the vertical alignment laser to the center of the set screw between the tissue vials with R_1 values of 0.67 and 0.94s^{-1} and ensuring that the phantom was level. At each scanner, one set of three short-term measurements was acquired to establish same day variance, which included an initial acquisition per protocol with VTI R_1 measurements followed by two more full acquisitions without the VTI acquisition. Performing these measurements was accomplished by setting the scanner to research mode, which allowed varying the user control variables *opflip* and *opti* for VFA and VTI acquisitions, respectively. To ensure machine settings, such as center frequency, RF gains, and gradient shims remained constant for serial relaxometry measurements, a manual pre-scan was

performed after downloading each new scan flip angle or *TI* value and before acquiring any images. Scan parameters are summarized in **Table 3-2**.

Multi-slice diffusion weighted imaging and diffusion tensor imaging data were acquired weekly from the ADC phantom. Phantom set up consisted of situating the phantom in the RF coil such that the cylinder's edge rested on the support pad of the coil, rotating the 20 %(w/w) sucrose vial such that the compartment was roughly aligned midway between 12 and 3 o'clock, and ensuring that the phantom was level. Diffusion-weighted images were then acquired in three orthogonal planes aligned with the

	VFA	DCE	VTI	DWI	DTI
Coil:	8CH Brain (32 CH Brain)	8CH Brain (32 CH Brain)	8CH Brain (32 CH Brain)	8CH Brain (32 CH Brain)	8CH Brain (32 CH Brain)
Sequence:	3D FSPGR	3D FSPGR	FSE-XL	EPI	EPI
User CVs:	Turbo=0	Turbo=2	---	ASSET=2, DSE	ASSET=2, DSE
TE (ms):	1.4 (1.5)	1.2 (1.3)	8.8	82.7 (81.7)	82.7 (81.7)
TI (ms):	---	---	50, 75, 100, 125, 150, 250, 500, 1000, 1500, 2000, 3000	---	---
TR (ms):	5.8 (5.4)	5.7 (4.2)	6000	10000	10000
ETL:	N/A	N/A	4	N/A	N/A
BW (kHz):	+/-62.5	+/-62.5	+/-31.25	+/-250	+/-250
NEX:	4	1	1	4	1
Flip Angle (deg.):	2, 5, 10, 15, 20, 25, 30	30	90	90	90
<i>b</i>-value (s/mm²):	---	---	---	0, 250, 500, 750, 1000, 1250	0, 1200
FOV (cm):	24	24	24	24	24
Phase FOV (%):	1	0.8	1	1	1
Sl. Th. (mm):	5	3	8 (7)	5 (3.5)	5 (3.0)
Gap Th. (mm):	---	---	---	1.5 (0.0)	1.5 (0.0)
Num. of Slices:	16	16	1	16	16
Matrix:	256x160	256x160	256x160	96x128	96x128

Table 3-2. DCE and ADC phantom protocol acquisition parameters. Parenthetical values represent the acquisition parameters at 3.0T. For the DWI and DTI acquisitions, the number of diffusion encoding directions was 3 and 27, respectively. Abbreviations: Sl. = Slice, Th. = Thickness, DSE = dual spin echo.

major axes of the encoding gradients (*i.e.*, R-L, A-P, and S-I). The diffusion tensor acquisitions used a 27 direction-encoding schema. Neither pulse sequence required user input to acquire multiple b -values as was the case with R_1 measurements. Protocol parameters for the acquisition of ADC phantom diffusion data are summarized in **Table 3-2**.

Laubach *et al.* [153] noted a rapid decrease in T_2 values with increasing concentration of sucrose, making T_2 characterization of the ADC phantom desirable. In an effort to confirm this result, T_2 measurements were performed on the 3.0T system described previously using a multiple echo time ($n=4$) acquisition. In addition to these measurements, T_2 was also quantified using a 4.7T Bruker BioSpec® scanner with S057 gradients and a 35 mm inner diameter birdcage coil with a 32 echo Carr-Purcell Meiboom-Gill (CPMG) SE acquisition.

3.3.2 Equipment Variance Quantitation

Following image acquisition as described in section 3.3.1, all images were transferred via file transfer protocol, in DICOM format, to a Linux workstation with an 8 core Intel® Xeon® E5640 CPU operating at 2.67GHz. Images from the individual series (*e.g.*, VFA or DWI) were then imported into the QUATTRO environment, which automatically detected the series type (*e.g.*, VFA or DWI), sorted the images by slice location into stacks containing each of the varied imaging parameters (*e.g.*, flip angle or b -value), and extracted, from the DICOM headers, pertinent acquisition parameters necessary for modeling the data and computing the image biomarker results. Uniquely named, elliptical ROIs were then manually defined within the individual phantom compartments and voxel signal intensities were averaged for each ROI at each series point (*e.g.*, at each flip angle). For 3D or multi-slice acquisitions, these ROIs were propagated to all slices, ensuring that the ROIs remained centered within the compartment of interest by manually adjusting the ROI placement as needed to account for minor rotations that inevitably occurred during setup. Care was also taken to ensure image artifacts, such as partial volume averaging, Gibb's ringing, and N/2 ghosting, were avoided. The resulting mean signal

intensities for each acquisition in the series were fitted to an appropriate model using a QUATTRO script to estimate and track model parameters and a goodness of fit parameter, the coefficient of determination. For the models discussed in the following subsections, the LOA and CCC values, as discussed in section 2.2.1, were computed to assess agreement between estimated and theoretical values where possible.

3.3.2.1 R_1 Variance

VFA estimation of R_1 was performed by fitting the averaged signal intensities from each ROI as a function of flip angle to the well-known analytical Bloch equation solution, as discussed in section 2.1.1, for a spoiled gradient echo sequence given by equation 2.19. Bland-Altman [144] and correlation plots were produced to compare R_1 estimates with the theoretical calculations. Since these measurements were acquired with a volume acquisition, variation of R_1 was also characterized as a function of distance from the magnet isocenter in the slice direction by averaging longitudinal measurements and normalizing those measurements to the corresponding R_1 estimate at isocenter.

VTI estimation of R_1 was accomplished by a two-step procedure. First, to avoid introducing bias in the modeling caused by artificially increased signal intensities near the null point [12], polarity restoration was performed on the magnitude image signal intensities averaged from each ROI using a simple algorithm described in section 4.2.1. The resulting polarity corrected signal intensities were then fitted to the three-parameter signal intensity relationship given by equation 2.22, and Bland-Altman and correlation plots were produced for comparison with theory.

3.3.2.2 DCE Variance

Estimating PK parameter variance from the static DCE phantom data was accomplished by fitting simulated tissue uptake curves using a VIF derived from patient brain data [162]. One-hundred fifty unique combinations of K^{trans} , v_e , and v_p were chosen to generate the tissue uptake curves from the

three parameter general kinetic model [110]. The values for those three parameters were selected from the sets $\{0.02, 0.05, 0.1, 0.2, 0.35\} \text{ min}^{-1}$, $\{0.05, 0.1, 0.2, 0.5\}$, and $\{0.05, 0.1, 0.2, 0.5\}$, respectively. To map meaningful estimates of variance in T_1 measurements to the respective PK parameter, the means and standard deviations of T_1 measurements for each compartment were provided to a random number generator to produce random fluctuations in the synthetic VIF and tissue uptake curve. Because the simulated curves represent discrete samples from a continuous distribution of ΔR_1 values, and because the number of ΔR_1 samples from the phantom is limited, a binning of ΔR_1 was performed to provide a distribution for all values of ΔR_1 . In other words, for synthesized [Gd] values, x , on a given tissue uptake curve, the values falling between the i^{th} and $i^{\text{th}}+1$ tissue vials' ΔR_1 values were randomly sampled from a normal distribution of mean x and standard deviation σ_i , where σ_i is the estimated standard deviation from the T_1 measurement distribution of the i^{th} vial and $i = \{1, 2, 3, \dots, 7\}$. This process was repeated for the VIF to produce noisy VIF values for simulations. To ensure that the random number generator was appropriately configured, longitudinal T_1 measurement distributions for every compartment were tested for normality using a single sample Kolmogorov-Smirnov goodness-of-fit hypothesis test [163] with significance level of 5%. All tests failed to reject the null hypothesis, suggesting that the data were, to a good approximation, normally distributed.

For each combination of PK parameters, 10,000 unique random tissue uptake curves and VIFs were simulated. Finally, these noisy synthetic tissue and VIF [Gd] concentration measurements were fitted using a three-parameter generalized kinetic model to estimate the PK parameters. To avoid introducing known effects, namely additional bias and variance of temporally undersampled dynamic curves [164], [165], allowing isolation of equipment related variance, high temporal resolution (50 ms sampling) versions of the tissue uptake and VIF concentration curves were used. LOA and CCC values were computed for the estimated pharmacokinetic parameters to assess the agreement between estimated and actual values.

3.3.2.3 ADC Variance

Estimation of the ADC values was performed by fitting the averaged signal intensities from each ROI to the intravoxel incoherent motion model equation 2.16 as described in section 2.1.1. This simplified version of the IVIM model ignores the pseudo-diffusion coefficient, perfusion fraction, and diffusion kurtosis terms of the full model since the physical processes represented by these terms are not captured in the simple sucrose solution ADC phantom. The exponential term containing the transverse relaxation constant was absorbed into the S_0 parameter since TE was held constant for all b -value acquisitions. This model was used to perform parameter estimation from the DWI images, where the images were first averaged over all three orthogonal acquisitions before modeling. DTI images transferred to a GE Advantage Windows Workstation running FuncTool (GE Healthcare, Waukesha, WI, version 4.5), which was used to compute ADC and FA maps. Both acquisition strategies were used to analyze the effect of off-isocenter parameter estimates as described previously for the VFA acquisitions.

3.4 Results

3.4.1 Equipment Variance

The ADC and DCE phantoms were manufactured as described in Section 3.2 and the acquisition protocols were created and stored on the respective 1.5T or 3.0T scanner as described in Section 3.3.1. MATLAB scripts were created to sort and assemble images into directories by date, phantom name, and series number, which allowed the import utility of QUATTRO to easily assemble exams for quantitation.

3.4.1.1 R_1 Variance

Longitudinal R_1 results estimated from a single central slice are summarized in **Table 3-3**, showing substantially better repeatability for R_1 measurements at 1.5T compared with the corresponding measurements at 3.0T, with coefficients of variation ranging from 2 to 3.2% and from 6.4

to 10.2% at 1.5 and 3.0T, respectively. VFA-estimated R_1 values measured from a single slice in the center of the phantom exhibited minimal bias compared with the theoretical values for measurements made at 1.5T as shown in **Figure 3-2**, while those same measurements at 3.0T were consistently over-estimated, particularly in the VIF compartments.

	1.5T VFA		3.0T VFA		1.5T VTI		3.0T VTI	
	R_1 (s ⁻¹)	COV (%)	R_1 (s ⁻¹)	COV (%)	R_1 (s ⁻¹)	COV (%)	R_1 (s ⁻¹)	COV (%)
VIF-1	1.0	2.4	0.7	7.2	1.0	0.6	1.1	1.6
VIF-2	2.7	2.4	2.1	8.3	2.9	0.6	3.0	2.3
VIF-3	6.4	2.3	5.3	7.2	6.8	0.8	7.2	1.5
VIF-4	11.5	2.2	9.4	6.4	10.5	2.4	11.8	3.4
VIF-5	17.4	2.0	14.0	6.7	17.5	0.8	18.8	1.0
VIF-6	24.7	2.0	18.9	7.4	24.6	1.0	27.1	2.7
VIF-7	32.1	2.5	24.4	7.6	32.8	1.3	36.5	4.0
VIF-8	40.3	3.2	30.8	7.6	41.7	1.8	48.5	8.2
Tissue-1	0.9	2.7	0.8	10.2	1.0	1.0	1.0	1.4
Tissue-2	1.2	2.6	1.0	9.4	1.2	0.6	1.3	3.0
Tissue-3	1.6	2.8	1.3	8.1	1.6	0.5	1.7	1.2
Tissue-4	2.0	2.6	1.7	8.0	2.1	0.6	2.3	0.5
Tissue-5	2.8	2.3	2.3	8.2	2.9	0.8	3.0	0.7
Tissue-6	3.8	2.5	3.4	7.8	3.9	0.9	4.2	1.1
Tissue-7	5.2	2.7	4.9	7.0	5.5	1.1	5.8	0.9
Tissue-8	7.6	2.3	7.5	7.5	7.6	0.7	8.2	1.3

Table 3-3. Summary of the DCE phantom longitudinal R_1 measurements. Measured R_1 values in s⁻¹ averaged over all repeatability measurements using a VFA technique (*left*) and VTI technique (*right*). Values for the centrifuge tubes containing the VIF sequence are in the top half of the table, while the representative tissue values are in the bottom half of the table.

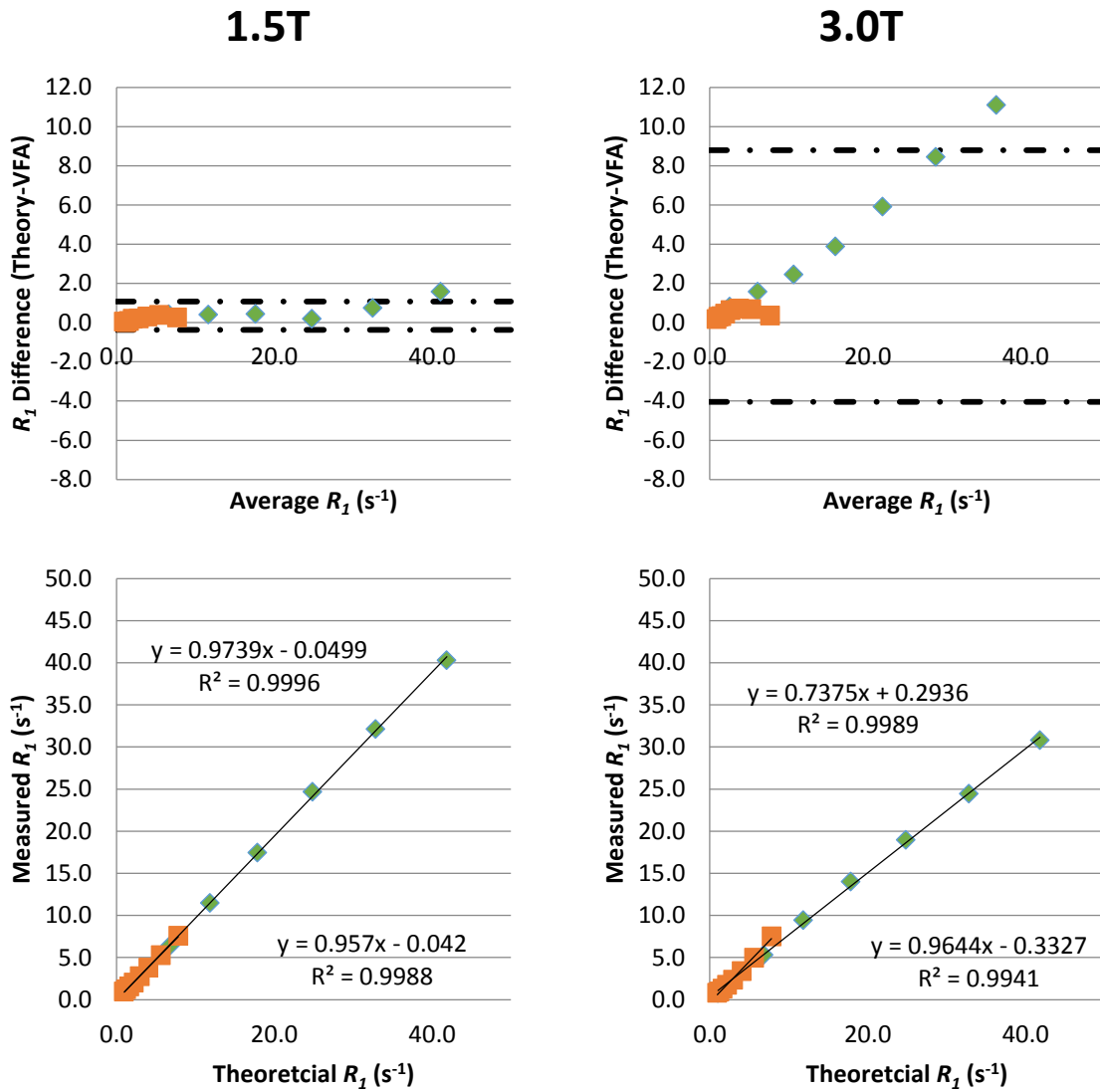


Figure 3-2. Comparison of VFA R_1 estimation with theory. Bland-Altman plots (*top*) of the longitudinally averaged VFA R_1 measurements vs. theoretical R_1 values at 1.5T (*left*) and 3.0T (*right*). The 8 VIF (*squares*) and tissue (*diamonds*) compartments were given different markers for clarity and the dashed lines are the limits of agreement. Correlation plots (*bottom*) are shown for comparison with the linear regression and corresponding coefficient of determination for 1.5T (*left*) displayed above the data and that for 3.0T (*right*) displayed below the data.

VFA R_1 relaxometry techniques are known to suffer inaccuracies in parameter estimation as a result of flip angle calibration errors and/or spatial heterogeneity in the transmitted B_1 field [104], [166], [167]. At higher field strengths, spatial heterogeneity of the B_1 fields, in particular, increasingly biases R_1 measurements as a function of position due to reduced RF penetrability and dielectric resonance effects. In brain applications, dielectric resonance is especially pronounced in objects roughly the size of a human head (*i.e.*, the DCE phantom) [12]. This would imply larger biases in R_1 estimates at the center of the phantom, where the VIF centrifuge vials are positioned, as is clearly demonstrated in **Figure 3-2**. In combination, these effects explain the larger observed bias and variance in longitudinal measurements at 3.0T when considering uncertainties such as phantom setup and flip angle calibrations.

Single slice, VTI-estimated R_1 values exhibited excellent longitudinal repeatability, with coefficients of variation ranging from 0.5 to 2.4% at 1.5T and from 0.5 to 6.9% at 3.0T as summarized in **Table 3-3**. Note that by neglecting a single set of outlier data acquired at 3.0T on February 24, 2013, COV ranges between different field strengths become equivalent. The cause of this deviation was most likely an incorrect acquisition setting; a BW of +/-62.5 kHz was used as opposed to +/-31.25 kHz as prescribed in the protocol. In addition to improved repeatability over VFA parameter estimation, the VTI R_1 estimates agreed with theory for the VIF $[\text{NiCl}_2]$ sequence at both field strengths (see **Figure 3-3**), although a trend is still discernible in the VIF $[\text{NiCl}_2]$ sequence at 3.0T. This trend, which is consistent in the VIF and tissue vials, might suggest a slight field strength dependence of the $[\text{NiCl}_2]$ relaxivity, r_1 . Additionally, these results further corroborate the notion that the less repeatable and more biased R_1 estimates using a VFA acquisition at 3.0T were the result of transmit RF inhomogeneities. While these issues would still be present in VTI acquisitions, the resulting bias and variance in R_1 measurements are effectively decoupled from the flip angle because long TR values were used, which allows near full recovery of the longitudinal magnetization.

In addition to single slice measurements, estimates of R_1 were tracked on each slice of the volume acquisition. **Figure 3-4** through **Figure 3-7** displays the relationship between R_1 and the slice distance from isocenter in the superior-inferior direction, omitting the first and last slices of the volume. In this range of slices, R_1 values deviated from the corresponding measurements on the slice at isocenter from -11.5 to 7.2% at 1.5T and -15.4 to 26.6% at 3.0T for all vials. Including all slices, however, creates substantially larger deviations: -76.8% at 1.5T and -74.6% at 3.0T. The additional deviation introduced by including the end slices is the result of non-ideal RF excitation of the imaging volume in the slice encoding direction and off-resonance effects.

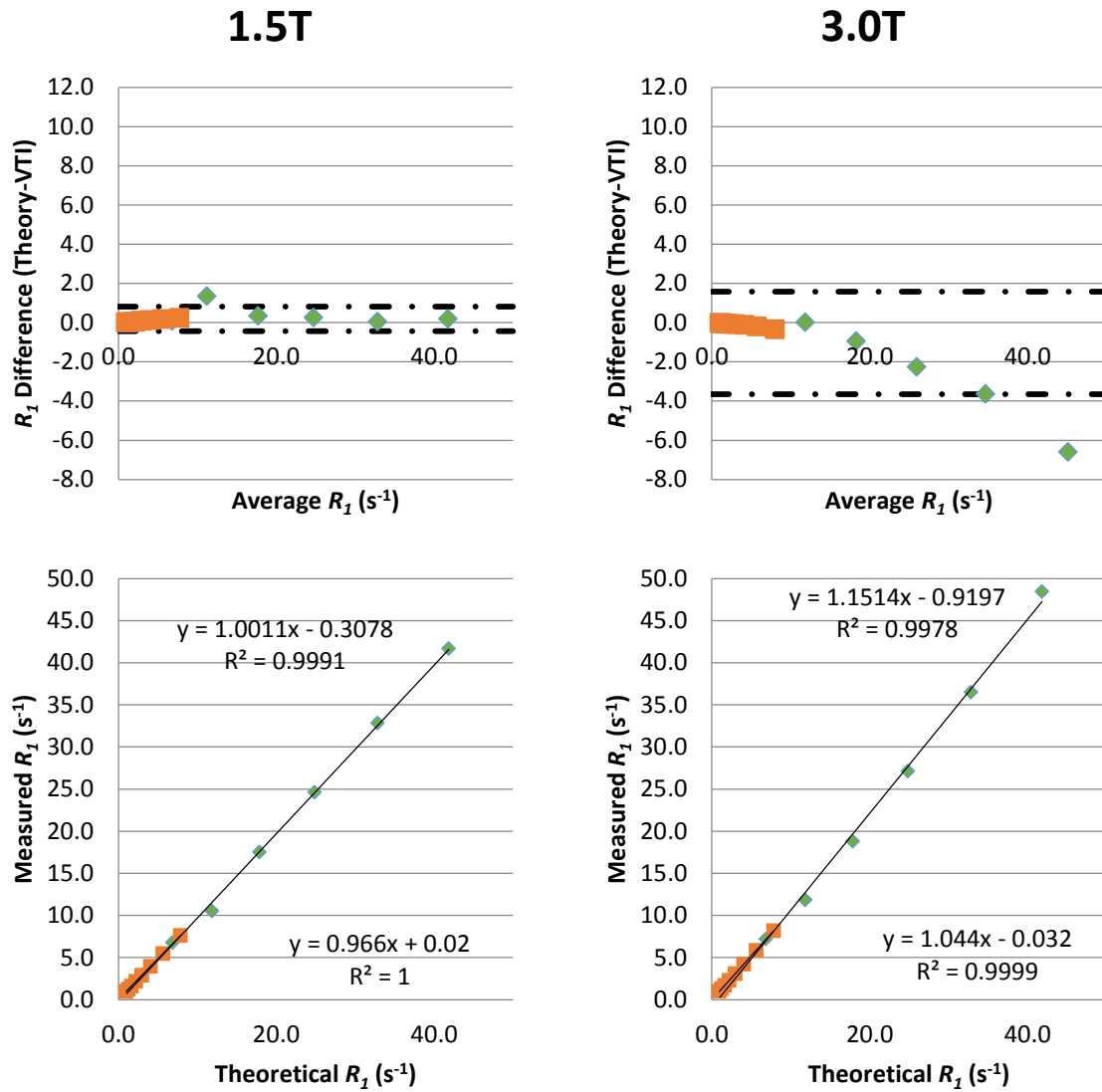


Figure 3-3. Comparison of VTI R_1 estimation with theory. Bland-Altman plots (*top*) of the longitudinally averaged VTI R_1 measurements vs. theoretical R_1 values at 1.5T (*left*) and 3.0T (*right*). The 8 VIF (*squares*) and tissue (*diamonds*) compartments were given different markers for clarity and the dashed lines are the limits of agreement. Correlation plots (*bottom*) are shown for comparison with the linear regression and corresponding coefficient of determination for 1.5T (*left*) displayed above the data and that for 3.0T (*right*) displayed below the data.

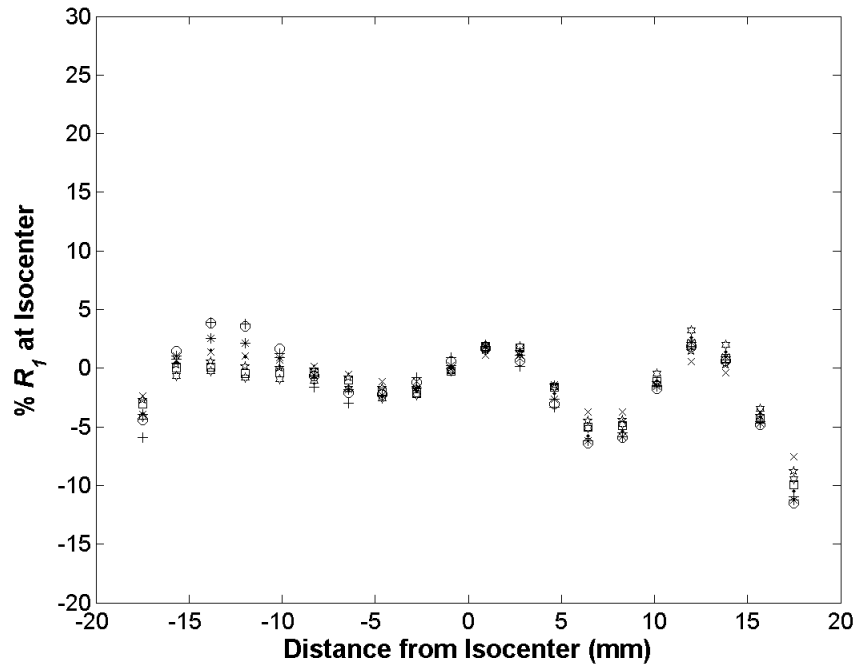


Figure 3-4. 1.5T VIF R_1 dependence as a function of distance from isocenter in the I-S direction.

Plots of the VIF [NiCl₂] sequence R_1 estimates normalized to the corresponding measurements on the slice at isocenter vs. distance from isocenter in the I-S (negative-positive) direction averaged over all longitudinal acquisitions at 1.5T. Data from the two most superior and two most inferior slices were omitted to ensure the imaging volume was matched to the corresponding 3T data. For visualization purposes, a shape-preserving smooth interpolant was used. Error bars were omitted because limits were so small that the error markers were indistinguishable from the data markers.

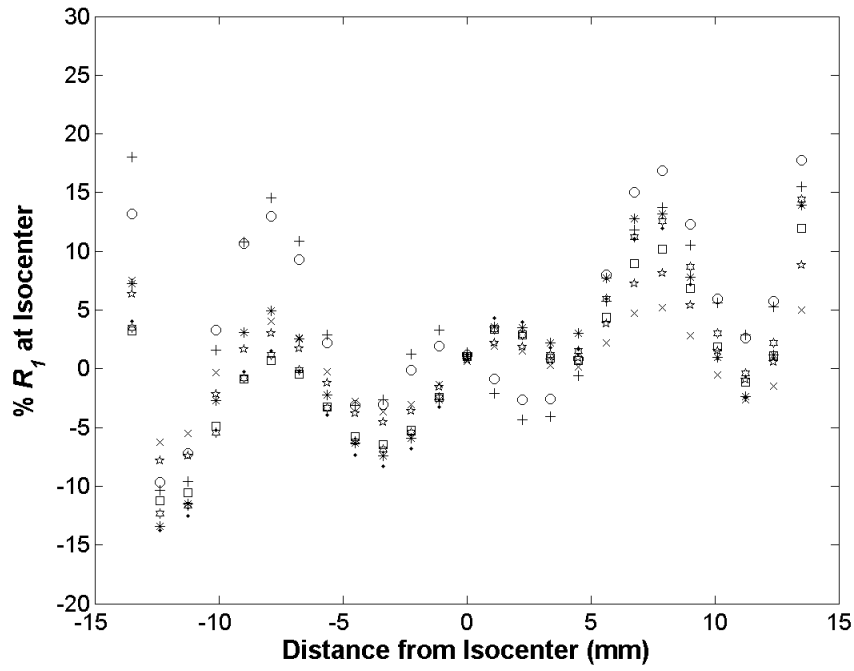


Figure 3-5. 3.0T VIF R_1 dependence as a function of distance from isocenter in the I-S direction.

Plots of the VIF [NiCl₂] sequence R_1 estimates normalized to the corresponding measurements on the slice at isocenter vs. distance from isocenter in the I-S (negative-positive) direction averaged over all longitudinal acquisitions at 3.0T. For visualization purposes, a shape-preserving smooth interpolant was used. Error bars were omitted because limits were so small that the error markers were indistinguishable from the data markers.

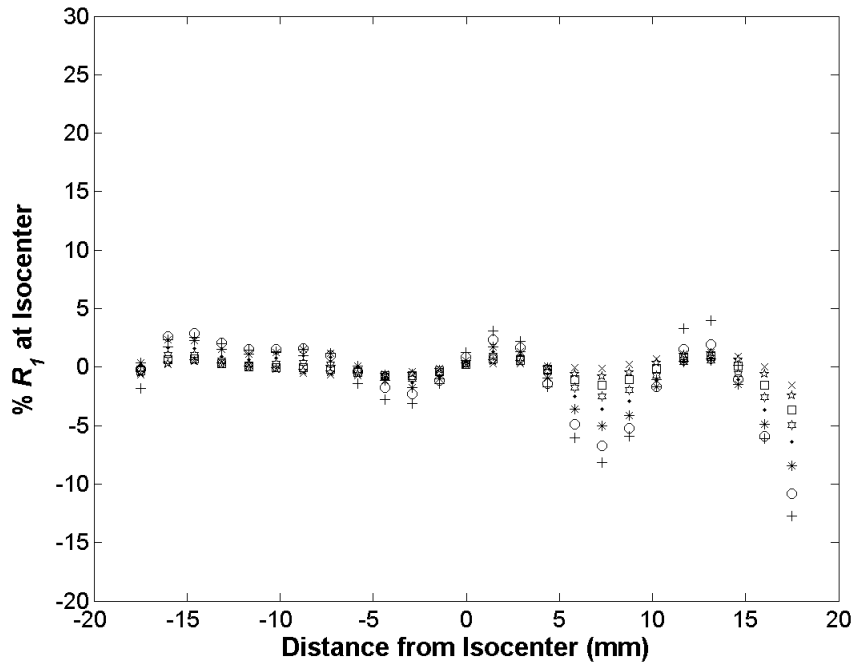


Figure 3-6. 1.5T tissue R_1 dependence as a function of distance from isocenter in the I-S direction.

Plots of the tissue [NiCl₂] sequence R_1 estimates normalized to the corresponding measurements on the slice at isocenter vs. distance from isocenter in the I-S (negative-positive) direction averaged over all longitudinal acquisitions at 1.5T. Data from the two most superior and two most inferior slices were omitted to ensure the imaging volume was matched to the corresponding 3T data. For visualization purposes, a shape-preserving smooth interpolant was used. Error bars were omitted because limits were so small that the error markers were indistinguishable from the data markers.

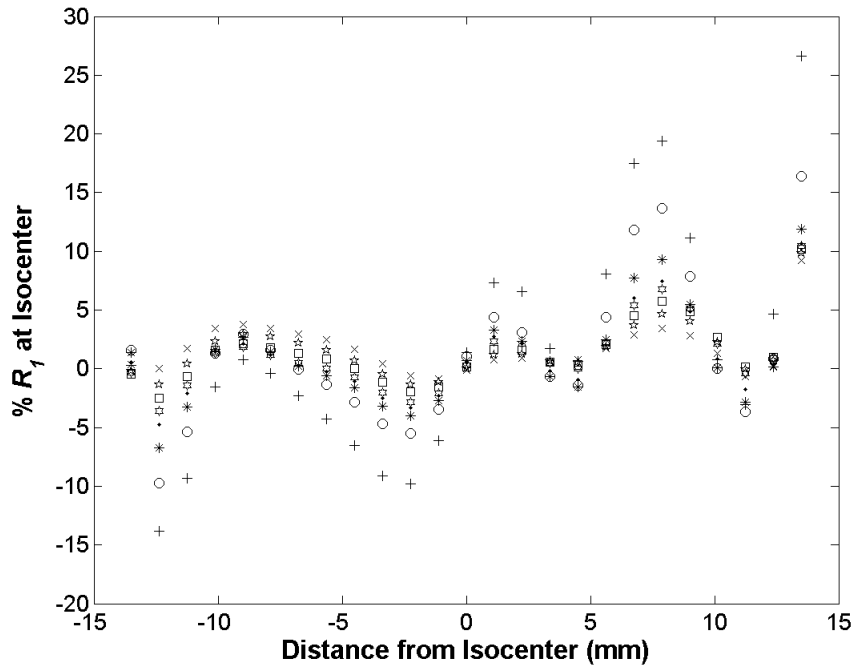


Figure 3-7. 3.0T tissue R_1 dependence as a function of distance from isocenter in the I-S direction.

Plots of the tissue [NiCl₂] sequence R_1 estimates normalized to the corresponding measurements on the slice at isocenter vs. distance from isocenter in the I-S (negative-positive) direction averaged over all longitudinal acquisitions at 3.0T. For visualization purposes, a shape-preserving smooth interpolant was used. Error bars were omitted because limits were so small that the error markers were indistinguishable from the data markers.

3.4.1.2 DCE Variance

LOA and CCC values calculated from the simulations performed using the R_1 variance estimated from the VFA and VTI data are summarized in **Table 3-4**. In all cases, the CCC was, for all intents and purposes, unity, suggesting excellent agreement between the estimated parameters and the actual values. Moreover, the limits of agreement, essentially two standard deviations, for the estimated

values of K^{trans} , v_e , and v_p , were essentially zero. With respect to expected changes in tissue resulting from treatment, the parameter noise introduced by the reproducibility of relaxometry measurements represents insignificant variation in T_1 -weighted DCE biomarkers. For example, Hawighorst *et al* [70], in a study of intracranial meningiomas treated with conformal radiation therapy, found a significant decrease in k_{ep} (K^{trans}/v_e) on the order of 50% in patients responding to the treatment (n=13) and a median increase greater than 500% in non-responders (n=4). Gutin *et al.* [168] found a significant decrease in K^{trans} (approximately 31%) following the administration of a single cycle of bevacizumab in patients exhibiting some radiographic response. The symmetry observed in all LOA values suggests negligible bias introduced by noise, which is confirmed by mean percentage error of approximately $7.6 \times 10^{-3} \%$ for all pharmacokinetic parameters using the variance estimated from VFA repeatability and $6.06 \times 10^{-3} \%$ for the VTI simulations. Overall, the results of these simulations suggest that the level of error introduced by variation in R_1 measurements is insignificant with respect to expected changes resulting from treatment.

		VFA Simulations			VTI Simulations		
		K^{trans} (min^{-1})	v_e	v_p	K^{trans} (min^{-1})	v_e	v_p
1.5T	Lower LOA	-8.77×10^{-4}	-3.17×10^{-3}	-2.14×10^{-4}	-9.61×10^{-4}	-3.48×10^{-3}	-2.36×10^{-4}
	Upper LOA	8.06×10^{-4}	3.21×10^{-3}	2.31×10^{-4}	8.72×10^{-4}	3.53×10^{-3}	2.56×10^{-4}
	CCC	1.000	1.000	1.000	1.000	1.000	1.000
3.0T	Lower LOA	-8.65×10^{-4}	-3.20×10^{-3}	-2.14×10^{-4}	-9.66×10^{-4}	-3.47×10^{-3}	-2.37×10^{-4}
	Upper LOA	7.95×10^{-4}	3.24×10^{-3}	2.30×10^{-4}	8.78×10^{-4}	3.53×10^{-3}	2.58×10^{-4}
	CCC	1.000	1.000	1.000	1.000	1.000	1.000

Table 3-4. Analysis of PK parameters simulated from R_1 variance. Lower/upper limits of agreement

and concordance correlation coefficient as calculated from the simulated PK parameter estimates for levels of ΔR_1 variance at 1.5T (*top*) and 3.0T (*bottom*).

