

UCSF

UC San Francisco Electronic Theses and Dissertations

Title

Application of Resampling Statistics in Diffusion Tensor MRI Analysis

Permalink

<https://escholarship.org/uc/item/0q96w1dm>

Author

Chung, Chris SungWon

Publication Date

2008-05-02

Peer reviewed|Thesis/dissertation

Application of Resampling Statistics in Diffusion Tensor MRI Analysis

by

Chris SungWon Chung

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

BIOENGINEERING

in the

GRADUATE DIVISIONS

of the

UNIVERSITY OF CALIFORNIA SAN FRANCISCO

and

UNIVERSITY OF CALIFORNIA BERKELEY

Copyright 2008

by

Chris SungWon Chung

To SooYeon

Acknowledgements

I would like to thank my advisor, Roland Henry, for his enthusiastic support and guidance. All the discussions and his advice during my graduate studies were crucial in the initiation, development and completion of this dissertation project. He made sure that I would have access to all the resources and people that I needed for this project. Most importantly, I would not have learned and enjoyed so much during my PhD training without him. Daniel Pelletier and Pratik Mukherjee helped me to work on the exciting projects on multiple sclerosis and traumatic brain injury, and their insights on the diseases and imaging as clinical scientists were invaluable. They also gladly served as my qualifying exam and dissertation committee members. Ying Lu helped me as a biostatistician and my qualifying exam committee member, and his advice steered my statistical theories into the right direction. Dan Vigneron guided me to get on the track toward PhD as my academic advisor and qualifying exam committee chair. I am also grateful to Sharmila Majumdar, who was in my dissertation committee, and Steve Conolly, who was in my qualifying exam committee.

Seung-Shik Yoo first introduced me to the world of MRI research. His advice had a strong influence on my decision to be a graduate student. Jeffrey Berman started his PhD works on DTI a few years earlier, and he gladly shared his experiences in DTI and in being a graduate student as a member of the same lab. Programming code that he wrote for his research was extremely valuable and

they made my own programming much easier. Michael Sdika wrote a powerful non-linear image registration program that became an essential tool for my research. David Newitt made a versatile image manipulation and visualization software which benefited my project. Caroline Rae, Blandine Courcot, and Kirsten Moffat acquired data for testing cardiac gating and helped me shaping and completing the project of cardiac artifact quantification. Numerous people were involved in recruiting multiple sclerosis and traumatic brain injury patients and acquiring MRI data from them, and I greatly benefited from all their efforts.

Grant support for this dissertation project was provided by the National Multiple Sclerosis Society (RG3240A1). I was personally supported by a UCSF Dean's Health Science Fellowship, an NIH Training Grant, and funds from Roland.

I would like to thank my parents and siblings for their support and encouragement. Finally, I would like to thank my wonderful wife, SooYeon. Her unending love and support during my graduate works were simply indescribable.

Abstract

Application of Resampling Statistics in Diffusion Tensor MRI Analysis

by

Chris SungWon Chung

Diffusion Tensor MRI (DTI) is an imaging technique that can probe the properties of microstructures in the brain tissue. DTI has proved to be useful in detecting brain pathologies invisible by other techniques and in delineating specific white matter pathways. Analyzing DTI data from a single subject at multiple time points is desirable when the disease of interest presents with large inter-subject variability. However no such analysis framework exists due to the lack of proper tools to accurately estimate uncertainties of DTI derived parameters in a single subject.

In this dissertation, novel DTI resampling statistical analysis frameworks were developed. Bootstrap algorithms were investigated by Monte Carlo simulation in order to better understand the properties of bootstrap and determine the optimal bootstrap algorithm for DTI. Then, DTI bootstrap was used to compare the uncertainties of DTI derived parameters with or without cardiac gating in order to determine the necessity of cardiac gating for DTI with single-shot EPI acquisitions. Bootstrap was also used to perform voxel-wise T-testing between

DTI data from two time points of a single subject, a technique called BLADE (Bootstrap-based Longitudinal Analysis of Diffusion Estimates). An alternative analysis framework based on permutation testing, another resampling technique, was developed as well. This framework called PERVADE (Permutation Voxel-wise Analysis of Diffusion Estimates) was designed to overcome some of the limitations of BLADE.

Monte Carlo simulation studies of bootstrap algorithms showed that the residual bootstrap algorithm developed in this work can estimate the uncertainties of DTI derived parameters with the highest accuracy. A cardiac gating scheme with minor increase in the scan time turned out to reduce the bootstrap-estimated uncertainties enough to justify the longer scan time, making cardiac gating necessary with partial Fourier acquisition. BLADE and PERVADE were able to detect FA changes in the normal appearing white matter as well as lesions of patients with traumatic brain injury and multiple sclerosis. Resampling techniques have shown great potential in subject-specific DTI analyses and quantification of artifacts, and are anticipated to play even bigger roles in the next generation diffusion MRI studies.

Table of Contents

Acknowledgements	iv
Abstract	vi
Table of Contents	viii
List of Figures	xi
List of Equations	xiii
 Chapter 1: Introduction	 1
 Chapter 2: Background	 5
2.1 Disease Populations Investigated	5
2.1.1 Multiple Sclerosis	7
2.1.1.1 Overview	7
2.1.1.2 Pathology & Imaging	9
2.1.2 Traumatic Brain Injury	12
2.1.2.1 Overview	12
2.1.2.2 Pathology & Imaging	14
2.1.3 Axonal Transection and Wallerian Degeneration	16
2.1.3.1 Pathology	16
2.1.3.2 Imaging	18
2.2 Diffusion Tensor MRI (DTI)	20
2.2.1 Diffusion	20
2.2.2 MR Measurement of Diffusion	22
2.2.3 Diffusion in the Biological Tissue	24
2.2.4 Basic Concepts of DTI	25
2.2.5 DTI Derived Parameters	27
2.2.6 Fiber Tracking	29
2.2.7 Uncertainty Estimation of DTI Derived Parameters	30
2.2.7.1 Analytical Approach	31
2.2.7.2 Numerical Approach	32
2.3 Reference	35
 Chapter 3: Comparison of DTI Bootstrap Approaches	 38
3.1 Abstract	38
3.2 Introduction	39
3.3 Methods	42
3.3.1 Under-estimation of Standard Errors by Repetition Bootstrap	42
3.3.2 Residual bootstrap and wild bootstrap	45
3.3.3 Monte Carlo Simulation	51
3.4 Results	54

3.4.1	Bias of Repetition Bootstrap.....	54
3.4.2	Bootstrap methods in the diffusion signals from single tensor	58
3.4.3	Bootstrap methods in the diffusion signals from tensor mixture	63
3.5	Discussion.....	65
3.6	Conclusion	71
3.7	Reference.....	72
Chapter 4: Bootstrap Quantification of Cardiac Pulsation Artifact in DTI		75
4.1	Abstract.....	75
4.2	Introduction	76
4.3	Methods	78
4.3.1	Subjects / Acquisition	78
4.3.1.1	Group A	78
4.3.1.2	Group B	81
4.3.2	Analysis.....	82
4.4	Results	85
4.4.1	Group A.....	85
4.4.2	Group B.....	93
4.5	Discussion.....	94
4.6	Conclusion	102
4.7	Reference.....	103
Chapter 5: Bootstrap Voxel-wise Serial DTI Analysis		105
5.1	Abstract.....	105
5.2	Introduction	106
5.3	Methods	106
5.3.1	Subjects / Acquisitions	106
5.3.2	Analysis.....	107
5.4	Results	110
5.4.1	Control Subjects.....	110
5.4.2	Patients with Cortical Contusion.....	112
5.4.3	Patients with Diffuse Axonal Injury	113
5.5	Discussion.....	116
5.6	Conclusion	120
5.7	Reference.....	120
Chapter 6: Permutation Voxel-wise Serial DTI Analysis		121
6.1	Abstract.....	121
6.2	Introduction	122
6.3	Methods	126
6.3.1	DTI Permutation Testing	126
6.3.1.1	Permutation Testing in Group Comparison	126
6.3.1.2	Permutation Testing in Intra-subject Serial Comparison.....	127

6.3.1.3	Permutation Testing for Multiple Comparisons	132
6.3.2	Non-linear Registration	134
6.3.3	Monte Carlo Simulation	136
6.3.4	Human Brain Data.....	138
6.3.4.1	Subjects / Acquisitions.....	138
6.3.4.2	Analysis	139
6.4	Results	142
6.4.1	Monte Carlo Simulation	142
6.4.2	Healthy Volunteer.....	145
6.4.3	Patient #1	146
6.4.4	Patient #2.....	148
6.4.5	Patient #3.....	149
6.4.6	Gradient tables rotation and gain factors	151
6.5	Discussion.....	151
6.6	Conclusion	157
6.7	Reference.....	158
Chapter 7:	Main Findings and Conclusions	162

List of Figures

Figure 2-1: MRI images from a MS patient.....	10
Figure 2-2: Wallerian degeneration of spinal cord axons.....	17
Figure 2-3: Pulse sequence diagram for a diffusion-weighted acquisition	22
Figure 2-4: Examples of 3D rendered fiber tracking results.....	30
Figure 2-5: Schematics of basic concept of bootstrap.	33
Figure 3-1: Standard error (SE) of FA estimated by repetition bootstrap and repetition bootknife	57
Figure 3-2: Bias, SD, and RMSE of the 95 th percentile confidence interval of the the angle of primary eigenvector estimated by bootstrap methods with different number of repetitions.....	59
Figure 3-3: Bias, SD, and RMSE of the 95 th percentile confidence interval of the angle of the primary eigenvector estimated by bootstrap methods with different numbers of directions	61
Figure 3-4: Comparison of the RMSE of the 95 th percentile confidence interval of the angle of the primary eigenvector estimated by bootstrap methods for varying numbers of repetitions and directions.....	62
Figure 3-5: Bias, SD, and RMSE of the 95 th percentile confidence interval of the angle of the primary eigenvector estimated by bootstrap methods with varying angles between two tensors	64
Figure 3-6: Same as Figure 3-5 except that b value used is 3000 s/mm ²	65
Figure 4-1: Schematics of cardiac gated DTI acquisition.....	79
Figure 4-2: Nominal acquisition times for each subject in group A.	86
Figure 4-3: Typical diffusion weighted images with or without cardiac gating.....	87
Figure 4-4: Typical maps of DTI derived parameters and their bootstrap-estimated uncertainty maps.....	88
Figure 4-5: Maps of R_U (gated to non-gated uncertainty ratio) and R_E (gated to non-gated efficiency ratio) for the 95 th percentile confidence intervals of the primary eigenvector directions, with partial Fourier acquisition.....	89
Figure 4-6: The same maps as Figure 4-5, except that data are from full Fourier acquisition.	90
Figure 4-7: Maps of R_U and R_E with partial or full Fourier acquisitions. In addition to $CI(\text{vec})$, standard errors of scalar parameters (FA, Ev1, MD, EvT) are shown as well for comparison.....	91
Figure 4-8: Spatial projections of R_U for partial and full Fourier acquisition toward sagittal, axial and coronal planes.....	92
Figure 4-9: Plots of R_U and R_E in each subject averaged in the high anisotropy white matter regions of the whole brain	93
Figure 4-10: Plots of R_U and R_E in each group B subject averaged in the high anisotropy white matter regions of the whole brain.....	94

Figure 5-1: Schematics of procedures in BLADE.	108
Figure 5-2: The plot showing the largest cluster with $\text{abs}(T) > 2.6$ in the whole brain of each healthy subject.....	111
Figure 5-3: FA maps in three different slices, and T statistics maps.....	112
Figure 5-4: BLADE analysis between 1 week and 1 month post injury in a focal cortical contusion patient.	113
Figure 5-5: BLADE analysis between 1 week and 1 month, and 1 week and 1 year in a diffuse axonal injury patient.	114
Figure 5-6: BLADE analysis in the different slice of the same patient shown in Figure 5-5	115
Figure 5-7: BLADE analysis between 1 week and 1 month in a diffuse axonal injury patient (different from the patient shown in Figure 5-5 and Figure 5-6). .	116
Figure 6-1: Overall schematics of permutation testing to calculate voxel-wise p-values and multiple-comparison-corrected cluster-wise p-values in a serial DTI study of a single subject	130
Figure 6-2: Detailed schematics of permutation testing to estimate voxel-wise p-values in a serial DTI study of a single subject.....	131
Figure 6-3: Qunatile-quantile plots for expected p-values following the uniform distribution between 0 and 1 versus observed p-values by permutation testing under diverse simulation conditions.....	144
Figure 6-4: Results of Healthy volunteer	146
Figure 6-5: Results of Patient #1	147
Figure 6-6: Results of Patient #2	149
Figure 6-7: Results of Patient #3	150