3.4.1.3 ADC Variance

When measured at a single slice at the center of the imaging volume, DWI-based ADC estimates exhibited good longitudinal reproducibility at both 1.5 and 3.0T, with coefficients of variation ranging from 3.4 to 6.8% and 1.5 to 19.4%, respectively (see Table 3-5). Again, neglecting the lone outlier datum of the longitudinal ADC estimates at 3.0T brings the maximum range of coefficients of variation to 4.8%, comparable with corresponding measurements at 1.5T. This discrepancy in reproducibility most likely stems from the differences in the signal-to-noise ratio of the two data sets. More specifically, the observed lower signal-to-noise ratio of data acquired at 3.0T. In addition to acquisition differences, sucrose concentration dependent T_2 values also created SNR dependencies of the repeatability as seen in the trend of increasing coefficients of variation with increasing sucrose concentration, although not as pronounced in the measurements made at 1.5T because of the higher overall SNR in those data. Laubach *et al.* [153] reported a rapid decrease in the solution T_2 value with increasing concentration of sucrose, with T_2 values as low as 50ms for a 30%(w/w) sucrose solution. T_2 values of the ADC phantom sucrose solutions were quantified with a multiple-echo spin echo acquisition at 3.0T and 4.7T to extend the results of that previous work, and to confirm the rapid signal loss as a function increasing sucrose concentration; results are summarized in **Figure 3-8**. With large effective TE values ($\sim 80ms$), the dramatic decrease in T_2 can clearly be considered a dominate factor in the decreased repeatability of higher concentration vials.

Sucrose (%w/w)	1.5T DWI		3.0T DWI		1.5T DTI		3.0T DTI	
	ADC ($\times 10^{-3}$ mm ² /s)	COV (%)	ADC ($\times 10^{-3}$ mm ² /s)	COV (%)	ADC ($\times 10^{-3}$ mm ² /s)	COV (%)	ADC ($\times 10^{-3}$ mm ² /s)	COV (%)
10	1.71	3.4	1.69	1.5	1.75	3.6	1.68	1.2
20	1.39	4.1	1.38	1.6	1.40	3.0	1.38	1.6
25	1.17	3.7	1.12	2.0	1.17	3.3	1.11	3.5
30	0.96	3.7	0.85	3.0	0.97	3.6	0.83	5.3
35	0.79	4.1	0.61	3.7	0.78	4.0	0.59	7.1
40	0.64	4.7	0.46	4.8	0.64	4.8	0.47	11.2
45	0.49	6.8	0.29	19.4	0.51	6.6	0.33	18.3

Table 3-5. Summary of the diffusion phantom longitudinal ADC measurements. Measured ADC

($\times 10^{-3}$ mm²/s) values averaged over all repeatability measurements using a DWI technique (*left*) and DTI technique (*right*) for sucrose concentrations ranging from 10% to 45% (top to bottom).

Phantom temperatures were measured prior to each set of image acquisitions. At 1.5T, the average temperature was 20.5°C, ranging from 16.0°C to 23.9°C with a COV of 7.5%, whereas the average temperature at 3.0T was 20.9°C with a minimum of 20.1°C and maximum of 22.6°C with a COV of 2.4%. This larger difference in the phantom temperature variability prior to measurements at 1.5T explains the less repeatable measurements, as clearly shown in the COVs, for vials containing low concentrations of sucrose ($\leq 35\%$) compared with the corresponding 3.0T measurements.

Reproducibility of ADC estimates acquired with DTI in a single central slice exhibited similar trends as those estimates made with a DWI technique and the results are summarized in the right half of **Table 3-5**. The range of coefficients of variation at 1.5T was 3.0 to 6.6% and at 3.0T was 1.2% to 18.3%, where the trend of increasing coefficients of variation with increasing sucrose concentration is again apparent. Fractional anisotropy estimates in a single central slice were less reproducible compared with ADC values, with coefficients of variation ranging from 6.4 to 25.5% at 1.5T and 18.2 to 26.3% at 3.0T. These results, while suggestive of poor fractional anisotropy reproducibility, more likely represent the fact that there were miniscule fluctuations in a material with isotropic diffusion (i.e., $FA=0$), which results in disproportionately large apparent variance. In addition to comparable reproducibility, the two

methods provided good agreement according to a Bland-Altman analysis of data measured at a single central slice where no regression coefficients of lines fitted to the difference between DWI and DTI measurements deviated significantly from 0 ($p > 0.20$ for coefficients at 1.5 and 3.0T).

Analysis of axial ADC measurements as a function of slice location relative to isocenter showed deviations in magnitude from 7.9% at 1.5T and 35.2% at 3.0T, as illustrated in **Figure 3-9** and **Figure 3-10**. For volumes acquired at 1.5T, only the portion that overlapped with the 3.0T acquisitions were considered in this analysis since, in a clinical environment, this region would contain the pertinent imaging biomarker information. DTI estimates of ADC, summarized in **Figure 3-11** and **Figure 3-12**, deviated from the isocenter slice normalized value by 4.5% at 1.5T and 26.8% at 3.0T, respectively. Fractional anisotropy deviated substantially more compared with ADC estimates (as easily seen in **Figure 3-13** and **Figure 3-14**) as the distance from isocenter increased, with deviations as large as 66.2% at 1.5T and 93.3% at 3.0T.

Bland-Altman analysis was also performed to compare the ADC and FA estimates between the two scanners, which yielded significant non-zero regression coefficients ($P < 0.01$) for ADC estimates, both DWI and DTI, and non-zero regression coefficients approaching significance ($P < 0.04$) for FA estimates. As stated previously, the differences in signal-to-noise ratios resulting from thinner slices at 3.0T and the already low signal, particularly in vials with high sucrose concentrations (*i.e.*, $> 25\%$), results in noise-dominated signal, which is known to underestimate ADC values [169]–[171]. When considering the lower concentration compartments (*i.e.*, $\leq 25\%$), the percent difference between 1.5T and 3.0T data was approximately 5% or less for all ADC estimates.

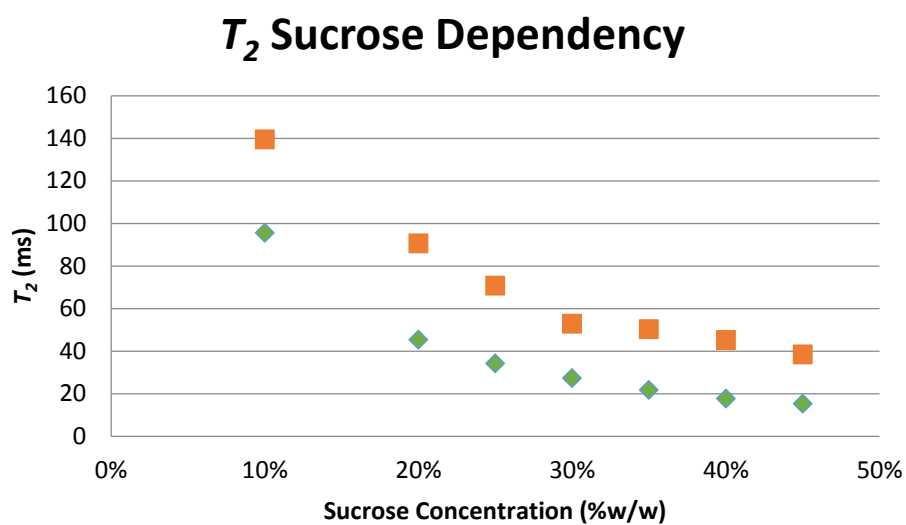


Figure 3-8. Sucrose solution T_2 values. Measured T_2 values at 3.0T (*diamonds*) and 4.7T (*squares*).

The large difference in values between the two field strengths stems for the acquisition details, namely, only 4 echoes were used to acquire measurements at 3.0T compared to the 24 echoes used at 4.7T. The number of echoes was limited by the acquisition pulse sequence at 3.0T. Also, the T_2 value at 4.7T for the 45% vial was extrapolated since no data were acquired for that vial.

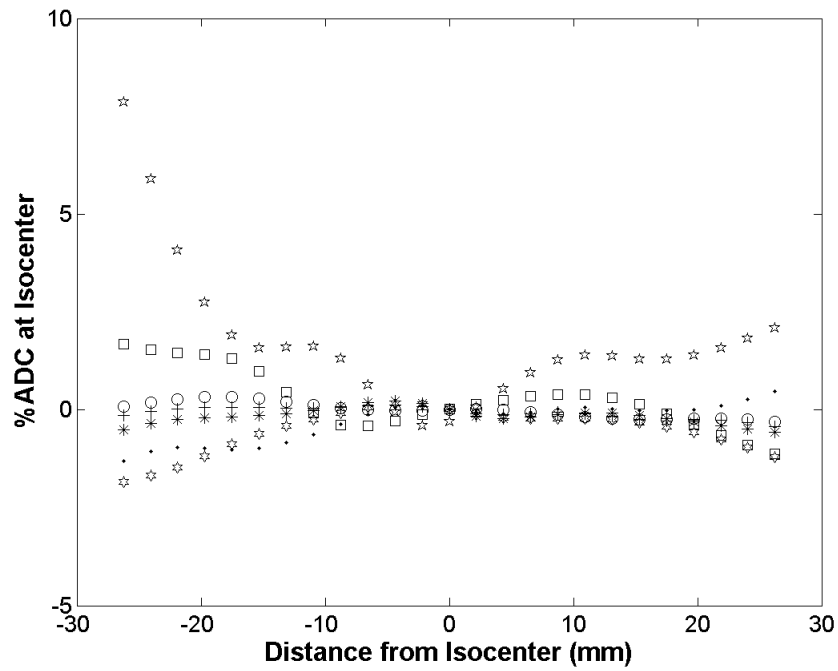


Figure 3-9. 1.5T DWI ADC dependence as a function of distance from isocenter in the I-S direction.

Plots of ADC measured at 1.5T using DWI normalized to the corresponding measurements on the slice at isocenter vs. distance from isocenter in the I-S (negative-positive) direction averaged over all longitudinal acquisitions. Data volumes were matched to 3.0T acquisition coverage and a shape-preserving smooth interpolant was used for visualization purposes. Error bars were omitted because the error markers were indistinguishable from the data markers.

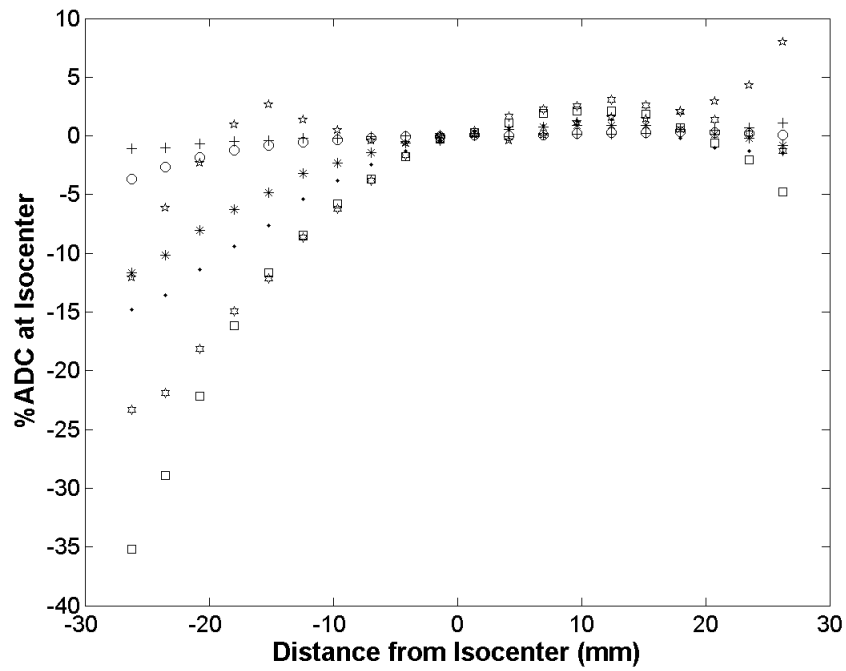


Figure 3-10. 3.0T DWI ADC dependence as a function of distance from isocenter in the I-S direction. Plots of ADC measured at 3.0T using DWI normalized to the corresponding measurements on the slice at isocenter vs. distance from isocenter in the I-S (negative-positive) direction averaged over all longitudinal acquisitions. Data volumes were matched to 3.0T acquisition coverage and a shape-preserving smooth interpolant was used for visualization purposes. Error bars were omitted because the error markers were indistinguishable from the data markers.

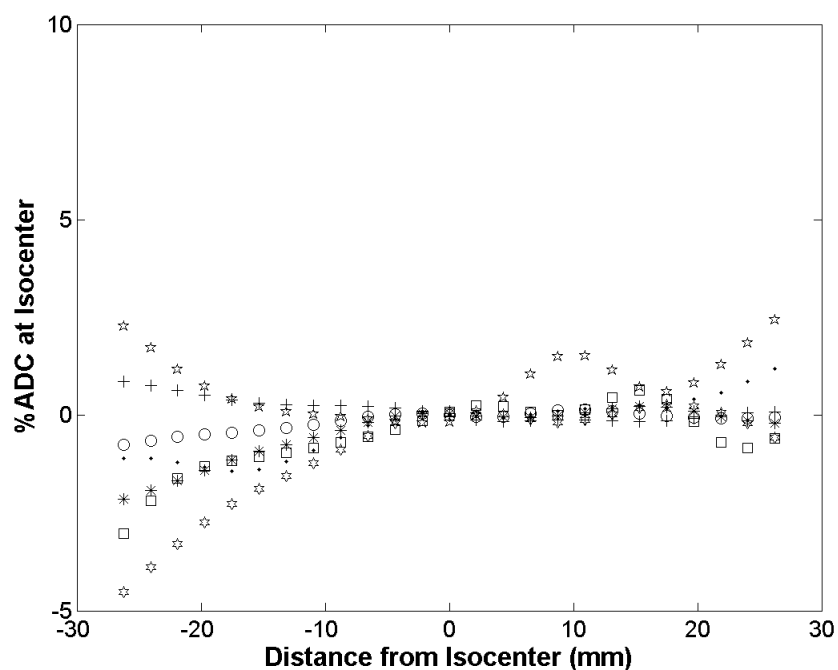


Figure 3-11. 1.5T DTI ADC dependence as a function of distance from isocenter in the I-S direction.

Plots of ADC measured at 1.5T using DTI normalized to the corresponding measurements on the slice at isocenter vs. distance from isocenter in the I-S (negative-positive) direction averaged over all longitudinal acquisitions. Data volumes were matched to 3.0T acquisition coverage and a shape-preserving smooth interpolant was used for visualization purposes. Error bars were omitted because the error markers were indistinguishable from the data markers

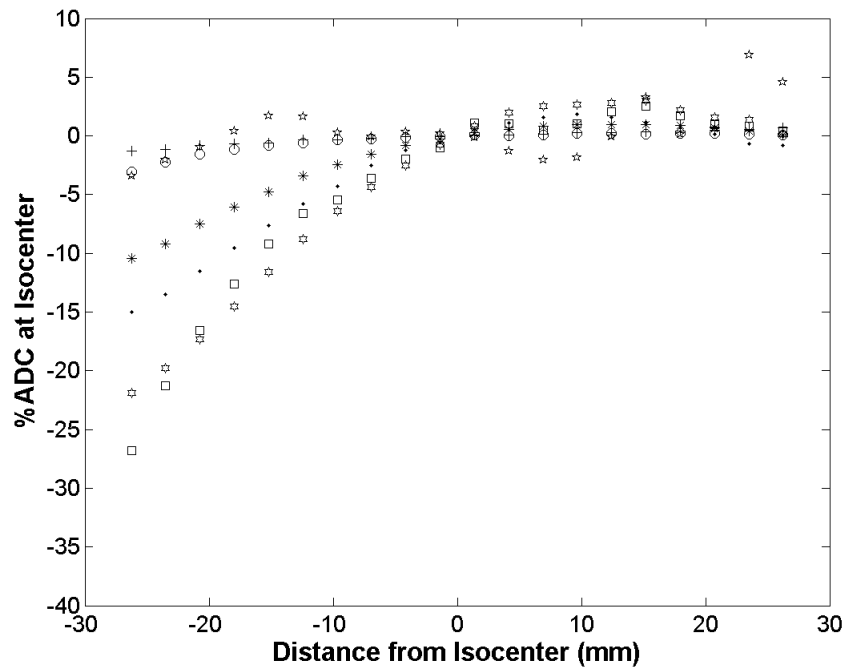


Figure 3-12. 3.0T DTI ADC dependence as a function of distance from isocenter in the I-S direction.

Plots of ADC measured at 3.0T using DTI normalized to the corresponding measurements on the slice at isocenter vs. distance from isocenter in the I-S (negative-positive) direction averaged over all longitudinal acquisitions. Data volumes were matched to 3.0T acquisition coverage and a shape-preserving smooth interpolant was used for visualization purposes. Error bars were omitted because the error markers were indistinguishable from the data markers

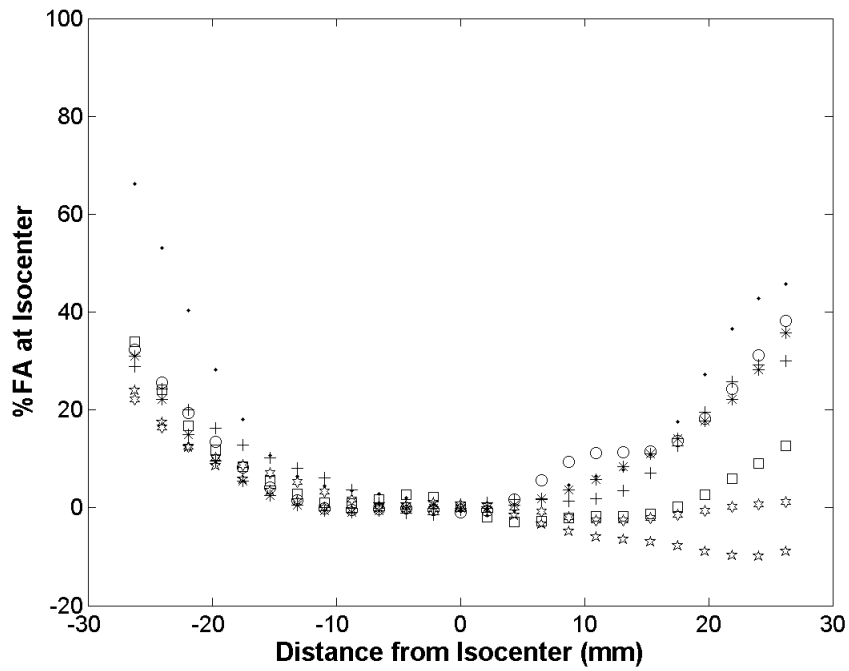


Figure 3-13. 1.5T FA dependence as a function of distance from at isocenter in the I-S direction.

Plots of FA measured at 1.5T normalized to the corresponding isocenter slice measurement as a function of distance from isocenter in the I-S (negative-positive) direction averaged over all longitudinal acquisitions. Data volumes were matched to 3.0T acquisition coverage and a shape-preserving smooth interpolant was used for visualization purposes. Error bars were omitted because limits were so small that the error markers were indistinguishable from the data markers.

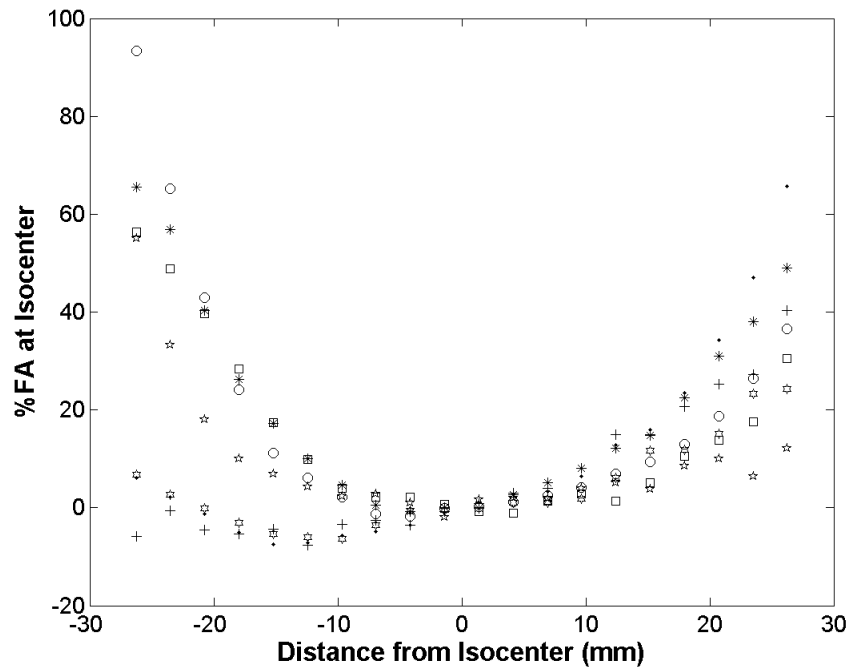


Figure 3-14. 3.0T FA dependence as a function of distance from at isocenter in the I-S direction.

Plots of FA measured at 3.0T normalized to the corresponding isocenter slice measurement as a function of distance from isocenter in the I-S (negative-positive) direction averaged over all longitudinal acquisitions. Data volumes were matched to 3.0T acquisition coverage and a shape-preserving smooth interpolant was used for visualization purposes. Error bars were omitted because limits were so small that the error markers were indistinguishable from the data markers.

3.5 Discussion

Quality assurance, particularly on machines used for the assessment of treatment response, remains a crucial component of any study employing quantitative imaging data. Documenting quantitative stability, especially following hardware/software upgrades, provides confirmation that longitudinal patient measurements are minimally impacted by scanner bias and variance. Additionally, performing such measurements at multiple sites provides the framework for designing and executing

data harmonization from multi-center trials. This study aimed to quantify the longitudinal reproducibility of quantities derived from diffusion and DCE imaging acquisitions at scanner field strengths commonly used, namely 1.5 and 3.0T, and under normal operating conditions.

Two scanners, one used for routine clinical imaging (1.5T) and one used for research purposes only (3.0T), were monitored over the course of 10 months, during which weekly acquisitions of relaxometry and diffusion data were performed and the resulting quantitative imaging biomarkers were tracked. Axial slice position dependences were also tracked for these quantities. In general, T_1 measurements agreed well with theoretical calculations based on published relaxivities [156], [157] and ADC estimates were consistent with previously published results [153]. In addition to providing longitudinal stability, these phantoms were easily manufactured requiring only a scale with milligram accuracy, a stirrer, and a source of degassed, deionized water, making these phantoms ideal candidates for assessing equipment variance for head applications.

DCE phantom longitudinal analysis resulted in reproducible R_1 estimates at 1.5T and 3.0T with maximum coefficients of variation 3.0% and 9.4%, respectively. Results of the pharmacokinetic simulations suggest that the equipment contributed bias and variance is negligible for these levels of R_1 . To this point, Hawighorst *et al.* [70] found significant changes in k_{ep} (k^{trans}/v_e) on the order of 50-142% compared with pre-treatment measurements, which is comparable with the results of Hayes *et al.* [84]. Reviews of work investigating DCE-MRI as a means of assessing antiangiogenic and vascular disruption treatments by Jackson *et al.* [71] and O'Connor *et al.* [72] suggest treatment effect changes of 40% or greater. These results, in combination with the work in this aim, suggest that treatment changes are easily distinguishable from equipment variance for the equipment monitored in this study. Even including additional error for setup, as the volume analyses might suggest a level of at most 15%, the expected treatment changes still exceed these deviations.

Longitudinal analysis of the diffusion data showed that equipment variances for the ADC estimates were well below thresholds of change that have been observed in previous work. For example the University of Michigan group [36] uses a threshold of approximately 30% change in the ADC while others, such as Mardor *et al.* [33], have noted changes on the order of 30-70% in responding tissues. A recent validation of the functional diffusion map [36] by Ellingson *et al.* [22] found minimum thresholds of change required, based on significant changes in cellularity, for the ADC values of normal appearing white matter ($0.25 \times 10^{-3} \text{ mm}^2/\text{s}$) and gray matter ($0.31 \times 10^{-3} \text{ mm}^2/\text{s}$). These thresholds are smaller by a factor of approximately two compared to those employed in the analysis by Hamstra *et al.* [36]. Even for changes of this magnitude ($\sim 15\%$), the equipment variance is still smaller by a factor of approximately three. The substantial deviations in ADC values as a function of position from isocenter, particularly at the edges of the volume, reinforce the necessity of reproducible patient setup. Moreover, the drastic differences observed between field strengths strongly encourage the use of a single field strength when acquiring longitudinal data. Recent work by Malyarenko *et al.* [172] suggested that evaluation and correction of the gradient non-linearities can reduce bias in ADC estimates. However, for the purposes of this work, not only were images acquired at 1.5T for which the bias in DTI ADC estimates was within 5% across the field of view, but patient setup consistently placed the tumor within approximately 10 mm of the slice at isocenter, suggesting that the treatment-induced changes will far exceed the deviations in ADC estimates.

This study has provided, over an extended period of time, characterization of bias and variance associated with quantitative imaging biomarkers for diffusion and DCE applications, elucidating several areas where bias can be and is introduced as a result of acquisition and/or analysis techniques. However, several limitations of this work exist, specifically assessment of bias and assessment of machine variance with regard to DSC exams. Future studies (addressed in detail in Chapter 6), should

involve careful characterization of temperature and chemical concentration dependencies, allowing a more thorough assessment of bias and acquisition.

Chapter 4 A Unified Framework for Assessing Treatment Response

4.1 Introduction

Quantitative imaging biomarkers offer significant promise as indicators for diagnosing, monitoring, and assessing tissue response for a plethora of treatment modalities, for which the latter appears to provide much needed improvements to current techniques such as RECIST [4], [7], [9], [14], [75], [76], [173]–[175]. One significant obstacle to utilizing quantitative imaging biomarkers is the labor and time intensive image analysis pipeline, requiring image registration between and within multiple data sets for each patient, identification and selection of image regions (including VIFs for DCE- and DSC-MRI studies), extraction of pixel or region information, and incorporation of this information into an appropriate statistical framework. These tasks are performed on multiple series within an imaging exam and across multiple imaging exams for a single patient, often including in excess of 10,000 images. While numerous large scale collaborative efforts have manifested in solutions to the various aforementioned components, such as the Insight Toolkit [176]–[178] and the Medical Image Processing, Analysis, and Visualization (MIPAV) [179] application, a unified analysis infrastructure for performing these computations remains unavailable. Other available software packages such as AFNI [180] or ImageJ [181], again, incorporate major components of treatment response analysis as plug-ins, including the DCE quantification utility TOPPCAT [182] or the AFNI image registration utility *3dvolreg*, but creating a platform from these software packages that links the necessary built-in tools into a broad multi-parametric analysis tool capable of performing longitudinal analyses is infeasible. Additionally, several groups have produced in-house software packages to accomplish similar analyses [183]–[185], but rarely are such efforts made available to researchers outside of those groups. An additional constraint imposed by available software is the lack of transparency in algorithms, which frequently necessitates a

plethora of validation steps and, in cases of failed validation, software updates by the developer are required (a potentially time consuming and sometimes altogether impossible task).

To address this need, the MATLAB® programming environment was used to develop open source code based on object oriented programming (OOP) principles, resulting in an advanced programming interface (API) and graphical user interface (GUI) – **Quantitative Utility for Assessing Treatment Response** or QUATTRO for short. This code implements a flexible modeling API (linked to the widely developed statistical libraries of R [186]) and links the highly developed image registration libraries of ITK [178] to a corresponding API for performing modeling and image registration, respectively, in a unified workspace.

4.2 Software Implementation, Validation, and Optimization

MATLAB was chosen as the development environment for the programming and graphical interfaces because of the need for quick development and linking of the ITK libraries. A special class within MATLAB, known as the handle class, was used as the basis for the API. Handle classes reference objects (similar to pointers in C/C++) and allow simple operations, such as setting a property, to be modified into complex computations, eradicating a substantial portion of the data bookkeeping required by the user. Properties and methods that control data operations are fully accessible to the user, but the debugged classes remove nearly all of the programming overhead when designing advanced features (e.g., image registration, computation of parameter maps, or visualization tools). Moreover, OOP principles, modeled after the ITK [176] class template design, were used to allow rapid prototyping of additional tools, such as the registration API, other visualizations, and development of new quantitative modeling framework. A full list of the necessary design components and corresponding descriptions of the respective functionality can be found in **Table 4-1**.

Classes were developed to represent common data types for quantitative imaging exams, which include an image representation, ROI (or point) representation, and a common workspace for interaction of those data. Using these classes a programmatic GUI was developed with tools for ROI- and voxel-based analysis, visualization of images, map overlays, and histograms, and standard input/output such as reporting analysis results or writing newly calculated maps as DICOM images. Additionally, development of an efficient scripting interface was added to assist in performing analysis of large quantities of data.

Essential to quantitative imaging, in particularly longitudinal quantitation, two other important analysis components required development: quantitative modeling and image registration. Additional details of these tools are addressed in detail in the following subsections. Briefly, a modeling class was needed to provide an engine for performing quantitative analysis with many of the standard models and additional flexibility for adding user-defined models. Image registration is a necessary component of

	Description
<i>qt_image</i>	Image storage class, providing functionality for reading/writing images and meta data, displaying images, and applying image transformations and filters
<i>qt_roi</i>	Region of interest storage class, providing functionality for reading/writing, creating and displaying, extracting voxels from, and transforming regions
<i>qt_exam</i>	Exams storage class, providing tools for creating interactions between <i>qt_image</i> , <i>qt_roi</i> , <i>qt_models</i> , <i>qt_options</i> , and <i>qt_reg</i> classes
<i>qt_options</i>	Options storage class (used only with graphical user interface)
<i>qt_models</i>	Modeling class that performs quantitative modeling with available model and fitting customization. Subclasses for specific model types (<i>e.g.</i> , DCE or VFA) were also developed
<i>qt_response</i>	Modeling class that performs statistical analysis of longitudinal quantitative imaging data
<i>qt_reg</i>	Image registration class that bridges an ITK-developed executable with MATLAB

Table 4-1. Description of QUATTRO classes. These classes, with the exception of *qt_options* were designed to perform as stand-alone command prompt software, meaning that users can utilize those classes even beyond the scope of the QUATTRO GUI (*e.g.*, in scripts or as an API).

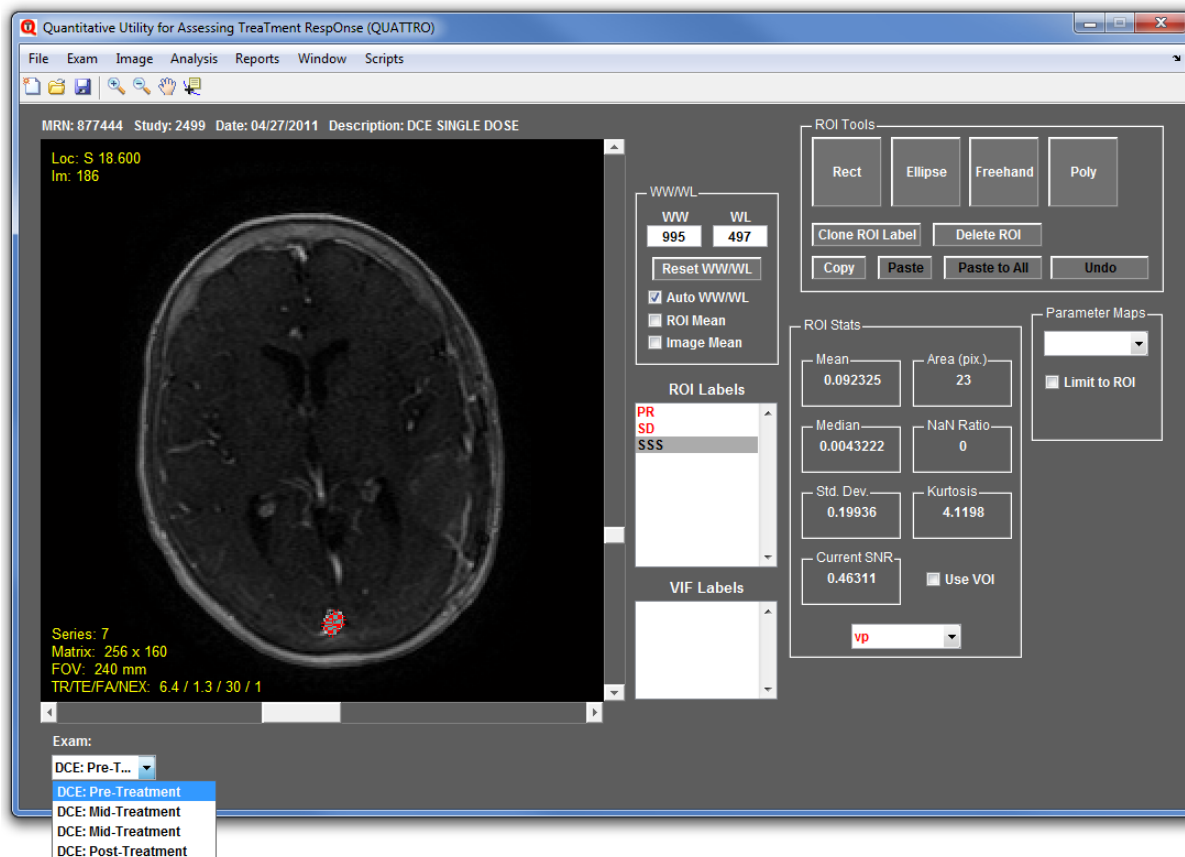


Figure 4-1. QUATTRO GUI screenshot. This is a view of the QUATTRO GUI, showing the main features and primary use as an image analysis tool. The common exam workspace can be seen in the expanded dropdown box at the lower left.

serial imaging, whether from the same series or different days. The image registration needed to link the highly developed, widely supported, open source ITK registration libraries [177] with MATLAB, providing substantial flexibility in image registration options, offering support for 3D rigid transformations.

4.2.1 Models and Model Validation

Standard models matched to common quantitative imaging strategies were incorporated into the QUATTRO modeling API (*qt_models*), including relaxometry, DCE, DSC, and diffusion models, manifested as individual *qt_models* sub-classes (*multiflip*, *multiti*, *dce*, *dsc*, and *dwi*, respectively). A

standard non-linear least squares fitting routine (*lsqcurvefit*) in MATLAB (MathWorks, Natick, MA) based on trust-region-reflective method [187], [188] or Levenberg-Marquardt [189], [190] numerical optimization algorithms was incorporated into the modeling API. Experience showed that both algorithms produced nearly identical results for relaxometry and diffusion data, but DCE and DSC applications often required the ability to bound specific parameters, limiting the search space to physically interpretable results. Therefore, unless stated explicitly, the reader should assume that the trust-region-reflective method was used for all fitting procedures. An additional feature of the modeling API is the cleaning procedure in which, after initial parameter estimates are estimated, voxels with suspect fits are refitted using, as an initial guess, parameter estimates from the surrounding good fits (defined by $R^2 > 0.5$, an adjustable parameter). In noisy data sets, this additional processing step mitigated the occurrence of single voxels or small patches of voxels for which unreasonable parameter estimates were made. Specific details of individual model implementations and validations are described in the following subsections, with discussions of the physical model and associated parameters contained in the respective subsection of Chapter 2.

All model validation was performed using digital reference objects (DRO) – essentially synthetic image acquisitions – to simulate a typical voxel-by-voxel image analysis setting. Because magnitude MR images were used in this work, stated SNR values were calculated using

$$SNR = \frac{S_0}{\sqrt{2 - \frac{\pi}{2} \sigma_{noise}}} \quad 4.1$$

where σ_{noise} is the standard deviation of ROI voxels either in air or in a difference image (*i.e.*, noise voxels), S_0 is the phantom estimated proton density (used for VFA and VTI test objects) or ROI-based signal estimate (used for DWI, DCE, and DSC test objects), and the numerical factor in the denominator accounts for the Rician distribution of the noise [105]. For each model type, the standard deviation of

noise voxels was used to estimate the standard deviation of the Gaussian distribution from which random numbers were drawn. To synthesize MR signal, two random numbers were generated from the same distribution and added to each synthetic signal datum in the following way:

$$S_N = \sqrt{(S + n_1)^2 + n_2^2} \quad 4.2$$

In equation 4.2, S_N is the noisy MR signal generated from the model signal S and two random numbers (n_1 and n_2) drawn from the same Gaussian distribution and the subscript N simply denotes a noisy signal.

4.2.1.1 Relaxometry Models

All implemented relaxometry models fitted for the parameter T_1 instead of R_1 , as might be suggested by equations 2.19 and 2.22, to provide a better scaled parameter search space; recall, $T_1 = 1/R_1$. FSPGR data were fitted directly according to equation 2.19, whereas the VTI data processing required an additional step, *i.e.*, signal intensity polarity restoration, which was performed prior to performing any modeling.