List of Equations

Equation 2-1: Probabilistic distribution function of free diffusion.....	21
Equation 2-2: Diffusion coefficient in case of free diffusion	21
Equation 2-3: MR signal attenuation by free diffusion	23
Equation 2-4: b-value	23
Equation 2-5: Calculation of diffusion coefficient by MR experiment	24
Equation 2-6: Three-direction-averaged apparent diffusion coefficient.....	25
Equation 2-7: Apparent diffusion coefficient in one direction	25
Equation 2-8: Probabilistic distribution function of anisotropic diffusion.....	26
Equation 2-9: Diffusion tensor	26
Equation 2-10: MR signal attenuation by anisotropic diffusion	26
Equation 2-11: Diffusion tensor diagonalization.....	27
Equation 2-12: Diagonal matrix and rotation matrix.....	27
Equation 2-13: D_{av} (averaged apparent diffusion coefficient)	28
Equation 2-14: FA (fractional anisotropy)	28
Equation 2-15: $\lambda_{parallel}$	28
Equation 2-16: $\lambda_{perpendicular}$	28
Equation 3-1: Biased estimator of population variance.....	43
Equation 3-2: Unbiased estimator of population variance	43
Equation 3-3: Mean squared error.....	43
Equation 3-4: Bias	43
Equation 3-5: Standard deviation	43
Equation 3-6: Diffusion weighted signal in DTI	47
Equation 3-7: Log transform of DTI signal attenuation equation.....	47
Equation 3-8: Multiple linear regression	47
Equation 3-9: Weighted least square estimation of diffusion tensor elements....	48
Equation 3-10: Model of variance after log transformation	48
Equation 3-11: Studentization of residuals	49
Equation 3-12: Residual bootstrap resampling	49
Equation 3-13: Bootstrap estimated standard error.....	50
Equation 3-14: Wild bootstrap resampling	50
Equation 3-15: Wild bootstrap resampled errors	50
Equation 3-16: DTI signal attenuation with mixture of fibers.....	53
Equation 5-1: T statistic in BLADE.....	109
Equation 5-2: T statistic in general	118
Equation 5-3: T statistic as the studentized difference	118
Equation 6-1: p-value in the permutation testing	127
Equation 6-2: Approximation of p-value with Monte Carlo permutation testing.	127
Equation 6-3: p-value of clusters	133

Chapter 1: Introduction

Diffusion tensor MRI (DTI) is a diffusion MRI technique with modeling of water self diffusion characteristics in each voxel using a 3×3 2nd rank diffusion tensor. DTI has been proved to be useful in investigating the integrity of brain tissue microstructure, such as axonal membranes and myelin. It has also become an indispensable tool in delineating axonal bundles that connect different parts of the brain through techniques collectively called DTI fiber tracking or tractography.

Most reported DTI studies so far were designed as group comparisons for regions of interests. This does not necessarily imply that DTI lacks the statistical power to draw meaningful conclusion from an individual subject data. Rather, it is simply a consequence of absence of proper analysis technique that takes into consideration the uncertainties of DTI derived parameters in any dataset, including one from a single subject. It would be extremely useful to be able to carry out statistical inference on single subject DTI data as is done for group studies. This is especially true when diseases of interest present with large inter-subject variability, such as multiple sclerosis (MS) and traumatic brain injury (TBI). In addition to the regional variability in the pathologies of these disorders, there is also strong regional variability in some diffusion parameters like fractional anisotropy (FA).

In this bioengineering PhD dissertation project, new statistical frameworks were developed that can measure DTI uncertainties with high accuracy and precision and that can make statistical inferences in the DTI dataset from a single subject. Resampling statistics, a group of non-parametric statistical techniques based on data-driven simulation, form the basis of developed analyses. Modeling error propagation and probability distributions of DTI parameters in analytic forms are unknown, limited or difficult, and resampling statistics can replace analytical methods by non-parametric, empirical approaches. This project involved implementing diffusion tensor estimation, fiber tracking and resampling statistics algorithms, evaluating different algorithms by Monte Carlo simulation, and applying resampling statistics algorithms to healthy subjects and patients with multiple sclerosis and traumatic brain injury.

Chapter 2 describes the biomedical and technical background behind this project. Multiple sclerosis and traumatic brain injury are briefly reviewed with the focus on the pathological and radiological aspects. The concepts and mathematics of DTI and its derived parameters are presented, and the uncertainty estimation problem in DTI is summarized with the emphasis on the rationales of utilizing resampling techniques.

In chapter 3, I propose new DTI bootstrap algorithms that can overcome the limitations of previously proposed DTI bootstrap methods. Monte Carlo simulation experiments prove that repetition bootknife corrects the under-

estimation bias present in the previously proposed repetition bootstrap, and that model-based resampling, especially residual bootstrap, can estimate the uncertainties of DTI metrics with the best accuracy and precision. Based on the results from this chapter, I use residual bootstrap whenever the bootstrap technique is required.

Chapter 4 is the first demonstration of the utility of DTI bootstrap. The necessity of cardiac gating in DTI studies with single-shot EPI sequences has been controversial. In this chapter, I provide evidence that cardiac gating is beneficial and actually more efficient in terms of achieved level of uncertainties per unit time. The residual bootstrap algorithm makes it possible to accurately quantify the effect of cardiac gating on DTI derived parameters. The possibility that full Fourier acquisition can be a better choice compared to partial Fourier acquisition when cardiac artifact is taken into consideration is presented as well.

In chapter 5, I use residual bootstrap to propose a new voxel-wise statistical testing method in the serial DTI study of a single subject called BLADE (Bootstrap-based Longitudinal Analysis of Diffusion Estimates). BLADE is essentially a voxel-wise T-testing method where pooled standard error values are provided by bootstrap. BLADE is performed on patients with TBI in order to detect local pathological changes with statistical confidence. BLADE provides evidence that while lesions visible in the conventional MRI techniques get better with increasing FA, white matter regions near the lesion (possibly connected to

the lesion) get worse over time in terms of microstructure integrity with decreasing FA.

Chapter 6 describes an alternative approach to voxel-wise serial DTI analysis using permutation testing, another resampling technique. This approach called PERVADE (Permutation Voxel-wise Analysis of Diffusion Estimates), has advantages over BLADE in that PERVADE includes a multiple comparison correction scheme and that permutation testing is a truly non-parametric statistical inference while BLADE relies on an assumed T distribution for testing. I describe how permutation testing can be applied in the statistical testing of DTI data and present results of PERVADE detecting possible Wallerian degeneration in MS patients.

Chapter 7 summarizes my major contributions and possible future directions. DTI and next-generation diffusion MRI collectively called HARDI (High Angular Resolution Diffusion Imaging) enjoys rich information expressed through diverse scalar or vector metrics and fiber tracking, and it is expected that the need for more sophisticated statistical analyses will only increase.

Chapter 2: Background

2.1 Disease Populations Investigated

In this dissertation, I applied resampling statistical techniques to two major clinical conditions in the central nervous system (CNS), multiple sclerosis and traumatic brain injury. While these two conditions might seem very different at first glance, they share a few features that make them suitable for my investigation.

First, both include non-diffuse multi-focal brain lesions with the large inter-subject variability. Since each patient presents with a unique temporal and spatial distribution of pathologies in the brain, patients with these conditions make good candidates for single-subject serial voxel-wise analysis, less so for multiple-subject (group-wise) and / or whole-brain analysis.

Second, conventional (three-direction) diffusion weighted MRI (DWI) and DTI is more sensitive than conventional MRI in detecting brain pathologies invisible to conventional MRI techniques. There are numerous studies reported to have DWI or DTI abnormalities in the normal appearing white matter (NAWM), which is a white matter region with no visible lesions by conventional MRI [1, 2].

Lastly, patients with both conditions can suffer from irreversible deterioration of brain functions due to axonal transection followed by Wallerian degeneration (WD). While other pathologies in these diseases, such as hematoma / hemorrhage in TBI and inflammation in MS, usually subside over time leaving little or no sequela if patients can recover, regeneration of transected axons in the CNS is rare and slow if not altogether absent. Detection and quantification of axonal transection / WD is important in clinical assessment and prognosis, yet it is a challenging task since conventional MRI techniques rarely reveal the ongoing degeneration process directly.

DTI has a potential to directly assess WD. It is sensitive to changes in the white matter microstructure and it can reconstruct white matter tracts along which the degeneration is likely to happen. The ability of DTI to detect abnormalities in NAWM may partially be explained by its sensitivity to Wallerian degeneration, though this was not obvious in previous studies since other pathologies can coincide. I attempted to increase the confidence in the detection of Wallerian degeneration via incorporating both temporal (serial study) and spatial (voxel-wise study with fiber tracking) information.

2.1.1 Multiple Sclerosis

2.1.1.1 Overview

Multiple sclerosis (MS) is a chronic, inflammatory, demyelinating disease of the central nervous system (CNS). It is estimated that around 350,000 people in the United States are affected by MS, and more than 2.5 million in the world. Though survival of MS patients is only slightly below normal, MS causes considerable neurological disability in the working age group. Patients typically present with CNS lesions (or plaques) that appear as indurated (hardened) areas in pathologic specimens, after which the disease was named (sclerosis is a medical term for hardening). These lesions, characterized by combinations of inflammation, demyelination, gliosis (scarring of CNS), and axonal injury and transection [3], are readily seen in MRI, and our ability to diagnose and monitor MS has improved dramatically with the advent of MRI. MR techniques also have a huge role in monitoring the disease's natural course as well as its modification by treatment.

MS patients can suffer from a wide variety of clinical features [4]. Symptoms can arise from spinal cord (muscle weakness and painful spasm, bladder dysfunction), brain stem (double vision, dizziness), cerebellum (tremor, clumsiness), optic nerve (unilateral painful loss of vision), or cerebrum (cognitive impairment, depression). General symptoms are also common, such as pain, fatigue, and intolerance to hot temperature. These symptoms are not only diverse, but they also follow a somewhat unpredictable course, often with certain

recurring pattern. This variety of symptoms and time courses reflects the diverse spatial and temporal evolutions of MS, and the diagnosis of MS can be challenging due to the resemblance to many other neurological disorders.

Several subtypes based on the patterns of progressions has been described in order to better predict the prognosis and make therapeutic decisions in each patient. Eighty percent of MS patients are initially diagnosed with the *relapsing-remitting* subtype (RRMS), characterized by unpredictable attacks (sudden onset of symptoms, relapse), which may or may not leave permanent deficits after remission. The remaining 20% can be described as *primary progressive* subtype (PPMS), having no remission after their initial MS symptoms. Around 80% of RRMS patients later become *secondary progressive* subtype (SPMS), experiencing continuous decline in the neurologic functions between attacks and ultimately suffering from severe disability such as inability to walk. While not fulfilling the diagnostic criteria for MS, the earliest clinical presentation of RRMS is in the form of patients with *clinically isolated syndromes* (CIS) [5]. Individuals with CIS suffer from an attack suggestive of CNS demyelination, often in the form of optic neuritis or sudden one-sided muscle weakness. CIS is clinically important since treatment during the initial attack can decrease the possibility of developing MS. Subjects with CIS are investigated in the chapter 6 to find any DTI changes in the normal appearing white matter.

Huge efforts have been put into establishing the cause of MS, but it still remains elusive. The most commonly held theory is that MS is caused by our body's own immune system attacking the CNS, triggered by genetic and environmental factors. The most likely injured structures are the myelin sheaths and oligodendrocytes, a type of neuroglial cell that produces the myelin sheath and support the axons. Since the cause is not clear, there is no available prevention or cure either. Nevertheless, several therapies proved to be useful in reducing relapses, slowing progression and managing neurologic deficits. The prognosis varies highly among the patients. Factors associated with a worse prognosis, i.e. faster accumulation of disability and shorter time before progression, include having PPMS (as supposed to RRMS), old age of onset and male (more likely to have PPMS), incomplete recovery after the initial attack, and frequent relapses.

2.1.1.2 Pathology & Imaging

Conventional MRI such as T_1 , and T_2 , and proton density weighted images are sensitive for the detection MS lesions (Figure 2-1). However, studies of lesion measures yielded generally disappointing correlations with the development of clinical disability, possibly due to inability to differentiate the heterogeneous pathologic substrates of individual lesions, and to estimate tissue damage that is known to occur outside the macroscopic lesions, i.e. in the normal-appearing white matter (NAWM) and gray matter (NAGM). Furthermore, even if the degree of degeneration is determined, the associated clinical deficit may not be predictable due to the unknown functional relevance of the degenerating tissue.

This prompted the utilization of new quantitative MR techniques that can overcome these limitations, and diverse new techniques such as MR spectroscopy, diffusion MRI, cell-specific imaging, myelin water imaging, ultra-high field MRI, and functional MRI are currently under active research with the hope of becoming more specific markers of various MS-related pathologies. Imaging for three key pathologic aspects of MS, i.e. inflammation, demyelination, and axonal injury, is described below.