A simple algorithm was developed to perform the polarity restoration. The minimum signal intensity and corresponding TI , say SI_{min} and TI_{min} , were found in addition to the signal intensity corresponding to the next smallest TI value (SI_{min-1}) and the next largest TI value (SI_{min+1}). These three signal intensity values were used to calculate the equation of secant lines based on for four unique cases of inverting or retaining the signal polarity: (1) none of the SI values required inversion, (2) only the SI_{min-1} point required inversion, (3) the SI_{min-1} and SI_{min} points needed to be inverted, and (4) all three points needed inversion. For each of these four cases, the Euclidean distances between the SI_{min} point and the line were calculated, with the line producing the smallest distance determining the best estimation of the TI_{null} . For each TI value smaller than TI_{null} , the signal intensities were inverted. Noisy data occasionally caused algorithm failures, in which case the values were manually inverted. Since VTI data were only used in estimation of phantom parameters using ROI-based analysis, which are easily

verified manually, performance quantification of the signal conversion algorithm was not considered in this work.

To quantify the performance of these relaxometry models, two DROs were created, one for each acquisition strategy. Each DRO was synthesized using a single S_0 and a range of R_1 values, summarized in **Table 4-2**, distributed in rectangular patches measuring 10 pixels in height and 20 pixels in width. Unique R_1 values were distributed across eight rows and six columns of patches with the first row containing the first eight values, the second row containing the next 8, and so in order of increasing R_1 . An example image of the VTI DRO is shown in **Figure 4-2** for visual reference. Parameter values were chosen to cover the range of R_1 values in the DCE phantom in addition to numerous intermediate values.

	S_0	SNR Range	Model Parameters	Simulation Parameters
VFA	5000	500-6000*	R_1 : 0.5, 0.75, 1, 2, ..., 45, 46 s ⁻¹	$TR=5\text{ms}$
VTI	800	250-2500*	R_1 : 0.5, 0.75, 1, 2, ..., 45, 46 s ⁻¹	$\theta=180^\circ$
DWI	1300	10-2000	ADC : 0.3, 0.35, ..., 2.60, 2.65x10 ⁻³ mm ² /s	$TE=85\text{ms}$
DCE	5000	500-3500*	K^{trans} : 0.05, 0.1, 0.15, 0.25, 0.35 min ⁻¹ v_e : 0.05, 0.1, 0.15, 0.25, 0.5 v_p : 0.01, 0.05, 0.1, 0.15, 0.2 Temporal resolution: 1 and 5s	$TR=5\text{ms}$ $FA=30^\circ$ $T_{10,blood}=1440\text{ms}$ $T_{10,tissue}=1000\text{ms}$ $r_1=4.9 \text{ (s}\cdot\text{mM)}^{-1}$ $HCT=0.45$
DSC	1000	5-150	α : 1, 1.2, 1.4,..., 2.8 β : 1, 1.2, 1.4,..., 2.8 $rCBV$: 0.002-0.48 Temporal resolution: 1s	$TE=35\text{ms}$

Table 4-2. Digital reference objects construction parameters. Parameters used to construct the

DROs used for software performance evaluation. Where possible, values were estimated (as described in the text) from phantom and/or patient data. The last column, “simulation parameters”, contains all other necessary model parameters used to perform the simulations. *denotes proton density estimated SNR (**not** image SNR). For example, the lowest SNR value used for VFA simulations corresponds to an FSPGR proton density image SNR of approximately 15.

The value of S_0 was also matched to phantom measurements by fitting the VFA or VTI model and averaging the proton density estimate over all vials and acquisitions of the respective image sets. From these DROs, new unique noisy reference objects were generated over a range of SNR values. Image SNR is a function of acquisition parameters (*e.g.*, a function of inversion time or flip angle) because of the signal dependence on the respective parameter, while noise is invariant. To circumvent this SNR dependency, image noise was calculated using the standard deviation of an ROI (approximately 600 voxels) placed in air, while carefully avoiding imaging artifacts, and calculated as the ratio of S_0 to this value. SNR minimum and maximum values were calculated for each image on the central slice of the VFA and VTI series, and used to determine the synthesized SNR range. Similarly, noise and S_0 were estimated in patient VFA data sets, extending the lower range of SNRs accordingly (see **Table 4-2**). Gaussian noise was then added to the simulated images as per the discussion in the introduction of this section. For each SNR level, parameter maps were generated, and limits of agreement and concordance correlation coefficients were calculated for the voxel-by-voxel estimated values (see section 3.3.2.2 for a detailed discussion of these metrics) in addition to bias and variance maps.

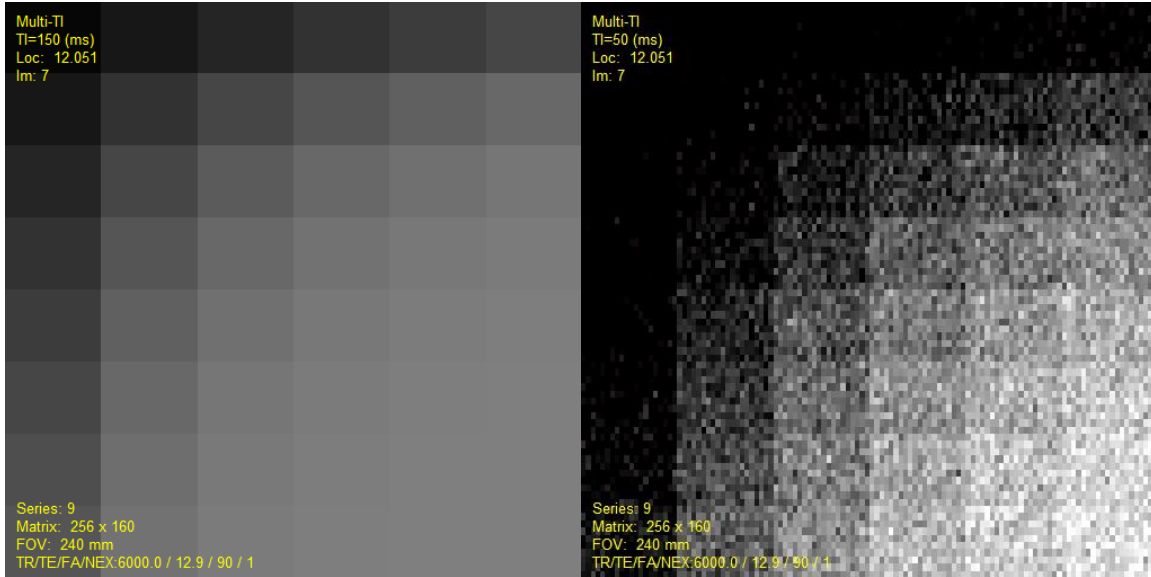


Figure 4-2. Example variable inversion time digital reference images. An example image of the noiseless (left) and noisy (right) VTI DRO. The patches for most values of R_1 can be easily seen with the top left square being the lowest value of R_1 and the bottom right corner the largest value of R_1 . To better illustrate the added noise, images simulated from two different TI values are displayed, hence the seemingly different test objects.

4.2.1.2 Diffusion Model

QUATTRO's modeling API only includes the IVIM model as DTI^2 modeling is currently available in an efficient, widely used implementation on available GE Advantage Windows workstations that provide utilities to export the resulting maps as DICOM images, and cross-validated with QUATTRO parameter estimates from DWI images (section 3.4.1.3). The IVIM model, in the simplest form (equation 2.15), was validated using a DRO with forty-eight unique, evenly spaced ADC values using the same image dimensions as the VFA and VTI DROs. ADC values were chosen to encompass the range of values in the ADC phantom and extended to include the self-diffusion coefficient of pure water reported in literature [93], [191]. For each ADC value, a corresponding T_2 value was estimated using a fitted cubic

² DTI^2 modeling is, in fact, the only component of this work, including that of specific aim 1 and 3, which was analyzed outside of QUATTRO's framework.

polynomial to interpolate the ADC phantom relaxometry data acquired at 4.7T (see **Figure 3-8**), effectively reducing the value S_0 . T_2 estimates at 4.7T were chosen over the corresponding 3.0T estimates because of the more robust acquisition method and the fact that T_2 values are relatively insensitive to field strength. SNR ranges were estimated from the ADC phantom by calculating the standard deviation an ROI placed in air (approximately 600 voxels) and average of individual vial ROIs, substituting these values into σ_{noise} and S_0 of equation 4.1, respectively. SNR estimated from phantom measurements was remained above approximately 150, however, the lower range was extended to include a generally accepted lower bound on image SNR for DTI acquisitions [99]. A unique DRO was created for each level of noise while all other model parameters were held constant. Simulation parameters are summarized in **Table 4-2**. Voxel-by-voxel parameter estimates were used to generate ADC maps, from which the LOA and CCC values were calculated to assess the performance of this fitting algorithm.

4.2.1.3 DCE Model

DROs were created to assess the three-parameter generalized kinetic model (equation 2.29) using the parameters summarized in **Table 4-2**. Each DRO contained five unique values of K^{trans} , v_e , and v_p , resulting in 125 unique combinations of those parameters. The five unique K^{trans} values, increasing from left to right, were distributed in vertical strips of twenty pixels, while unique values of v_p were placed in descending horizontal stripes (fifty pixels in height) starting at the bottom of the phantom. Values for v_e were also placed in descending horizontal stripes. However, the height of these stripes was on ten pixels and the group of five stripes recurred for each section of v_p values. **Figure 4-3** illustrates the distribution of parameters. For each combination of PK parameters, the concentration of Gd was calculated for these “tissue” uptake curves using the VIF described previously (section 3.3.2.2) using a temporal resolution of 1 and 5s. To simulate noisy images, the tissue uptake curves, expressed in mM of Gd generated from the PK parameter combinations, were converted to signal intensity assuming a perfectly spoiled FSPGR sequence with an S_0 of 5000, FA of 30° , and TR of 5ms (*i.e.*, equation 2.19).

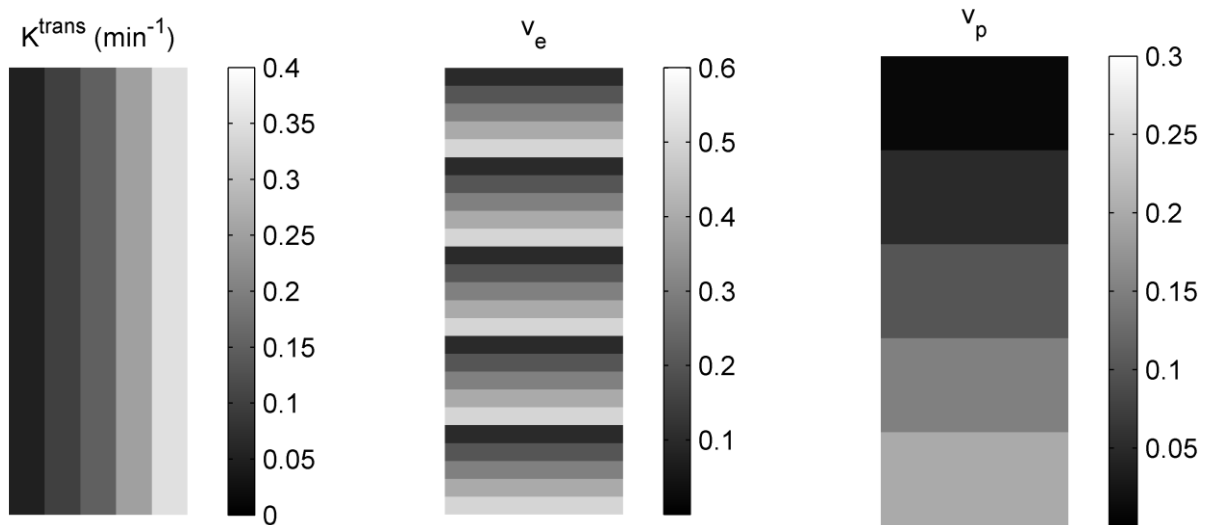


Figure 4-3. Actual pharmacokinetic values of the DCE digital reference object. Actual values of K^{trans} (left), v_e (middle), and v_p (right) used to generate synthetic tissue uptake curves. The DRO is 250 pixels in height and 100 pixels in width.

Noise was then added as discussed previously to these signal intensity values at each time point. Since the VIF was generated separately from the rectangular DRO, noise was simulated as described previously for 50 unique VIF signal intensity curves (all derived from the noiseless VIF) and the average of these curves was used as the “measured” VIF when estimating PK parameters from the noisy DRO, simulating averaging over a user-defined VIF ROI. As with the previous models, noise levels were estimated from patient and phantom data, image SNR in this case. For phantom data, subtraction images were created between the subsequent images in the serial DCE acquisition and ROIs for each of the sixteen vials were used to compute the SNR from the subtraction images using equation 4.1. Noise estimation from patient data utilized this same procedure for pre-contrast images in the DCE series, extracting noise from the physician-defined tumor ROIs (*i.e.*, tumor contours). S_0 was not estimated for these data, but rather assumed arbitrarily to be 5,000 (noise was scaled appropriately). Once maps were generated from each of these synthetic data sets, LOA and CCC values were calculated to assess the algorithm performance.

4.2.1.4 DSC Model

Modeling of the $rCBV$ relies on estimation of a gamma-variate fit (equation 2.45) for each voxel, since only the first-pass of the CA bolus was considered in this work. To determine the range of simulated fit parameters, estimates of the VIF ΔR_2^* integral and tissue $rCBV$ (equation 2.44) values were generated from the patient data sets. The ΔR_2^* integral (~ 340), averaged over all VIFs and patients, was approximated well by a gamma-variate with shape parameters of $\alpha=2.85$ and $\beta=3$; the scaling constant κ was set to 1. Various combinations of shape parameters were mapped to find those values of α and β that satisfied the patient-estimated $rCBV$ value ranges, as summarized in **Table 4-2**. Similarly to the DCE DRO, all nineteen unique α values were distributed in vertical stripes with a ten pixel width and unique β values were distributed in horizontal stripes ten pixels in height, resulting in a DRO of 190-by-190 pixels. Image SNR values were computed from patient data sets as described in previous section (4.2.1.3) using

a large ROI placed in the centrum semiovale of the frontal lobe. An arbitrarily chosen value of S_0 was used (though the value was close to the patient-averaged VIF S_0 , 876), scaling the noise appropriately. Using this set of parameters, arbitrary [Gd], derived from the gamma-variate fit, was generated and converted to signal intensity curves by the inverse of equation 2.50 on a patch-by-patch basis. Noise was then added to these simulated tissue signal intensity curves to create unique DROs each with a different level of noise, where the VIF simulation approach of the previous section was used, averaged over ten simulated time courses (similar to *in vivo* VIF ROI sizes). Voxel-by-voxel gamma-variate fitting and rCBV estimation was then performed, and LOA and CCC values were calculated to assess the algorithm performance.

4.2.2 Image Registration and Registration Validation

Routines for performing image registration were developed from the ITK libraries (version 4.3.2) and linked to the *qt_reg* API through an executable. 3D rigid image registration was performed using Mattes mutual information [192] as the similarity metric and an implementation of a regular step gradient descent optimizer [193]. In addition to Mattes mutual information, the more common Viola-Wells implementation of mutual information [194] and normalized correlation [195] were incorporated, but were otherwise unused in this work. In addition, other basic registration options such as the smallest/largest gradient step size, initial transformation guess, number of image pyramid levels, minimum signal intensity, and number of spatial samples to use in computation of the similarity metric were incorporated into the interface, which allowed fine-tuning of the algorithm's performance.

One of the major advantages of using ITK, although there are many, is the extensive validation and continued development, which provides the expansive user base the most up-to-date image analysis technologies [178], [196]. Therefore, the need for exhaustive validation of the individual software components is unnecessary, leaving only the need to select appropriate registration settings to optimize the task specific accuracy of the implemented algorithms. To accomplish this, 100 randomly

generated transformations were applied to a digital phantom matching the imaging volumes of the data analyzed in chapter 5 (see **Table 5-1**) as closely as possible based on a digital FSPGR simulation of the human head from the Montreal Neurological Institute [197]. Specifically, twelve slices with a thickness of 5mm were used with a FOV of 20cm. Image registration was then performed using the same generated transformations for several combinations of settings where the emphasis was placed on registration accuracy and not speed, although the latter was quantified. Once registered, displacement vectors were computed for each voxel within the non-zero voxels of the head and the average, median, and displacement ranges were computed for this vector field. Exclusion of zero-valued voxels avoided biasing the error estimate with voxels that would not be considered in the image analysis. These seemingly larger displacements occur as the result of being far from the center of rotation, not necessarily from misregistration. Finally, the set of registration options that offered the most accurate results were used to register an additional 10,000 random transformations to better estimate the accuracy of the algorithm for the synthetic data.

4.2.3 Code Optimization

Quantitative imaging produces immense data sets, the analyses of which can, in some cases, require several hundred to thousands of hours of total CPU time. Although the processing of these data is generally not time sensitive, substantial improvements can be easily gained through utilization of now ubiquitous multiple core processors. MathWorks offers, within the Parallel Computing Toolbox, tools for the distribution of computations across multicore processors, graphics processing units, and computer clusters. In particular, the computation of parameter maps had substantial potential for improvements in computational efficiency by distributing the workload, resulting in the development of parallel computation infrastructure. To evaluate the performance improvements, parameter maps for a single DCE image were computed using between one and four cores. Note that ITK utilizes multi-threaded computations whenever applicable. In addition to parallelization, special computationally efficient

functions were implemented for use with curve-fitting routines and other processing pipelines within the various APIs.

4.3 Results

4.3.1 Relaxometry Validation

R_1 maps were generated for the synthesized VFA and VTI digital phantoms for all levels of SNR. Bland-Altman plots calculated from all voxels of the resulting parameter maps, shown in **Figure 4-4**, clearly illustrate the dependence of bias and variance associated with noisy parameter estimations as a function of SNR. Because these values were computed from every voxel, encompassing the entire range of R_1 values, the results are not illustrative of the relationship between the bias/variance and R_1 . To investigate this relationship, maps of percentage error were computed (see **Figure 4-5**) by averaging the percent error over each R_1 patch, equivalent to averaging over a 100 voxel ROI. Bias was relatively small, even for the highest noise levels, not exceeding, in magnitude, 2.9% or 2.1% error for the VFA or VTI simulations, respectively. Comparing **Figure 4-4** and **Figure 4-5** elucidates the cause of bias in the low SNR regime of the VTI model. Essentially, the large, positively biased R_1 estimates outweigh the computation of the average bias, when taken across all DRO pixels. Similarly, variance investigations were performed to explore the R_1 dependence by computing maps of the COV of relaxation rate differences for each R_1 patch at all SNR levels, showing (**Figure 4-6**), as expected, increasing variance with decreasing SNR. The COVs for VFA estimation were roughly 0.75% on average, reaching a maximum of 12.6%. For the VTI simulations, the average COV was 0.29% and maximum COVs of 3.0% were seen. Finally, the CCC was greater than 0.998 for both DROs evaluated at all noise levels, suggesting excellent agreement between the actual and estimated parameters.

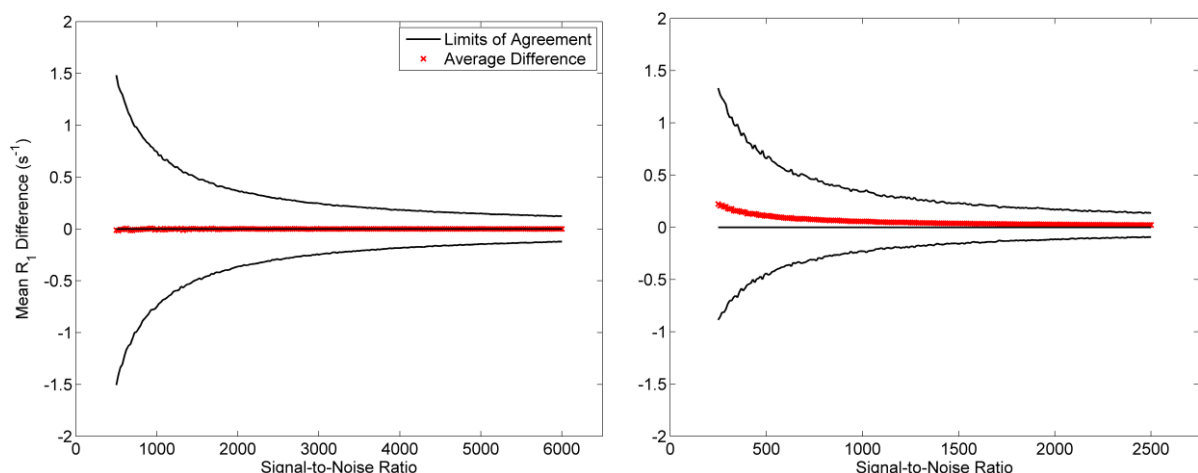


Figure 4-4. R_1 Bland-Altman plots for the VFA and VTI digital phantoms. Summary plots of the Bland-Altman analysis performed (averaged over all DRO voxels) for the VFA (*left*) and VTI (*right*) test objects. The red “x” symbols, which appear as a solid line because of the high density, denote the DRO averaged bias, where the black lines above and below these symbols represent the limits of agreement. For comparison visual comparisons, a line of zero bias was also plotted.

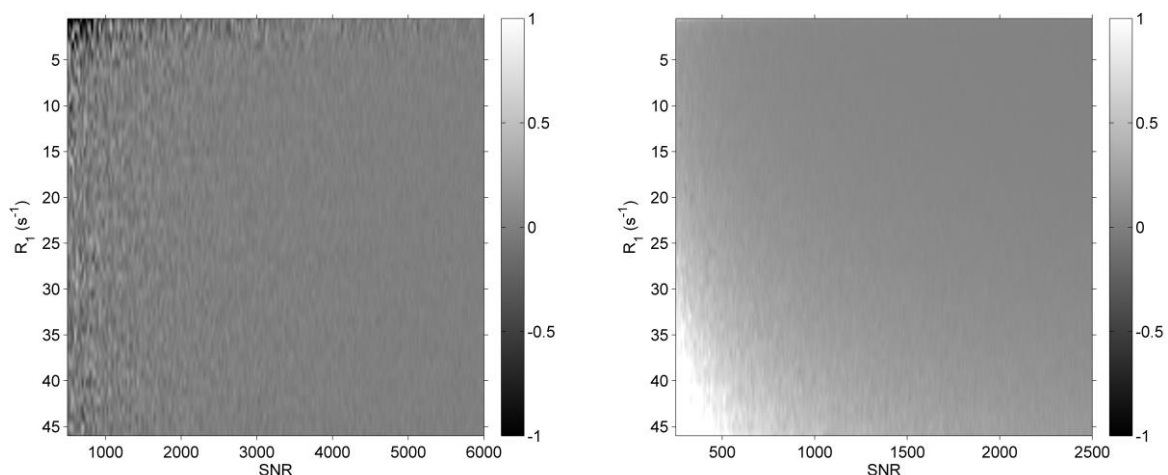


Figure 4-5. R_1 percent error maps for the VFA and VTI digital phantoms. Percent error from the true value as a function of SNR and R_1 averaged over all voxels of each respective relaxation rate for VFA (*left*) and VTI (*right*) simulations. The color bar shows the percent difference from the actual parameter used to simulate the MR signal.

Both DROs exhibited similar trends in the bias and variance associated with fitting noisy data. Namely, decreases in SNR were associated with increases in the level of bias and variance. This trend was especially pronounced at higher R_1 values in the VTI DRO and low R_1 in the VFA DRO, illustrating the parameter space region where the respective acquisition method becomes most sensitive to noise. This sensitivity, for VFA acquisitions, is the result of substantially reduced signal (a byproduct of the short TR) for low relaxation rates, resulting in noise dominated signal. VTI acquisitions exhibit increased bias for high relaxation rates as a result of the rapid signal recovery from $-S_0$. In practice, lower bounds often exist on the values of TI (50ms for the scanners used in this study), resulting in inadequate sampling of the dependence of the signal intensity on TI . In this study, for R_1 values greater than approximately 14s¹, the signal remains above the null point, yielding a poor data representation of the physical model (equation 2.22).

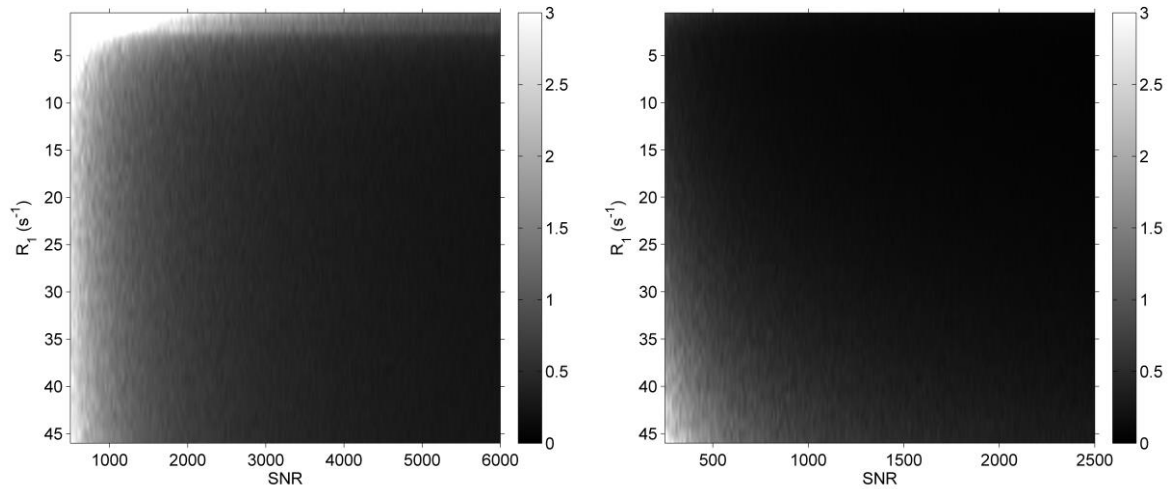


Figure 4-6. R_1 difference coefficient of variation maps for VFA and VTI digital phantoms. Maps

showing the coefficient of variation of R_1 bias for VFA (*left*) and VTI (*right*) simulations averaged over all voxels for a given R_1 value. The color bar shows the COV in %. COV was chosen over other measures of spread because of the normalization, allowing quick comparison of different parameter values.

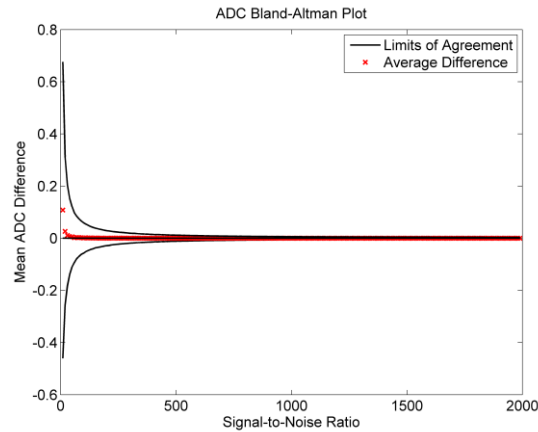


Figure 4-7. ADC Bland-Altman plot for the diffusion digital reference object. Summary plot of the Bland-Altman analysis performed (averaged over all DRO voxels) on the ADC test objects are shown. The red “x” symbols, which appear as a solid line because of the high density, denote the DRO averaged bias, where the black lines above and below these symbols represent the limits of agreement. For comparison visual comparisons, a line of zero bias was also plotted, although this line is not readily visible.

4.3.2 Diffusion Validation

ADC maps were generated for the simulated DWI acquisition for the range of SNR values listed in **Table 4-2**. **Figure 4-7** illustrates the SNR dependence of bias/variance (Bland-Altman plot) for differences between the estimated and true ADC values averaged all DRO voxels at each SNR level. These results suggest excellent agreement between ADC estimates and the true values, corroborated by a CCC of greater than 0.88 for all SNR levels (1.00 for SNR levels found in phantom data). To illustrate the parameter value dependence of bias and variance, maps for the percentage error and COV were computed and are shown in **Figure 4-8**. A maximum absolute bias of 13.0% was found when considering all SNR levels, limiting these levels to those measured in the phantom yielded a maximum bias of 0.87%. The COVs, on average, were approximately 0.55%, reaching a maximum of 29%. Again, when

considering only those SNR levels measured in the ADC phantom, the average and maximum COVs were 0.33% and 5.26%, respectively. Bias and variance results for the diffusion model suggest minimal influence of the model fitting routine, even when considering SNR levels substantially (by an order of magnitude) below that of a standard acquisition.

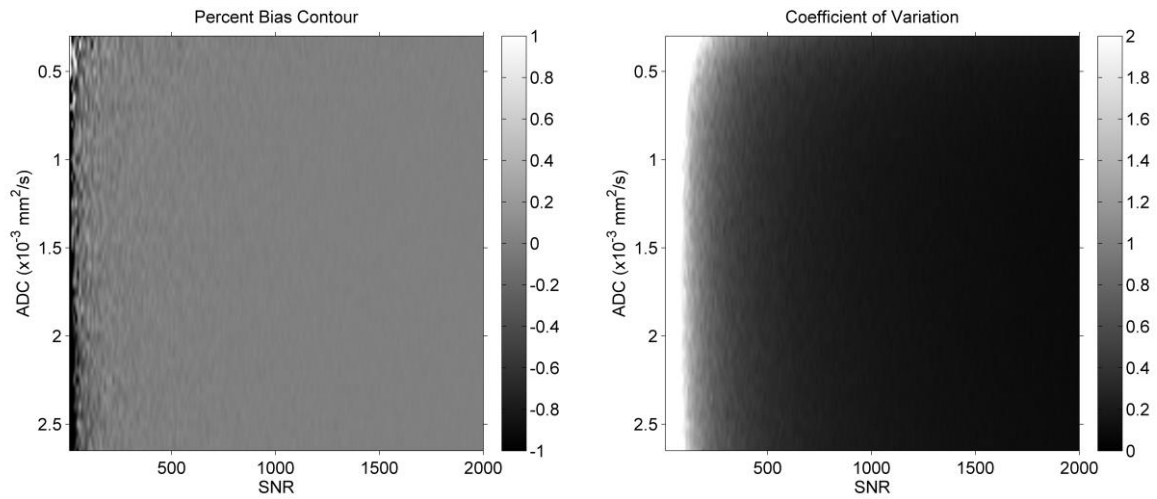


Figure 4-8. ADC percentage bias and coefficient of variation maps for the diffusion DRO. Percentage error (*left*) and coefficient of variation (*right*) maps for ADC as a function of SNR. The color bars show the percent error (*left*) and the percent COV (*right*).

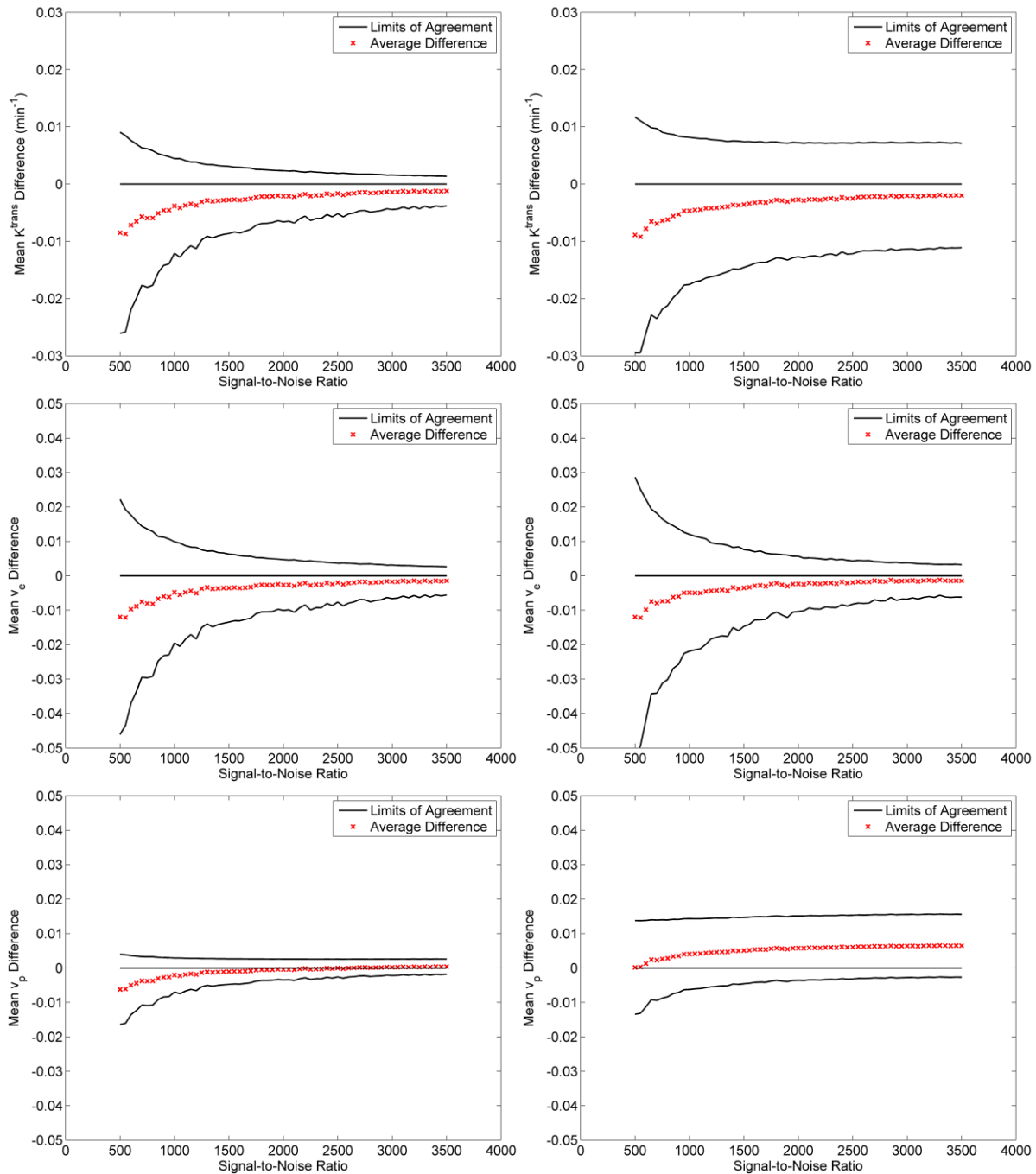


Figure 4-9. Bland-Altman plots of PK parameters as a function of SNR. Summary plots of the Bland-Altman analysis performed (averaged over all DRO voxels) on the DCE test objects are shown. The red “x” symbols denote the DRO averaged bias, where the black lines above and below these symbols represent the limits of agreement. For comparison visual comparisons, a solid black line of zero bias was also plotted.

4.3.3 DCE Validation

K^{trans} , v_e , and v_p maps were generated for the simulated DCE acquisition using the range of SNR values listed in **Table 4-2**. **Figure 4-9** illustrates the Bland-Altman analysis as a function of SNR for parameters estimated from one and five second temporal resolutions. Consistent underestimate of K^{trans} and v_e are seen at both temporal resolutions on average, with substantial differences in the variance between the two resolutions (increasing with temporal resolution). The v_p results exhibited the same trend in variance, though notable differences were seen in the bias. Namely, v_p was consistently overestimated with increasing temporal resolution. These results are consistent with the observation of increasing bias with temporal resolution of Aerts *et al.* [164] and Cao *et al.* [165]. Again, these results indicate, on the whole, performance of the algorithm averaged over all parameter values, neglecting the influence of SNR, temporal resolution, and the actual parameter values. Similarly to the relaxometry and diffusion model analysis, maps of bias and COV were generated as a function of SNR and model parameters. Additional model parameters (more than 2) inhibit straightforward visualization of these metrics as a function of model parameters and SNR, resulting in five unique K^{trans} percent difference 3D scatter plots, one for each value of v_p , at the high sampling rate (**Figure 4-10**), and five K^{trans} percent difference plots for the lower temporal resolution (**Figure 4-11**). Comparison of these bias figures confirms the general trend of increasing parameter estimates as temporal resolution decreases, and for all parameters, increasing bias with decreasing SNR. In addition to the trend of decreasing K^{trans} bias with increasing v_p , another trend of decreasing K^{trans} bias with decreasing v_e is clearly visible at both temporal resolutions, particularly for higher values of v_p . K^{trans} bias exhibited a slight increasing trend with decreasing K^{trans} , although slightly less apparent at the higher temporal resolution. Overall, the maximum absolute K^{trans} bias was 10.4% and 13.9% and averaged absolute bias was 1.4% and 1.8% for the 1s and 5s temporal resolutions, respectively. Absolute maximum and average percent errors for v_e were similar to those of K^{trans} , 23.1% and 1.1%, respectively, at the high temporal sampling rate and for

those at the lower sampling rate, 23.2% and 1.1 %, respectively. The v_e percent bias plots are shown in **Figure 4-13**, and increased bias can be easily seen at the edges of the parameter space (*i.e.*, low/high v_e , v_p , and K^{trans}). However, v_p exhibited substantially higher bias (**Figure 4-12**) in magnitude and on average at the lower sampling rate, 100% and 17.8%, respectively, while the values at the 1s resolution, 3.7% and 29.0%, respectively, were comparable with other PK parameters.