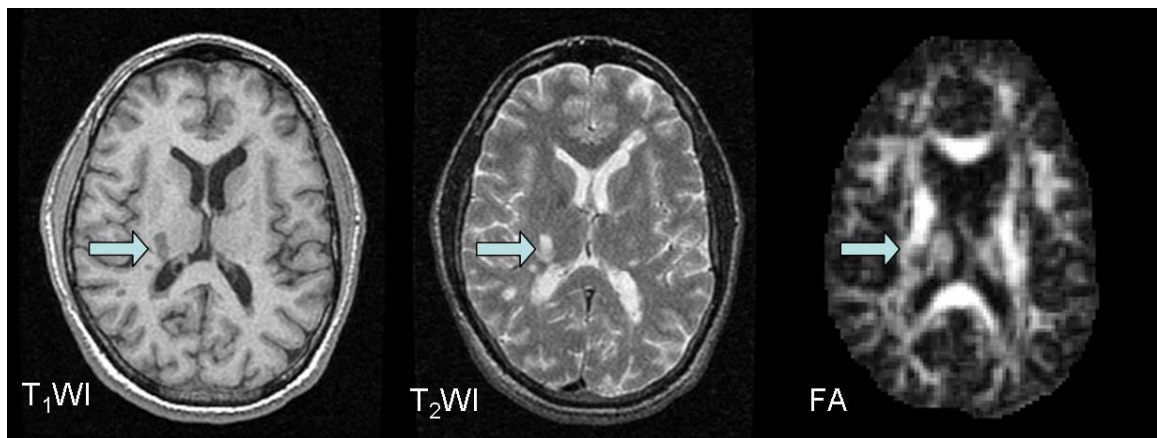


Figure 2-1. MRI images from a MS patient. Arrow: MS lesion. T₁WI: T1 weighted image. T₂WI: T2 weighted image. FA: fractional anisotropy of DTI.

Inflammation is a complex process involving multiple cellular and humoral components and cell trafficking through the blood-brain barrier (BBB). MRI has been able to observe inflammation only indirectly, most successfully in the disturbance of the BBB. Inflammatory cells induce loosening of the BBB, i.e. capillary endothelial cells connected together with tight junctions, in order to allow infiltration of inflammatory cells inside the brain parenchyma. Gadolinium (Gd) based MRI contrast agents infiltrate only disrupted BBBs and invade the

parenchyma allowing Gd contrast-enhanced MRI to be used as a surrogate inflammation marker. However, macroscopically visible inflammation does not represent the full inflammatory process that includes microscopic and perhaps more diffuse process. Cell-specific imaging based on ultra small particles iron oxide (USPIO) [6, 7] is a promising new technique that is sensitive to a different aspect of inflammation, macrophage migration.

Demyelination, loss of the myelin sheath insulating axons, is the hallmark pathology of MS and some other immune-mediated neurodegenerative diseases. It causes an array of neurological symptoms by disrupting efficient nerve conduction. While there are several MR techniques that are sensitive to demyelination, such as magnetic transfer MRI, diffusion MRI, proton MR spectroscopy, and conventional MR relaxation times, these methods are also sensitive to other pathologies, making myelin-specific measurements challenging or unlikely. Myelin water fraction imaging has been reported to distinguish water signals coming from CSF, intra / extra-cellular, and myelin by analyzing T_2 relaxation, and it has the potential to provide more specific quantification of degree of myelination [8, 9].

Axonal injury, previously under-emphasized, has recently drawn attention as an early cause of neurologic disability in MS. Pathological studies have found axonal loss in the NAWM as well as inside lesions. Axonal injuries within lesions are likely to be caused by inflammation, but the mechanism behind NAWM axonal

damage remains unclear. It simply might be Wallerian degeneration from axonal transection occurred in lesions some distance away. However, some investigators raised the possibility of unidentified global axonopathy directly damaging NAWM axons independent of lesions [10]. Thus, non-invasive imaging techniques that can accurately localize and quantify axonal injury and degeneration is critical in elucidating the pathogenesis of NAWM axonal damage and better estimating overall irreversible neurological damage. The pathology and imaging of axonal injury and Wallerian degeneration are further discussed in section 2.1.3.

2.1.2 Traumatic Brain Injury

2.1.2.1 Overview

Traumatic brain injury (TBI) is a common and often devastating problem in modern society, especially for the population under age 45. It is estimated that TBI occurs in 1.5 to 2 million people in the United States every year. Around 2% of the entire population of the USA are thought to live with disabilities caused by TBI [11]. Both young (teenagers through twenties) and old (older than 70) age groups have especially high incident rates. Motor vehicle accidents are the most common cause with young patients, while elderly population suffer from accidental falls most frequently. Abuse and neglect are also common causes with children.

TBI is typically classified by the severity and the nature of injury. The severity of TBI is usually based on the Glasgow Coma Scale (GCS), a clinical scoring system between 3 (worst) and 15 (best) that combines eye, verbal and motor responses. Eighty percent of the patients are mild (GCS 13 to 15), 10-13% moderate (GCS 9 to 12), and 7-10% severe (GCS 3 to 8) [12]. TBI can also be classified into primary and secondary injuries [11]. Primary injuries occur as direct result of the trauma and they include hemorrhage / hematoma, cortical contusion, axonal shearing injury and vascular injury. Secondary injuries are caused by complications of primary injuries and they are preventable. Acute / subacute secondary injuries include cerebral edema, ischemia, and brain herniation. Hydrocephalus, cerebrospinal fluid (CSF) leak, leptomeningeal cyst and encephalomalacia can also follow as chronic secondary injuries.

TBI presents with symptoms and complications of a wide variety in type, severity and duration, ranging from mild headaches to coma and brain death. Possible outcomes are impairments in movement (e.g. difficulty walking), sensation, cognition (e.g. memory loss, poor judgment), mood (e.g. depression, agitation, impulsiveness), and sexual function. Patients can also suffer from seizures and psychotic symptoms such as hallucination and delusion. Since little can be done to reverse the initial brain damage, initial treatments are focused on stabilizing patients and preventing secondary injuries, such as brain swelling. Surgery might be required to decrease the intracranial pressure and to remove / repair hematomas or contusions.

2.1.2.2 Pathology & Imaging

As described above, diverse types of injuries can occur with TBI, such as fractures and hemorrhages as primary injuries and cerebral swelling and herniation as secondary injuries. While current clinical imaging techniques such as CT and conventional MRI are capable of detecting most of these pathologies, their sensitivity to diffuse axonal injury (DAI) is limited. As such, DAI is often under-diagnosed and accurate evaluation of DAI using more advanced imaging techniques is of special interest. Chapter 5 includes studies of patients suspected of having cortical contusions and DAI and therefore these pathologies are described in detail below.

Cortical contusion is a focal brain lesion mainly in the cortical gray matter. The parts of the brain in close contact with the rough surface on the inner skull table are commonly involved. Cortical contusions are thought to be caused by the brain tissue colliding against the inner surface of the skull. On histology, hemorrhage and edema are the seen in the earliest stage, followed by signs of injuries in the neuronal cell bodies and axons. On CT, contusions associated with hemorrhages are easily identified but non-hemorrhagic contusions can be difficult to detect until edema develops. On MRI, contusions appear as ill-defined regions of variable intensity. Contusions usually have better prognoses than DAIs.

As the word 'diffuse' suggests, diffuse axonal injury (DAI) occurs in a more widespread area than focal injuries such as contusion. The affected areas are often distant from the site of direct impact, and two thirds of DAI lesions occur near the junctions of gray and white matter. DAI is caused by rotational acceleration and deceleration forces that produce shearing deformation of the brain leading to axonal transection. It was once believed that traumatic mechanical forces themselves are the main cause of the axonal transection. However, recent studies show that secondary reactions occurring hours to days after the trauma in response to the primary injury are largely responsible for the transection [13]. Disruption of the cytoskeleton structures during the stretching of axons are thought to initiate the biochemical processes that eventually separate the axons.

On histology, axonal swelling and transection are observed initially. Axonal swelling occurs at the site of damaged cytoskeleton due to failed axonal transport. Large enough swellings tear the axon and form a bulb called the retraction ball. Transected axons are followed by Wallerian degeneration. Small focal hemorrhages are sometimes accompanied in DAI. On imaging, CT is limited for the diagnosis of DAI since CT can only detect hemorrhages, not axonal injuries directly. Gradient-echo T_2^* -weighted MRI depicts more DAI lesions than CT and has higher sensitivity to hemorrhage. FLAIR MRI can detect many non-hemorrhagic foci of DAI. Unfortunately, FLAIR MRI still underestimates the true extent of DAI since it is not sensitive to the some axonal pathologies [11].

2.1.3 Axonal Transection and Wallerian Degeneration

2.1.3.1 Pathology

Once a neuronal fiber is severed either by traumatic or non-traumatic injury, the distal (i.e. away from the neuronal cell body) part of the axon / myelin unit degenerates, a process called Wallerian degeneration (WD). It is also known as anterograde or secondary degeneration. First, the axonal membrane swells and breaks apart. The axonal cytoskeleton and inner organelles undergo disintegration as well. This axonal degeneration is a rapid process, taking around 24 hours in the peripheral nervous system (PNS) and somewhat longer in the CNS. The second step is the clearance of the myelin sheath. Schwann cells that are responsible for producing and supporting the myelin sheath in the PNS are actively involved in this process, and the myelin clearance in the PNS is very fast, usually completed within a few weeks. On the other hand, oligodendrocytes that assume similar roles in the CNS are thought to be less involved in this process. The BBB in the CNS also hinders the recruitment of macrophages away from the transected nerve. Hence, non-functioning axon-less myelin can remain for an extended time in the CNS, as long as years. Third, regeneration follows degeneration, but only in the PNS. Regeneration is either absent or very slow in the CNS, making any axonal transection an irreversible and permanent neurologic damage.

Inefficient myelin clearance in the CNS is significant in a few aspects. First, it is suspected to be one of major reasons for the lack of regeneration following WD [14]. Second, it contributes to a grossly normal appearance of the affected white matter, both in pathologic and radiologic studies [15]. Though the microstructure of myelin sheath is altered, forming hollow (missing axons inside) and segmented structures called myelin ovoids, these features are not easily appreciated either in routine histologic studies or by most MR techniques.

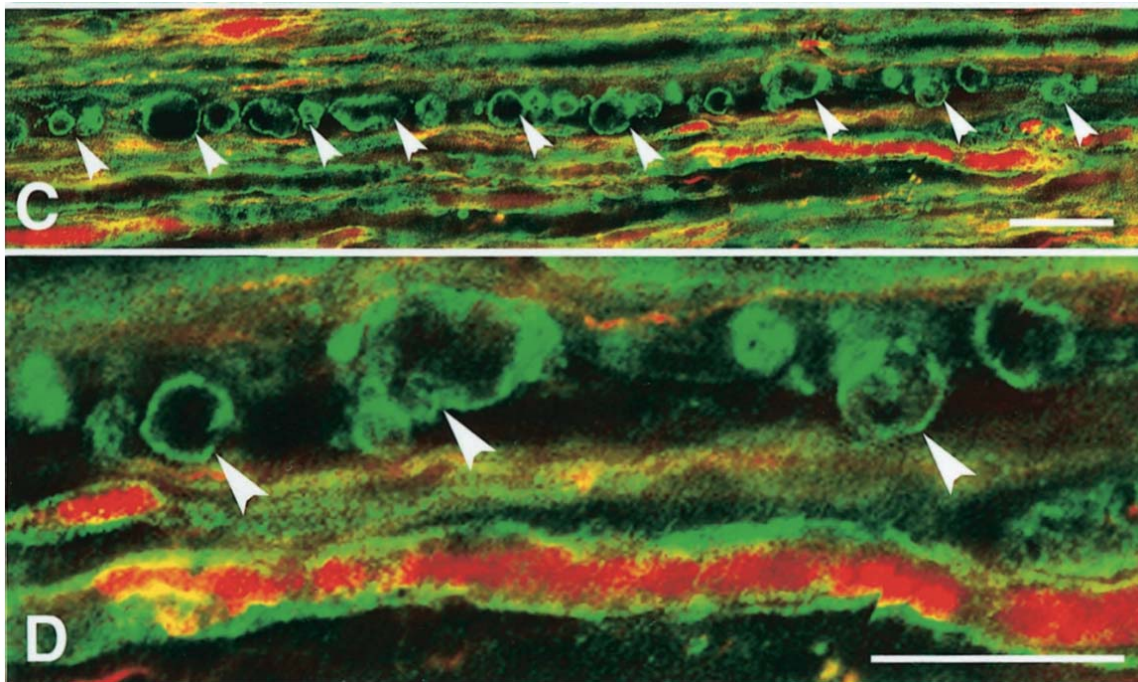


Figure 2-2. Wallerian degeneration of spinal cord axons. Confocal images of longitudinal sections of neuronal fibers. Myelin is stained as green, and axonal neurofilaments are stained as red. In the upper image (C), fibers undergoing denervation appears as rows of myelin ovoids (arrows) between intact myelinated axons. In the magnified image below (D), it is clear that myelin ovoids lack axonal structures indicated by lack of red staining. Scale bar = 25 μ m. This image is taken from [15].