Parameter variance was minimal for K^{trans} with COVs less than 4% in magnitude and on average less than 0.75% at both temporal resolutions. The v_e results varied substantially more in magnitude at the lower sampling rate (9.6% max COV), but, on average, COVs were less than 1%. However, variance in the estimate of vascular plasma volume was substantially more. The COVs for the 1s sampling interval were found to be less than 1%, on average, and nearly 10% in magnitude, with those values for the 5s interval reaching 27% and greater than 1400%. Most of this additional variance, not surprisingly, can be attributed to the lowest value of v_p , as removal of these data yields COVs of approximately 0.6% on average and 5.9% in magnitude for the lower temporal resolution. These results clearly indicate the need for careful assessment of parameter estimates, particularly in regions of low v_p .

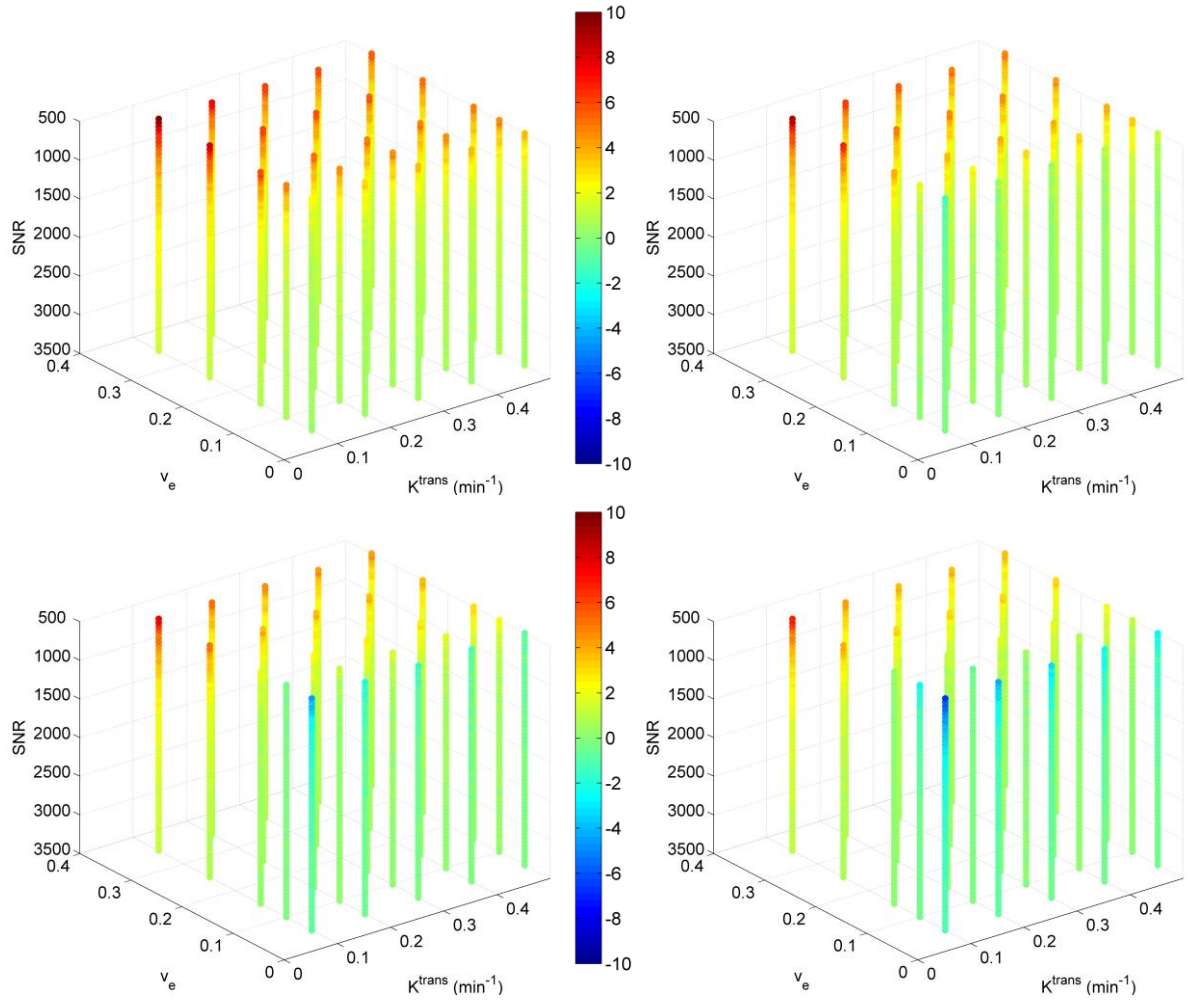


Figure 4-10. DCE DRO percent bias maps for estimated K^{trans} at 1s temporal resolution. Color maps of the percent bias as a function of the actual value of v_p (0.05-*top left*, 0.1-*top right*, 0.15-*middle left*, 0.2-*middle right*), v_e on the y-axis, K^{trans} on the x-axis, and SNR on z-axis. The color bar represents the percentage difference between the estimated parameter and the actual parameter value.

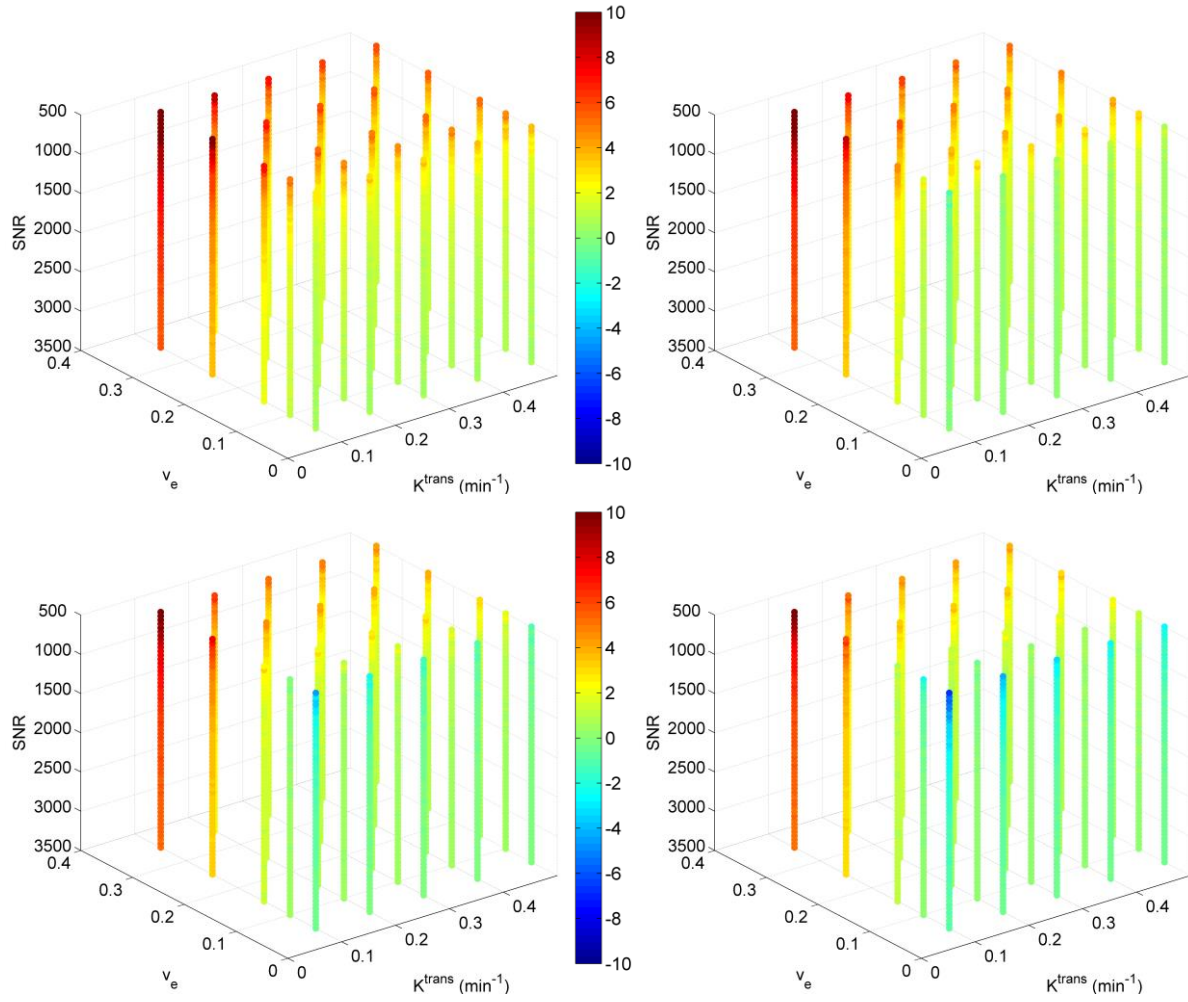


Figure 4-11. DCE DRO percent bias maps for estimated K^{trans} at 5s temporal resolution. Color maps of the percent bias as a function of increasing v_p (0.05-*top left*, 0.1-*top right*, 0.15-*middle left*, and 0.2-*middle right*), v_e on the y-axis, K^{trans} on the x-axis, and SNR on z-axis. The color bar represents the percentage difference between the estimated parameter and the actual parameter value.

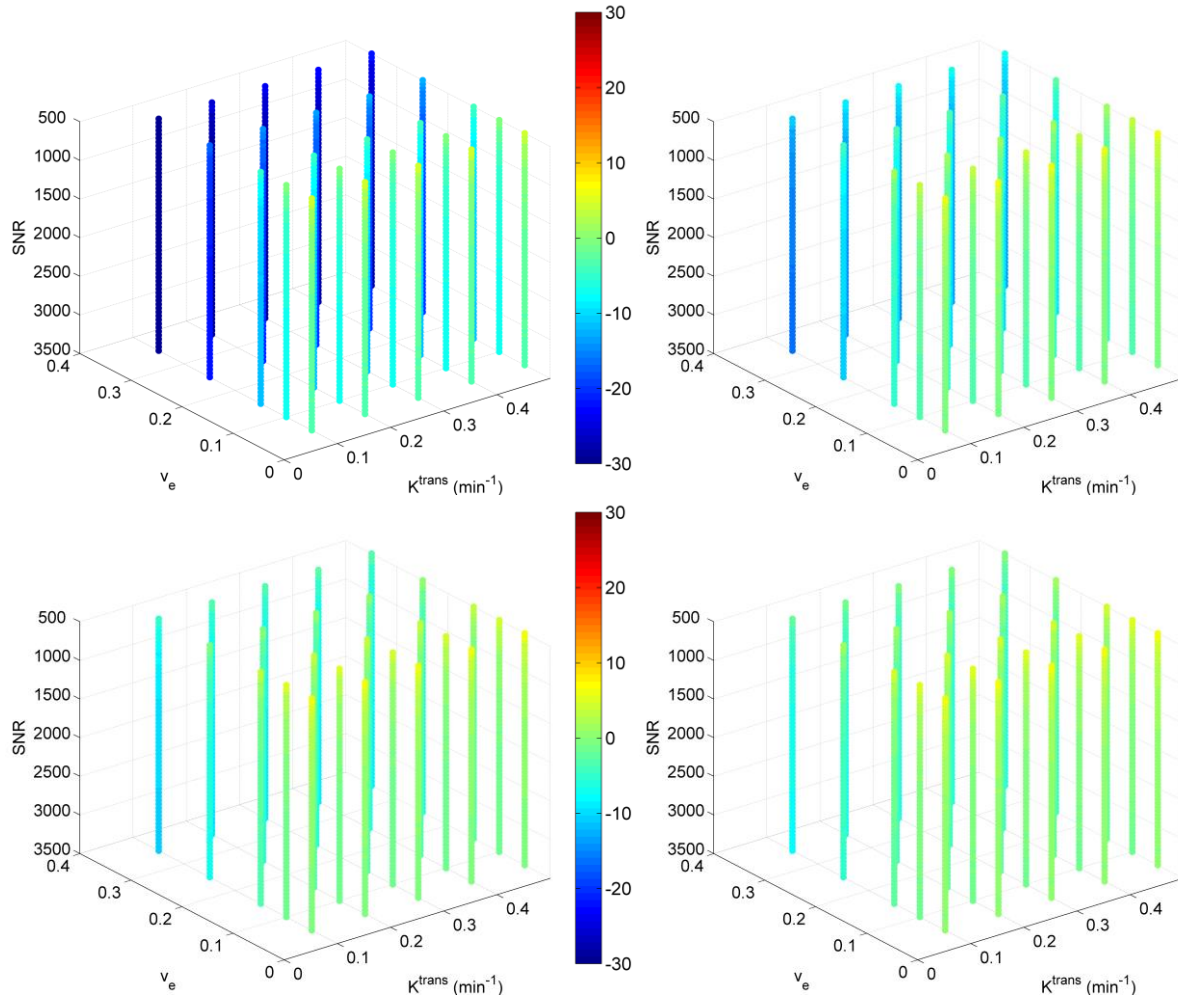


Figure 4-12. DCE DRO percent bias maps for estimated v_p at 5s temporal resolution. Color maps of the percent bias as a function of increasing v_p (0.51-*top left*, 0.1-*top right*, 0.15-*middle left*, and 0.2-*middle right*), v_e on the y-axis, K^{trans} on the x-axis, and SNR on z-axis. The color bar represents the percentage difference between the estimated parameter and the actual parameter value.

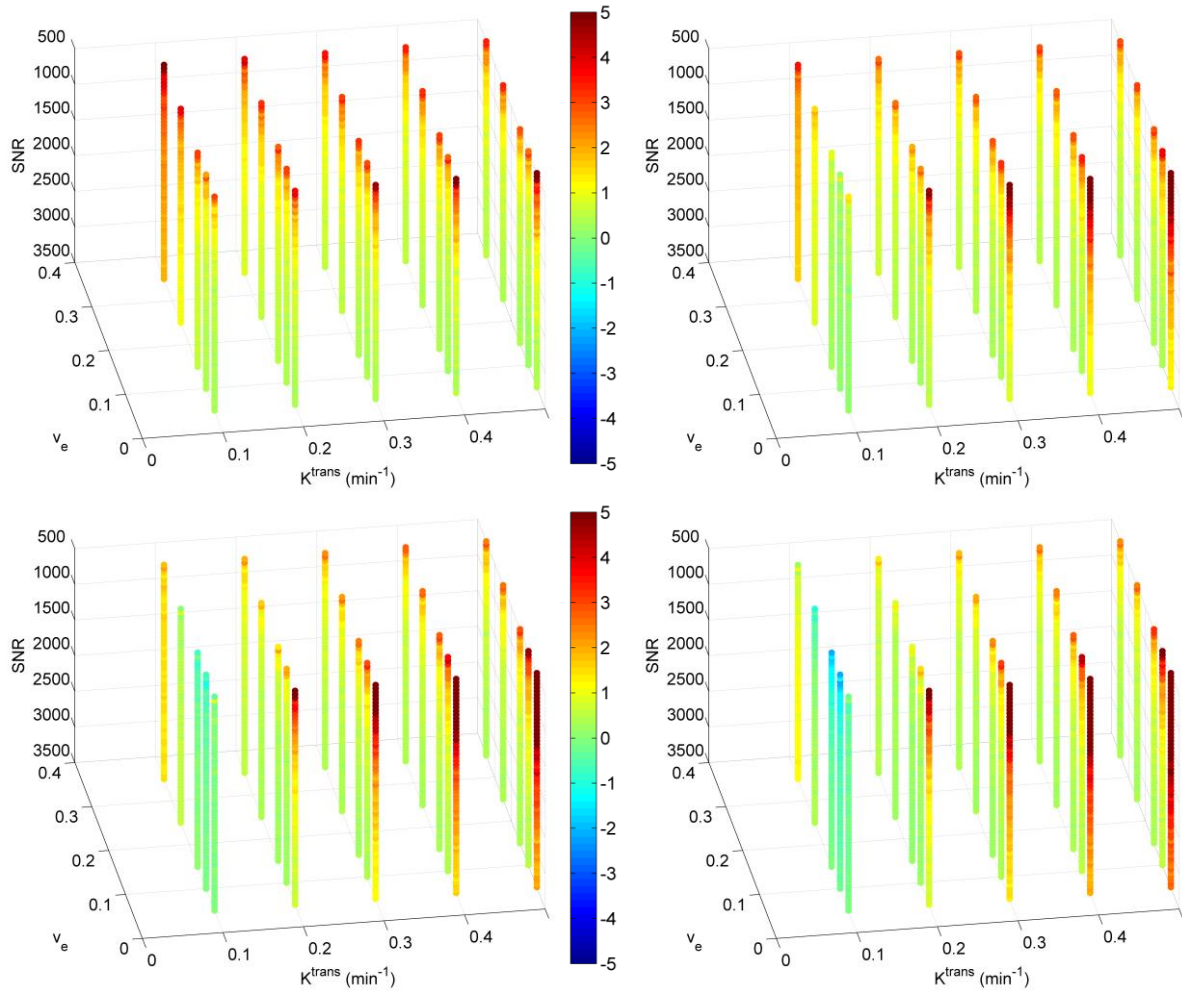


Figure 4-13. DCE DRO percent bias maps for estimated v_e at 5s temporal resolution. Color maps of the percent bias as a function of increasing v_p (0.05-*top left*, 0.1-*top right*, 0.15-*middle left*, and 0.2-*middle right*), v_e on the x-axis, K^{trans} on the y-axis, and SNR on z-axis. The color bar represents the percentage difference between the estimated parameter and the actual parameter value.

4.3.4 DSC Validation

$rCBV$ maps were generated for the simulated DSC acquisition for each SNR level used to simulate the MR acquisition, as listed in **Table 4-2**. Results of a Bland-Altman analysis and CCC computations for all voxels as a function of SNR are summarized in **Figure 4-14**. These results suggest, as expected, increased bias and variance as the noise levels increase (*i.e.*, SNR decreases), albeit slight

trends. The CCC results as a function of SNR suggests a drastic departure in the estimated values from the true values. However, this trend is likely the result of a small number of highly biased $rCBV$ values. To further investigate the cause, the SNR dependence of bias, specifically percent bias, was mapped and is shown in **Figure 4-15**, in addition to a map of the COV. A maximum absolute bias of approximately 1000% was found, however, comparison with the bias map suggests that these large deviations from the true $rCBV$ occur at very low SNR/ $rCBV$ values and are few in number. In fact, more than 80% of the SNR/ $rCBV$ combinations shown in **Figure 4-15** are within 5% of the true $rCBV$. Similarly, the average and maximum COV values were meaninglessly large (greater than 1000%). This again occurs in an isolated region of the parameter space and is most likely the result of small variations in values close to zero.

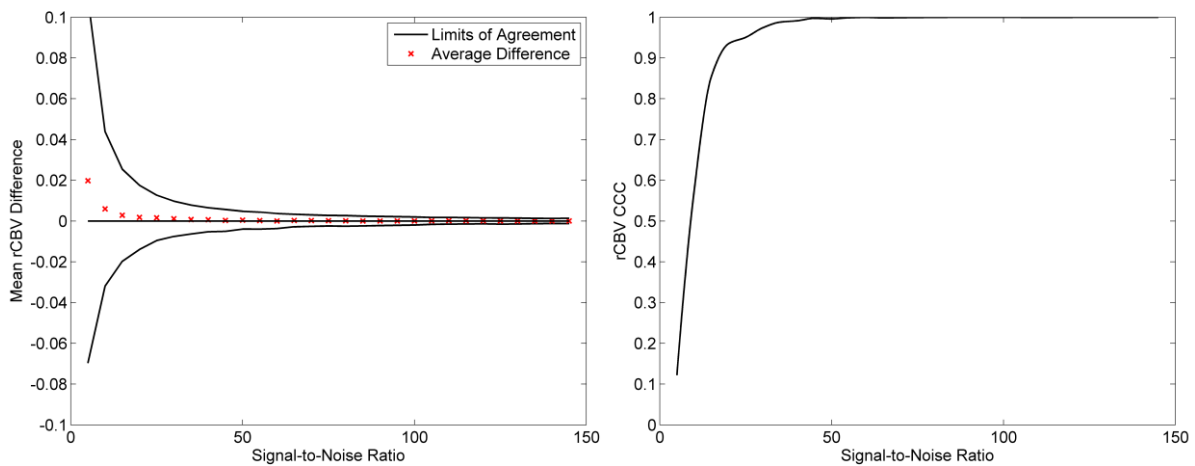


Figure 4-14. $rCBV$ Bland-Altman plot and CCC plot as a function of SNR for the DSC DRO. Summary plots of the Bland-Altman analysis performed (averaged over all DRO voxels) on the DSC test objects are shown. The red “x” denotes the DRO averaged bias, where the black lines above and below these symbols represent the limits of agreement. For comparison visual comparisons, a solid black line of zero bias was also plotted.

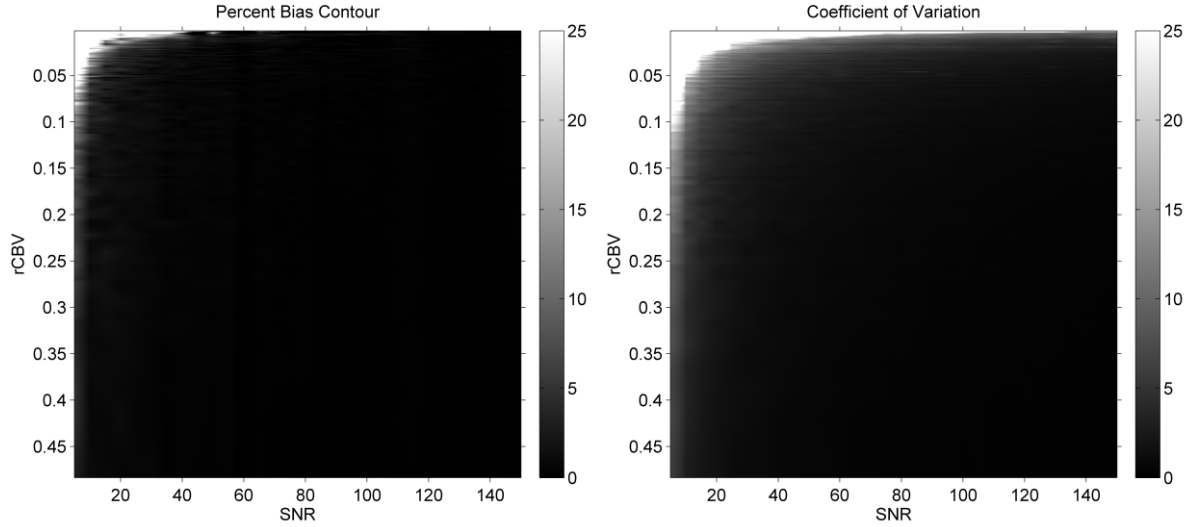


Figure 4-15. *rCBV* percentage bias and COV maps for the DSC DRO. Percentage error (*left*) and coefficients of variation (*right*) maps for *rCBV* as a function of SNR. The color bars show the percent error (*left*) and the percent COV (*right*).

4.3.5 Image Registration

Numerous combinations of the minimum signal intensity, minimum gradient step size, and number of pyramid levels were tested using the same 100 randomly generated transformations. For the expected magnitude of transformations required in this work, a single level (the original image resolution) was found to provide the most accurate results when signal intensities greater than or equal to zero were used. No additional improvements in accuracy were found by decreasing the minimum step size below a value of 1×10^{-4} . This set of registration settings provided sub-voxel accuracy when registering noiseless simulated FSPGR images, and was used in all subsequent work. Using the fine-tuned settings, 10,000 random transformations were applied to the same image set and the images were re-registered. The average voxel displacement from the original position, averaged over all random transformations, was approximately 0.73mm with a standard deviation of 0.074mm. These results suggest excellent rigid alignment accuracy.

4.3.6 Code Optimization

Parallel computations were implemented for parameter map calculations. The computations were benchmarked using a single 250x100 voxel DCE image data set. Total computation time for a single processor was measured at 866s (approximately 14min). Using two and four processors resulted in improvements (reductions in processing time) of 16% and 27%, respectively. Much of the computation time contributing to this relatively small gain in computational efficiency was the result of the parameter re-estimation algorithm, which accounted for 633s of the total computation time stated for a single processor. Removing this value from the total processing time resulted in gains of 203% and 484% for two and four cores, respectively, results comparable to the expected improvements in processing time. Unfortunately, the post-processing re-estimation algorithm depends on neighboring parameter estimates during each iteration, rendering multi-threaded processing incompatible with the task.

4.4 Discussion

Quantitative analysis of multi-parametric MR imaging data is a multifaceted, time-intensive process requiring a plethora of tools for quantitation and meaningful representation of imaging information. These tools, along with the equipment used to acquire data, require rigorous assessment to ensure that bias and variance are quantified, and, where possible, mitigated in an effort to avoid propagating unnecessary errors through the image analysis and, ultimately, the response assessment framework. This study aimed to develop the tools necessary for providing a unified environment, in addition to assessing/mitigating bias and variance through the use of synthesized digital reference objects representing values commonly measured *in vivo* and analyzed in the same manner as patient data sets.

QUATTRO was developed in the MATLAB environment using object oriented programming principles, linking other commonly used and highly developed programmatic tools, specifically ITK and R. The flexibility afforded by MATLAB allows further customization and development of tools associated and/or derived from any of the developed classes summarized in **Table 4-1**, in addition to providing a unified workspace for performing analysis of quantitative images. While these basic tools provide a relatively comprehensive image analysis framework, the linkage of QUATTRO with the libraries of ITK and R provides innumerable possibilities for more advanced image registration procedures (*e.g.*, deformable image registration) and statistical analysis, respectively. Parallelization tools were also incorporated to improve performance of parametric map computations using a framework that can be extended easily to incorporate other amenable computational tasks. Although these tools were benchmarked using a single multi-core CPU, the increasing availability of machines supporting GPU computations and high performance computing clusters (*e.g.*, cloud computing) affords opportunities for further expansions of parallelization capabilities.

Each of the pre-programmed modeling modes and the image registration algorithm was validated using QUATTRO's tools to quantify inherent bias and variance of the algorithms' performance. Model validation showed no substantial bias and variance introduced for the relaxometry and diffusion models. For temporal resolutions comparable to clinical acquisition schema, DCE validation results suggested increased bias, when compared with faster sampling rates (*e.g.*, 1 frame/s). In addition to increased bias, substantial variance was observed in the lower range of tested v_p (<0.02). All other PK parameters were relatively robust with respect to bias and variance. Similarly, validation of the gamma-variate fitting for DSC models resulted in substantially increased bias and variance for small values of $rCBV$ (<0.02 and <0.05 , respectively).

QUATTRO was the manifestation of considerable efforts focused on developing a unified workspace for performing quantitative image analysis in the context of treatment response, the goal of

this specific aim. The performance of the various built-in models was validated and characterized using simulated data sets. Future work should involve development and incorporation of advanced modeling and image registration techniques (see **Chapter 6** for additional details).

Chapter 5 Radiological Response Maps

5.1 Introduction

Evaluation of novel targeted or combined therapeutics using traditional measures of response, such as RECIST [1] and RANO [2], has faced a number of challenges. Increasingly, functional MR QIBs, many of which are potentially suitable for use as surrogate endpoints, are being incorporated into clinical trials, as evidence supporting the ability of these markers to assess treatment response is growing [7]–[18]. Myriad studies have demonstrated correlations between QIBs and histological samples [20]–[32], [37], [43], [56], [62]–[64], and, moreover have established a relationship between patterns of QIB changes and treatment response [18], [33]–[39], [70]–[72]. As increasingly complex therapies are performed, establishing methods for elucidating locoregional tissue response offers promise as a means of assessing the heterogeneity associated with treatment induced changes.

In an effort to capture these biological changes associated with treatment, a multi-parametric statistical framework was developed that provided a means of producing voxel-by-voxel predictions of treatment response. A small cohort of patients having undergone advanced imaging studies was retrospectively evaluated, including acquisition of DTI, DCE, and DSC imaging data sets, in an effort to preliminarily identify an appropriate model and evaluate the generalizability of that model.

5.2 Patient Acquisitions

Retrospective analysis was performed on a cohort of patients ($n=13$) with surgically resected recurrent glioblastomas. All 13 patients received concurrent radiation therapy and temozolomide plus or minus bevacizumab. Of these 13 patients, 6 were treated without bevacizumab, 6 were treated with bevacizumab, and the treatment status, with regard to bevacizumab, 1 patient is unconfirmed. This last

patient, in the professional opinion of our expert radiation oncologists, is thought to have been treated without bevacizumab. Patients underwent advanced MRI imaging four times during this study. The first study was performed approximately one week prior to chemoradiation (and/or bevacizumab) therapy (4 weeks post-surgery), two studies were performed 3-4 weeks mid-treatment, and one post-treatment study was performed 3-5 weeks after treatment was complete. Pre-treatment scans, which occurred 5.6 ± 0.8 (mean \pm SEM) days pre-treatment (24.6 ± 1.8 days post-surgery), and both mid-treatment scans, occurring at 21.5 ± 0.3 days and 24.8 ± 1.1 days after treatment was initiated. **Table 5-1** summarizes the quantitative acquisition protocol.

5.3 Methods and Materials

All work described in the following sections was performed using QUATTRO. A combination of

	DCE	DSC	DTI
Coil:	8CH Brain	8CH Brain	8CH Brain
Sequence:	3D FSPGR	GRE-EPI	SE-EPI
User CVs:	Turbo=2	---	ASSET=2
TE (ms):	1.3	35	91.8
TR (ms):	6.4	1500	10000
BW (kHz):	+/-62.5	+/-250	+/-250
NEX:	1	1	1
Flip Angle (deg.):	30	90	90
b-value (s/mm²):	---	---	0, 1200
FOV (cm):	24	24	24
Phase FOV (%):	0.75	1	1
Sl. Th. (mm):	5	5	5
Gap Th. (mm):	---	---	1
Num. of Slices:	12	12	25
Num. of Frames:	56	120	27*
Matrix:	256x160	128x128	128x128

Table 5-1. Patient DCE, DSC, and DTI phantom protocol acquisition parameters. The temporal

resolution of the DCE scan was approximately 6.3s. *Denotes the number of diffusion encoding

directions (one b_0 image was acquired). Abbreviations: Num. = Number, Sl. = Slice, Th. = Thickness

GUI and command prompt functionality was used in addition to numerous QUATTRO and MATLAB scripts. The only portion of the following analyses to be performed outside of QUATTRO was the computation of the DTI quantities (ADC and FA). Images were transferred and analyzed on the Linux workstation described in section 3.3.2.

5.3.1 Image Registrations

Rigid image registrations executed with the `qt_reg` class were used to align all imaging volumes. As discussed in section 4.2.2, the image registration parameters were fine-tuned for the DCE, DSC, and DTI data sets. Briefly, a single pyramid level (native image resolution), tri-linear interpolation, and a minimum step size of 1×10^{-4} were used to perform all alignments. In circumstances where the voxel sizes between the template and moving imaging volumes were disparate, the larger voxel size was resampled to match the higher resolution volume. Intra-series (same day serial acquisitions), inter-series (same day, different series), and longitudinal (different dates, same series) were performed independently.

Intra-series alignment was performed for DCE, DSC, and DTI data sets prior to generating any parametric maps. For a given serial DCE acquisition, all serial images were registered to the first frame in the series. A similar methodology was used for the DSC series, but the third frame in the series was used *in lieu* of the first frame to avoid non-steady state signal intensities. All non-zero b -value images in the DTI series were registered to the b_0 image. Following estimating of the image registration transformations, voxel displacements averaged across the volume were calculated and transformations were only applied to images for which the average displacement was greater than the minimum voxel dimension. Intra-series registrations were only used to adjust for slight patient motion during the serial acquisitions prior to performing acquisition appropriate quantitative analysis; these transformations were otherwise unused.

Voxel-by-voxel analysis of radiographic response required co-registration of all parametric maps for a given patient, a two part procedure. For a given patient and imaging study date, any inter-series

patient motion was corrected for by aligning the DSC and DTI imaging volumes to that of the DCE series. Although the DCE and DSC volumes were identical, the acquired imaging matrices were different by a factor of four, which was corrected by first resampling the DSC images to 256x256 (original matrix: 128x128) prior to aligning the volumes. As with the series registrations, the first DCE series was chosen as the template volume and the third frame of the DSC time series selected as the moving volume. Differences between the DCE and DTI volume extent and slice spacing, required additional considerations. Prior to co-registering the DCE and DTI data sets, DTI slice spacing was resampled to that of the DCE image set (*i.e.* 5mm thickness and no gaps) and the DCE image volumes was padded with zeroes so as to be commensurate with the diffusion image volumes. Following alignment of the b_0 DTI volume to the first DCE series volume, the DTI imaging volume extending beyond that of the DCE volume was removed, resulting in equivalent voxel dimensions and volume extent. All volume transformations were stored as ITK iteration files, a *qt_reg* specific text-based file format storing algorithm setup details, the transformation and similarity value at each iteration, and the algorithm stop criteria. Similarly, longitudinal imaging data were co-registered. DCE, DSC, and DTI longitudinal volumes were registered to the corresponding volume of the pre-treatment date using the native resolution. As described previously, the first series volume of the DCE acquisitions, the third serial DSC imaging volume, and the b_0 image of the DTI data sets were used for co-registration, and all resulting transformations were stored as ITK iteration files.

Image transformations were calculated and stored, but the application of those transformations was performed simultaneously to avoid introducing image interpolation artifacts. The *qt_reg* class provides a method for generating a single transformation from serial rigid transformations. To generate the new imaging volumes used for radiographic assessment and subsequently extraction values from all parametric maps, the inter-series transformations were imported from the ITK iteration files into *qt_reg* followed by the longitudinal transformations. A single transformation matrix was automatically

calculated and applied to each of the DCE, DSC, and DTI imaging volumes and parametric maps. By generating the registered volumes in this way, as opposed to creating intermediate transformed imaging volumes, propagation of interpolation artifacts was avoided. These transformed images were used for visual assessment of registration success and the images of the transformed DCE series provided a volume on which the oncologist drew all response ROIs.

5.3.2 Quantitative Image Analysis

As discussed in section 2.1.3, DCE modeling was performed using the three parameter general kinetic model (equation 2.29) on a voxel-by-voxel basis. VIFs were manually defined on a central slice of the imaging volume using a polygon ROI by an iterative selection process. All three longitudinal DCE exams were visualized simultaneously. Candidate blood vessels were delineated and the time courses for the manually defined VIFs of each study date were compared qualitatively. Blood vessels for which in-flow enhancement was minimal and for which qualitative features such as peak enhancement, bolus shape, and washout were most similar were used to define the VIF for that respective DCE study. During the VIF selection process, the number of pre-contrast images was determined by visually inspecting the baseline signal, identifying the last time point before substantial trending increases in signal intensity were observed which denoted the number of baseline images. After defining the VIFs and determining the number of pre-contrast imaging phases, the DCE settings of QUATTRO were updated and parametric maps for K^{trans} , v_e , v_p , and R^2 were calculated using the change in signal intensity (equation 2.41) and the intra-series registered images as discussed in section 5.3.1, noting the discrimination criteria for application of the image transformations. DCE modeling performed in QUATTRO averages the signal intensity of all pre-contrast images before calculating the change in signal intensity with no additional image processing.

DSC data were analyzed in a similar fashion to the DCE data. Since only first pass kinetics were considered, a manually defined recirculation cut-off was determined during the VIF selection process in

addition to the number of pre-contrast images. Determination of the number of pre-contrast images was identical to that of the DCE analysis and the recirculation cut-off was chosen to be the frame after the peak [Gd] concentration in the VIF time course for which a visually perceptible change in the negative slope was present. **Figure 5-1** illustrates two examples of pre-enhancement and recirculation cut-off selections. QUATTRO settings for the DSC exam were updated with these parameters, and computation of a voxel-by-voxel gamma-variate fit was performed. Prior to fitting each voxel CA time course, a gamma-variate is fit to the VIF and the estimated bolus arrival time (*BAT*) is used, in conjunction with the recirculation cut-off, to provide bounds for the individual voxel *BAT* estimation. In practice, this was found to prevent, at least in part, a vast majority of gamma-variate fits with unreasonably large integrals (on the order of 10^8). Following these initial processing steps, gamma-variate parameter maps and corresponding R^2 maps were computed. A Gaussian-Kronrod method,

implemented in the MATLAB function *quadgk* [198], was used to numerically integrate the individual voxel gamma-variate and the VIF fits, taking the ratio of the two integrals yielding the *rCBV* map. Again, the intra-series DSC registrations were used where appropriate.

DTI data sets, following application of intra-series registration, were transferred to a GE Advantage Windows Workstation running FuncTool (GE Healthcare, Waukesha, WI, version 4.5) using FTP. Images were initialized in FuncTool and signal thresholds were set to the smallest allowable minimum and largest allowable maximum values to ensure diffusion parameter computations were performed in all voxels. *ADC* and *FA* maps were computed and stored in DICOM format, after which the data sets were transferred back to the Linux workstation for further analysis. QUATTRO scripts were developed to import the *ADC* and *FA* maps into the corresponding DTI QUATTRO save files.

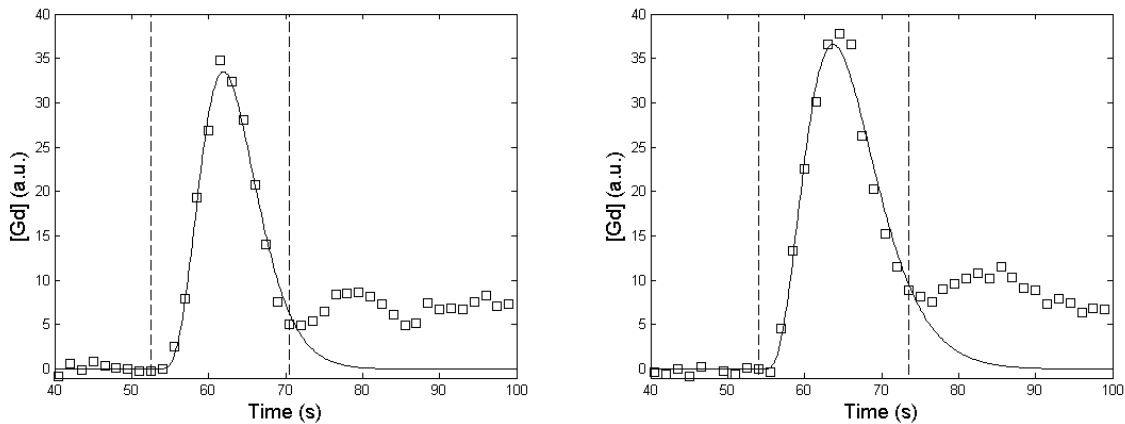


Figure 5-1. Example pre-contrast and recirculation cut-offs for DSC VIF. Two examples of a DSC VIF demonstrating the manually defined pre-enhancement (first dotted vertical line) frames and the bolus recirculation cut-off (second vertical line). The small squares are the measured data, converted from signal intensities to arbitrary units of $[Gd]$, and the solid line is the fitted gamma-variate. Note that even though the pre-contrast frame is manually defined a value for the bolus arrival is actually one of the fitting parameters.

5.3.3 Radiological Assessment

Two experienced radiation oncologists provided consensus radiological assessment of each patient. Both oncologists were instructed to contour areas surrounding the surgical cavity, using all clinical data available at the conclusion of the quantitative imaging study, with labels of partial response (PR), stable disease (SD), and progressive disease (PD). Initial imaging studies were reviewed, tracking changes throughout subsequent quantitative imaging studies. In addition to imaging information, consideration was given to the radiation therapy treatment plan to potentially discriminate treatment effects from progressive disease. All ROIs were drawn on the co-registered (between series and imaging study dates), during which time the oncologists verified the success of the individual rigid registrations.

5.3.4 Statistical Classification and Response Maps

Ordinal regression was performed on a voxel-by-voxel basis using the individually and consensus defined radiographic assessments (PR, SD, and PD) as the response variable. Thresholds were defined for each of the quantitative parameters, except *ADC*, to prevent modeling physically unrealizable model parameters, excluding voxels that exceeded those bounds. Bounds on the *ADC* performed the additional task of discriminating between tissue and fluid filled cavities remaining from the surgical resection.

Table 5-2 summarizes those physical thresholds. Scripts were created in MATLAB to perform the longitudinal regression analysis, importing co-registered parameter maps, response ROIs, and extracting all pertinent data using the *qt_exam* class. ROIs defined by the radiation oncologists representing the final patient outcome were loaded through the *qt_roi* class, and masks were created and applied to all registered parametric maps, from which voxels were extracted. After applying the parameter thresholds, resultant voxels extracted in this manner were imported and stored in a *qt_response* object as the predictor variable, from which point, generation of visualizations such as histograms and probability histograms was easily performed.

A basic model was created to perform initial investigation of subsequent differences (*i.e.* first mid-treatment minus baseline and second mid-treatment minus first mid-treatment) of the form

$$\begin{aligned} & \left(\Delta ADC_1 + \Delta FA_1 + \Delta K^{trans}_1 + \Delta v_{e1} + \Delta v_{p1} + \Delta rCBV_1 \right)^2 \\ & + \left(\Delta ADC_2 + \Delta FA_2 + \Delta K^{trans}_2 + \Delta v_{e2} + \Delta v_{p2} + \Delta rCBV_2 \right)^2 \\ & + I[Treatment = bevacizumab] \end{aligned} \quad 5.1$$

where the subscripts 1 and 2 refer to the baseline/first mid-treatment parameter changes and differences in the mid-treatment parameters, respectively, and $I[...]$ is an indicator variable defined as,

$$Ind[Treatment = bevacizumab] = \begin{cases} 0 & \text{if } Treatment = placebo \\ 1 & \text{if } Treatment = bevacizumab \end{cases} \quad 5.2$$

The necessity of this indicator variable arose because the two treatments (placebo and bevacizumab) were expected to affect different changes in functional QIB, and the incorporation of this additional term provided a means of specifying a single model for all patients in this cohort. Expansion of the parenthetical statements included interaction terms between QIBs. Additional models were also investigated after qualitative analysis of the data was performed, and are discussed in further detail in the following section (5.4). For each of these covariates (including the treatment indicator), the sample mean and standard deviation was calculated and used to standardize the covariate, creating covariates with a mean of zero and standard deviation of one. Regressions were performed as discussed in section

	$ADC (x10^{-3} mm^2/s)$	FA	$K^{trans} (min^{-1})$	v_e	v_p	$rCBV$
Minimum	0	0	0	0	0	0
Maximum	2	1	3	1	1	1

Table 5-2. QIB modeling thresholds. Voxels exceeding the thresholds defined in this table were

excluded from the regression analyses, as these thresholds nominally represent physically realizable parameter quantities of biological tissue.

2.2.2 using the *ordinal* package (version 2013.9-30) developed in R [199], estimating unique regression coefficients for each model term, including a regression intercept and category cut-offs.

Leave-one-out (LOO) cross-validation, as discussed in section 2.2.3, was incorporated to provide an exploratory assessment of the generalizability of the developed regression classifiers. When performing cross-validation with adequate data sets, the data can more easily be partitioned to form training, validation, and testing data sets. By separating the data in this manner, both components of cross-validation – model selection and model validation – can be optimally performed. However, the limited size of the data set in this work was limiting, despite the number of observations (*i.e.*, voxels) was large. Voxel measurements from the same patient are expected to be highly correlated, requiring that the cross-validation be performed at the patient level to avoid substantially overestimating the prediction accuracies. To this end, patients were excluded one at a time, leaving the remaining patient data sets to train the regression models. For each LOO patient, every combination of the six unique QIBs, ADC , FA , K^{trans} , v_e , v_p , and $rCBV$, was fitted to a unique ordinal regression model. For example, one model might include only the ADC and $rCBV$ parameters, by dropping all other terms from the model defined in equation 5.1. Of these models, with the accuracies averaged across all LOO cases, the multi-parameter model with the highest average accuracy (average of the individual categories' accuracies) was compared to the accuracies of the corresponding single parameter models to determine whether the multi-parameter model was a better predictor of response compared with the single parameter counterpart.

In addition to the LOO cross-validation, the generalized coefficient of determination was calculated according to the formulation in section 2.2.2 (equation 2.62) for each model trained during this process. “Response” maps representing the predicted category on a voxel-by-voxel basis were generated for individual category probabilities. To avoid confusion in the interpretation of these maps, each map was masked by the original consensus ROIs.

Initial results of the ordinal regression (from both qualitative and quantitative analysis) suggested that the use of three categories (PR, SD, and PD) overextended the capabilities of this particular data set. Training data from the consensus ROIs exhibited similar behavior, except for the poor accuracies were present in prediction of the partial response category. As a result, the PR and SD categories were combined, and a logistic model was also considered with the two categories PR/SD and PD, similar to the approach taken in the development of the functional diffusion map by Moffat *et al.* [34]. Analyses were performed in the same manner as described for the ordinal regression, except there was no need to estimate category cut-offs since logistic models contain only a single fixed cut-off that is estimated as the regression intercept. In addition to poor prediction capabilities, initial qualitative investigations also revealed a non-negligible portion of some ROIs overlapped with the fluid-filled surgical category. A maximum *ADC* threshold was determined, as defined in **Table 5-2**, to provide a means of rejecting these cavities from the regression model.

5.4 Results

5.4.1 Data Preparation

All image series were successfully co-registered, as verified by the radiation oncologists, according to the prescription described in section 5.3.1. Inter-series and inter-study image transformations were applied to a single series of the DCE, DSC, and DTI images sets, in addition to transforming the respective parametric maps, and were subsequently stored in a separate QUATTRO ROI save file. The single series images extracted from each of the DCE, DSC, and DTI series were used only for quick reference when preparing the ordinal regression script. Response ROIs were then drawn by the two experience radiation oncologists, including the set of consensus ROIs. Finally, calculated intra-series image transformations were applied to the individual imaging series prior to modeling and

stored in an appropriate format for modeling (QUATTRO save for DCE and DSC data and DICOM format for DTI data).

Patient #	DCE VIF	DSC VIF
1	Superior Sagittal Sinus	Middle Cerebral Artery
2	Superior Sagittal Sinus	Middle Cerebral Artery
3	Superior Sagittal Sinus	Posterior Cerebral Artery*
4	Superior Sagittal Sinus	Middle Cerebral Artery*
5	Superior Sagittal Sinus	Posterior Cerebral Artery*
6	Inferior Sagittal Sinus	Middle Cerebral Artery
7	Superior Sagittal Sinus	Posterior Cerebral Artery*
8	Superior Sagittal Sinus	Posterior Cerebral Artery
9	Inferior Sagittal Sinus	Internal Carotid Artery
10	Middle Cerebral Artery	Posterior Cerebral Artery
11	Superior Sagittal Sinus	Middle Cerebral Artery

Table 5-3. DCE and DSC VIF locations. Artery/vein used in defining the vascular input functions used for quantitation of the DCE and DSC studies.

	Number of Voxels			
Patient #	PR	SD	PD	Totals
1	-	606	51	657
2	-	325	-	325
5*	-	-	2435	2435
6	-	2202	-	2202
7	-	148	352	500
8	-	5336	1695	7031
11	-	357	-	357
Placebo	-	8974	4533	13507
3	-	1229	3134	4363
4	1297	-	167	1464
9	-	21	1465	1486
10	-	821	569	1390
Bevacizumab	1297	2071	5335	8703

Table 5-4. Voxel counts for the consensus ROI definitions of both radiation oncologists. This table summarizes the voxel counts of all ROI data defined in consensus between the oncologists gathered from all patients and separated by patient, response category, and treatment type. QIB thresholds were applied prior to calculating these voxel counts. *indicates that the true treatment was unknown.

VIFs and the corresponding number of pre-enhancement frames were determined for the DCE studies; the former are summarized in **Table 5-3**. Following VIF definition, all DCE parameter maps (K^{trans} , v_e , and v_p) and the associated coefficient of determination map were calculated and transformed using the inter-series registration results. Resultant QUATTRO workspace data were then saved in a separate file for ease of access. Results of the DSC analysis, namely VIF locations are also summarized in **Table 5-3**. VIF gamma-variate fits were verified visually and corresponding R^2 values (>0.919 , except one VIF with $R^2=0.82$) were computed and inspected to ensure reasonable VIF fits. Calculated $rCBV$ and R^2 maps were successfully transformed and saved as QUATTRO files, in the same manner as the DCE data.

Similarly, DTI maps (*ADC* and *FA*) were computed successfully on a GE workstation, imported into the corresponding DTI QUATTRO exam file, transformed, and stored.

5.4.2 Qualitative Inspection

Using the oncologist defined ROI data, voxels were extracted from each of the parametric map, in addition to tracking the response category and treatment type (placebo or bevacizumab). During this process, a non-negligible number of R^2 values, the criteria used for assessing GKM fits, were found to be below zero. Further investigation into the source of these erroneous values revealed voxels exhibiting minimal, if any, enhancement during the course of the DCE acquisition. Nominally, R^2 lies in the closed interval from zero to one, but the use of non-linear fitting in these particular voxels consisting of primarily noise resulted in the total sum of squares (SS_{tot}) falling below one (but greater than zero) and the residual sum of squares (SS_{res}) attaining a non-negligible value. Since the coefficient of determination is defined as one minus the ratio of SS_{res} to SS_{tot} , the value fell below zero. These voxels were excluded from further analysis. Resultant voxel counts after applying the parameter and R^2 thresholds for these consensus ROI definitions are summarized in **Table 5-4**. Immediately obvious from these tabulated data is the large variation in the number of voxels for each patient and the overall underrepresentation of the partial response category. In fact, the single collection of PR voxels renders prediction accuracy assessment using LOO cross-validation impossible, and provides a strong argument for performing logistic regression by combining the PR/SD into a single category. Another glaringly apparent characteristic is the large variability in the number of voxels between individual patients. This high degree of variation and skewedness in the voxel counts is a likely source of bias and variance in the regression models.

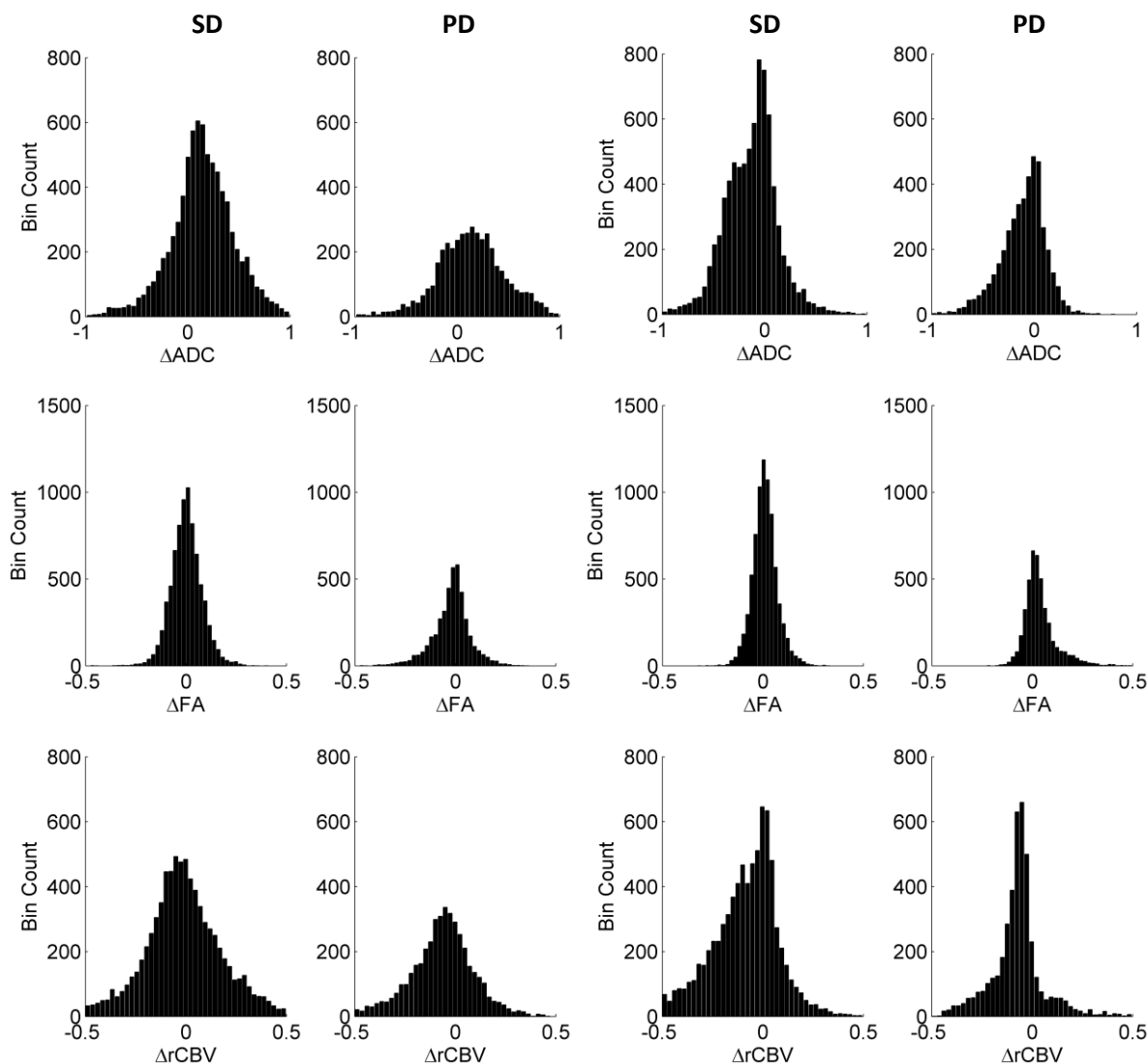


Figure 5-2. DTI and DSC quantitative imaging biomarker difference histograms for patients

receiving the placebo. Histograms displaying the distributions of QIB differences between baseline and the first mid-treatment study (*left*) and between the first and second mid-treatment studies (*right*). For a given parameter (label shown under each plot), the x and y limits are held constant, but vary between parameters.

Difference histograms were also generated based on data grouped by voxel response for the patients treated with the placebo. DSC and DTI QIB histograms for the placebo group are shown in **Figure 5-2**, while the parameter histograms for the DCE analysis are displayed in **Figure 5-3**. The

corresponding summary statistics (means and standard deviations) for each histogram were compiled and are displayed in **Table 5-5**. Visual inspection of the histograms in **Figure 5-2** and **Figure 5-3** provides some potential insight into longitudinal changes in these QIB. For example, between the first and second mid-treatment imaging studies, the distribution of *rCBV* in the SD category became skewed towards decreasing values, suggesting a net decrease in the *rCBV*, despite the lack of corroboration in the corresponding summary statistics, which conforms to the expected patterns of tissues response [86]. Similarly, expected response patterns are seen in the PD histograms, which in this case are in agreement with the summary statistics, as *FA* values appear to increase, consistent with the response pattern of progressing disease [43]. Despite these reassuring changes, other counter-intuitive trends were also present. For example, although ADC values initially increased after the initiation of treatment in SD voxels, a negative overall shift in ADC was observed between the two mid-treatment time points, suggesting potential cellular proliferation associated with progressive disease. Despite some potentially interesting visual trends, in general, the large spread of the distributions obscured any large scale trends in QIB changes, likely resulting from the varying magnitude of responses in the patient population.

		ΔADC	ΔFA	ΔK^{trans}	Δv_e	Δv_p	$\Delta rCBV$
Placebo	PR	-	-	-	-	-	-
	SD	0.15(0.31)	0.003(0.081)	0.027(0.160)	0.023(0.339)	-0.001(0.024)	-0.008(0.22)
	PD	0.13(0.33)	-0.018(0.096)	0.009(0.051)	-0.096 (0.42)	-0.002(0.013)	-0.081(0.19)
Bevacizumab	PR	-0.048(0.20)	0.014(0.079)	-0.006(0.030)	-0.085(0.36)	-0.004(0.014)	0.017(0.15)
	SD	-0.008(0.27)	0.017(0.065)	0.014(0.11)	-0.013(0.42)	0.004(0.018)	0.004(0.17)
	PD	0.023(0.19)	0.010(0.079)	0.033(0.17)	-0.055(0.40)	-0.002(0.013)	0.001(0.062)
		ΔADC	ΔFA	ΔK^{trans}	Δv_e	Δv_p	$\Delta rCBV$
Placebo	PR	-	-	-	-	-	-
	SD	-0.12(0.25)	0.008(0.062)	-0.014(0.10)	0.043(0.32)	-0.003(0.015)	-0.10(0.20)
	PD	-0.11(0.22)	0.038(0.081)	-0.005(0.043)	0.009(0.31)	-0.002(0.011)	-0.051(0.18)
Bevacizumab	PR	0.001(0.074)	0.007(0.045)	-0.013(0.022)	0.39(0.31)	-0.008(0.006)	-0.12(0.089)
	SD	-0.009(0.19)	0.003(0.054)	-0.015(0.11)	0.002(0.31)	-0.003(0.015)	-0.031(0.14)
	PD	-0.013(0.13)	-0.008(0.063)	0.017(0.11)	-0.033(0.26)	0.000(0.006)	0.02(0.072)

Table 5-5. Voxel summary statistics grouped by response category and treatment type. This table shows summary statistics, mean (standard deviation), for differences in QIBs between at the baseline/first mid-treatment (top) and the second/first mid-treatment (bottom) time points for voxels from all patients separated by response category and treatment type. All mean values have physical units corresponding to the parameter of interest.

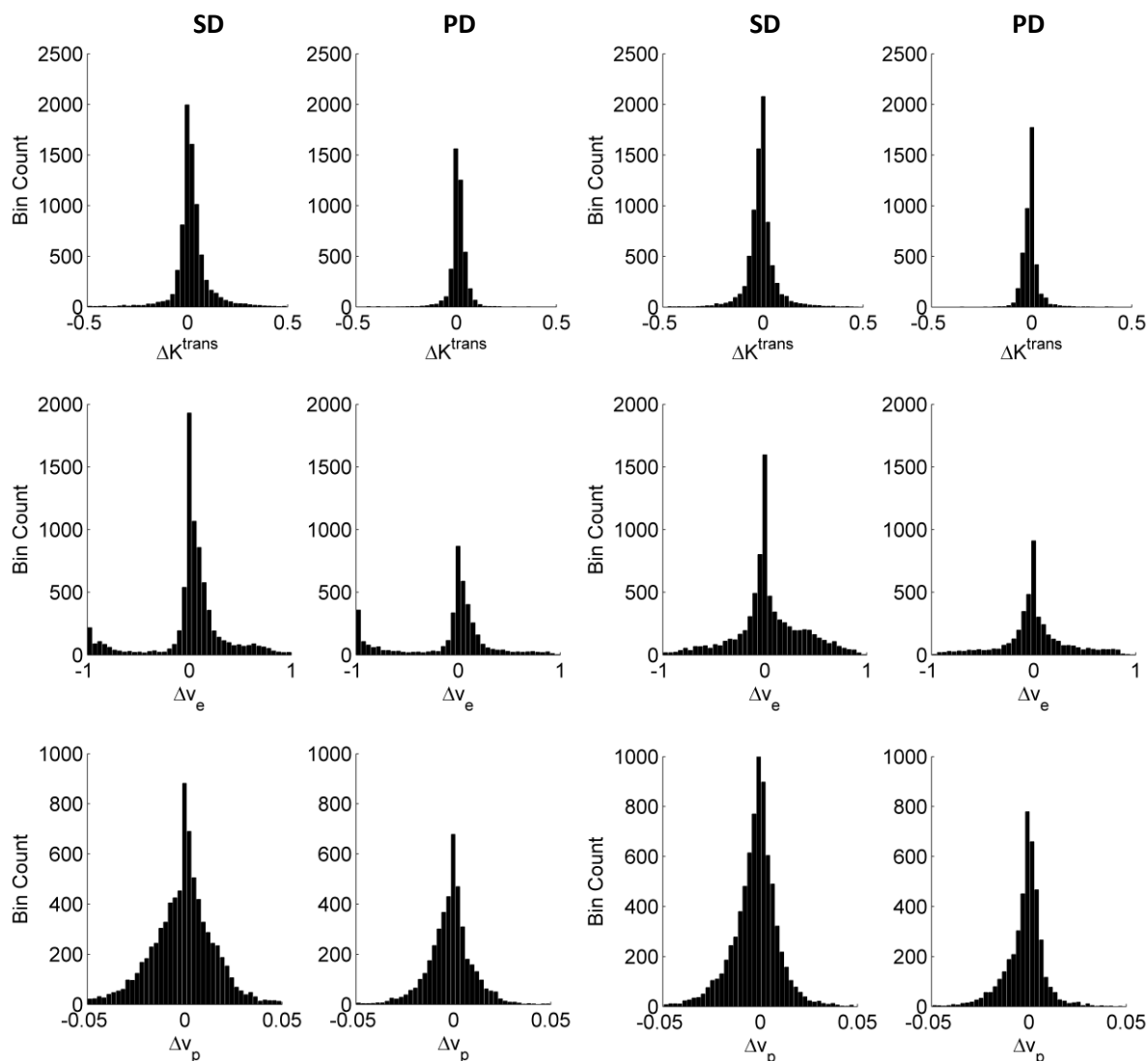


Figure 5-3. DCE quantitative imaging biomarker difference histograms for patients receiving the placebo. Histograms displaying the distributions of QIB differences between baseline and the first mid-treatment study (*left*) and between the first and second mid-treatment studies (*right*). For a given parameter (label shown under each plot), the x and y limits are held constant, but vary between parameters.

Surprisingly, the EES parameter, v_{e^*} , exhibited a large degree of dispersion and an apparent bimodal distribution for difference histograms created with baseline and the first mid-treatment voxel

data. Nominally, the fractional EES volume can assume values between zero and one, but little, if any, physical significance can be attributed to voxels attaining v_e values near one. Further investigation into the collection of voxels values near one revealed, not surprisingly, fitting errors. More specifically, in noisy voxels with relatively low enhancement, the term describing the plasma fraction (first term of equation 2.29) is sufficient to explain the majority of the tissue signal enhancement, resulting in an over-parameterized, unidentifiable model, arising from the correlation between the two fractional space terms [200]. These abnormalities are likely the result of poor permeability caused by only partial BBB disruption. In this regime, K^{trans} is usually forced to a relatively small value and v_e becomes arbitrarily large within the parameter bounds [0, 1]. More interesting than the poor fit is the dominance of this phenomenon in the baseline scan. Although similarly large v_e values were found in the two mid-treatment scans, the number of voxels found in these later scans was substantially smaller than the voxel counts of the baseline maps. Approximately 3,300 voxels (of over 22,000) were found to have v_e values greater than 0.9, with only 292 and 579 found in the first and second mid-treatment data sets, respectively. One possible cause is increased permeability in the BBB resulting from treatment effects.

In addition to creating difference histograms for patients treated with the placebo, an analogous similar set of histograms was also generated for patients treated with bevacizumab. Selected QIBs ($rCBV$, v_e , and v_p) were chosen to qualitatively illustrate trends as seen in **Figure 5-4**, though the other quantities were investigated. As with the previous voxel data, summary statistics can be found in **Table 5-5**. The $rCBV$ exhibited trends consistent with expected biological change, especially in describing differences in the first and second difference histogram. Specifically, between the two mid-treatment studies, the $rCBV$ distribution in PR and SD categories changed from a widely spread set of values to a more compact, negative mean distribution, while voxels in the PD category exhibited a generally positive shift, which might be understood intuitively as increasing vascularization. These observations were corroborated by corresponding changes in v_p . For v_e , large changes were observed from a well-

defined collection of values distributed primarily around zero to a widely spread distribution, suggesting an increase in the EES space associated with positive treatment response. Little change was observed in the v_e values for voxels categorized as SD and PD. Other parameters such as ADC and FA exhibited qualitative changes similar to those described for the voxels extracted from the QIB maps of patients treated without bevacizumab.

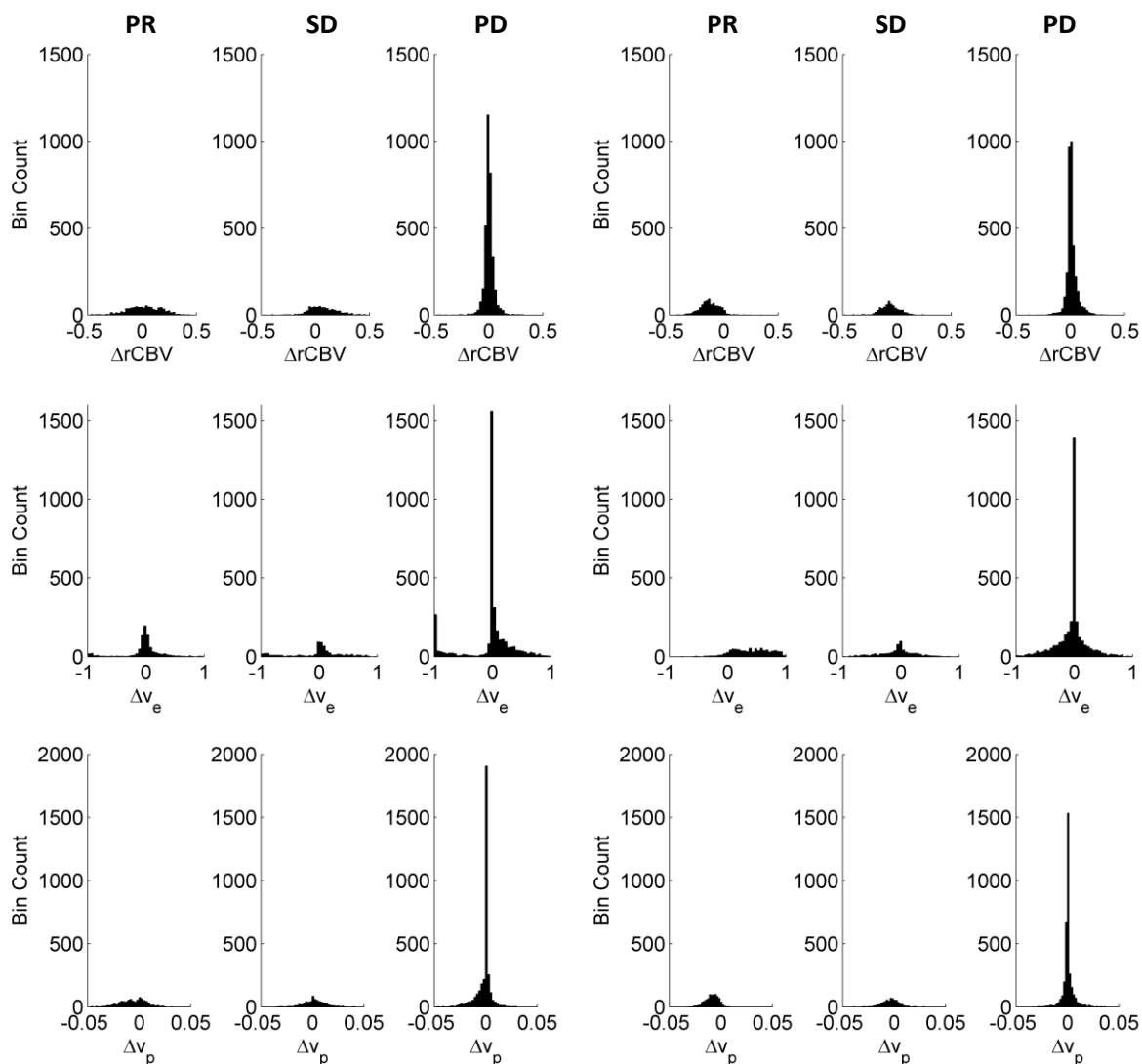


Figure 5-4. Selected DCE and DSC quantitative imaging biomarker difference histograms for patients treated with bevacizumab. Histograms displaying the distributions of QIB differences between baseline and the first mid-treatment study (*left*) and between the first and second mid-treatment studies (*right*). For a given parameter (label shown under each plot), the x and y limits are held constant, but vary between parameters.

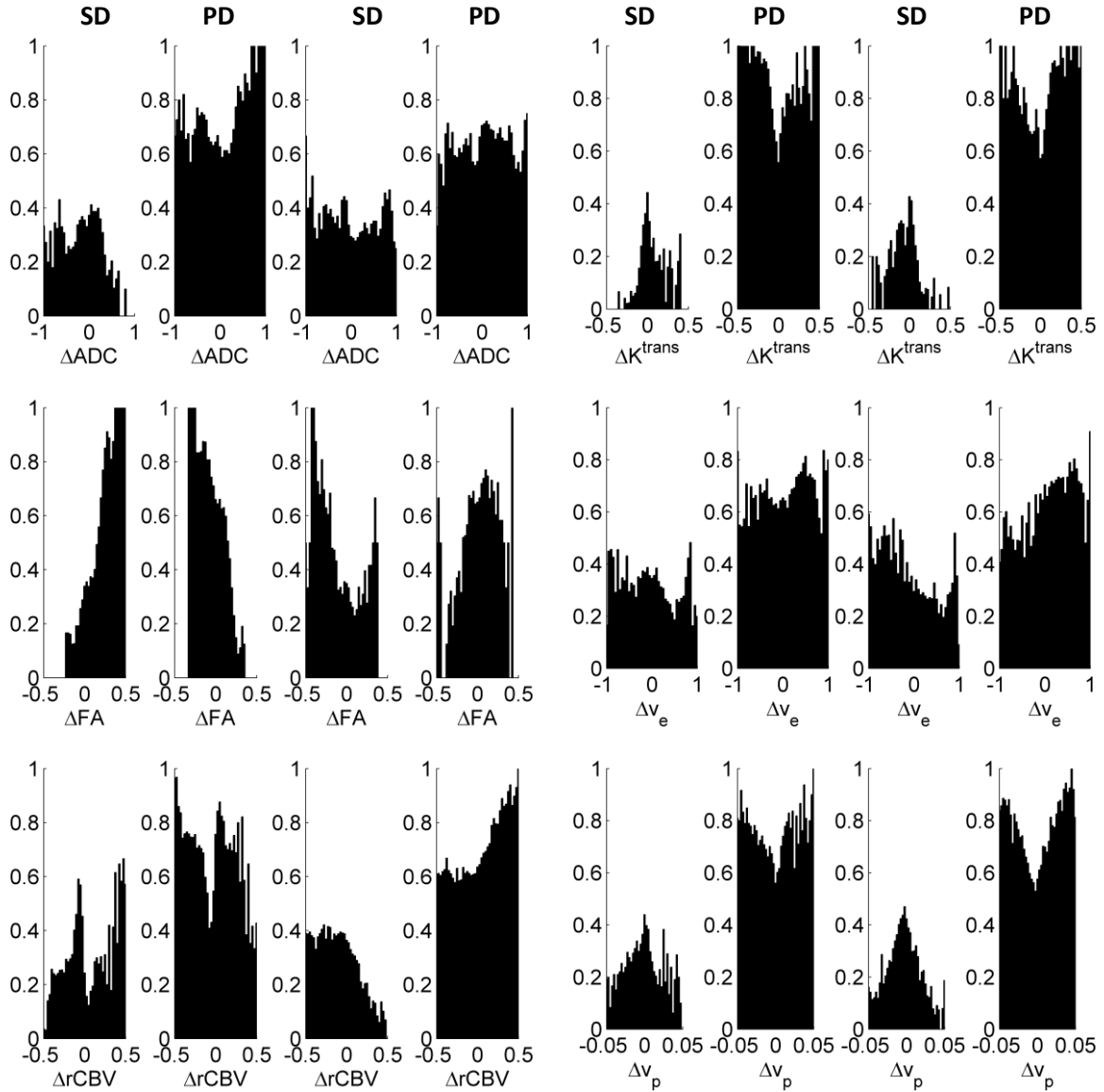


Figure 5-5. Probability histograms for patients receiving the placebo. Histograms displaying the probability of being a specific category, either SD or PD, are shown for DTI and DSC parameters (*left*) and DCE parameters (*right*). Each half of this figure (left/right) represents the data from baseline/first mid-treatment differences (*left*) and differences between the two mid-treatment data sets (*right*).