2.1.3.2 Imaging

As discussed above, axonal transection and WD is difficult to detect, unless in later stages when the myelin debris is eventually cleared and white matter tracts undergo atrophy. Atrophy measurement based on conventional MRI such as high resolution T_1 weighted images is currently the most common approach to study CNS degeneration. With advancement in data acquisition and analysis techniques, atrophy has been shown to start very early in MS, even at presentation with a CIS [16] (and Roland's VBM paper at JNNP) and that atrophy measurement have a better correlation with clinical disability compared to lesion loading [17]. However, atrophy measurement 1) cannot measure degeneration when myelin debris still remains, 2) is less than optimal in localizing the atrophy, though some progress has been made in measuring atrophy in specific gray or white matter regions, 3) can easily be confounded by other factors such as fluid content, proliferation of other cells such as gliosis, and demyelination without degeneration, and 4) the predictive power for prognoses is still weak.

N-Acetyl Aspartate (NAA) measured by Proton (^1H) MR Spectroscopy (MRS) reflects the neuronal metabolic integrity as well as neuroaxonal density. Reduced NAA level in the NAWM of MA patients has been used as evidence of widespread pathology outside lesions [18, 19], possibly neurodegeneration either primary or secondary. However, this NAA alteration can resolve with time [20], suggesting that NAA reduction can also be explained by reversible axonal dysfunction or metabolic changes. Limited spatial resolution is another drawback

of NAA measurement. Gadofluorine M, a novel gadolinium-based MRI contrast agent, selectively accumulates in nerve fibers undergoing WD [21], has the potential to provide even more specific information about axonal injury, highlighting only irreversible damages with high spatial resolution. However, safety issue needs to be addressed and more human trials are needed to demonstrate efficacy.

Lastly, DTI also has the potential to be useful in the characterization of neurodegeneration. DTI is sensitive to changes in the microstructure, possibly including the myelin ovoids in the earlier stage of WD. DTI fiber tracking can help narrow down the search region where WD is most likely (i.e. white matter tracts connected to a lesion) in order to increase the power of detection via either DTI derived parameters or other measurements such as NAA. In this dissertation, only DTI is utilized, but combining DTI with other techniques would be ideal considering that DTI is not highly specific to WD.

2.2 Diffusion Tensor MRI (DTI)

2.2.1 Diffusion

Diffusion is a result of thermally driven random translational motions of molecules called Brownian motion. These motions cannot be predicted since they are random events governed by thermodynamics, collisions with other molecules and the structures in the medium. Still, we can describe the probabilistic distribution of the translation of a molecule after certain time, and collective behaviors of these molecules in large numbers become predictable simply as a statistical outcome. This macroscopic aspect of diffusion is usually defined as the spontaneous net movement of particles from an area of high concentration to an area of low concentration in a given volume of fluid (medium) down the concentration gradient, a process that will continue until the concentrations are equalized. An example is a drop of ink within a glass of water which will diffuse until the ink is evenly mixed with the water.

In diffusion MRI, though, we are not looking at a certain species of solute diffusing within a water media. Even without the presence of certain solutes, the water molecules are diffusing amongst themselves, under the same principle of Brownian motion as ink molecules within the water. This process is called self diffusion. Unlike a typical diffusion process involving separate solute and medium, self diffusion does not demonstrate any net effect that can be observed, such as

the spread of ink droplet inside water. Even with water molecules moving around within water, they are still collectively water in the same state. To make water self diffusion an observable phenomenon in the diffusion MRI experiments, we *tag* water protons so that water molecules become heterogeneous and concentration gradients are created. The diffusion MRI concept is further described in the section 2.2.2. From now on, the word *diffusion* and *self diffusion* will be used interchangeably.

It has been proved that in case of free diffusion, i.e. diffusion without hindrance or restriction, the probabilistic distribution function (*pdf*) of the 3x1 position vector \mathbf{r} after time τ ($\mathbf{r} = \mathbf{r}_0$ when $\tau = 0$) is a Gaussian function as below.

$$p(\mathbf{r} | \mathbf{r}_0, \tau) = \frac{1}{\sqrt{(4\pi D\tau)^3}} \exp\left(-\frac{(\mathbf{r} - \mathbf{r}_0)^T (\mathbf{r} - \mathbf{r}_0)}{4D\tau}\right) \quad \text{Equation 2-1}$$

In this equation, D is the diffusion coefficient, a scalar value that determines the amount of displacement in a time τ . D can be estimated by an equation proved by Einstein as below,

$$D = \frac{RT}{6\pi kPN} \quad \text{Equation 2-2}$$

where R is the universal gas constant, T is the absolute temperature, N is Avogadro's number, k is the viscosity, and P is the radius of molecules.

2.2.2 MR Measurement of Diffusion

Diffusion is typically measured by adding a pair of diffusion-sensitive gradients to the spin echo pulse sequence, termed pulsed-gradient spin echo (PGSE). A typical 90° and 180° radiofrequency pulse spin echo sequence partially undoes the signal attenuation at the echo time TE by reversing the proton spin dephasing caused by time-independent field inhomogeneity. Signals from the spin echo sequence are T_2 weighted, meaning that signals are reduced only by irreversible spin-spin interactions (T_2 effect) but not by both T_2 effect and the reversible effect of field inhomogeneity on stationary spins (T_2' effect). Even with the spin echo sequence, though, signals can be attenuated by the T_2' effect since the diffusion of protons results in incomplete rephasing in an inhomogeneous field. Diffusion was once considered an artifact of the spin echo experiment, but is now exploited to measure diffusion by artificially creating field inhomogeneities.

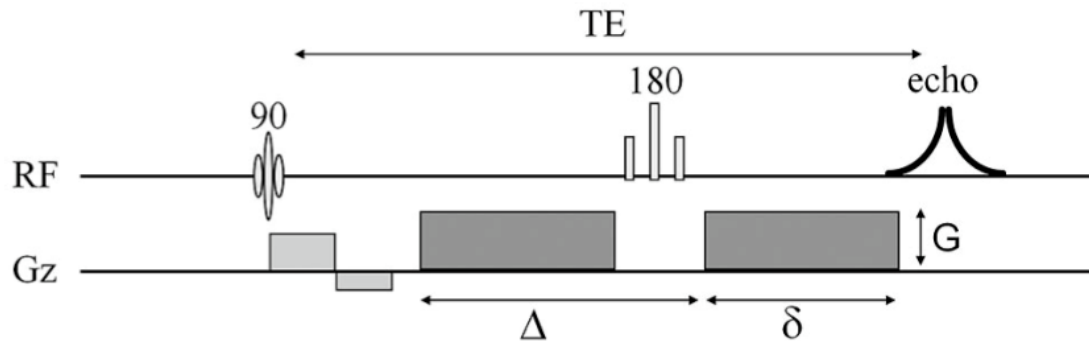


Figure 2-3. Pulse sequence diagram for a diffusion-weighted acquisition. A pair of diffusion-sensitizing gradients (dark gray) are added to a spin echo sequence. This image is taken from [22].

The mechanism of diffusion weighting by PGSE is the following (Figure 2-3). A 90° pulse excites the proton spins and puts the spins in phase with each other.

Immediately after the pulse, they start dephasing by T_2^* effects, which is a combination of T_2 and T_2' effects. Next, the first of the pair of diffusion-sensitive gradients is turned on, artificially creating a strong field inhomogeneity that increases the T_2' effect. This inhomogeneity can be made reversible by applying the second of the pair of gradients after a 180° pulse, with exactly the same properties (strength, duration and shape) as the first gradient. Just like other static unintended field inhomogeneities (caused by improper shimming, for instance), spin dephasing caused by diffusion gradients can be undone assuming that there is no diffusion (or other movement of the spins). However, diffusion will cause dephasing that will not be reversed by 180° pulse, and the amount of irreversible dephasing (and signal attenuation due to dephasing) depends on the *pdf* of water diffusion (including the amount of diffusion and any other incoherent intravoxel motion) and on the properties of the diffusion gradients.

In the case of imaging free diffusion with a PGSE experiment, the MR signal attenuation is expressed as,

$$S(b) = S_0 \exp(-bD) \quad \text{Equation 2-3}$$

where $S(b)$ is the signal with diffusion weighting defined by b , S_0 is the signal without diffusion weighting, D is the diffusion coefficient. The b -value, b in the equation above, can be calculated by the following equation,

$$b = |\gamma G \delta|^2 (\Delta - \delta / 3) \quad \text{Equation 2-4}$$

where γ is the Lamour constant, G is the diffusion gradient strength, δ is the diffusion gradient duration, and Δ is the time from the start of the first diffusion gradient to the start of the second one (for δ and Δ , see Figure 2-3). From Equation 2-3, we can simply calculate the diffusion coefficient D as the following.

$$D = -b^{-1} \ln(S(b)/S_0) \quad \text{Equation 2-5}$$

From D and $\tau = \Delta - \delta/3$, we can completely describe the *pdf* of free diffusion in the isotropic 3D Gaussian function as shown in Equation 2-1.

2.2.3 Diffusion in the Biological Tissue

Unlike free diffusion, the diffusion in the biological tissues is hindered and / or restricted by various microstructures such as axonal membrane and myelin sheath. The *pdf* of diffusion no longer follows the Gaussian formula shown in Equation 2-1, and the diffusion coefficient D is not strictly defined. Nevertheless, we can perform MR experiments of diffusion as described above by approximating the *pdf* of hindered diffusion by Equation 2-1. By this, we can dramatically reduce the complexity of the problem of *pdf* measurement to a simple estimation of D as with free diffusion, and we can come up with a single value that roughly represents the amount of hindered / restricted diffusion. This simplification made diffusion-weighted MRI feasible. On the other hand, a complete description of the *pdf* with q-space formalism [23] has not been used in in-vivo imaging until recently due to more rigorous technical requirements.

The diffusion coefficient in the measurement of hindered / restricted diffusion with Gaussian approximation is called the *apparent* diffusion coefficient (ADC) so that it can be distinguished from the diffusion coefficient defined under conditions of free diffusion. ADC values are smaller than what is predicted by Equation 2-2, reflecting the effects of microstructures that act as barriers of diffusion. As much as the utility of ADC values, it was recognized soon after the emergence of DWI that measurement of ADC by two images (corresponding to $S(b)$ and S_0 in the Equation 2-2) was problematic. The diffusion coefficient D in the Equation 2-1 is isotropic, and the measured D should be independent of measured direction of diffusion. On the other hand, the measured ADC turned out to be strongly anisotropic in nervous tissue [24]. Thus, it has been a common practice to acquire four images (one S_0 and three $S(b)$ in x , y , and z directions of MRI coordinate system) and calculate a rotationally invariant ADC value by averaging ADC measured in x , y , z directions as below, undoing the anisotropic property of measured ADC.

$$ADC = (ADC_x + ADC_y + ADC_z) / 3 \quad \text{Equation 2-6}$$

$$\text{where } ADC_i = -b^{-1} \ln(S(b_i) / S_0) \text{ (i=x, y or z)} \quad \text{Equation 2-7}$$

2.2.4 Basic Concepts of DTI

Measurement of anisotropic diffusion [25] within a mathematically rigorous framework was made feasible with the introduction of DTI [26, 27]. DTI explicitly

models anisotropic diffusion with a more generic multivariate (x, y, and z) Gaussian function shown below (extension of Equation 2-1).

$$p(\mathbf{r} | \mathbf{r}_0, \tau) = \frac{1}{\sqrt{|\mathbf{D}|}(4\pi\tau)^3} \exp\left(-\frac{(\mathbf{r}-\mathbf{r}_0)^T \mathbf{D}^{-1}(\mathbf{r}-\mathbf{r}_0)}{4\tau}\right) \quad \text{Equation 2-8}$$

In this equation, \mathbf{D} is a diffusion tensor, 3x3 diagonally symmetric matrix.

$$\mathbf{D} = \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{xy} & D_{yy} & D_{yz} \\ D_{xz} & D_{yz} & D_{zz} \end{bmatrix} \quad \text{Equation 2-9}$$

\mathbf{D} describe ADC values separately along x, y, and z directions (diagonal elements, i.e. D_{xx} , D_{yy} and D_{zz}) as well as the correlation between these (off-diagonal elements, i.e. D_{xy} , D_{xz} and D_{yz}). In fact, the covariance matrix \mathbf{C} of the pdf of diffusion in the Equation 2-8 is $\mathbf{C} = 2\mathbf{D}\tau$.