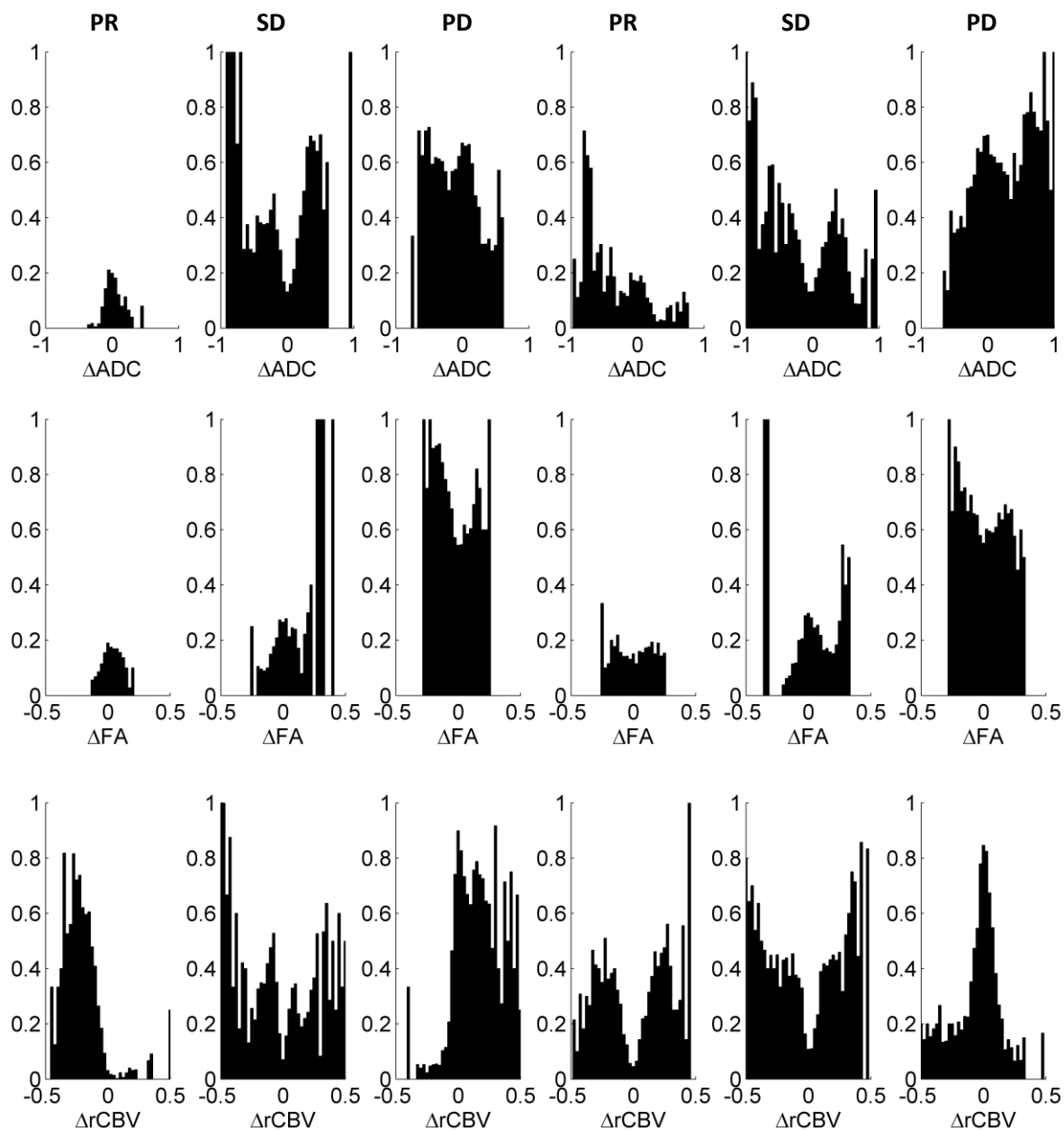


Figure 5-6. DTI and DSC QIB probability histograms for voxels from patients receiving bevacizumab. Histograms displaying the distributions of DTI and DSC derived QIB differences between baseline and the first mid-treatment study (*left*) and between the first and second mid-treatment studies (*right*)

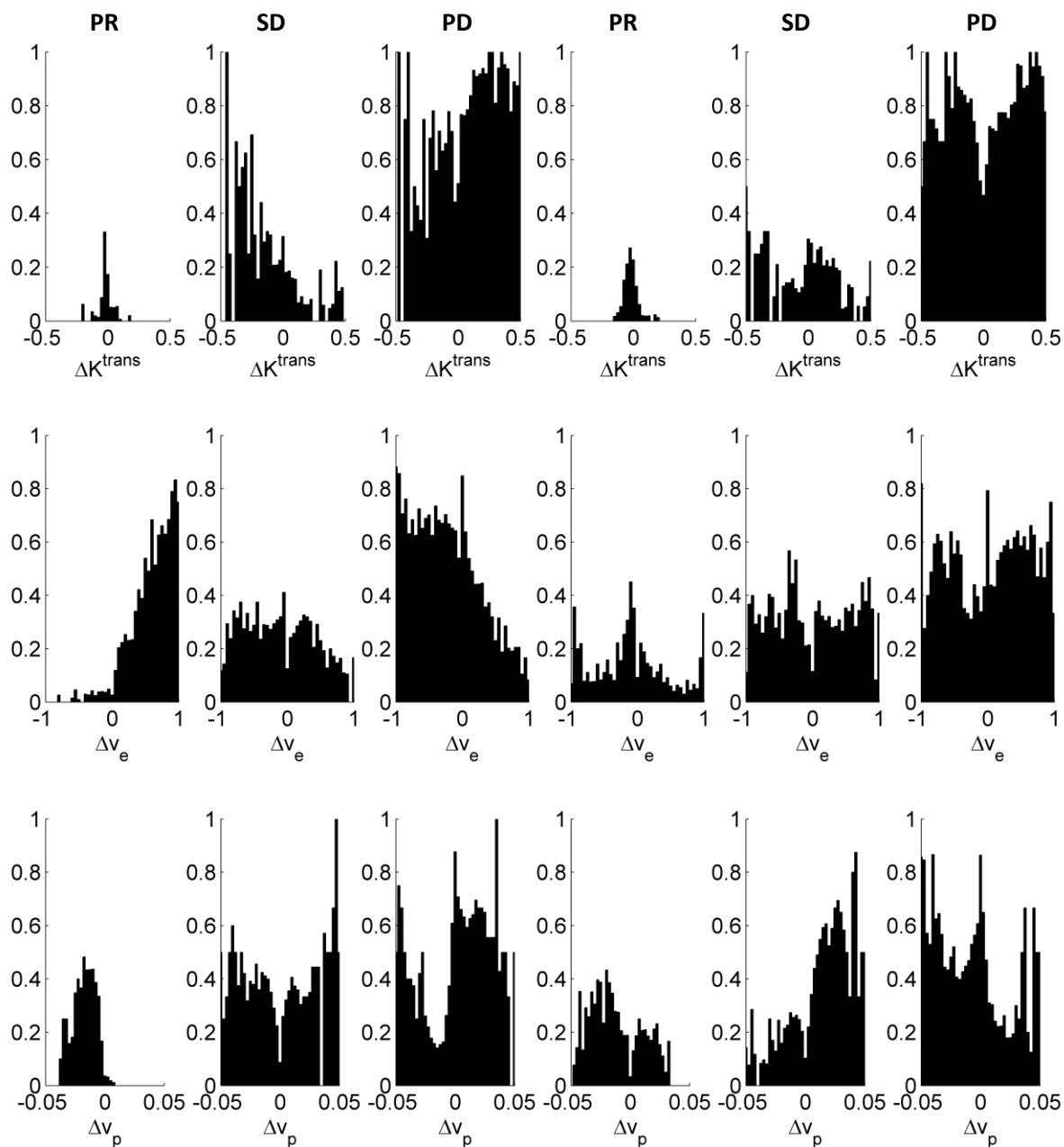


Figure 5-7. DCE QIB probability histograms for voxels from patients receiving bevacizumab.

Histograms displaying the distributions of DTI and DSC derived QIB differences between baseline and the first mid-treatment study (*left*) and between the first and second mid-treatment studies (*right*)

Probability histograms are another visualization that are perhaps more instructive for assessing data patterns in the context of ordinal regression, and consist of the probability of an observation plotted as a function of the individual covariates. Because of the manner in which ordinal regression is performed, namely that the probability is the dependent variable, these plots can assist in elucidating the functional form of the covariates (e.g., linear or quadratic), if any exists. Visualization of these histograms for voxel data from patients receiving the placebo is provided in **Figure 5-5**. The *FA* baseline/mid-treatment difference histograms show an example of data that would be modeled appropriately with a linear covariate. In this case, increasing differences in *FA* are associated with an increased probability of exhibiting SD, and, conversely, decreasing *FA* values appear to be an indicator of increasing chances of a voxel belonging to the PD category. K^{trans} and v_p both exhibit interesting trends in that the probability for a voxel to be in the PD category increase as the differences in the magnitude of those parameters becomes different from zero. A second order term would be a likely candidate for modeling trends of this nature. Similar plots are provided for voxels extracted from patients undergoing bevacizumab treatment, and are shown in **Figure 5-6** and **Figure 5-7**. Although these data are substantially noisier compared with patients on the other treatment arm, some potential trends still exist. For example, changes in *rCBV* between the baseline and first mid-treatment scan seem to be a good discriminator between the PR and SD categories, which would likely be most optimal if modeled with an indicator variable such as

$$Ind[\Delta rCBV_1 < 0] = \begin{cases} 0 & \text{if } \Delta rCBV_1 < 0 \\ 1 & \text{if } \Delta rCBV_1 \geq 0 \end{cases} \quad 5.3$$

as opposed to a continuous variable. The *rCBV* difference between the two mid-treatment data sets might be modeled well as the absolute value of the *rCBV*. Another potential linear trend can be seen in v_e differences from baseline, but such a relationship would provide little information about the SD.

This qualitative investigation of the data provided substantial incentive to reduce the proposed ordinal classifier (categories: PR, SD, and PD) to a simple logistic regression model by combining the PR and SD classes. With this limited data set, a full range of responses was not exhibited, especially apparent in the lack of patients exhibiting a partial response. In fact, using an ordinal model with all voxels in the PR category extracted from a single patient precludes accuracy assessment of that particular category using LOO cross-validation. In addition, there was little evidence in the probability histograms that voxels categorized as SD were distinguishable from PR categories. In the clinical setting, prediction of progressive disease carries substantially more importance than either stable disease or partial response, as knowledge of progression could potentially inform adaptation treatment regimens. For these reasons, logistic models were also used to build a classifier for this cohort of patients.

5.4.3 Statistical Classification and Response Maps

Initially, the model proposed in equation 5.1, and all unique combinations of the QIB

Model Parameters	Stable Disease	Progressive Disease	Overall	R^2
ADC	72.9(13.0)	25.3(14.4)	46.8(11.1)	0.040(0.010)
FA	67.3(13.3)	40.0(15.4)	51.3(10.7)	0.060(0.012)
K^{trans}	69.7(12.7)	30.9(16.2)	48.8(10.6)	0.046(0.009)
v_e	61.9(8.7)	34.3(13.0)	48.2(8.4)	0.12(0.007)
v_p	70.6(11.4)	25.5(12.8)	50.8(10.1)	0.054(0.008)
$rCBV$	52.2(10.2)	43.2(13.8)	45.7(8.5)	0.090(0.010)
$v_e/rCBV$	46.4(6.7)	50.2(13.2)	41.0(5.9)	0.17(0.009)
FA/v_p	69.0(11.1)	35.2(11.7)	52.3(9.7)	0.08(0.010)

Table 5-6. Ordinal classifier leave-one-out accuracies for linear, first-order interaction, and

squared terms. This table summarizes the prediction accuracies for the multi-parameter models exhibiting the highest categorical and overall accuracies. Table cells with a black background and white text represent the highest accuracy for the respective category, identifying both single and multi-parameter maximum accuracies.

parameters, was fitted to the ordinal regression model. Accuracies for these models for all combinations of QIBs and LOO patient data sets were computed, noting that patient 4 was excluded from the ordinal LOO cross-validation because creation of a training data set incorporating the PR category required the inclusion of voxel data from this patient. Results from multi-parameter ordinal models with the highest categorical and overall prediction accuracies, including model accuracies for the corresponding single parameter models, are summarized in **Table 5-6**. For the single parameter models, overall accuracies were approximately 50%, with SD accuracies approaching 70% (except for $rCBV$) and PD accuracies on the order of 30% ($rCBV$ is again an exception). Single parameter models incorporating ADC , FA , K^{trans} , and v_p were comparable in SD accuracies, but the corresponding PD accuracies were abysmal. Conversely, $rCBV$ seemed to perform best as a predictor for the PD category with accuracies near 50%. R^2 values (<0.116) suggest that fitted single parameter models were inadequate in improving the fitted models compared with the null model. Moreover, incorporating multiple QIB parameters did little to improve either the fits or prediction accuracies with the slight increases in R^2 values likely attributed to increasing the number of model parameters. Although omitted from these tables, the highest R^2 value (0.250) was found to correspond to the model including all QIBs, an unsurprising result consistent with R^2 inflation.

Corresponding logistic models were also fit in the same manner as these ordinal models by combining the PR and SD categories (*i.e.*, no progression, abbreviated as NP) as was motivated by the observations discussed in the preceding subsection (5.4.3). Accuracies for all single parameter models and multiple parameter models exhibiting the highest classification prediction accuracies in each of the two response categories and overall accuracies are tabulated in **Table 5-7**. Similarly to the ordinal classifier, minimal variation was observed across the categories of the single parameter models, with prediction accuracies for NP hovering around 60%, PD approximately 45%, and overall accuracies also around 60%. In all cases, R^2 seemed to indicate that by combining the PR and SD categories the resultant models better explained the data. Inflation of R^2 resulting from over-fitting the data was also observed

Model Parameters	No Progression	Progressive Disease	Overall	R^2
<i>ADC</i>	60.7(15.2)	49.3(17.4)	65.2(10.9)	0.11(0.18)
<i>FA</i>	65.8(13.3)	44.3(14.8)	61.2(10.5)	0.13(0.020)
k^{trans}	69.3(12.3)	34.9(14.2)	56.4(10.1)	0.12(0.016)
v_e	61.5(11.3)	47.3(15.7)	59.2(9.1)	0.15(0.014)
v_p	66.2(12.3)	46.8(16.2)	62.7(10.3)	0.13(0.016)
<i>rCBV</i>	57.8(12.2)	47.4(16.3)	57.4(9.5)	0.17(0.018)
<i>ADC/v_p</i>	64.0(12.1)	46.4(16.2)	60.8(10.0)	0.14(0.016)
<i>ADC/k^{trans}</i>	68.7(11.7)	35.4(14.1)	55.5(9.9)	0.12(0.017)
<i>FA/v_e/v_p/rCBV</i>	58.8(8.4)	53.8(12.7)	53.9(7.9)	0.28(0.016)

Table 5-7. Logistic classifier leave-one-out accuracies and R^2 values for linear, first-order

interaction, and squared terms. This table summarizes the prediction accuracies (standard error) for the multi-parameter models exhibiting the highest categorical and overall accuracies. Table cells with a black background and white text represent the highest accuracy for the respective category, identifying both single and multi-parameter maximum accuracies.

for the logistic regression models where, again, the six QIB model produced the highest R^2 (0.314) averaged over all LOO cases.

While the average cross-validation accuracies for the SD or, in the case of the logistic models, NP categories seemed promising using a relatively straightforward regression model, those same accuracies for the PD category appeared disproportionately low. Investigation of individual LOO cases revealed that accuracies for patient 5, who exhibited only PD, were on the order of 1-2% for logistic and ordinal, and, interestingly, this is also the patient for whom the exact treatment details were unknown. Incorrect treatment assignment is one potential cause for these low prediction accuracies. Investigation of the training accuracies, the accuracy with which the model predicts the correct category for the data used to generate the model, showed some of the highest overall accuracies, suggesting that patient 5 QIBs were potentially deleteriously impacting the fitted model. As a quick test, the cross-validation was re-run for the same model, excluding patient 5 from consideration. The results (not shown here) for the ordinal

model showed negligible (a few percent) improvements in the overall and SD prediction accuracies, but decreases in PD accuracies by a factor of two in most cases. Curiously, the R^2 values, on average, increased by a factor of nearly 2 for the models listed in **Table 5-6**. Similar observations were noted in the logistic model when performing LOO cross-validation without using patient 5 data. A similar occurrence of highly skewed accuracies occurs when patient 10 is the LOO case. However, under these circumstances, the SD (or logistic equivalent, NP) accuracies exhibit abnormally low prediction accuracies (<30%) that depend heavily on the QIBs used in the model. Other patient data sets (patients 8 and 9), when removed as an LOO case, exhibited low PD prediction accuracies on the order of 10% for the logistic models, but not the ordinal models. For each of the overall accuracies listed in these tables, there was a wide range of accuracies for the associated categories.

As discussed in section 5.4.2, the probability of classifying a given voxel in a specific category as a function of the $rCBV$ differences between baseline and the first mid-treatment imaging study exhibited behavior more suggestive of a binary response as opposed to a continuous variable. To investigate this potential improvement in both the ordinal and logistic regression models, LOO was re-run replacing all instances of $rCBV_1$ in equation 5.1 with an indicator variable; the squared term was excluded from the model. Resultant ordinal and logistic prediction accuracies exhibited negligible change over the original model.

Although the prediction accuracies of the models developed in this specific aim were less than ideal, illustration of the resultant maps for certain combinations of parameters suggests that these models can potentially detect heterogeneous response. To that end, several axial slices from the imaging data of patient 8, which included categorization of SD and PD voxels on the same slice, provided an excellent qualitative example of response heterogeneity. Individual maps were generated for each of the categories with the values of these color maps representing the probability of a model specific vector of covariates belonging to the category specified by the map. **Figure 5-8** and **Figure 5-9** show an

example slice exhibiting SD (left half of the colored region) and PD (right half of the colored region) as defined by the oncologist for the ordinal and logistic models, respectively, containing the QIBs *ADC*, *FA*, and *rCBV*. For the ordinal model, a large “hotspot” was seen on the PR probability map within the area categorized as SD by the oncologists, indicating, according to the model, an area that will likely exhibit PR. Similar hotspots were seen in areas of PD on the PD probability map, suggesting correct categorization. Of the three maps, the SD probability map exhibits a more uniform pattern, categorizing voxels from both the SD/PD regions as SD. For the logistic model, the combined PR/SD map shows a relatively large region of high probability extending well into the PD region defined by the oncologists, consistent with the combined probability maps of the ordinal model. Hotspots similar to those exhibited by the ordinal models are also seen in the PD probabilities of the logistic PD map. For comparison, the models producing the best multi-parameter overall accuracies are shown in **Figure 5-10** and **Figure 5-11** (best ordinal model) and **Figure 5-12** and **Figure 5-13** (best logistic model). These maps exhibit more uniform probabilities compared with the maps of **Figure 5-8** and **Figure 5-9**.

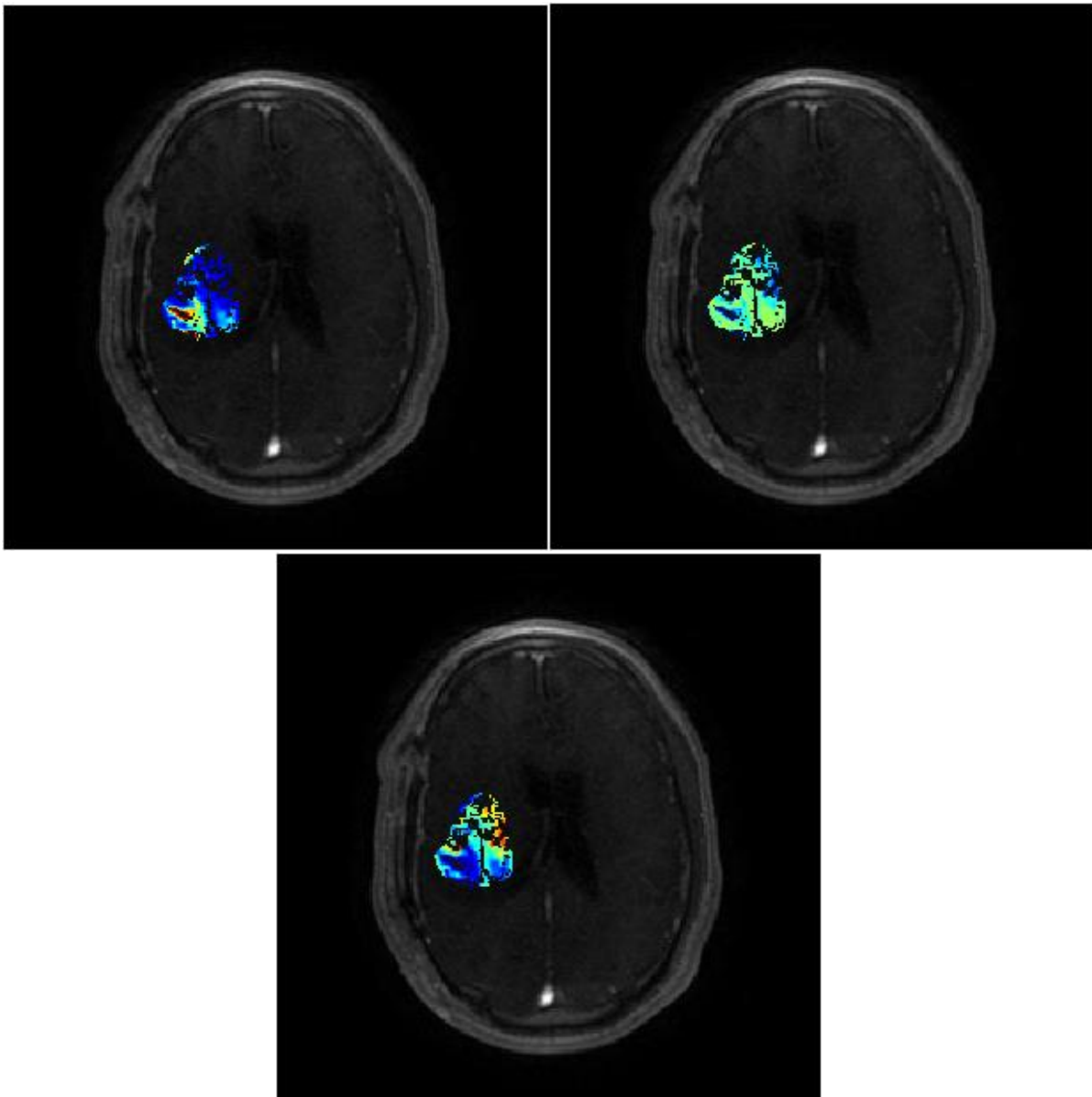


Figure 5-8. Example ordinal regression response category scores. Response category scores for PR (*top left*), SD (*top right*), and PD (*bottom*) for the ordinal model incorporating *ADC*, *FA*, and *rCBV*. The left half of the region displayed in color was categorized as SD and the other half as PD.

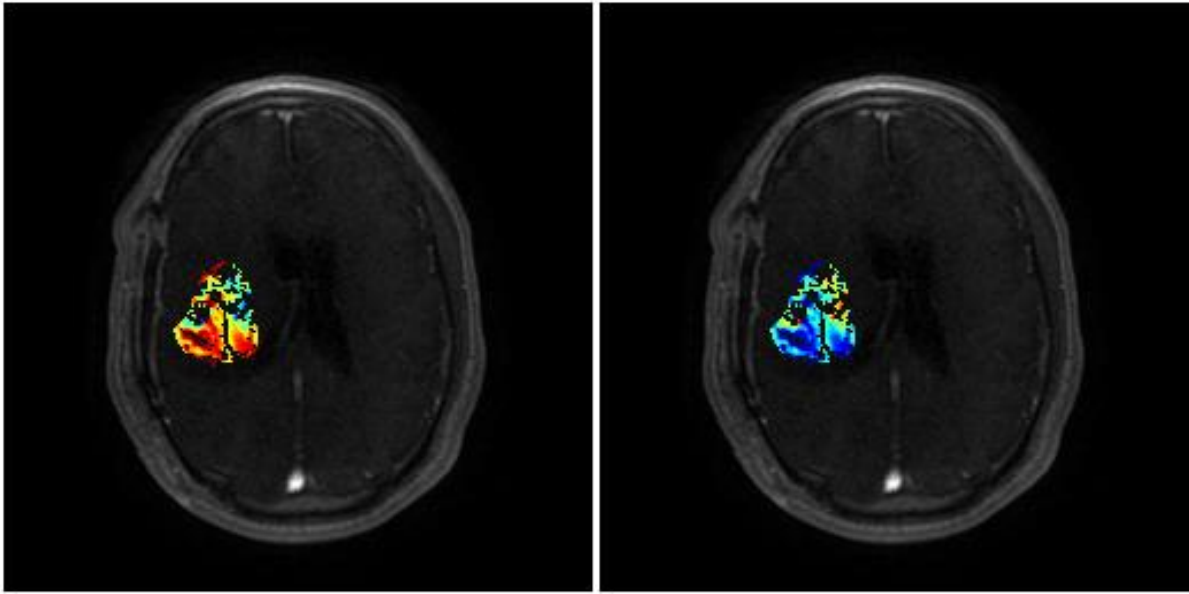


Figure 5-9. Example logistic regression response category scores. Response category scores for PR/SD (*left*) and PD (*right*) for the logistic model incorporating *ADC*, *FA*, and *rCBV*. The left half of the region displayed in color was categorized as SD and the other half as PD.

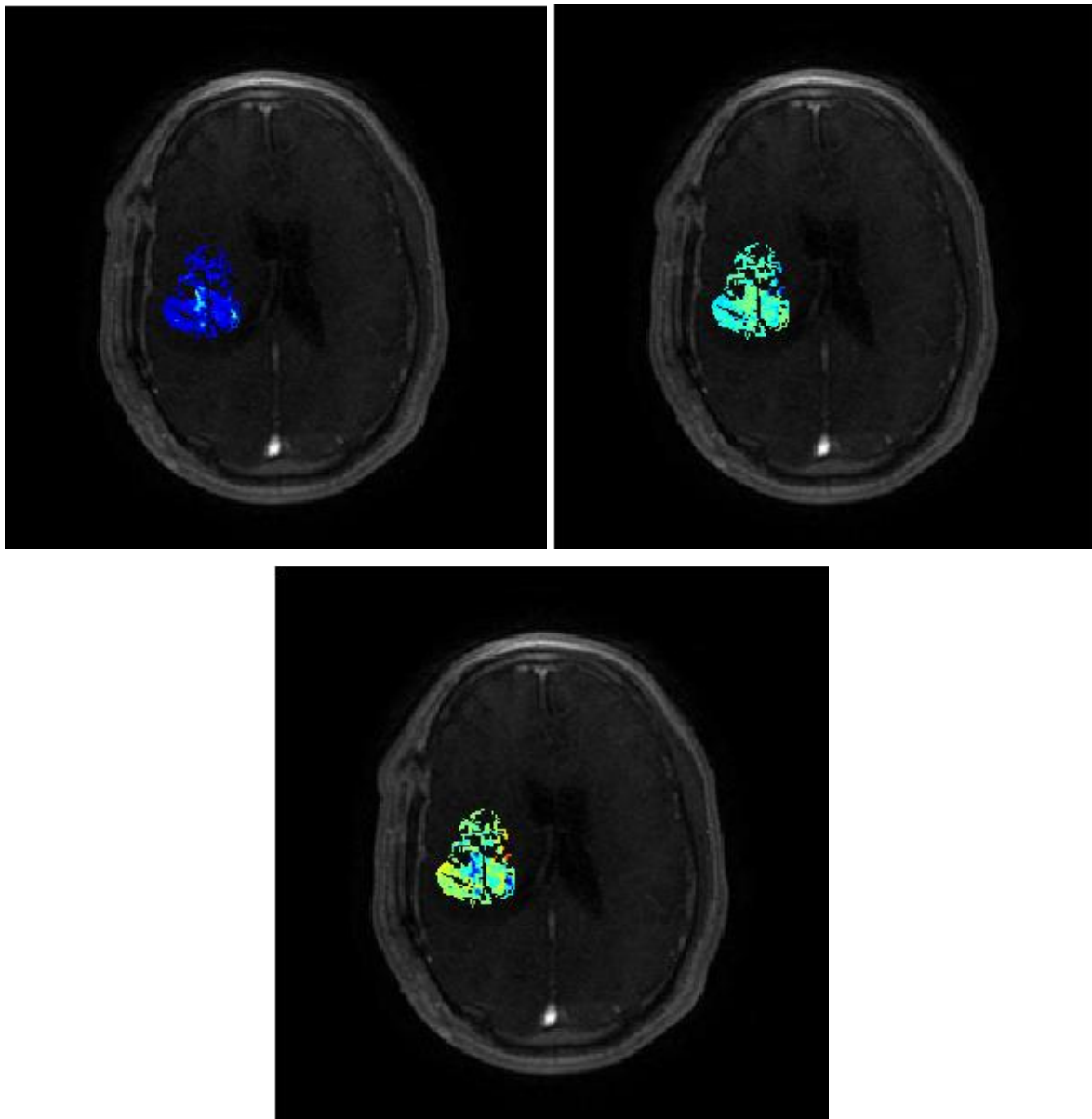


Figure 5-10. Example ordinal regression response category scores. Response category scores for PR (*top left*), SD (*top right*), and PD (*bottom*) for the ordinal model incorporating FA and v_p . The left half of the region displayed in color was categorized as SD and the other half as PD.

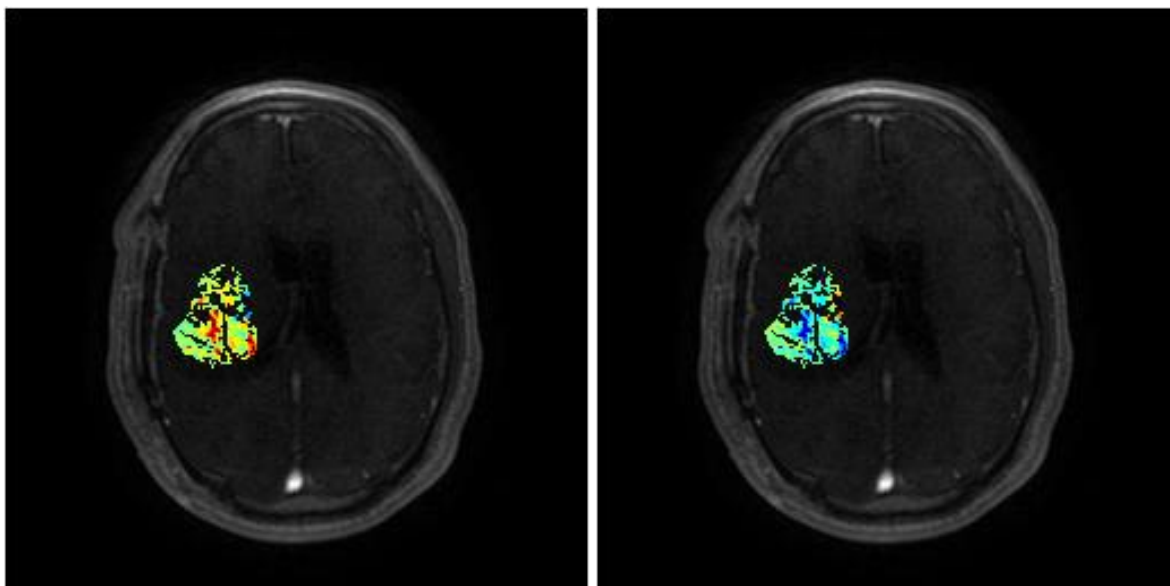


Figure 5-11. Example logistic regression response category scores. Response category scores for PR/SD (*left*) and PD (*right*) for the logistic model incorporating FA and v_p . The left half of the region displayed in color was categorized as SD and the other half as PD.

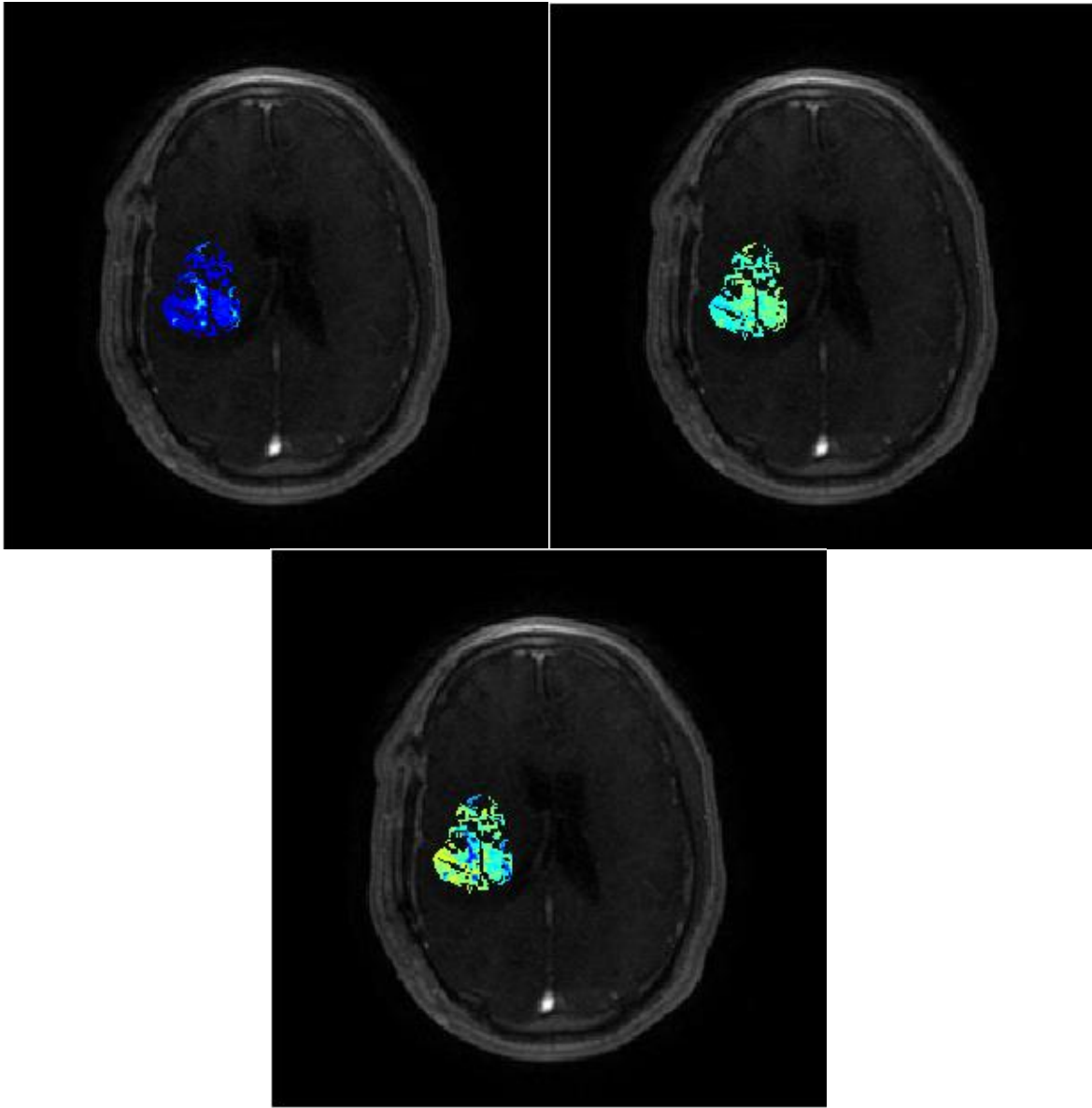


Figure 5-12. Example ordinal regression response category scores. Response category scores for PR/SD (*left*) and PD (*right*) for the logistic model incorporating ADC and v_p . The left half of the region displayed in color was categorized as SD and the other half as PD.

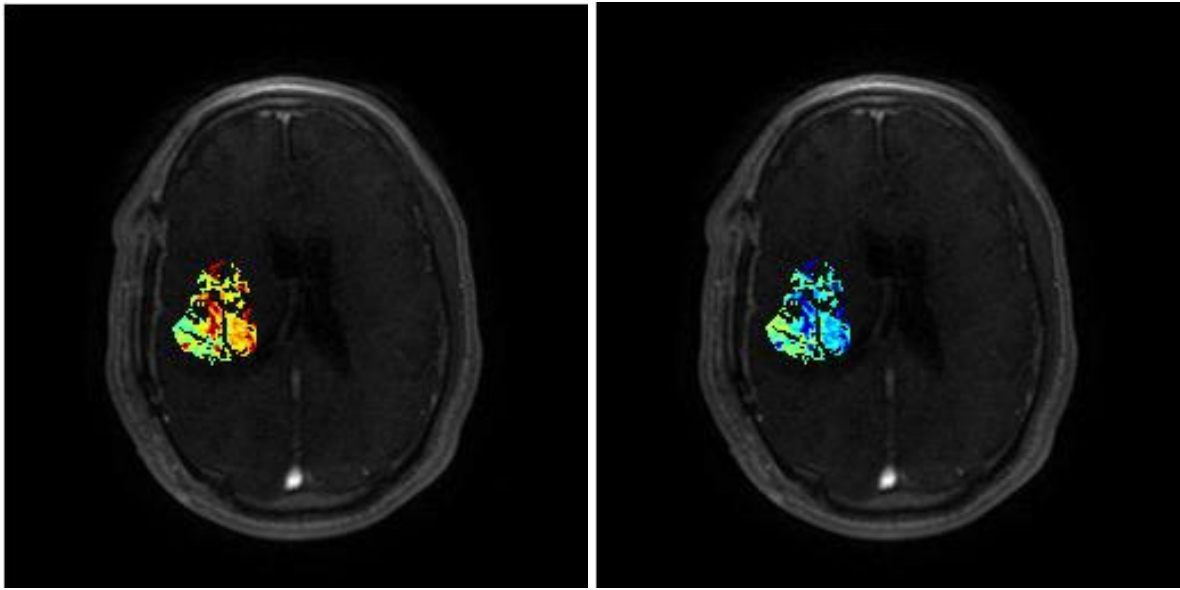


Figure 5-13. Example logistic regression response category scores. Response category scores for PR/SD (*left*) and PD (*right*) for the logistic model incorporating ADC and v_p . The left half of the region displayed in color was categorized as SD and the other half as PD.

5.5 Discussion

Incorporating multiple quantitative imaging biomarkers into a framework that allows predictions of treatment response based on radiological response is a promising avenue for the early detection of treatment response. More importantly, the development and validation of methods for developing multi-parametric analysis techniques which provide means of preserving the inherent spatial heterogeneity of underlying disease processes is paramount. The work performed in this specific aim utilized the software developed in specific aim 2, with confidence that variance introduced by the analysis tools was below that expected in the QIB treatment related changes, to show that ordinal regression models could be used to incorporate multiple parameters resulting in maps that provided predictions of treatment response based on radiological response. Although the accuracy of these

techniques, estimated in a severely limited data set, was relatively low, the framework has been shown to generate spatial maps encoding treatment response information.

Perhaps the biggest limitation of this current study, although there are several, is in the very limited nature of the data sets. Despite the large numbers of voxels that were available for training the classification models, only 9 patient data sets were suitable for use in generating these ordinal regression models (10 for the logistic models). This limited number of patients likely resulted in a highly imprecise and inaccurate estimate of the generalization accuracy. In addition to the limited number of patients, the limited number of voxels surrounding the surgical cavity that represented original disease was also very limited. Voxels beyond this region are more difficult to model with DCE PK models because of the intact, or partially so, BBB, effectively introducing noise into the models (as was shown in the model validation of specific aim 2). A problem that was further compounded by the low resolution and poor fits exhibited in the analysis of the DSC data.

Perhaps another limitation of the current study can be found in the definition of response. Despite the vast experience of the two expert radiation oncologists, misclassifications of voxels at this level would introduce large variances and potential biases in the trained models. Ideally, validation of the framework developed herein would require biological correlates such as histopathology samples, which could potentially be guided by these response maps. For example, patient 5 exhibited prediction accuracies on the order of 1-2% for a large number of voxels (~2500) if these voxels were assumed to correspond to PD. If in fact, these voxels represented should have been classified as SD, the model might be heavily biased toward classifying SD voxels as PD. Thus, the lack of “ground truth” data makes the clinical interpretation of these response maps difficult, if not impossible, but this framework, with further validation of QIBs and the models used to amalgamate those parameters may potentially introduce meaningful clinical information, possibly to the extent of informing the adaptation of ineffective treatments.

Chapter 6 Conclusions and Future Directions

As discussed at length in Chapter 1, quantitative MR imaging biomarkers are becoming increasingly incorporated into early phase clinical trials as a means of non-invasively assessing treatment response and the associated heterogeneity inherent in those responses. Substantial evidence exists to support the investigation of numerous promising applications for utilizing these QIBs as a means of assessing treatment response. Moreover, the increasing complexity of anticancer therapies and associated heterogeneity in tissue responses necessitates tools capable of detecting locoregional treatment associated changes. Incorporation of multiple complimentary QIBs can potentially elucidate these spatially heterogeneous treatment associated changes, potentially guiding clinical decisions, if incorporated into a classification model. The goal of this work was to address investigate ordinal regression as a classification method for incorporating multiple QIBs on a voxel-by-voxel basis.

6.1 Evaluation of the Hypothesis

The hypothesis of this work was that by incorporating multiple quantitative imaging biomarkers into an ordinal regression model, better prediction of locoregional radiological response in post-surgical glioblastoma patients treated with intensity-modulated radiation therapy and/or bevacizumab would result compared with the response predicted by a single quantitative imaging biomarker. Three specific aims were developed and completed to provide an assessment of the individual framework components (*i.e.*, equipment and analysis methods) and to generate response maps allowing evaluation of the central hypothesis.

Development of quality control phantoms and ongoing acquisitions to assess and mitigate equipment bias and variance was the focus of specific aim 1. Two phantoms were developed: a multiple

compartment phantom with a range of R_1 values for assessing DCE acquisitions and a multiple compartment phantom with multiple ADC values for assessing diffusion acquisitions (DWI and DTI). As discussed in Chapter 3, longitudinal variance, and even biases associated with measurements made away from isocenter, equipment level variance is well below the level of expected changes associated with *in vivo* QIB measurements. These results suggest that equipment related sources of bias and variance will minimally impact models developed to account for longitudinal changes *in vivo*.

Specific aim 2 focused on the development and validation of a comprehensive software package for both modeling quantitative MR acquisitions and incorporating the resultant QIBs into an ordinal regression analysis. The Quantitative Utility for Assessment of Treatment Response (QUATTRO) was developed, featuring advanced programming and graphical user interfaces. Object oriented programming principles were employed to handle the numerous bookkeeping tasks such as importing, storing, and manipulating images, ROIs, and treatment planning information with support for a number of common data storage formats (*e.g.*, DICOM 3.0, Pinnacle, and Analyze). Included in this software package are a collection of classes for performing modeling of common quantitative MR imaging acquisition strategies and specific interfaces for modeling longitudinal data in the context of treatment response. Furthermore, these quantitative modeling tools were validated with digital reference objects encompassing a range of model parameter space matched to expected *in vivo* values and over a range of acquisition parameters and noise levels. As with the results of specific aim 1, the software validation steps suggested levels of variance well below the anticipated changes resulting from treatment.

Work performed in specific aim 2 provided a means of programmatically calculating QIBs, co-registering resultant QIB maps, extracting QIB parameters from patient data sets on which response was defined, and generating ordinal regression models used for classification. These tasks (or more succinctly, the generation of response maps) were the goal of specific aim 3. By performing leave-one-out cross-validation, an estimate, albeit a very preliminary estimate, of the generalization error for the

trained ordinal (and logistic) classifiers was attained. Little, if any, distinction was found between those multi-parametric models with the highest accuracies and the corresponding single parameter models, let alone statistically significant differences.

Unfortunately, the proposed hypothesis was not supported by the data collected and analyzed in this work. In large part, this lack of support likely stems from the very limited data set that was retrospectively analyzed, with only 9 (10 for logistic models) patient data sets for developing and assessing the ordinal regression models. Of these data sets, an insufficient number of partial response data prevented assessing the classification accuracies for the PR category, and concentrated data distributions (*e.g.*, patient 8) as discussed in chapter 5 may have deleteriously impacted the training of regression models during the LOO process, resulting in an underestimation of the generalization prediction accuracies. Another potential confounding factor is the absence of ground truth data. Some of the voxels categorized as SD may, in actuality, be exhibiting partial response or progressive disease, and similarly for other categories. Limitations of the routine clinical imaging used to assess treatment response are likely central to these initial misclassifications, which would effectively obscure detectable trends in QIBs across the response categories.

Although the hypothesis was not confirmed, the work performed to address this hypothesis resulted in a number of substantive and novel contributions. Phantoms and associated acquisition protocols were developed for brain applications, demonstrating the ability to longitudinally monitor equipment related sources of bias and variance in relation to diffusion and DCE QIBs. A comprehensive software package, QUATTRO, was designed and implemented in the MATLAB programming environment providing all necessary tools for perform voxel-by-voxel assessments of response to anticancer therapies using multiple QIB. More generally, a methodology for incorporating multiple QIBs, not limited to quantitative MRI acquisitions, into an ordinal regression classifier was developed.

6.2 Future Directions

Although self-evident, the acquisition of quality control data that provides a means of assessing equipment related bias and variance should be ongoing. One notably flaw in the work performed in specific aim 1 is the lack of DSC acquisition assessments. Future studies should incorporate, in addition to the diffusion and DCE acquisitions, an additional assessment of DSC parameter reproducibility. Like the DCE QIBs, those derived from DSC data are incapable of being measured using a static phantom. However, by measuring the signal stability with the DCE phantom designed in this study using a T_2 - or T_2^* -weighted dynamic scan, simulations similar to the PK simulations could be performed to assess $rCBV$ reproducibility. In addition to incorporating an assessment of DSC derived QIBs, a more thorough assessment of external stimuli such as temperature could potentially provide a means of correcting quantitative measurements for experimental conditions to better assess equipment bias and variance. To that end, assessments of bias and variance might be potentially improved by better characterizing design parameters, allowing better estimates of expected measurement quantities. This is particularly true for the ADC phantom, as the ADC values used to design the phantom were crudely estimated from previously published data.

With regard to the phantom studies, another interesting avenue of investigation is to determine the frequency with which these acquisitions should be performed, as weekly phantom studies can become quite time consuming. Moreover, at institutions with other ongoing quality control procedures in place that occur with some frequency, for example the daily quality control that is performed M.D. Anderson Cancer Center, complimentary or surrogate information derived from these acquisitions might provide a means of reducing the frequency of quantitative acquisitions.

Because of the developmental nature of this project and the limited patient cohort used for testing, many of the tools developed herein represent relatively rudimentary techniques. For example, the image registration implemented in QUATTRO as a link through the ITK libraries performs

straightforward rigid registration. Under normal circumstances, the rigidity of the brain with respect to head movements (or in the case of longitudinal studies, patient setup inaccuracies) allows very reasonable registration of images using only rigid transformations. However, the cohort used in this consisted of post-surgical patients, in which small tissue deformations were easily observed. Moving forward, these small deformations might be accounted for through the implementation of a deformable image registration tool. For studies investigating extracranial applications, such deformable techniques become an absolute necessity.

Other tools, such as the quantitative image modeling, might benefit from more advanced developments. For example, the DCE module of QUATTRO, which uses a standard least squares fitting routine, performed well when applied to the pharmacokinetic DRO. However, as was observed in the analysis of patient data sets, a number of suspiciously large values for the parameter v_e were found. Several techniques that have shown promise in improving the fitting of DCE-MRI data are the Bayesian methods proposed by Schmid *et al.* [201], the use of Gaussian Process inference as proposed by Wang *et al.* [202], or the incorporation of a spatial model [203]. Interestingly, the work performed by Kelm and Menze [203] provided an example fit in which the model was also unidentifiable, the same issue that plagued the analysis of patient data. Although DCE data sets were used as an example in this discussion, these methods could presumably be used for any of the other supported quantitative image models.

Potential model improvements also extend to the classifier developed in the last specific aim of this work. Specifically, a prominent source of bias and variance exists in the uncertainties associated with expert readers. Initial development of the models presented in this work relied solely on consensus data defined between two expert observers. The responses defined in this manner were taken as the “true” response, but the introduction of a second observer, or multiple observers, for that matter, would very likely produce different results. In the context of ordinal models, a multi-rater model can be developed to estimate the variance associated with a group expert readers. While this is useful for

investigating variance associated with individual raters, perhaps more useful is the coupling of such an analysis with the ordinal regression model. By constructing the regression model in this manner, a comparison between the variance associated with individual raters and the fitted regression models can be made. In addition to this more appropriate model, the expected spatial correlations might also be modeled using a spatial model.

Validation of the framework and models developed herein would ideally involve acquisition of image guided multi-site biopsies and corresponding histology to confirm the predictions made by these models. By performing validation in this manner, a large source (expert reader) of variance, and possibly bias, may be avoided, potentially providing improved model SNR characteristics. In addition to biological correlates, further performance assessment of these response maps might involve prediction of overall patient survival (the gold standard for clinical trials).

References

- [1] P. Therasse, S. G. Arbuck, E. A. Eisenhauer, J. Wanders, R. S. Kaplan, L. Rubinstein, J. Verweij, M. Van Glabbeke, T. Van Oosterom, M. C. Christian, and S. G. Gwyther, "New Guidelines to Evaluate the Response to Treatment," *J. Natl. Cancer Inst.*, vol. 92, no. 3, pp. 205–216, 2000.
- [2] P. Y. Wen, D. R. Macdonald, D. a Reardon, T. F. Cloughesy, a G. Sorensen, E. Galanis, J. Degroot, W. Wick, M. R. Gilbert, A. B. Lassman, C. Tsien, T. Mikkelsen, E. T. Wong, M. C. Chamberlain, R. Stupp, K. R. Lamborn, M. a Vogelbaum, M. J. van den Bent, and S. M. Chang, "Updated response assessment criteria for high-grade gliomas: response assessment in neuro-oncology working group.," *J. Clin. Oncol.*, vol. 28, no. 11, pp. 1963–72, Apr. 2010.
- [3] J. E. Husband, L. H. Schwartz, J. Spencer, L. Ollivier, D. M. King, R. Johnson, and R. Reznick, "Evaluation of the response to treatment of solid tumours - a consensus statement of the International Cancer Imaging Society.," *Br. J. Cancer*, vol. 90, no. 12, pp. 2256–60, Jun. 2004.
- [4] C. C. Jaffe, "Measures of response: RECIST, WHO, and new alternatives.," *J. Clin. Oncol.*, vol. 24, no. 20, pp. 3245–51, Jul. 2006.
- [5] R. Lencioni and J. M. Llovet, "Modified RECIST (mRECIST) assessment for hepatocellular carcinoma.," *Semin. Liver Dis.*, vol. 30, no. 1, pp. 52–60, Feb. 2010.
- [6] B. Definitions and W. Group, "Biomarkers and surrogate endpoints: preferred definitions and conceptual framework.," *Clin. Pharmacol. Ther.*, vol. 69, no. 3, pp. 89–95, Mar. 2001.
- [7] A. Waldman, A. Jackson, and S. Price, "Quantitative imaging biomarkers in neuro-oncology," *Nat. Rev.*, vol. 6, no. 8, pp. 445–54, Aug. 2009.
- [8] S. Galbraith, "MR in oncology drug development," *NMR Biomed.*, vol. 19, no. 6, pp. 681–689, 2006.

- [9] A. R. Padhani and A. A. Khan, "Diffusion-weighted (DW) and dynamic contrast-enhanced (DCE) magnetic resonance imaging (MRI) for monitoring anticancer therapy.," *Target. Oncol.*, vol. 5, no. 1, pp. 39–52, Mar. 2010.
- [10] A. Jackson, D. Buckley, and G. Parker, *Dynamic Contrast-Enhanced Magnetic Resonance Imaging in Oncology*. Berlin/Heidelberg: Springer-Verlag, 2005.
- [11] D. M. Koh and H. C. Thoeny, Eds., *Diffusion-Weighted MR Imaging*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010.
- [12] P. Tofts, *Quantitative MRI of the Brain*. Chichester, UK: John Wiley & Sons, Ltd, 2003.
- [13] J. M. Provenzale, S. Mukundan, and D. P. Barboriak, "Diffusion-weighted and perfusion MR imaging for brain tumor characterization and assessment of treatment response.," *Radiology*, vol. 239, no. 3, pp. 632–49, Jun. 2006.
- [14] E. R. Gerstner and a G. Sorensen, "Diffusion and diffusion tensor imaging in brain cancer.," *Semin. Radiat. Oncol.*, vol. 21, no. 2, pp. 141–6, Apr. 2011.
- [15] S. Walker-Samuel, M. O. Leach, and D. J. Collins, "Evaluation of response to treatment using DCE-MRI: the relationship between initial area under the gadolinium curve (IAUGC) and quantitative pharmacokinetic analysis.," *Phys. Med. Biol.*, vol. 51, no. 14, pp. 3593–602, Jul. 2006.
- [16] D. Collins and A. Padhani, "Dynamic magnetic resonance imaging of tumor perfusion," *Eng. Med. Biol.*, vol. 23, no. 5, pp. 65–83, 2004.
- [17] R. M. Stephen and R. J. Gillies, "Promise and progress for functional and molecular imaging of response to targeted therapies.," *Pharm. Res.*, vol. 24, no. 6, pp. 1172–85, Jun. 2007.
- [18] D. M. Patterson, A. R. Padhani, and D. J. Collins, "Technology insight: water diffusion MRI--a potential new biomarker of response to cancer therapy.," *Nat. Clin. Pract. Oncol.*, vol. 5, no. 4, pp. 220–33, Apr. 2008.

- [19] A. Padhani and G. Liu, "Diffusion-weighted magnetic resonance imaging as a cancer biomarker: consensus and recommendations," *Neoplasia*, vol. 11, no. 2, pp. 102–125, 2009.
- [20] T. Sugahara, Y. Korogi, M. Kochi, I. Ikushima, Y. Shigematu, T. Hirai, T. Okuda, L. Liang, Y. Ge, Y. Komohara, Y. Ushio, and M. Takahashi, "Usefulness of diffusion-weighted MRI with echo-planar technique in the evaluation of cellularity in gliomas.," *J. Magn. Reson. imaging*, vol. 9, no. 1, pp. 53–60, Jan. 1999.
- [21] E. R. Gerstner, M. P. Frosch, and T. T. Batchelor, "Diffusion magnetic resonance imaging detects pathologically confirmed, nonenhancing tumor progression in a patient with recurrent glioblastoma receiving bevacizumab.," *J. Clin. Oncol.*, vol. 28, no. 6, pp. e91–3, Mar. 2010.
- [22] B. M. Ellingson, M. G. Malkin, S. D. Rand, J. M. Connelly, C. Quinsey, P. S. LaViolette, D. P. Bedekar, and K. M. Schmainda, "Validation of functional diffusion maps (fDMs) as a biomarker for human glioma cellularity.," *J. Magn. Reson. imaging*, vol. 31, no. 3, pp. 538–48, Mar. 2010.
- [23] Y. Hayashida, T. Hirai, S. Morishita, M. Kitajima, R. Murakami, Y. Korogi, K. Makino, H. Nakamura, I. Ikushima, M. Yamura, M. Kochi, J. Kuratsu, and Y. Yamashita, "Diffusion-weighted imaging of metastatic brain tumors: comparison with histologic type and tumor cellularity.," *Am. J. Neuroradiol.*, vol. 27, no. 7, pp. 1419–25, Aug. 2006.
- [24] G. Manenti, M. Di Roma, S. Mancino, D. a Bartolucci, G. Palmieri, R. Mastrangeli, R. Miano, E. Squillaci, and G. Simonetti, "Malignant renal neoplasms: correlation between ADC values and cellularity in diffusion weighted magnetic resonance imaging at 3 T.," *Radiol. Med.*, vol. 113, no. 2, pp. 199–213, Mar. 2008.
- [25] B. Zelhof, M. Pickles, G. Liney, P. Gibbs, G. Rodrigues, S. Kraus, and L. Turnbull, "Correlation of diffusion-weighted magnetic resonance data with cellularity in prostate cancer.," *BJU Int.*, vol. 103, no. 7, pp. 883–8, Apr. 2009.

- [26] M. I. Yoshikawa, S. Ohsumi, S. Sugata, M. Kataoka, S. Takashima, T. Mochizuki, H. Ikura, and Y. Imai, "Relation between cancer cellularity and apparent diffusion coefficient values using diffusion-weighted magnetic resonance imaging in breast cancer.," *Radiat. Med.*, vol. 26, no. 4, pp. 222–6, May 2008.
- [27] A. Goyal, R. Sharma, A. S. Bhalla, S. Gamanagatti, A. Seth, V. K. Iyer, and P. Das, "Diffusion-weighted MRI in renal cell carcinoma: a surrogate marker for predicting nuclear grade and histological subtype.," *Acta radiol.*, vol. 53, no. 3, pp. 349–58, Apr. 2012.
- [28] M. E. Loveless, D. Lawson, M. Collins, M. V. P. Nadella, C. Reimer, D. Huszar, J. Halliday, J. C. Waterton, J. C. Gore, and T. E. Yankeeelov, "Comparisons of the efficacy of a Jak1/2 inhibitor (AZD1480) with a VEGF signaling inhibitor (cediranib) and sham treatments in mouse tumors using DCE-MRI, DW-MRI, and histology.," *Neoplasia*, vol. 14, no. 1, pp. 54–64, Jan. 2012.
- [29] T. M. Zaw, W. B. Pope, T. F. Cloughesy, A. Lai, P. L. Nghiemphu, and B. M. Ellingson, "Short-interval estimation of proliferation rate using serial diffusion MRI predicts progression-free survival in newly diagnosed glioblastoma treated with radiochemotherapy.," *J. Neurooncol.*, vol. 116, no. 3, pp. 601–8, Feb. 2014.
- [30] N. Gahr, K. Darge, G. Hahn, B. W. Kreher, M. von Buiren, and M. Uhl, "Diffusion-weighted MRI for differentiation of neuroblastoma and ganglioneuroblastoma/ganglioneuroma.," *Eur. J. Radiol.*, vol. 79, no. 3, pp. 443–6, Sep. 2011.
- [31] M. P. Aryal, T. N. Nagaraja, K. a Keenan, H. Bagher-Ebadian, S. Panda, S. L. Brown, G. Cabral, J. D. Fenstermacher, and J. R. Ewing, "Dynamic contrast enhanced MRI parameters and tumor cellularity in a rat model of cerebral glioma at 7 T.," *Magn. Reson. Med.*, vol. 00, pp. 1–9, Jul. 2013.
- [32] N. Sadeghi, N. D’Haene, C. Decaestecker, M. Levivier, T. Metens, C. Maris, D. Wikler, D. Baleriaux, I. Salmon, and S. Goldman, "Apparent diffusion coefficient and cerebral blood volume in brain

- gliomas: relation to tumor cell density and tumor microvessel density based on stereotactic biopsies,” *Am. J. Neuroradiol.*, vol. 29, no. 3, pp. 476–82, Mar. 2008.
- [33] Y. Mardor, “Early Detection of Response to Radiation Therapy in Patients With Brain Malignancies Using Conventional and High b-Value Diffusion-Weighted Magnetic Resonance Imaging,” *J. Clin. Oncol.*, vol. 21, no. 6, pp. 1094–1100, Mar. 2003.
- [34] B. a Moffat, T. L. Chenevert, T. S. Lawrence, C. R. Meyer, T. D. Johnson, Q. Dong, C. Tsien, S. Mukherji, D. J. Quint, S. S. Gebarski, P. L. Robertson, L. R. Junck, A. Rehemtulla, and B. D. Ross, “Functional diffusion map: a noninvasive MRI biomarker for early stratification of clinical brain tumor response,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 15, pp. 5524–9, Apr. 2005.
- [35] T. Chenevert, P. McKeever, and B. Ross, “Monitoring early response of experimental brain tumors to therapy using diffusion magnetic resonance imaging,” *Clin. Cancer Res.*, vol. 3, no. 9, pp. 1457–1466, 1997.
- [36] D. a Hamstra, C. J. Galbán, C. R. Meyer, T. D. Johnson, P. C. Sundgren, C. Tsien, T. S. Lawrence, L. Junck, D. J. Ross, A. Rehemtulla, B. D. Ross, and T. L. Chenevert, “Functional diffusion map as an early imaging biomarker for high-grade glioma: correlation with conventional radiologic response and overall survival,” *J. Clin. Oncol.*, vol. 26, no. 20, pp. 3387–94, Jul. 2008.
- [37] S. Higano, X. Yun, T. Kumabe, M. Watanabe, S. Mugikura, A. Umetsu, A. Sato, T. Yamada, and S. Takahashi, “Malignant astrocytic tumors: clinical importance of apparent diffusion coefficient in prediction of grade and prognosis,” *Radiology*, vol. 241, no. 3, pp. 839–46, Dec. 2006.
- [38] M. I. Schubert, M. Wilke, S. Müller-Weihrich, and D. P. Auer, “Diffusion-weighted magnetic resonance imaging of treatment-associated changes in recurrent and residual medulloblastoma: preliminary observations in three children,” *Acta radiol.*, vol. 47, no. 10, pp. 1100–4, Dec. 2006.

- [39] H. C. Thoeny, F. De Keyzer, F. Chen, Y. Ni, W. Landuyt, E. K. Verbeken, H. Bosmans, G. Marchal, and R. Hermans, "Diffusion-weighted MR imaging in monitoring the effect of a vascular targeting agent on rhabdomyosarcoma in rats.," *Radiology*, vol. 234, no. 3, pp. 756–64, Mar. 2005.
- [40] B. D. B. Ross, B. B. A. Moffat, T. S. T. Lawrence, S. K. Mukherji, S. S. Gebarski, D. J. Quint, T. D. Johnson, L. Junck, P. L. Robertson, K. M. Muraszko, Q. Dong, C. R. Meyer, P. H. Bland, P. McConville, H. Geng, A. Rehemtulla, and T. L. Chenevert, "Evaluation of Cancer Therapy Using Diffusion Magnetic Resonance Imaging," *Mol. Cancer Ther.*, vol. 2, no. June, pp. 581–587, Jun. 2003.
- [41] R. J. Theilmann, R. Borders, T. P. Trouard, G. Xia, E. Outwater, J. Ranger-Moore, R. J. Gillies, and A. Stopeck, "Changes in Water Mobility Measured by Diffusion MRI Predict Response of Metastatic Breast Cancer to Chemotherapy," *Neoplasia*, vol. 6, no. 6, pp. 831–837, Nov. 2004.
- [42] P. J. Basser and C. Pierpaoli, "Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI.," *J. Magn. Reson. Ser. B*, vol. 111, pp. 209–219, Dec. 1996.
- [43] A. Stadlbauer, O. Ganslandt, R. Buslei, T. Hammen, S. Gruber, E. Moser, M. Buchfelder, E. Salomonowitz, and C. Nimsky, "Gliomas: histopathologic evaluation of changes in directionality and magnitude of water diffusion at diffusion-tensor MR imaging.," *Radiology*, vol. 240, no. 3, pp. 803–10, Sep. 2006.
- [44] E. Goebell, S. Paustenbach, O. Vaeterlein, X.-Q. Ding, O. Heese, J. Fiehler, T. Kucinski, C. Hagel, M. Westphal, and H. Zeumer, "Low-grade and anaplastic gliomas: differences in architecture evaluated with diffusion-tensor MR imaging.," *Radiology*, vol. 239, no. 1, pp. 217–22, Apr. 2006.
- [45] S. Lu, D. Ahn, G. Johnson, and S. Cha, "Peritumoral diffusion tensor imaging of high-grade gliomas and metastatic brain tumors.," *Am. J. Neuroradiol.*, vol. 24, no. 5, pp. 937–41, May 2003.

- [46] T. Beppu, T. Inoue, Y. Shibata, A. Kurose, H. Arai, K. Ogasawara, A. Ogawa, S. Nakamura, and H. Kabasawa, "Measurement of fractional anisotropy using diffusion tensor MRI in supratentorial astrocytic tumors.," *J. Neurooncol.*, vol. 63, no. 2, pp. 109–16, Jun. 2003.
- [47] S. Mori, K. Frederiksen, P. C. M. van Zijl, B. Stieltjes, M. A. Kraut, M. Solaiyappan, and M. G. Pomper, "Brain white matter anatomy of tumor patients evaluated with diffusion tensor imaging.," *Ann. Neurol.*, vol. 51, no. 3, pp. 377–80, Mar. 2002.
- [48] J. Folkman, "New perspectives in clinical oncology from angiogenesis research.," *Eur. J. Cancer*, vol. 32A, no. 14, pp. 2534–9, Dec. 1996.
- [49] H. J. Aronen, F. S. Pardo, D. N. Kennedy, J. W. Belliveau, S. D. Packard, D. W. Hsu, F. H. Hochberg, a J. Fischman, and B. R. Rosen, "High microvascular blood volume is associated with high glucose uptake and tumor angiogenesis in human gliomas.," *Clin. Cancer Res.*, vol. 6, no. 6, pp. 2189–200, Jun. 2000.
- [50] N. Weidner, "Intratumoral vascularity as a prognostic factor in cancers of the urogenital tract.," *Eur. J. Cancer*, vol. 32A, no. 14, pp. 2506–12, Dec. 1996.
- [51] R. K. Jain, "Transport of molecules across tumor vasculature.," *Cancer Metastasis Rev.*, vol. 6, no. 4, pp. 559–593, Dec. 1987.
- [52] R. K. Jain, "Determinants of tumor blood flow: a review.," *Cancer Res.*, vol. 48, no. 10, pp. 2641–58, May 1988.
- [53] M. V Knopp, F. L. Giesel, H. Marcos, H. von Tengg-Kobligk, and P. Choyke, "Dynamic contrast-enhanced magnetic resonance imaging in oncology.," *Top. Magn. Reson. imaging*, vol. 12, no. 4, pp. 301–8, Aug. 2001.
- [54] P. Carmeliet and R. K. Jain, "Angiogenesis in cancer and other diseases.," *Nature*, vol. 407, no. 6801, pp. 249–57, Sep. 2000.

- [55] L. D. Buadu, J. Murakami, S. Murayama, N. Hashiguchi, S. Sakai, K. Masuda, S. Toyoshima, S. Kuroki, and S. Ohno, "Breast lesions: correlation of contrast medium enhancement patterns on MR images with histopathologic findings and tumor angiogenesis.," *Radiology*, vol. 200, no. 3, pp. 639–49, Sep. 1996.
- [56] H. J. Aronen, I. E. Gazit, D. N. Louis, B. R. Buchbinder, F. S. Pardo, R. M. Weisskoff, G. R. Harsh, G. R. Cosgrove, E. F. Halpern, and F. H. Hochberg, "Cerebral blood volume maps of gliomas: comparison with tumor grade and histologic findings.," *Radiology*, vol. 191, no. 1, pp. 41–51, Apr. 1994.
- [57] H. Hawighorst, P. G. Knapstein, W. Weikel, M. V Knopp, I. Zuna, a Knof, G. Brix, U. Schaeffer, C. Wilkens, S. O. Schoenberg, M. Essig, P. Vaupel, and G. van Kaick, "Angiogenesis of uterine cervical carcinoma: characterization by pharmacokinetic magnetic resonance parameters and histological microvessel density with correlation," *Cancer Res.*, vol. 57, no. 21, pp. 4777–86, Nov. 1997.
- [58] M. Haris, R. K. Gupta, A. Singh, N. Husain, M. Husain, C. M. Pandey, C. Srivastava, S. Behari, and R. K. S. Rathore, "Differentiation of infective from neoplastic brain lesions by dynamic contrast-enhanced MRI.," *Neuroradiology*, vol. 50, no. 6, pp. 531–40, Jun. 2008.
- [59] M. Haris, N. Husain, A. Singh, R. Awasthi, R. K. Singh Rathore, M. Husain, and R. K. Gupta, "Dynamic contrast-enhanced (DCE) derived transfer coefficient (k_{trans}) is a surrogate marker of matrix metalloproteinase 9 (MMP-9) expression in brain tuberculomas.," *J. Magn. Reson. imaging*, vol. 28, no. 3, pp. 588–97, Sep. 2008.
- [60] O. Tynninen, H. J. Aronen, M. Ruhala, A. Paetau, K. Von Boguslawski, O. Salonen, J. Jääskeläinen, and T. Paavonen, "MRI enhancement and microvascular density in gliomas. Correlation with tumor cell proliferation.," *Invest. Radiol.*, vol. 34, no. 6, pp. 427–34, Jun. 1999.

- [61] J. Strugar, D. Rothbart, W. Harrington, and G. R. Criscuolo, "Vascular permeability factor in brain metastases: correlation with vasogenic brain edema and tumor angiogenesis.," *J. Neurosurg.*, vol. 81, no. 4, pp. 560–6, Oct. 1994.
- [62] M. Maeda, S. Itoh, H. Kimura, T. Iwasaki, N. Hayashi, K. Yamamoto, Y. Ishii, and T. Kubota, "Tumor vascularity in the brain: evaluation with dynamic susceptibility-contrast MR imaging.," *Radiology*, vol. 189, no. 1, pp. 233–8, Oct. 1993.
- [63] E. A. Knopp, S. Cha, G. Johnson, A. Mazumdar, J. G. Golfinos, D. Zagzag, D. C. Miller, P. J. Kelly, and I. I. Kricheff, "Glial neoplasms: dynamic contrast-enhanced T2*-weighted MR imaging.," *Radiology*, vol. 211, no. 3, pp. 791–8, Jun. 1999.
- [64] T. Sugahara, Y. Korogi, M. Kochi, I. Ikushima, T. Hirai, T. Okuda, Y. Shigematsu, L. Liang, Y. Ge, Y. Ushio, and M. Takahashi, "Correlation of MR imaging-determined cerebral blood volume maps with histologic and angiographic determination of vascularity of gliomas.," *Am. J. Roentgenol.*, vol. 171, no. 6, pp. 1479–86, Dec. 1998.
- [65] J. O. Barentsz, O. Berger-Hartog, J. A. Witjes, C. Hulsbergen-van der Kaa, G. O. Oosterhof, J. A. VanderLaak, H. Kondacki, and S. H. Ruijs, "Evaluation of chemotherapy in advanced urinary bladder cancer with fast dynamic contrast-enhanced MR imaging.," *Radiology*, vol. 207, no. 3, pp. 791–7, Jun. 1998.
- [66] M.-L. W. Ah-See, A. Makris, N. J. Taylor, M. Harrison, P. I. Richman, R. J. Burcombe, J. J. Stirling, J. A. D'Arcy, D. J. Collins, M. R. Pittam, D. Ravichandran, and A. R. Padhani, "Early changes in functional dynamic magnetic resonance imaging predict for pathologic response to neoadjuvant chemotherapy in primary breast cancer.," *Clin. Cancer Res.*, vol. 14, no. 20, pp. 6580–9, Oct. 2008.
- [67] A. R. Padhani, C. Hayes, L. Assersohn, T. Powles, A. Makris, J. Suckling, M. O. Leach, and J. E. Husband, "Prediction of clinicopathologic response of breast cancer to primary chemotherapy at

- contrast-enhanced MR imaging: initial clinical results.," *Radiology*, vol. 239, no. 2, pp. 361–74, May 2006.
- [68] J. A. Lancaster, B. M. Carrington, J. R. Sykes, A. P. Jones, S. M. Todd, R. Cooper, D. L. Buckley, S. E. Davidson, J. P. Logue, R. D. Hunter, and C. M. L. West, "Prediction of radiotherapy outcome using dynamic contrast enhanced MRI of carcinoma of the cervix.," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 54, no. 3, pp. 759–67, Nov. 2002.
- [69] M. L. George, A. S. Dzik-Jurasz, A. R. Padhani, G. Brown, D. M. Tait, S. A. Eccles, and R. I. Swift, "Non-invasive methods of assessing angiogenesis and their value in predicting response to treatment in colorectal cancer.," *Br. J. Surg.*, vol. 88, no. 12, pp. 1628–36, Dec. 2001.
- [70] H. Hawighorst, R. Engenhart, M. V Knopp, G. Brix, M. Grandy, M. Essig, P. Miltner, I. Zuna, M. Fuss, and G. van Kaick, "Intracranial meningiomas: time- and dose-dependent effects of irradiation on tumor microcirculation monitored by dynamic MR imaging.," *Magn. Reson. Imaging*, vol. 15, no. 4, pp. 423–32, Jan. 1997.
- [71] A. Jackson, J. P. B. O'Connor, G. J. M. Parker, and G. C. Jayson, "Imaging tumor vascular heterogeneity and angiogenesis using dynamic contrast-enhanced magnetic resonance imaging.," *Clin. cancer Res.*, vol. 13, no. 12, pp. 3449–59, Jun. 2007.
- [72] J. P. B. O'Connor, A. Jackson, G. J. M. Parker, and G. C. Jayson, "DCE-MRI biomarkers in the clinical evaluation of antiangiogenic and vascular disrupting agents.," *Br. J. Cancer*, vol. 96, no. 2, pp. 189–95, Jan. 2007.
- [73] M. O. Leach, B. Morgan, P. S. Tofts, D. L. Buckley, W. Huang, M. a Horsfield, T. L. Chenevert, D. J. Collins, A. Jackson, D. Lomas, B. Whitcher, L. Clarke, R. Plummer, I. Judson, R. Jones, R. Alonzi, T. Brunner, D. M. Koh, P. Murphy, J. C. Waterton, G. Parker, M. J. Graves, T. W. J. Scheenen, T. W. Redpath, M. Orton, G. Karczmar, H. Huisman, J. Barentsz, and A. Padhani, "Imaging vascular

- function for early stage clinical trials using dynamic contrast-enhanced magnetic resonance imaging.," *Eur. Radiol.*, vol. 22, no. 7, pp. 1451–64, Jul. 2012.
- [74] L. Clarke, B. Croft, and R. Nordstrom, "Quantitative imaging for evaluation of response to cancer therapy," *Transl. Oncol.*, vol. 2, no. 4, pp. 195–197, 2009.
- [75] E. Jackson and D. Barboriak, "Magnetic resonance assessment of response to therapy: tumor change measurement, truth data and error sources," *Transl. Oncol.*, vol. 2, no. 4, pp. 211–215, 2009.
- [76] C. R. Meyer, S. G. A. Iii, C. P. Fenimore, G. McLennan, M. F. McNitt-gray, and P. E. Kinahan, "Quantitative Imaging to Assess Tumor Response to Therapy: Common Themes of Measurement, Truth Data, and Error Sources," *Transl. Oncol.*, vol. 2, no. 4, pp. 198–210, 2009.
- [77] M. F. McNitt-Gray, L. M. Bidaut, S. G. Armato, C. R. Meyer, M. A. Gavrielides, C. Fenimore, G. McLennan, N. Petrick, B. Zhao, A. P. Reeves, R. Beichel, H.-J. G. Kim, and L. Kinnard, "Computed tomography assessment of response to therapy: tumor volume change measurement, truth data, and error.," *Transl. Oncol.*, vol. 2, no. 4, pp. 216–22, Dec. 2009.
- [78] R. Bosca, E. Ashton, G. Zahlmann, and E. F. Jackson, "RSNA Quantitative Imaging Biomarker Alliance (QIBA) DCE-MRI Phantom: Goal Design, and Initial Results," 2012.
- [79] D. Malyarenko, C. J. Galbán, F. J. Londy, C. R. Meyer, T. D. Johnson, A. Rehemtulla, B. D. Ross, and T. L. Chenevert, "Multi-system repeatability and reproducibility of apparent diffusion coefficient measurement using an ice-water phantom.," *J. Magn. Reson. imaging*, vol. 37, no. 5, pp. 1238–46, May 2013.
- [80] T. Zhu, R. Hu, X. Qiu, M. Taylor, Y. Tso, C. Yiannoutsos, B. Navia, S. Mori, S. Ekholm, G. Schifitto, and J. Zhong, "Quantification of accuracy and precision of multi-center DTI measurements: a diffusion phantom and human brain study.," *Neuroimage*, vol. 56, no. 3, pp. 1398–411, Jun. 2011.