The MR signal attenuation is the extension of the Equation 2-3 as the following.

$$S(b, \mathbf{g}) = S_0 \exp(-b\mathbf{g}^T \mathbf{D} \mathbf{g}) \quad \text{Equation 2-10}$$

Here, \mathbf{g} is the 3x1 unit vector of the direction of diffusion gradient, usually a combination of x, y and z gradients. The gradient strength term is included in the b-value as shown in Equation 2-4. In Equation 2-10, there are seven unknown parameters, the six unique elements of \mathbf{D} and S_0 . Thus, in order to estimate \mathbf{D} , we need at least seven different equations that can be acquired by varying either b or \mathbf{g} (typically one image with $b \approx 0$, and the remaining six images with $b > 0$ and non-collinear \mathbf{g}), though often many more images are acquired in order to better estimate \mathbf{D} . \mathbf{D} is commonly solved by linear regression with weighted least

squares after log transformation of the above equations (this procedure is further described in the chapter 3).

2.2.5 DTI Derived Parameters

D is an ellipsoidal approximation to the diffusion pdf and therefore can be rotated so that ADC values along the new three principal axes are not correlated with each other. The rotated matrix will have zero values for all off-diagonal elements, and thus this process is called diagonalization shown in the equation below.

$$\mathbf{\Lambda} = \mathbf{R} \cdot \mathbf{D} \cdot \mathbf{R}^T \quad \text{Equation 2-11}$$

$$\text{where } \mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}, \mathbf{R} = \begin{bmatrix} \varepsilon_{1x} & \varepsilon_{2x} & \varepsilon_{3x} \\ \varepsilon_{1y} & \varepsilon_{2y} & \varepsilon_{3y} \\ \varepsilon_{1z} & \varepsilon_{2z} & \varepsilon_{3z} \end{bmatrix} \quad \text{Equation 2-12}$$

$\mathbf{\Lambda}$ is the diagonal matrix with three eigenvalues, and \mathbf{R} is the rotation matrix where each column is the eigenvector associated with the corresponding eigenvalues ($\lambda_1 > \lambda_2 > \lambda_3$). These three eigenvectors and eigenvalues describe the directions and ADC of three principle axes of the diagonalized **D**. Eigenvectors and eigenvalues are rotationally invariant and do not depend on the MR coordinate system. Therefore, eigenvectors and eigenvalues themselves or any other parameters derived from these (described below) can be estimated quite robustly regardless of relative positioning of subjects inside MR system.

Various scalar parameters have been defined using eigenvalues, and the parameters that are most commonly used are the following.

$$D_{av} = (\lambda_1 + \lambda_2 + \lambda_3) / 3 \quad \text{Equation 2-13}$$

D_{av} is an average ADC, and it is equivalent to the average ADC in the conventional three-directional DWI experiments.

$$FA = \sqrt{\frac{3}{2}} \sqrt{\frac{(\lambda_1 - D_{av})^2 + (\lambda_2 - D_{av})^2 + (\lambda_3 - D_{av})^2}{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}} \quad \text{Equation 2-14}$$

Fractional anisotropy (FA) is the most popular index of the degree of anisotropy. FA is a unitless entity ranging between 0 (isotropic) and 1 (completely anisotropic). FA is conceptually similar to coefficient of variation between the three eigenvalues.

$$\lambda_{\parallel} = \lambda_1 \quad \text{Equation 2-15}$$

λ_{\parallel} (lamda parallel, also known as axial or longitudinal diffusivity, is the ADC along the primary (largest) eigenvector direction.

$$\lambda_{\perp} = (\lambda_2 + \lambda_3) / 2 \quad \text{Equation 2-16}$$

λ_{\perp} (lamda perpendicular, also known as radial or perpendicular diffusivity) is the average of the ADCs in the directions perpendicular to primary eigenvector.

In coherent neuronal fiber bundles, the primary eigenvector tends to align parallel to the axons while the secondary and tertiary eigenvectors are perpendicular to the axons. Thus, the eigenvectors are helpful in inferring the course of white matter pathways and techniques that use this information to delineate this course across voxels is called fiber tracking, and is discussed in the following section. Regarding parameters derived from eigenvalues, numerous studies on MS and

TBI patients have shown that D_{av} is increased and FA is decreased in the NAWM as well as lesions, proving that DTI is able to detect the alterations of microstructure that conventional MRI is missing [1, 2]. Recently, there have been efforts to investigate axial and radial diffusivity as well. Studies on mice brain suggests that λ_{\parallel} is specific to axonal degeneration and λ_{\perp} is specific to demyelination in the highly ordered axonal bundle structures [28], and overall increase of λ_{\perp} is observed in the MS patients [29].

2.2.6 Fiber Tracking

Fiber tracking [30] is a technique of reconstructing plausible white matter pathways using the 3D diffusion profile in each voxel provided by DTI or more sophisticatedly by HARDI (High Angular Resolution DWI). The basic idea is to connect from voxel to voxel following the direction indicative of axonal bundle (such as primary eigenvector for DTI). There are diverse approaches in the actual implementation of this concept, such as either deterministic (do not consider uncertainty in the direction) or probabilistic, how to follow the tensor field (for DTI), the minimum conditions for tracking, etc. Since this technique is the only way to estimate white matter pathways in vivo, validation is difficult yet crucial. With supports from direct (such as comparing with the electrostimulation in human [31]) and indirect (prior neuroanatomy knowledge) evidence, fiber tracking is generally accepted as a valuable tool in studying connectivity and delineating specific tracts. In this dissertation, fiber tracking was an essential tool

to select regions likely to be connected to lesion, or more generally, regions that are likely to be closely connected.

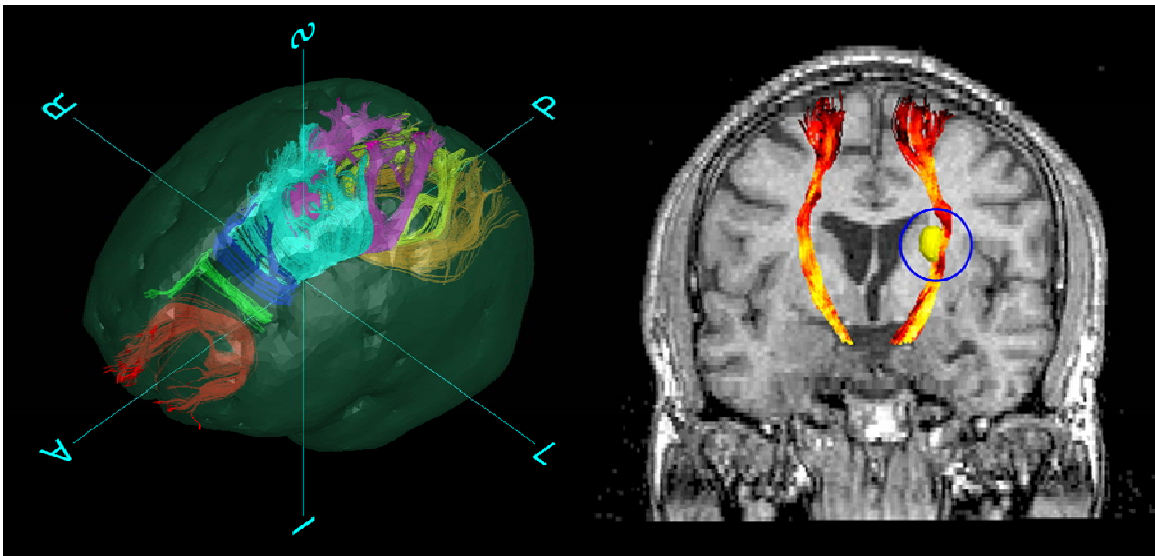


Figure 2-4. Examples of 3D rendered fiber tracking results. Left image shows corpus callosum fibers segmented by connectivity to distinct Brodmann areas (antero-left-superior view). Right image shows corticospinal tract (CST) fibers in a MS patient (anterior view). A MS lesion including a part of CST is also rendered (a yellow mass inside the blue circle). By fiber tracking, we can selectively study white matter areas connected to a known pathologic region.

2.2.7 Uncertainty Estimation of DTI Derived Parameters

In order to properly interpret experimental data, we need to know how different our measurements could be to the unknown true value. This degree of potential difference is typically referred to as uncertainty or error. It can simply be described by a single number (mostly commonly standard error) or it can take more complicated forms such as confidence intervals or probability distributions, depending on the intended use of the uncertainty information. Multiple methods for estimating the uncertainties of diverse DTI derived parameters have been

reported, and they can be roughly classified as either analytical or numerical approaches.

2.2.7.1 Analytical Approach

The analytical approach attempts to derive closed-form expressions for errors in DTI derived parameters [32-35]. Noise in the measured diffusion signals are related to uncertainties in the diffusion tensor elements by propagation of errors [36]. These errors of the tensor elements are then related to DTI derived scalar or vector parameters by combinations of perturbation and propagation of errors methods. While the analytical approach provides a way to simply calculate the error with minimal computation effort as long as the equations are available, there are several limitations that make this approach impractical.

First, the analytical approach relies on the assumption that the only source of error is the Gaussian noise in the MR signals. In reality, MR images are affected by various artifacts such as subject head motion, cardiac pulsation artifacts, ghosting, and eddy current distortion, that contribute significantly to the overall uncertainties. Even if we assume that thermal Gaussian noise is the only source of error, this noise is spatially dependent (more pronounced with surface coils) and there is a compelling need to estimate this noise voxel-wise, which is an issue that has been neglected in the studies utilizing the analytical approach. Second, perturbation theory relies on a small angle approximation, which makes the estimated errors unreliable when these angles are relatively large. Third, the

analytical approach can only estimate errors with single values (such as standard error) since no analytic formula exist for the *pdf* of DTI derived parameters except for D_{av} .

2.2.7.2 Numerical Approach

The numerical approach depends on modern computer power and algorithms rather than mathematical derivations. If the probabilistic distribution of the source of noise is known (or is assumed to be known) then Monte Carlo (MC) simulation studies can reproduce the probabilistic distributions for any DTI derived parameters. MC simulation works even if no closed analytic formulas exist for the distributions and the accuracy of results depends on the number of trials. MC studies have been used extensively to analyze how the noise affects DTI derived parameters [37-39], and they also serve as the gold standard in evaluating other uncertainty estimation approaches under the assumption of known distributions of noise sources [40-42].

When the assumption of a known noise source distribution no longer holds, which is the case with any real in vivo studies, the numerical approach cannot directly replicate the noise propagation process as is done with MC studies. In these cases the computer simulation process has to be based on the data themselves, and two quite different techniques are available, Bayesian and bootstrap. Both methods need the data to be over-sampled, a requirement easily met with the recent trend of DTI acquisitions with many diffusion-encoding

directions. Bayesian methods [40, 43] can accurately estimate the *pdf* of DTI derived parameters but it is very computationally intensive and the noise has to be mathematically modeled (strictly speaking, the source of noise is not unknown completely but only the parameters defining the model of the source of noise is unknown). Bootstrap [41, 44-46], an empirical, non-parametric, resampling statistical technique, does not require modeling of the source of noise unlike the Bayesian method. Implementing the bootstrap algorithm is straightforward and the computational burden is relatively light, especially compared to Bayesian approaches. For these reasons, bootstrap was further developed and applied in this dissertation.