- [81] S. Eida, M. Sumi, and T. Nakamura, "Multiparametric magnetic resonance imaging for the differentiation between benign and malignant salivary gland tumors.," *J. Magn. Reson. imaging*, vol. 31, no. 3, pp. 673–9, Mar. 2010.
- [82] B. F. Jordan and B. Gallez, "Surrogate MR markers of response to chemo- or radiotherapy in association with co-treatments: a retrospective analysis of multi-modal studies.," *Contrast Media Mol. Imaging*, vol. 5, no. 6, pp. 323–32, 2010.
- [83] N. C. Atuegwu, J. C. Gore, and T. E. Yankeelov, "The integration of quantitative multi-modality imaging data into mathematical models of tumors.," *Phys. Med. Biol.*, vol. 55, no. 9, pp. 2429–49, May 2010.
- [84] C. Hayes, A. Padhani, and M. Leach, "Assessing changes in tumour vascular function using dynamic contrast enhanced magnetic resonance imaging," *NMR Biomed.*, vol. 15, pp. 154–163, 2002.
- [85] E. M. Haacke, C. L. Filletti, R. Gattu, C. Ciulla, A. Al-Bashir, K. Suryanarayanan, M. Li, Z. Latif, Z. DelProposto, V. Sehgal, T. Li, V. Torquato, R. Kanaparti, J. Jiang, and J. Neelavalli, "New algorithm for quantifying vascular changes in dynamic contrast-enhanced MRI independent of absolute T1 values.," *Magn. Reson. Med.*, vol. 58, no. 3, pp. 463–72, Sep. 2007.
- [86] C. J. Galbán, T. L. Chenevert, C. R. Meyer, C. Tsien, T. S. Lawrence, D. a Hamstra, L. Junck, P. C. Sundgren, T. D. Johnson, D. J. Ross, A. Rehemtulla, and B. D. Ross, "The parametric response map is an imaging biomarker for early cancer treatment outcome.," *Nat. Med.*, vol. 15, no. 5, pp. 572–6, May 2009.
- [87] A. Fick, "Ueber Diffusion," *Ann. der Phys. und Chemie*, vol. 170, no. 1, pp. 59–86, 1855.
- [88] A. Einstein, *Investigations on the Theory of the Brownian Movement*. 1956.
- [89] E. O. Stejskal and J. E. Tanner, "Spin Diffusion Measurements: Spin Echoes in the Presence of a Time-Dependent Field Gradient," *J. Chem. Phys.*, vol. 42, no. 1, p. 288, 1965.

- [90] H. Torrey, "Bloch Equations with Diffusion Terms," *Phys. Rev.*, vol. 104, no. 3, pp. 563–565, Nov. 1956.
- [91] J. Mattiello, P. J. Basser, and D. LeBihan, "Analytical Expressions for the b Matrix in NMR Diffusion Imaging and Spectroscopy," *J. Magn. Reson. Ser. A*, vol. 108, no. 2, pp. 131–141, Jun. 1994.
- [92] J. Mattiello, P. J. Basser, and D. Le Bihan, "The b matrix in diffusion tensor echo-planar imaging," *Magn. Reson. Med.*, vol. 37, no. 2, pp. 292–300, Feb. 1997.
- [93] D. Le Bihan, E. Breton, D.ALLEMAND, P. Grenier, E. Cabanis, and M. Laval-Jeantet, "MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders.," *Radiology*, vol. 161, no. 2, pp. 401–7, Nov. 1986.
- [94] D. Le Bihan, E. Breton, D.ALLEMAND, M. L. Aubin, J. Vignaud, and M. Laval-Jeantet, "Separation of diffusion and perfusion in intravoxel incoherent motion MR imaging.," *Radiology*, vol. 168, no. 2, pp. 497–505, Aug. 1988.
- [95] Y. Pang, B. Turkbey, M. Bernardo, J. Kruecker, S. Kadoury, M. J. Merino, B. J. Wood, P. a Pinto, and P. L. Choyke, "Intravoxel incoherent motion MR imaging for prostate cancer: An evaluation of perfusion fraction and diffusion coefficient derived from different b-value combinations.," *Magn. Reson. Med.*, vol. 000, Apr. 2012.
- [96] D. a Yablonskiy and A. L. Sukstanskii, "Theoretical models of the diffusion weighted MR signal.," *NMR Biomed.*, vol. 23, no. 7, pp. 661–81, Aug. 2010.
- [97] P. W. Schaefer, "Applications of DWI in clinical neurology.," *J. Neurol. Sci.*, vol. 186 Suppl , pp. S25–35, May 2001.
- [98] P. J. Basser, J. Mattiello, and D. LeBihan, "Estimation of the effective self-diffusion tensor from the NMR spin echo.," *J. Magn. Reson. B*, vol. 103, no. 3, pp. 247–54, Mar. 1994.

- [99] M. E. Bastin, P. A. Armitage, and I. Marshall, "A theoretical study of the effect of experimental noise on the measurement of anisotropy in diffusion imaging," *Magn. Reson. Imaging*, vol. 16, no. 7, pp. 773–785, Sep. 1998.
- [100] P. A. Armitage and M. E. Bastin, "Utilizing the diffusion-to-noise ratio to optimize magnetic resonance diffusion tensor acquisition strategies for improving measurements of diffusion anisotropy," *Magn. Reson. Med.*, vol. 45, no. 6, pp. 1056–65, Jun. 2001.
- [101] D. K. Jones, M. A. Horsfield, and A. Simmons, "Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging," *Magn. Reson. Med.*, vol. 42, no. 3, pp. 515–25, Sep. 1999.
- [102] A. H. Poonawalla and X. J. Zhou, "Analytical error propagation in diffusion anisotropy calculations," *J. Magn. Reson. imaging*, vol. 19, no. 4, pp. 489–98, Apr. 2004.
- [103] a P. Crawley and R. M. Henkelman, "A comparison of one-shot and recovery methods in T1 imaging," *Magn. Reson. Med.*, vol. 7, no. 1, pp. 23–34, May 1988.
- [104] G. J. Parker, G. J. Barker, and P. S. Tofts, "Accurate multislice gradient echo T(1) measurement in the presence of non-ideal RF pulse shape and RF field nonuniformity," *Magn. Reson. Med.*, vol. 45, no. 5, pp. 838–45, May 2001.
- [105] H. Gudbjartsson and S. Patz, "The Rician distribution of noisy MRI data," *Magn. Reson. Med.*, vol. 34, no. 6, pp. 910–914, Dec. 1995.
- [106] P. Caravan, J. J. Ellison, T. J. McMurry, and R. B. Lauffer, "Gadolinium(III) Chelates as MRI Contrast Agents: Structure, Dynamics, and Applications," *Chem. Rev.*, vol. 99, no. 9, pp. 2293–352, Sep. 1999.
- [107] D. G. Mitchell, "MR imaging contrast agents — what's in a name?," *J. Magn. Reson. Imaging*, vol. 7, no. 1, pp. 1–4, Jan. 1997.

- [108] S. S. Kety, "The theory and applications of the exchange of inert gas at the lungs and tissues.," *Pharmacol. Rev.*, vol. 3, no. 1, pp. 1–41, Mar. 1951.
- [109] P. S. Tofts and a G. Kermode, "Measurement of the blood-brain barrier permeability and leakage space using dynamic MR imaging. 1. Fundamental concepts.," *Magn. Reson. Med.*, vol. 17, no. 2, pp. 357–67, Feb. 1991.
- [110] P. Tofts and G. Brix, "Estimating kinetic parameters from dynamic contrast-enhanced T 1-weighted MRI of a diffusable tracer: standardized quantities and symbols," *J. Magn. Reson. imaging*, vol. 10, pp. 223–232, 1999.
- [111] J. U. Harrer, G. J. M. Parker, H. a Haroon, D. L. Buckley, K. Embelton, C. Roberts, D. Balériaux, and A. Jackson, "Comparative study of methods for determining vascular permeability and blood volume in human gliomas.," *J. Magn. Reson. imaging*, vol. 20, no. 5, pp. 748–57, Nov. 2004.
- [112] E. M. Renkin, "Transport of potassium-42 from blood to tissue in isolated mammalian skeletal muscles.," *Am. J. Physiol.*, vol. 197, pp. 1205–10, Dec. 1959.
- [113] K.-L. Li and A. Jackson, "New hybrid technique for accurate and reproducible quantitation of dynamic contrast-enhanced MRI data.," *Magn. Reson. Med.*, vol. 50, no. 6, pp. 1286–95, Dec. 2003.
- [114] K. S. St Lawrence and T.-Y. Lee, "An Adiabatic Approximation to the Tissue Homogeneity Model for water exchange in the brain," *J. Cereb. Blood Flow Metab.*, vol. 18, no. 12, pp. 1365–1377, 1998.
- [115] K. M. Donahue, D. Burstein, W. J. Manning, and M. L. Gray, "Studies of Gd-DTPA relaxivity and proton exchange rates in tissue.," *Magn. Reson. Med.*, vol. 32, no. 1, pp. 66–76, Jul. 1994.
- [116] G. J. Stanisz and R. M. Henkelman, "Gd-DTPA relaxivity depends on macromolecular content.," *Magn. Reson. Med.*, vol. 44, no. 5, pp. 665–7, Nov. 2000.

- [117] P. A. Bottomley, T. H. Foster, R. E. Argersinger, and L. M. Pfeifer, "A review of normal tissue hydrogen NMR relaxation times and relaxation mechanisms from 1-100 MHz: dependence on tissue type, NMR frequency, temperature, species, excision, and age.," *Med. Phys.*, vol. 11, no. 4, pp. 425–48, 1984.
- [118] P. A. Bottomley, C. J. Hardy, R. E. Argersinger, and G. Allen-Moore, "A review of ^1H nuclear magnetic resonance relaxation in pathology: are T1 and T2 diagnostic?," *Med. Phys.*, vol. 14, no. 1, pp. 1–37, 1987.
- [119] S. Naruse, Y. Horikawa, C. Tanaka, K. Hirakawa, H. Nishikawa, and K. Yoshizaki, "Significance of proton relaxation time measurement in brain edema, cerebral infarction and brain tumors," *Magn. Reson. Imaging*, vol. 4, no. 4, pp. 293–304, Jan. 1986.
- [120] H. J. Weinmann, M. Laniado, and W. Mützel, "Pharmacokinetics of GdDTPA/dimeglumine after intravenous injection into healthy volunteers.," *Physiol. Chem. Phys. Med. NMR*, vol. 16, no. 2, pp. 167–72, Jan. 1984.
- [121] G. Brix, W. Semmler, R. Port, L. R. Schad, G. Layer, and W. J. Lorenz, "Pharmacokinetic parameters in CNS Gd-DTPA enhanced MR imaging.," *J. Comput. Assist. Tomogr.*, vol. 15, no. 4, pp. 621–8, 1991.
- [122] E. Ashton, D. Raunig, C. Ng, F. Kelcz, T. McShane, and J. Evelhoch, "Scan-rescan variability in perfusion assessment of tumors in MRI using both model and data-derived arterial input functions.," *J. Magn. Reson. imaging*, vol. 28, no. 3, pp. 791–6, Sep. 2008.
- [123] G. J. M. Parker, C. Roberts, A. Macdonald, G. a Buonaccorsi, S. Cheung, D. L. Buckley, A. Jackson, Y. Watson, K. Davies, and G. C. Jayson, "Experimentally-derived functional form for a population-averaged high-temporal-resolution arterial input function for dynamic contrast-enhanced MRI.," *Magn. Reson. Med.*, vol. 56, no. 5, pp. 993–1000, Nov. 2006.

- [124] C. Karmonik and E. F. Jackson, "Monte-Carlo simulation study to determine the confidence limits for the plasma and tissue concentration time curves in dynamic Gd-DTPA enhanced MRI," 1999, p. 4.
- [125] E. Henderson, B. K. Rutt, and T.-Y. Lee, "Temporal sampling requirements for the tracer kinetics modeling of breast disease," *Magn. Reson. Imaging*, vol. 16, no. 9, pp. 1057–1073, Nov. 1998.
- [126] L. E. Kershaw and H.-L. M. Cheng, "Temporal resolution and SNR requirements for accurate DCE-MRI data analysis using the AATH model.," *Magn. Reson. Med.*, vol. 64, no. 6, pp. 1772–1780, Dec. 2010.
- [127] D. a Kovar, M. Lewis, and G. S. Karczmar, "A new method for imaging perfusion and contrast extraction fraction: Input functions derived from reference tissues," *J. Magn. Reson. Imaging*, vol. 8, no. 5, pp. 1126–1134, Sep. 1998.
- [128] T. E. Yankeelov, J. J. Luci, M. Lepage, R. Li, L. Debusk, P. C. Lin, R. R. Price, and J. C. Gore, "Quantitative pharmacokinetic analysis of DCE-MRI data without an arterial input function: a reference region model.," *Magn. Reson. Imaging*, vol. 23, no. 4, pp. 519–29, May 2005.
- [129] M. Heisen, X. Fan, J. Buurman, N. a W. van Riel, G. S. Karczmar, and B. M. ter Haar Romeny, "The influence of temporal resolution in determining pharmacokinetic parameters from DCE-MRI data.," *Magn. Reson. Med.*, vol. 63, no. 3, pp. 811–816, Mar. 2010.
- [130] H. B. Larsson, S. Rosenbaum, and T. Fritz-Hansen, "Quantification of the effect of water exchange in dynamic contrast MRI perfusion measurements in the brain and heart.," *Magn. Reson. Med.*, vol. 46, no. 2, pp. 272–81, Aug. 2001.
- [131] D. Buckley, "Uncertainty in the analysis of tracer kinetics using dynamic contrast-enhanced T1-weighted MRI," *Magn. Reson. Med.*, vol. 606, pp. 601–606, 2002.

- [132] P. Wedeking, C. H. Sotak, J. Telser, K. Kumar, C. a Chang, and M. F. Tweedle, "Quantitative dependence of MR signal intensity on tissue concentration of Gd(HP-DO3A) in the nephrectomized rat.," *Magn. Reson. Imaging*, vol. 10, no. 1, pp. 97–108, Jan. 1992.
- [133] C. S. Landis, X. Li, F. W. Telang, J. a Coderre, P. L. Micca, W. D. Rooney, L. L. Latour, G. Véték, I. Pályka, and C. S. Springer, "Determination of the MRI contrast agent concentration time course in vivo following bolus injection: effect of equilibrium transcytolemmal water exchange.," *Magn. Reson. Med.*, vol. 44, no. 4, pp. 563–74, Oct. 2000.
- [134] T. E. Yankeelov, W. D. Rooney, X. Li, and C. S. Springer, "Variation of the relaxographic 'shutter-speed' for transcytolemmal water exchange affects the CR bolus-tracking curve shape.," *Magn. Reson. Med.*, vol. 50, no. 6, pp. 1151–69, Dec. 2003.
- [135] D. L. Buckley, L. E. Kershaw, and G. J. Stanisiz, "Cellular-interstitial water exchange and its effect on the determination of contrast agent concentration in vivo: dynamic contrast-enhanced MRI of human internal obturator muscle.," *Magn. Reson. Med.*, vol. 60, no. 5, pp. 1011–9, Nov. 2008.
- [136] K. M. Donahue, R. M. Weisskoff, D. A. Chesler, K. K. Kwong, A. A. Bogdanov, J. B. Mandeville, and B. R. Rosen, "Improving MR quantification of regional blood volume with intravascularT1 contrast agents: Accuracy, precision, and water exchange," *Magn. Reson. Med.*, vol. 36, no. 6, pp. 858–867, Dec. 1996.
- [137] P. Meier and K. L. Zierler, "On the theory of the indicator-dilution method for measurement of blood flow and volume," *J. Appl. Physiol.*, vol. 6, no. 293, pp. 731–744, 1954.
- [138] L. Østergaard, A. G. Sorensen, K. K. Kwong, R. M. Weisskoff, C. Gyldensted, and B. R. Rosen, "High resolution measurement of cerebral blood flow using intravascular tracer bolus passages. Part II: Experimental comparison and preliminary results," *Magn. Reson. Med.*, vol. 36, no. 5, pp. 726–736, Nov. 1996.

- [139] a G. Sorensen, A. L. Tievsky, L. Ostergaard, R. M. Weisskoff, and B. R. Rosen, "Contrast agents in functional MR imaging," *J. Magn. Reson. Imaging*, vol. 7, no. 1, pp. 47–55, Jan. 1997.
- [140] H. J. Aronen and J. Perkiö, "Dynamic susceptibility contrast MRI of gliomas," *Neuroimaging Clin. N. Am.*, vol. 12, no. 4, pp. 501–523, Nov. 2002.
- [141] H. K. Thompson, C. F. Starmer, R. E. Whalen, and H. D. McIntosh, "Indicator Transit Time Considered As a Gamma Variate.," *Circ. Res.*, vol. 14, pp. 502–15, Jun. 1964.
- [142] A. Kassner, D. J. Annesley, X. P. Zhu, K. L. Li, I. D. Kamaly-Asl, Y. Watson, and A. Jackson, "Abnormalities of the contrast re-circulation phase in cerebral tumors demonstrated using dynamic susceptibility contrast-enhanced imaging: a possible marker of vascular tortuosity.," *J. Magn. Reson. imaging*, vol. 11, no. 2, pp. 103–13, Feb. 2000.
- [143] G. Johnson, S. G. Wetzel, S. Cha, J. Babb, and P. S. Tofts, "Measuring blood volume and vascular transfer constant from dynamic, $T(2)^*$ -weighted contrast-enhanced MRI.," *Magn. Reson. Med.*, vol. 51, no. 5, pp. 961–8, May 2004.
- [144] J. M. Bland and D. G. Altman, "Measuring agreement in method comparison studies.," *Stat. Methods Med. Res.*, vol. 8, no. 2, pp. 135–60, Jun. 1999.
- [145] L. I. Lin, "A concordance correlation coefficient to evaluate reproducibility.," *Biometrics*, vol. 45, no. 1, pp. 255–68, Mar. 1989.
- [146] V. E. Johnson and J. H. Albert, *Ordinal Data Modeling*. New York, NY: Springer-Verlag New York, Inc, 1999.
- [147] N. J. D. Nagelkerke, "A note on a general definition of the coefficient of determination," *Biometrika*, vol. 78, no. 3, pp. 691–692, 1991.
- [148] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. New York, NY: Springer New York, 2009.

- [149] R. Bosca and E. Jackson, "Stability of T1 Relaxation Time and DCE-MRI Measures," *Med. Phys.*, vol. 36, no. 6, p. 2705, 2009.
- [150] P. S. Tofts, D. Lloyd, C. a Clark, G. J. Barker, G. J. Parker, P. McConville, C. Baldock, and J. M. Pope, "Test liquids for quantitative MRI measurements of self-diffusion coefficient in vivo.," *Magn. Reson. Med.*, vol. 43, no. 3, pp. 368–74, Mar. 2000.
- [151] I. Delakis, E. M. Moore, M. O. Leach, and J. P. De Wilde, "Developing a quality control protocol for diffusion imaging on a clinical MRI system," *Phys. Med. Biol.*, vol. 49, no. 8, pp. 1409–1422, Apr. 2004.
- [152] T. Ihalainen, O. Sipilä, and S. Savolainen, "MRI quality control: six imagers studied using eleven unified image quality parameters.," *Eur. Radiol.*, vol. 14, no. 10, pp. 1859–65, Oct. 2004.
- [153] H. J. Laubach, P. M. Jakob, K. O. Loevblad, a E. Baird, M. P. Bovo, R. R. Edelman, and S. Warach, "A phantom for diffusion-weighted imaging of acute stroke.," *J. Magn. Reson. imaging*, vol. 8, no. 6, pp. 1349–54, 1998.
- [154] M. Freed, J. a de Zwart, P. Hariharan, M. R. Myers, and A. Badano, "Development and characterization of a dynamic lesion phantom for the quantitative evaluation of dynamic contrast-enhanced MRI," *Med. Phys.*, vol. 38, no. 10, pp. 5601–11, Oct. 2011.
- [155] F. B. Laun, S. Huff, and B. Stieltjes, "On the effects of dephasing due to local gradients in diffusion tensor imaging experiments: relevance for diffusion tensor imaging fiber phantoms.," *Magn. Reson. Imaging*, vol. 27, no. 4, pp. 541–8, May 2009.
- [156] W. D. Rooney, G. Johnson, X. Li, E. R. Cohen, S.-G. Kim, K. Ugurbil, and C. S. Springer, "Magnetic field and tissue dependencies of human brain longitudinal $1H_2O$ relaxation in vivo.," *Magn. Reson. Med.*, vol. 57, no. 2, pp. 308–18, Feb. 2007.
- [157] K. a Kraft, P. P. Fatouros, G. D. Clarke, and P. R. Kishore, "An MRI phantom material for quantitative relaxometry.," *Magn. Reson. Med.*, vol. 5, no. 6, pp. 555–62, Dec. 1987.

- [158] R. A. Lerski and J. D. de Certaines, "II. Performance assessment and quality control in MRI by Eurospin test objects and protocols," *Magn. Reson. Imaging*, vol. 11, no. 6, pp. 817–833, Jan. 1993.
- [159] S. T. Engelter, J. M. Provenzale, J. R. Petrella, D. M. DeLong, and J. R. MacFall, "The effect of aging on the apparent diffusion coefficient of normal-appearing white matter.," *Am. J. Roentgenol.*, vol. 175, no. 2, pp. 425–30, Aug. 2000.
- [160] T. L. Chenevert, "Diffusion Magnetic Resonance Imaging: an Early Surrogate Marker of Therapeutic Efficacy in Brain Tumors," *J. Natl. Cancer Inst.*, vol. 92, no. 24, pp. 2029–2036, Dec. 2000.
- [161] C. Pierpaoli, P. Jezzard, P. J. Basser, a Barnett, and G. Di Chiro, "Diffusion tensor MR imaging of the human brain.," *Radiology*, vol. 201, no. 3, pp. 637–48, Dec. 1996.
- [162] D. P. Barboriak, J. R. MacFall, B. L. Viglianti, and M. W. Dewhirst Dvm, "Comparison of three physiologically-based pharmacokinetic models for the prediction of contrast agent distribution measured by dynamic MR imaging.," *J. Magn. Reson. imaging*, vol. 27, no. 6, pp. 1388–98, Jun. 2008.
- [163] F. J. Massey, "The Kolmogorov-Smirnov Test for Goodness of Fit," *J. Am. Stat. Assoc.*, vol. 46, no. 253, pp. 68–78, Mar. 1951.
- [164] H. J. W. L. Aerts, K. Jaspers, and W. H. Backes, "The precision of pharmacokinetic parameters in dynamic contrast-enhanced magnetic resonance imaging: the effect of sampling frequency and duration.," *Phys. Med. Biol.*, vol. 56, no. 17, pp. 5665–78, Sep. 2011.
- [165] Y. Cao, D. Li, Z. Shen, and D. Normolle, "Sensitivity of quantitative metrics derived from DCE MRI and a pharmacokinetic model to image quality and acquisition parameters.," *Acad. Radiol.*, vol. 17, no. 4, pp. 468–78, Apr. 2010.

- [166] J. a Brookes, T. W. Redpath, F. J. Gilbert, a D. Murray, and R. T. Staff, "Accuracy of T1 measurement in dynamic contrast-enhanced breast MRI using two- and three-dimensional variable flip angle fast low-angle shot.," *J. Magn. Reson. imaging*, vol. 9, no. 2, pp. 163–71, Feb. 1999.
- [167] R. Venkatesan, W. Lin, and E. M. Haacke, "Accurate determination of spin-density and T1 in the presence of RF-field inhomogeneities and flip-angle miscalibration.," *Magn. Reson. Med.*, vol. 40, no. 4, pp. 592–602, Oct. 1998.
- [168] P. H. Gutin, F. M. Iwamoto, K. Beal, N. a Mohile, S. Karimi, B. L. Hou, S. Lymberis, Y. Yamada, J. Chang, and L. E. Abrey, "Safety and efficacy of bevacizumab with hypofractionated stereotactic irradiation for recurrent malignant gliomas.," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 75, no. 1, pp. 156–63, Sep. 2009.
- [169] C. Pierpaoli and P. J. Basser, "Toward a quantitative assessment of diffusion anisotropy.," *Magn. Reson. Med.*, vol. 36, no. 6, pp. 893–906, Dec. 1996.
- [170] P. J. Basser and S. Pajevic, "Statistical artifacts in diffusion tensor MRI (DT-MRI) caused by background noise.," *Magn. Reson. Med.*, vol. 44, no. 1, pp. 41–50, Jul. 2000.
- [171] A. Anderson, "Theoretical analysis of the effects of noise on diffusion tensor imaging," *Magn. Reson. Med.*, vol. 45, no. 6, pp. 1174–1188, 2001.
- [172] D. I. Malyarenko, B. D. Ross, and T. L. Chenevert, "Analysis and correction of gradient nonlinearity bias in apparent diffusion coefficient measurements," *Magn. Reson. Med.*, vol. 000, p. 0n/a–0n/a, May 2013.
- [173] C. D. Marcus, V. Ladam-Marcus, C. Cucu, O. Bouché, L. Lucas, and C. Hoeffel, "Imaging techniques to evaluate the response to treatment in oncology: current standards and perspectives.," *Crit. Rev. Oncol. Hematol.*, vol. 72, no. 3, pp. 217–38, Dec. 2009.

- [174] G. McLennan, L. Clarke, and R. J. Hohl, "Imaging as a biomarker for therapy response: cancer as a prototype for the creation of research resources," *Clin. Pharmacol. Ther.*, vol. 84, no. 4, pp. 433–6, Oct. 2008.
- [175] N. Hylton, "Dynamic contrast-enhanced magnetic resonance imaging as an imaging biomarker," *J. Clin. Oncol.*, vol. 24, no. 20, pp. 3293–8, Jul. 2006.
- [176] T. S. Yoo, M. J. Ackerman, W. E. Lorensen, W. Schroeder, V. Chalana, S. Aylward, D. Metaxas, and R. Whitaker, "Engineering and algorithm design for an image processing Api: a technical report on ITK--the Insight Toolkit.," *Stud. Health Technol. Inform.*, vol. 85, pp. 586–92, Jan. 2002.
- [177] T. S. Yoo and D. N. Metaxas, "Open science--combining open data and open source software: medical image analysis with the Insight Toolkit.," *Med. Image Anal.*, vol. 9, no. 6, pp. 503–6, Dec. 2005.
- [178] L. Ibanez, L. Ng, J. Gee, and S. Aylward, "Registration patterns: the generic framework for image registration of the insight toolkit," in *Proceedings IEEE International Symposium on Biomedical Imaging*, 2002, pp. 345–348.
- [179] M. J. McAuliffe, F. M. Lalonde, D. McGarry, W. Gandler, K. Csaky, and B. L. Trus, "Medical Image Processing, Analysis and Visualization in clinical research," in *Proceedings 14th IEEE Symposium on Computer-Based Medical Systems. CBMS 2001*, pp. 381–386.
- [180] R. W. Cox, "AFNI: software for analysis and visualization of functional magnetic resonance neuroimages.," *Comput. Biomed. Res.*, vol. 29, no. 3, pp. 162–73, Jun. 1996.
- [181] W. Rasband and T. Ferreira, *ImageJ User Guide*. .
- [182] D. P. Barboriak, J. R. MacFall, A. O. Padua, G. E. York, B. L. Viglianti, and M. W. Dewhirst, "Standardized software for calculation of Ktrans and vp from dynamic T1-weighted MR images," in *International Society for Magnetic Resonance in Medicine Workshop on MR in Drug Development: From Discovery to Clinical Therapeutic Trials*, 2004.

- [183] Y. Cao, "Development of Image Software Tools for Radiation Therapy Assessment," *Med. Phys.*, vol. 32, no. 6, p. 2136, 2005.
- [184] T. Neff, F. Kiessling, G. Brix, K. Baudendistel, C. Zechmann, F. L. Giesel, and R. Bendl, "An optimized workflow for the integration of biological information into radiotherapy planning: experiences with T1w DCE-MRI," *Phys. Med. Biol.*, vol. 50, no. 17, pp. 4209–23, Sep. 2005.
- [185] K. Covington, E. S. McCreedy, M. Chen, A. Carass, N. Aucoin, and B. A. Landman, "Interfaces and Integration of Medical Image Analysis Frameworks: Challenges and Opportunities.," *ORNL Biomed. Sci. Eng. Cent. Conf.*, vol. 2010, pp. 1–4, May 2010.
- [186] T. R Core, "A language and environment for statistical computing.," in *R Foundation for Statistical Computing*, 2013.
- [187] T. Coleman and Y. Li, "On the convergence of interior-reflective Newton methods for nonlinear minimization subject to bounds," *Math. Program.*, vol. 67, pp. 189–224, 1994.
- [188] T. F. Coleman and Y. Li, "An Interior Trust Region Approach for Nonlinear Minimization Subject to Bounds," *SIAM J. Optim.*, vol. 6, no. 2, pp. 418–445, May 1996.
- [189] K. Levenberg, "A Method for the Solution of Certain Problems in Least-Squares," *Q. Appl. Math.*, vol. 2, pp. 164–168, 1944.
- [190] D. W. Marquardt, "An Algorithm for Least-Squares Estimation of Nonlinear Parameters," *J. Soc. Ind. Appl. Math.*, vol. 11, no. 2, pp. 431–441, Jun. 1963.
- [191] T. L. James and G. G. McDonald, "Measurement of the self-diffusion coefficient of each component in a complex system using pulsed-gradient fourier transform NMR," *J. Magn. Reson.*, vol. 11, no. 1, pp. 58–61, Jul. 1973.
- [192] D. Mattes, D. R. Haynor, H. Vesselle, T. K. Lewellen, and W. Eubank, "PET-CT image registration in the chest using free-form deformations," *IEEE Trans. Med. Imaging*, vol. 22, no. 1, pp. 120–8, Jan. 2003.

- [193] F. Maes, D. Vandermeulen, and P. Suetens, "Comparative evaluation of multiresolution optimization strategies for multimodality image registration by maximization of mutual information," *Med. Image Anal.*, vol. 3, no. 4, pp. 373–86, Dec. 1999.
- [194] P. Viola and W. W. III, "Alignment by maximization of mutual information," *Int. J. Comput. Vis.*, vol. 24, no. 2, pp. 137–154, 1997.
- [195] A. Roche, G. Malandain, X. Pennec, and N. Ayache, "The correlation ratio as a new similarity measure for multimodal image registration," *Med. Image Comput. Comput. Interv.*, vol. 1496, pp. 1115–1124, 1998.
- [196] L. Ibanez, W. Schroeder, L. Ng, and J. Cates, *The ITK software guide*, Second. 2003.
- [197] B. Aubert-Broche, M. Griffin, G. B. Pike, A. C. Evans, and D. L. Collins, "Twenty new digital brain phantoms for creation of validation image data bases.," *IEEE Trans. Med. Imaging*, vol. 25, no. 11, pp. 1410–6, Nov. 2006.
- [198] L. F. Shampine, "Vectorized adaptive quadrature in MATLAB," *J. Comput. Appl. Math.*, vol. 211, no. 2, pp. 131–140, Feb. 2008.
- [199] R. H. B. Christensen, "ordinal - Regression Models for Ordinal Data R Package." 2013.
- [200] R. G. P. Lopata, W. H. Backes, P. P. J. van den Bosch, and N. a W. van Riel, "On the identifiability of pharmacokinetic parameters in dynamic contrast-enhanced imaging.," *Magn. Reson. Med.*, vol. 58, no. 2, pp. 425–9, Aug. 2007.
- [201] V. J. Schmid, B. Whitcher, A. R. Padhani, N. J. Taylor, and G.-Z. Yang, "Bayesian Methods for Pharmacokinetic Models in Dynamic Contrast-Enhanced Magnetic Resonance Imaging," *IEEE Trans. Med. Imaging*, vol. 25, no. 12, pp. 1627–1636, Dec. 2006.
- [202] S. Wang, P. Liu, B. Turkbey, P. Choyke, P. Pinto, and R. M. Summers, "Gaussian process inference for estimating pharmacokinetic parameters of dynamic contrast-enhanced MR images.," *Med. Image Comput. Comput. Interv.*, vol. 15, no. Pt 3, pp. 582–9, Jan. 2012.

- [203] B. Kelm and B. Menze, “Estimating kinetic parameter maps from dynamic contrast-enhanced MRI using spatial prior knowledge,” *Med. Imaging*, vol. 28, no. 10, pp. 1534–1547, 2009.

Vita

Ryan Joseph Bosca was born in Arlington, Texas on November 2nd, 1982, the son of Donald and Tonya Bosca. After graduating from Mansfield High School in Mansfield, Texas, he earned a Bachelor of Science degree in physics, graduating with Honors from the University of North Texas – Denton, Texas – in May of 2006. He later went on to complete a Bachelor of Science in Mathematics (2007) and Master of Science in Physics (2008) before matriculating to The University of Texas Health Science Center at Houston Graduate School of Biomedical Sciences in August of 2008.

Permanent address:

2514 Woodbridge Trail
Mansfield, Texas 76063