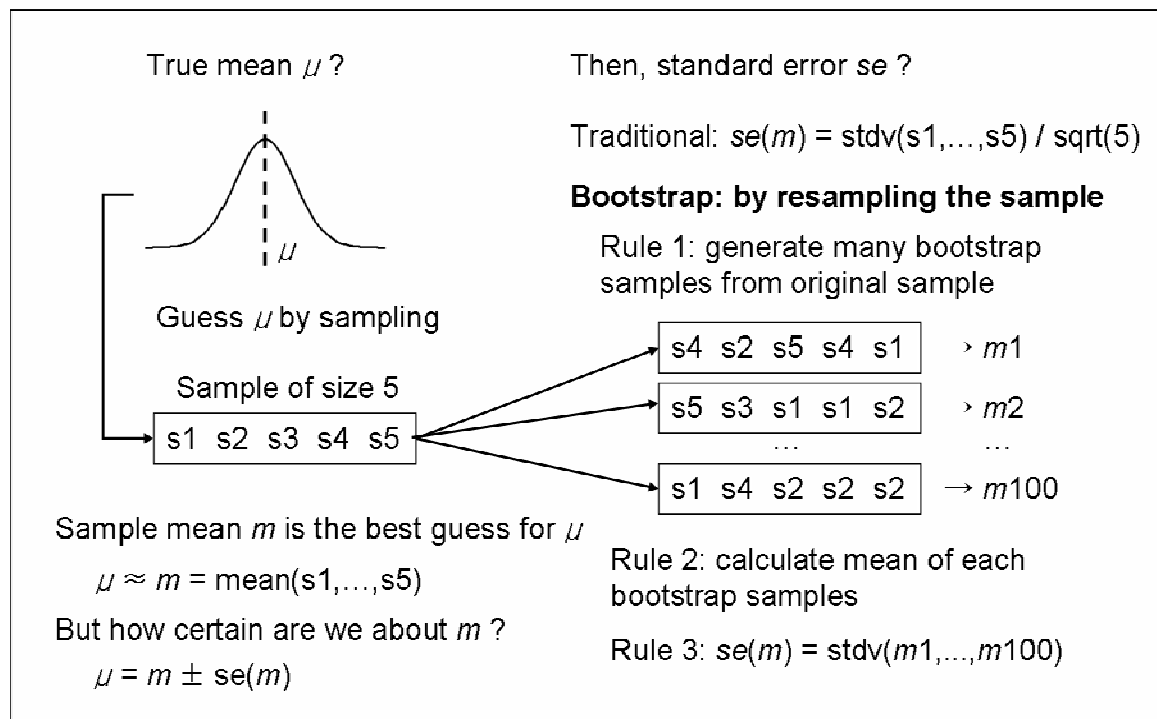


Figure 2-5. Schematics of basic concept of bootstrap.

In the simple scenario with a sample drawn from a population (Figure 2-5), bootstrap samples with the same size as the original sample are drawn from the original sample with replacement (i.e. same element of the original sample can be drawn multiple times) and the statistics of interest (such as mean) from each bootstrap samples are calculated. The distribution, then, is used to infer uncertainty for a given statistic of the original sample, such as standard errors, bias, confidence interval, and an empirical *pdf*. In reality, bootstrap is not defined as a single exact algorithm but rather exists as a set of algorithms that can be applied for specific data types. For DTI, implementation of bootstrap was originally proposed in terms of resampling separately for different samples of each direction [46], which I will refer to as repetition bootstrap. Alternative bootstrap approaches exist which are described in detail in chapter 3.

In the chapter 6, permutation testing, another resampling technique is developed for DTI serial analysis. Statistical testing based on bootstrap has to assume that the null distribution of our statistic of interest (such as difference in certain DTI derived parameters) closely follow a Gaussian distribution. Permutation testing estimates the null distribution directly from the data instead of making assumptions about this distribution and it may prove to be more accurate in the estimating the p-values. Lastly, probably the ultimate method of uncertainty estimation would be repeated measures, such as acquiring the same dataset a hundred times. Though it can serve as a gold standard, it is not feasible due to time constraints.

2.3 Reference

1. Belanger, H.G., et al., *Recent neuroimaging techniques in mild traumatic brain injury*. J Neuropsychiatry Clin Neurosci, 2007. **19**(1): p. 5-20.
2. Rovaris, M., et al., *Diffusion MRI in multiple sclerosis*. Neurology, 2005. **65**(10): p. 1526-32.
3. Frohman, E.M., M.K. Racke, and C.S. Raine, *Multiple sclerosis--the plaque and its pathogenesis*. N Engl J Med, 2006. **354**(9): p. 942-55.
4. Compston, A. and A. Coles, *Multiple sclerosis*. Lancet, 2002. **359**(9313): p. 1221-31.
5. Miller, D., et al., *Clinically isolated syndromes suggestive of multiple sclerosis, part I: natural history, pathogenesis, diagnosis, and prognosis*. Lancet Neurol, 2005. **4**(5): p. 281-8.
6. Rausch, M., et al., *MRI-based monitoring of inflammation and tissue damage in acute and chronic relapsing EAE*. Magn Reson Med, 2003. **50**(2): p. 309-14.
7. Dousset, V., et al., *MR imaging of relapsing multiple sclerosis patients using ultra-small-particle iron oxide and compared with gadolinium*. AJNR Am J Neuroradiol, 2006. **27**(5): p. 1000-5.
8. Laule, C., et al., *Water content and myelin water fraction in multiple sclerosis. A T2 relaxation study*. J Neurol, 2004. **251**(3): p. 284-93.
9. Laule, C., et al., *Magnetic resonance imaging of myelin*. Neurotherapeutics, 2007. **4**(3): p. 460-84.
10. Filippi, M. and M.A. Rocca, *MRI evidence for multiple sclerosis as a diffuse disease of the central nervous system*. J Neurol, 2005. **252** Suppl 5: p. v16-24.
11. Le, T.H. and A.D. Gean, *Imaging of head trauma*. Semin Roentgenol, 2006. **41**(3): p. 177-89.
12. Arciniegas, D.B., K. Held, and P. Wagner, *Cognitive Impairment Following Traumatic Brain Injury*. Curr Treat Options Neurol, 2002. **4**(1): p. 43-57.
13. Arundine, M., et al., *Vulnerability of central neurons to secondary insults after in vitro mechanical stretch*. J Neurosci, 2004. **24**(37): p. 8106-23.
14. George, R. and J.W. Griffin, *Delayed macrophage responses and myelin clearance during Wallerian degeneration in the central nervous system: the dorsal radicotomy model*. Exp Neurol, 1994. **129**(2): p. 225-36.
15. Bjartmar, C., et al., *Axonal loss in normal-appearing white matter in a patient with acute MS*. Neurology, 2001. **57**(7): p. 1248-52.
16. Miller, D.H., et al., *Measurement of atrophy in multiple sclerosis: pathological basis, methodological aspects and clinical relevance*. Brain, 2002. **125**(Pt 8): p. 1676-95.
17. Kalkers, N.F., et al., *Optimizing the association between disability and biological markers in MS*. Neurology, 2001. **57**(7): p. 1253-8.
18. De Stefano, N., et al., *Diffuse axonal and tissue injury in patients with multiple sclerosis with low cerebral lesion load and no disability*. Arch Neurol, 2002. **59**(10): p. 1565-71.

19. Pelletier, D., et al., *MRI lesion volume heterogeneity in primary progressive MS in relation with axonal damage and brain atrophy*. J Neurol Neurosurg Psychiatry, 2003. **74**(7): p. 950-2.
20. De Stefano, N., et al., *In vivo evidence for axonal dysfunction remote from focal cerebral demyelination of the type seen in multiple sclerosis*. Brain, 1999. **122** (Pt 10): p. 1933-9.
21. Wessig, C., M. Bendszus, and G. Stoll, *In vivo visualization of focal demyelination in peripheral nerves by gadofluorine M-enhanced magnetic resonance imaging*. Exp Neurol, 2007. **204**(1): p. 14-9.
22. Mukherjee, P., et al., *Diffusion Tensor MR Imaging and Fiber Tractography: Theoretic Underpinnings*. AJNR Am J Neuroradiol, 2008.
23. Callaghan, P.T., C.D. Eccles, and Y. Xia, *Nmr Microscopy of Dynamic Displacements - K-Space and Q-Space Imaging*. Journal of Physics E-Scientific Instruments, 1988. **21**(8): p. 820-822.
24. Moseley, M.E., et al., *Diffusion-weighted MR imaging of anisotropic water diffusion in cat central nervous system*. Radiology, 1990. **176**(2): p. 439-45.
25. Beaulieu, C., *The basis of anisotropic water diffusion in the nervous system - a technical review*. NMR Biomed, 2002. **15**(7-8): p. 435-55.
26. Bassar, P.J., J. Mattiello, and D. LeBihan, *Estimation of the effective self-diffusion tensor from the NMR spin echo*. J Magn Reson B, 1994. **103**(3): p. 247-54.
27. Bassar, P.J., J. Mattiello, and D. LeBihan, *MR diffusion tensor spectroscopy and imaging*. Biophys J, 1994. **66**(1): p. 259-67.
28. Song, S.K., et al., *Diffusion tensor imaging detects and differentiates axon and myelin degeneration in mouse optic nerve after retinal ischemia*. Neuroimage, 2003. **20**(3): p. 1714-22.
29. Henry, R.G., et al., *Directional diffusion in relapsing-remitting multiple sclerosis: a possible in vivo signature of Wallerian degeneration*. J Magn Reson Imaging, 2003. **18**(4): p. 420-6.
30. Mori, S. and P.C. van Zijl, *Fiber tracking: principles and strategies - a technical review*. NMR Biomed, 2002. **15**(7-8): p. 468-80.
31. Berman, J.I., et al., *Accuracy of diffusion tensor magnetic resonance imaging tractography assessed using intraoperative subcortical stimulation mapping and magnetic source imaging*. J Neurosurg, 2007. **107**(3): p. 488-94.
32. Chang, L.C., et al., *Variance of estimated DTI-derived parameters via first-order perturbation methods*. Magn Reson Med, 2007. **57**(1): p. 141-9.
33. Anderson, A.W., *Theoretical analysis of the effects of noise on diffusion tensor imaging*. Magn Reson Med, 2001. **46**(6): p. 1174-88.
34. Salvador, R., et al., *Formal characterization and extension of the linearized diffusion tensor model*. Hum Brain Mapp, 2005. **24**(2): p. 144-55.
35. Poonawalla, A.H. and X.J. Zhou, *Analytical error propagation in diffusion anisotropy calculations*. J Magn Reson Imaging, 2004. **19**(4): p. 489-98.
36. Bevington, P.R. and D.K. Robinson, *Data reduction and error analysis for the physical sciences*. 2nd ed. 1992, New York: McGraw-Hill. xvii, 328 p.
37. Jones, D.K., *The effect of gradient sampling schemes on measures derived from diffusion tensor MRI: a Monte Carlo study*. Magn Reson Med, 2004. **51**(4): p. 807-15.

38. Jones, D.K. and P.J. Basser, *"Squashing peanuts and smashing pumpkins": how noise distorts diffusion-weighted MR data*. Magn Reson Med, 2004. **52**(5): p. 979-93.
39. Pierpaoli, C. and P.J. Basser, *Toward a quantitative assessment of diffusion anisotropy*. Magn Reson Med, 1996. **36**(6): p. 893-906.
40. Behrens, T.E., et al., *Characterization and propagation of uncertainty in diffusion-weighted MR imaging*. Magn Reson Med, 2003. **50**(5): p. 1077-88.
41. Chung, S., Y. Lu, and R.G. Henry, *Comparison of bootstrap approaches for estimation of uncertainties of DTI parameters*. Neuroimage, 2006. **33**(2): p. 531-41.
42. O'Gorman, R.L. and D.K. Jones, *Just how much data need to be collected for reliable bootstrap DT-MRI?* Magn Reson Med, 2006. **56**(4): p. 884-90.
43. Sivia, D.S., *Data analysis : a Bayesian tutorial*. Oxford science publications. 1996, Oxford: Oxford University Press.
44. Jones, D.K., *Determining and visualizing uncertainty in estimates of fiber orientation from diffusion tensor MRI*. Magn Reson Med, 2003. **49**(1): p. 7-12.
45. Efron, B. and R. Tibshirani, *An Introduction to the bootstrap*. 1993, New York: Chapman & Hall. xvi, 436.
46. Pajevic, S. and P.J. Basser, *Parametric and non-parametric statistical analysis of DT-MRI data*. J Magn Reson, 2003. **161**(1): p. 1-14.

Chapter 3: Comparison of DTI Bootstrap Approaches

3.1 Abstract

Bootstrap is an empirical non-parametric statistical technique based on data resampling that has been used to quantify uncertainties of diffusion tensor MRI (DTI) parameters, which are very useful in tractography and in assessing DTI methods. The current bootstrap method (repetition bootstrap) used for DTI analysis performs resampling within the data sharing common diffusion gradients, which requires multiple acquisitions for each diffusion gradient. Recently, wild bootstrap was proposed which can be applied without multiple acquisitions. In this paper, two new approaches are introduced called residual bootstrap and repetition bootknife. We show that repetition bootknife corrects for a large bias present in the repetition bootstrap method and therefore, better estimates the standard errors. Like wild bootstrap, residual bootstrap is applicable to single acquisition scheme, and both are based on regression residuals (called model-based resampling). Residual bootstrap is based on the assumption that non-constant variance of measured diffusion-attenuated signals can be modeled, which is actually the assumption behind the widely used weighted least squares approaches for the calculation of diffusion tensor. The performances of these bootstrap approaches are compared in terms of bias, variance, and overall error of bootstrap-estimated standard error by Monte Carlo simulation. We demonstrate that residual bootstrap has smaller biases and overall errors, which

enables estimation of uncertainties with higher accuracy. Understanding the properties of these bootstrap procedures will help us to choose the optimal approach for estimating uncertainties that can benefit hypothesis testing based on DTI parameters, probabilistic fiber tracking, and optimizing DTI methods.

3.2 Introduction

Diffusion Tensor MRI (DTI) is a diffusion weighted MRI technique capable of accurately describing anisotropic diffusion properties within a voxel [1, 2]. DTI was a breakthrough in the studies of white matter microstructure through characterization with DTI parameters and delineation of white matter pathways with DTI fiber tracking. In order to conduct the statistical hypothesis tests on DTI parameters in different pathophysiologic conditions, especially for voxel-wise longitudinal studies, or to follow the white matter tracks in probabilistic sense, characterization of uncertainties associated with estimated DTI parameters is essential. One approach for this is an empirical, non-parametric statistical technique based on data resampling called bootstrap [3]. Bootstrap was designed to replace complicated and often inaccurate approximations to uncertainty measures, such as biases and variances, by computer simulation based on real data. The bootstrap approach can be very helpful in DTI where final parameters of interest are known to be complicated non-linear function of measured signals.

In DTI, a particular implementation of bootstrap was proposed [4] in which resampling was done within the data sharing common diffusion gradients. This approach makes no assumptions about the noise properties at the cost of requiring multiple acquisitions for each diffusion gradient; thus we call this approach repetition bootstrap. Applications of repetition bootstrap were reported for the fiber tracking [5-7], quality assessment [8], and comparison of DTI anisotropy indices [9]. Unfortunately, a substantial under-estimation bias in the degree of uncertainty was reported for this method, which degrades the reliability of bootstrap with small numbers of repeats [10]. Furthermore, a small number of samples is likely to be the case with the most applications, especially in clinical settings where acquisition time is limited.

In addition to the limitation of scan times, there is an interest in obtaining more diffusion sensitizing directions at the cost of repetitions. By definition, the repetition bootstrap approach cannot be used when only one measurement per each diffusion direction is made. Acquisition of a single measurement is becoming more common practice with evidence that DTI parameters can be estimated more robustly with more diffusion gradient directions and with increasing interest in high angular resolution diffusion-weighted MRI (HARDI). In order to deal with the desire to acquire more diffusion sensitizing directions instead of multiple repetitions of the same directions, implementing wild bootstrap in the DTI analysis was proposed [11]. Wild bootstrap is a model-based resampling technique designed to investigate the uncertainty in the linear

regression with heteroscedasticity, i.e. non-constant variance with different regressors, of unknown form [12, 13].

In this paper, we first describe the property of the downward bias of the estimated degree of uncertainty in the repetition bootstrap and propose to reduce this bias by implementing the bootknife approach [14], which we call repetition bootknife. Evidence of bias correction actually improving the overall error of estimation is presented as well. Then, we investigate the feasibility of another model-based resampling approach called residual bootstrap, a well-known resampling technique in the statistics. Using Monte Carlo simulation, we compare the performance of repetition bootstrap, repetition bootknife, wild bootstrap, and residual bootstrap in terms of accuracy of estimating the degrees of uncertainty in diverse conditions such as different number of gradients, number of repetitions, diffusion tensor anisotropy, and partial volume with multiple tensors. Particular attention is paid to DTI sampling conditions within clinically feasible range, since our ultimate goal is to establish the optimal bootstrap procedure that can be applied to clinical data. Based on the results, the optimal bootstrap approaches under various DTI sampling scheme are discussed.

3.3 Methods

3.3.1 Under-estimation of Standard Errors by Repetition Bootstrap

Standard errors estimated by bootstrap are known to be generally smaller than the ideal values (downward biased) due to the basic mechanism of bootstrap. Bootstrap assumes that the empirical probability distribution \hat{F} created by putting equal probabilities of $1/n$ to all the n elements of a sample, faithfully represents the unknown population probability distribution F from which the sample is drawn. Creating bootstrap samples from the sample \hat{F} can be thought of as replicating the process of drawing new samples from the unknown population. Thus an approximate distribution of some statistic $\hat{\theta}$ (some function of the sample \hat{F} as an estimate of the parameter θ of the population F) can be generated via the bootstrap algorithm. If many samples of size n had been drawn from the population F , the standard deviation of the distribution of the statistic $\hat{\theta}$ would indicate the precision of $\hat{\theta}$; this is defined as the standard error of the statistic $\hat{\theta}$ of the sample \hat{F} . Thus the standard error can be estimated by simply calculating the standard deviation of $\hat{\theta}^*$, which are the statistic of interest calculated from the bootstrap samples. [15].

When the original sample size n is small, bootstrap estimated uncertainties are noticeably downward biased because the original sample that bootstrap relies on is biased. This phenomenon is similar to the bias in the estimator of population variance σ^2 . It is well known that the estimator

$$s^2 = n^{-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad \text{Equation 3-1}$$

is biased while the estimator

$$s^2 = (n-1)^{-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad \text{Equation 3-2}$$

is unbiased. This downward bias is a factor of $(n-1)/n$. When n is sufficiently large this factor hardly makes any difference in the estimation, and the biased estimator is actually known to be slightly better than unbiased estimator in terms of mean squared error (MSE) defined by

$$MSE = E((\hat{\theta} - \theta)^2) = bias^2 + SD^2 \quad \text{Equation 3-3}$$

$$bias = E(\hat{\theta}) - \theta \quad \text{Equation 3-4}$$

$$SD = \sqrt{(E(\hat{\theta}^2) - E^2(\hat{\theta}))} \quad \text{Equation 3-5}$$

where $\theta = \sigma$ and $\hat{\theta} = \hat{\sigma}$. Note that how far away estimator $\hat{\theta}$ is distributed from the population parameter θ depends on both the bias and the variability of the estimate (shown as SD or standard deviation in Equation 3-5) and the mean squared error (MSE) reflects the average quadratic loss or distance.

For the bootstrap, the usual estimator of uncertainty such as standard error can be thought to correspond to Equation 3-1, the biased estimator. In particular, the bootstrap estimator of variance (squared standard error) of the sample mean is different from the unbiased estimator by the factor of $(n-1)/n$ [14, 15]. When bootstrap is performed on the samples from stratified random sampling, the

bootstrap bias depends on the size of individual strata (corresponding to number of repetitions for repetition bootstrap), not the size of total sample (corresponding to number of repetitions times the number of gradient directions), which makes the bias substantial in situations with many small strata [14]. For instance, when the statistic being bootstrapped is a linear function of the means from multiple strata, the degree of bias for the bootstrap estimated variance can be expressed as the scaling factor of $(n-1)/n$ just like the sample mean in the non-stratified sampling case described above, though now n is the number of samples in each stratum [16, 17].

Repetition bootstrap can be regarded as an extreme case of stratified bootstrap in the sense that measurements with the same diffusion gradients (including $b=0$) are treated as strata and bootstrap resampling is performed only within each strata. Since it is unlikely that acquisitions will be repeated more than a few times even in experimental studies, repetition bootstrap will generally underestimate the standard error of DTI parameters to a substantial degree. DTI parameters are not linear functions of the raw measurements, and the degree of bias for bootstrap estimated uncertainty is difficult to express analytically, though we might expect it to be somewhat around $\sqrt{(n-1)/n}$ for the standard error, where n is number of repetitions, not total number of measurements. Assuming that this is true, in repetitions of 2, 3, 4 and 5 we would expect the repetition bootstrap to estimate standard errors that are only about 71, 82, 87, and 89% of the true values.

Multiple algorithms have been proposed to correct this bias in the stratified sampling [16-18], and in this paper we propose a very simple modification of the conventional repetition bootstrap based on the bootknife algorithm [14]; thus we call this approach **repetition bootknife**. Bootknife is a resampling technique combining the features of jackknife and bootstrap as implied by name. Bootknife samples are created by first randomly omitting one sample from the original sample of size n in each stratum (jackknife) and drawing a bootstrap sample of size n with replacement from the remaining sample with size $n-1$ (bootstrap). The strata are the repeats for each diffusion gradient just like the repetition bootstrap originally proposed [4], and the rule that resampling does not mix the elements from different strata (gradients) is not violated just like repetition bootstrap (thus we will call these two algorithms collectively repetition-based or stratified resampling). Since bias correction might actually increase the MSE by increasing the variance of the standard error estimates more than the decrease in the bias, the total MSE needs to be compared with and without the bias correction. This will tell whether the bias correction is actually beneficial.

3.3.2 Residual bootstrap and wild bootstrap

Possible alternatives to repetition-based resampling are model-based resampling approaches such as the residual bootstrap and wild bootstrap. Implementation of the wild bootstrap was introduced [11] while there are no reports of residual bootstrap in DTI. Model-based resampling refers to the bootstrap resampling

technique applied to the linear regression model, where the residuals based on the initially fitted model are resampled instead of the raw sample values. One might choose to do resampling pairs (of certain regressors and response) instead of residuals, but this approach, called pair bootstrap, would not be suitable in DTI since uncertainties estimated by pair bootstrap includes variance generated due to different design (such as skipping some diffusion gradients) which does not reflect the fixed design of DTI. Thus pair bootstrap is not considered in this study. Another possibility is to assume symmetry in the distribution of residuals for a given data point, and resample based on randomly changing the signs of the residuals; this is an implementation of the wild bootstrap. A third alternative is to assume that all residuals have similar distributions and freely resample among them without stratification; this is called residual bootstrap.

Since model-based resampling is 'based on a model', the model (diffusion tensor) needs to be adequate in describing the measured diffusion signals so that the error terms at different design points (different gradient directions for example) will have a common mean of zero. Regarding the variance of errors, residual bootstrap can be used in the **homoscedastic** condition (constant variance of error terms for the different design points) and also in the **heteroscedastic** condition as long as the heteroscedasticity can be modeled. If heteroscedasticity cannot be described mathematically, wild bootstrap may be a better approach since it does not require homoscedasticity.

In DTI, it was recognized from the beginning that the degree of uncertainty of log transformed signals used for linear regression is the inverse of the raw signals, and this property has been widely used for constructing weighting factors in the weighted least squares solution of diffusion tensor [1]. Similarly the residual bootstrap of DTI can be carried out based on the propagation of variance in log transformed signals, and the details of residual bootstrap as well as the diffusion tensor calculation are described in the following.

In a DTI experiment, the diffusion-weighted signal S is modeled by

$$S(\mathbf{g}_j) = S_0 \exp(-b\mathbf{g}_j^T \mathbf{D} \mathbf{g}_j), \text{ with } j=1,2,\dots,N \quad \text{Equation 3-6}$$

where S_0 is the signal intensity without diffusion weighting, b is the diffusion weighting factor, \mathbf{D} is effective self-diffusion tensor in the form of 3x3 positive definite matrix, \mathbf{g} is 3x1 unit vector of the diffusion-sensitive gradient direction, and N is the total number of experiments, including repeated measurements. By log transform, the equation above becomes

$$\ln(S(\mathbf{g}_j)) = \ln(S_0) - b\mathbf{g}_j^T \mathbf{D} \mathbf{g}_j \quad \text{Equation 3-7}$$

which can be structured into well-known multiple linear regression form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad \text{Equation 3-8}$$

where $\mathbf{y} = [\ln(S(\mathbf{g}_1)), \ln(S(\mathbf{g}_2)), \dots, \ln(S(\mathbf{g}_N))]^T$ are the logarithm of measured signals,

$\boldsymbol{\beta} = [D_{xx}, D_{yy}, D_{zz}, D_{xy}, D_{xz}, D_{yz}, \ln S_0]^T$ are the unknown regression coefficients

including the 6 unique elements of \mathbf{D} , \mathbf{X} is a design matrix of different diffusion gradient directions,

$$\mathbf{X} = -b \begin{bmatrix} g_{1x}^2 & g_{1y}^2 & g_{1z}^2 & 2g_{1x}g_{1y} & 2g_{1x}g_{1z} & 2g_{1y}g_{1z} & 1/b \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ g_{Nx}^2 & g_{Ny}^2 & g_{Nz}^2 & 2g_{Nx}g_{Ny} & 2g_{Nx}g_{Nz} & 2g_{Ny}g_{Nz} & 1/b \end{bmatrix}$$

and $\boldsymbol{\varepsilon} = [\varepsilon_0, \varepsilon_1, \dots, \varepsilon_N]^T$ are error terms. The weighted least squares (WLS) estimate of $\boldsymbol{\beta}$ is the following.

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{y} \quad \text{Equation 3-9}$$

In order to determine the diagonal weighting matrix \mathbf{W} , the ordinary least squares (OLS) estimate is calculated first by $\hat{\boldsymbol{\beta}}_{OLS} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$ leading to OLS fitted log measurements $\hat{\boldsymbol{\mu}}_{OLS} = \mathbf{X} \hat{\boldsymbol{\beta}}_{OLS}$ and fitted diffusion signals $\hat{\mathbf{S}}_g = \exp(\hat{\boldsymbol{\mu}}_{OLS})$. Then $\mathbf{W} = \text{diag}(\hat{\mathbf{S}}_g^2)$, which is based on the property

$$\text{Var}(\varepsilon_j) = \sigma^2 / S_j^2 \quad \text{Equation 3-10}$$

where σ is the standard deviation of noise in the raw signal. σ is assumed to be constant for each voxel regardless of the measured signal intensity.

After $\hat{\boldsymbol{\beta}}$ is calculated, the WLS fitted log measurements $\hat{\boldsymbol{\mu}} = \mathbf{X} \hat{\boldsymbol{\beta}}$ are used to calculate the residual vector $\mathbf{e} = \mathbf{y} - \hat{\boldsymbol{\mu}}$. In order to resample the errors, error terms ε_j need to be i.i.d. (independent and identically distributed) to satisfy the basic assumption of bootstrap that the samples are i.i.d. However, generally the raw residuals \mathbf{e} do not satisfy this condition due to the effect of possible heterogeneous leverages for different points. Also, ε_j actually have non-constant

variance (heteroscedasticity) but for DTI this can be modeled as shown in Equation 3-10. Therefore, raw residuals needs to be modified to have constant variance by following equation

$$r_j = \frac{y_j - \hat{\mu}_j}{w_j^{-1/2}(1-h_j)^{1/2}} \quad \text{Equation 3-11}$$

where the weighting factor w_j is j^{th} diagonal element of \mathbf{W} and the leverage value h_j is the j^{th} diagonal element of the hat matrix \mathbf{H} defined by $\mathbf{H} = \mathbf{X}(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}$. Finally, residual bootstrap resampling is defined as

$$y_j^* = \mathbf{x}_j \hat{\boldsymbol{\beta}} + w_j^{-1/2} \varepsilon_j^* \quad \text{Equation 3-12}$$

where y_j^* is j^{th} element of resampled log measurements, x_j is j^{th} row of \mathbf{X} , and ε_j^* is randomly resampled with replacement from the set of centered modified residuals $r_1 - \bar{r}, r_2 - \bar{r}, \dots, r_N - \bar{r}$ [12].

A bootstrap sample set $\mathbf{y}^* = [y_1^*, y_2^*, \dots, y_N^*]^T$ undergoes the WLS fitting procedure described above which leads to \mathbf{D}^* , from which a DTI parameter $\hat{\theta}^*$ such as FA (fractional anisotropy) is calculated. Resampling $\boldsymbol{\varepsilon}^* = [\varepsilon_1^*, \varepsilon_2^*, \dots, \varepsilon_N^*]^T$ and calculating $\hat{\theta}^*$ are repeated for some fixed large number N_B (typically hundreds to thousands times) to acquire N_B independent bootstrap samples $\hat{\theta}^{*b}$, $b=1, 2, \dots, N_B$. Here, the sample statistic $\hat{\theta}$ is an estimation of the true unknown θ (such as the noise free FA of the voxel) using the original sample \mathbf{y} by WLS, and $\hat{\theta}^*$ are bootstrap replications of $\hat{\theta}$. The bootstrap estimated standard error of $\hat{\theta}$ is simply the standard deviation of the N_B replications

$$s\hat{e}_B = \left\{ \sum_{b=1}^B [\hat{\theta}^*(b) - \hat{\theta}^*(\cdot)]^2 / (N_B - 1) \right\}^{1/2} \quad \text{Equation 3-13}$$

where $\hat{\theta}^*(\cdot) = \sum_{b=1}^B \hat{\theta}^*(b) / N_B$.

As mentioned above, wild bootstrap is suitable when heteroscedasticity cannot be modeled. In DTI with least squares estimation, this means that we are not relying on Equation 3-10 to modify raw residuals and resample residuals gathered from total design. Instead of resampling residuals from the pool of modified residuals causing the residuals from diffusion weighting of specific direction to be randomly distributed on any other directions, wild bootstrap creates variability by simply multiplying the individual residuals with a mutually independent random function. Wild bootstrap resampling is defined as

$$y_j^* = \mathbf{x}_j \hat{\boldsymbol{\beta}} + \varepsilon_j^* \quad \text{Equation 3-14}$$

where the resampled error ε_j^* is

$$\varepsilon_j^* = \frac{y_j - \hat{\mu}_j}{(1 - h_j)^{1/2}} t_j \quad \text{Equation 3-15}$$

and t_j is i.i.d. random variables with $E(t_j) = 0$, $E(t_j^2) = 1$, and $E(t_j^3) = 0$. Commonly t_j is a two-point distribution, and in this study the Rademacher distribution F2 with the property of $\Pr(t_j = 1) = 0.5$ and $\Pr(t_j = -1) = 0.5$ was used due to its good performance [19]. Simply speaking, modified residuals are randomly multiplied by either +1 or -1 and then added back to the fitted point where they originated from,

without being distributed to other design points. All the other steps are equivalent to residual bootstrap.

3.3.3 Monte Carlo Simulation

The performance of the four bootstrap approaches in terms of bias, standard deviation and overall error (MSE) of bootstrap estimated standard error (or 95% confidence interval for the angle of primary eigenvector) were compared under diverse conditions by Monte Carlo simulation. We assumed that Johnson RF noise is the only source of uncertainty in the diffusion signals. Schemes with the number of diffusion encoding directions, ranging from 6 to 54 were investigated. The six directions cases were based on the dual gradient scheme while the other number of directions were based on the electrostatic repulsion scheme [20]. Two b value experiments were used, $b=0$ s/mm² and $b=1000$ s/mm² (or 3000 s/mm² when specified). Number of images for $b=0$ and $b>0$ were kept in the ratio of 1:6, such that for 54 directions there were 9 $b=0$ images. One to 9 numbers of repetitions were studied, since clinical DTI scans are rarely repeated 10 or more even with only 6 directions.

Simulation was performed in a similar manner as described elsewhere [21], using custom software in IDL 6.1 (Research Systems, Inc., Boulder, CO). After an ideal, noise-free diffusion tensor was derived based on the desired DTI parameters such as FA (0.2, 0.5, and 0.8 were considered) and D_{av} ($=\text{Tr}(\mathbf{D})/3$, fixed to 0.7×10^{-3} mm²/s), noise-free diffusion weighted signals along specific direction of

diffusion gradients were calculated according to Equation 3-6 (S_0 arbitrarily set to 100). Then, noise modeled as complex random number with real and imaginary parts following Gaussian distribution of zero mean and standard deviation σ ($=S_0/\text{SNR}$) was added to the noise-free signal and the magnitude of the noisy signal was calculated. SNR of each $b=0$ image was set to 25 for this study unless specified otherwise. After a complete set of noisy signals was acquired, noisy diffusion tensor and DTI parameters were calculated. These steps were repeated a large number of times (100,000 used in this study) and a gold standard version of the standard error (or confidence interval of angle of primary eigenvector) for the DTI parameter of interest was directly calculated from the standard deviation (or 95% range of angle of primary eigenvector) of all the noisy parameters.

We also investigated the conditions with partial volume effects (PVE) where signals are actually originating from a system more complicated than a single tensor. Multiple regions in the brain are known to have PVE due to intravoxel crossing of two distinct axonal bundles, and this can violate the assumption of appropriateness of the single tensor DTI model in model-based resampling. Model-based resampling may not perform optimally with PVE and therefore the performance of residual or wild bootstrap with PVE can be important when implementing bootstrap in clinical data. In this study, we focused on a mixture of white matter bundles [22] where the diffusion weighted signals comes from two compartments described as

$$S(\mathbf{g}_j) = S_0 f \exp(-b \mathbf{g}_j^T \mathbf{D}_1 \mathbf{g}_j) + S_0 (1-f) \exp(-b \mathbf{g}_j^T \mathbf{D}_2 \mathbf{g}_j) \quad \text{Equation 3-16}$$

where \mathbf{D}_1 and \mathbf{D}_2 represent the tensor from each compartment, f and $(1-f)$ are the signal fractions from \mathbf{D}_1 and \mathbf{D}_2 . We assumed no exchange between the compartments, which will make PVE most pronounced. \mathbf{D}_1 and \mathbf{D}_2 were assumed to be prolate tensors with FA=0.7, f was fixed to 0.5, and angles between primary eigenvectors of two tensors were varied. Then, the usual single tensor design matrix was used to fit the diffusion signals, in the calculation of gold standard or bootstrap estimates of SE. We also created the equivalent single tensor system as follows. The noise-free PVE data were fitted and the calculated tensor was used to define the equivalent single tensor. Noise was then added to this equivalent tensor for further analysis.

In order to evaluate the performance of the bootstrap approaches, the bootstrap procedures described above were performed either directly on the diffusion-weighted signals (repetition bootstrap and repetition bootknife) or on the residuals between measured signals and fitted signals (residual bootstrap and wild bootstrap) to create the bootstrap samples, and they were used to calculate the diffusion tensors and the DTI parameters of interest. This process was repeated N_B times (1000 was used for this study), and finally the bootstrap estimated standard errors were calculated by Equation 3-13. Since standard errors cannot be used for vector quantities such as the primary eigenvector, the 95 percentile confidence intervals of the minimum angle between each bootstrap estimate and the average primary eigenvector was used instead [5]. Though not

used in this study, there is an alternative measure of accuracy of the primary eigenvector based on the dispersion parameter of Watson distribution [23]. The experiment of bootstrap estimation of uncertainty was repeated 1000 times from which the variance of these estimations were calculated by Equation 3-5, and the biases determined by comparing the expectation of estimations with the gold standard value by Equation 3-4. The MSE of the bootstrap standard error estimates were computed from the biases and variances as shown in Equation 3-3, which reflects the overall degrees of error and is an objective index of performance.

3.4 Results

3.4.1 Bias of Repetition Bootstrap

The typical downward bias of repetition bootstrap for small numbers of repetitions is demonstrated in Figure 3-1a. Simulation results were acquired for varying number of repetitions while the number of gradient directions was fixed to 18 plus 3 $b=0$ images. Figure 3-1a show the mean and standard deviation of the bootstrap estimates of SE of FA (from 1000 experiments) as well as the gold standard value of SE (from 100,000 experiments). The mean of the repetition bootstrap SE estimates was substantially smaller than the gold standard, while the repetition bootknife estimates are nearly unbiased, though with slightly larger standard deviations than the repetition bootstrap. Figure 3-1d-f show the bias, standard deviation (SD), and square root of mean squared error (RMSE) of the

bootstrap estimates separately, displayed as a percent of the gold standard SE of FA value. This normalization allows the bootstrap performance to be compared to other conditions such as different DTI parameters, number of repetitions and number of directions. Note, again, that repetition bootstrap was substantially downward biased down to 30% while the repetition bootknife was nearly unbiased. The repetition bootknife proved to be a better estimator with the smaller RMSEs, especially for small numbers of repetitions.

Figure 3-1a and Figure 3-1d-f also show that both repetition methods are more accurate with more repetitions. This simply reflects the fact that bootstrap performs better with a larger sample pool and in particular the estimates do not improve due to the increasing total SNR associated with more repeated acquisitions. To illustrate this point, the total SNR was held fixed for the different numbers of repetitions and is shown in Figure 3-1b. For Figure 3-1a, on the other hand, SNR of each repetition is fixed to 25 resulting in increasing SNR with more repetitions (and subsequent reduction in the gold standard SE). Figure 3-1b shows that now the gold standard SE (FA) is constant instead of decreasing but the bootstrap bias, SD, and RMSE (% of gold standard) is almost identical to Figure 3-1d-f (thus result not shown in the format of Figure 3-1d-f), illustrating that the estimates are still more accurate with larger repetitions independent of total SNR. This clearly shows that SNR itself is not a factor influencing the bias of the repetition methods. For repetition bootstrap the RMSE is primarily influenced by the decrease in bias with increasing number of repetitions. Given that the

origin of the bias is due to small sample sizes, it is clear that increasing the number of samples (and not the SNR) determines the percent RMSE. For the repetition bootknife, the reduction in SD of the SE estimates with increasing repetitions is the strongest factor in the RMSE. In both repetition methods, the decreases in the SD of SE with increasing the number of repetitions are due to the increased sample size.

As expected, the under-estimation bias of repetition bootstrap is problematic when the number of directions rather than the number of repetitions is increased. Note from Figure 3-1c and Figure 3-1g-i that as more directions are acquired while the number of repetitions is fixed to 2, the degree of bias for repetition bootstrap hardly improves, which is expected since repetition bootstrap bias depends only on the number of repetitions. This property leads to a poor improvement of RMSE for repetition bootstrap even with a large number of directions. The repetition bootknife, on the other hand, has a small bias that becomes even smaller with more directions, leading to the similar trend of RMSE as Figure 3-1f. Thus the gap of performance between these two methods is more pronounced with a larger sample pool made from increasing the number of directions.

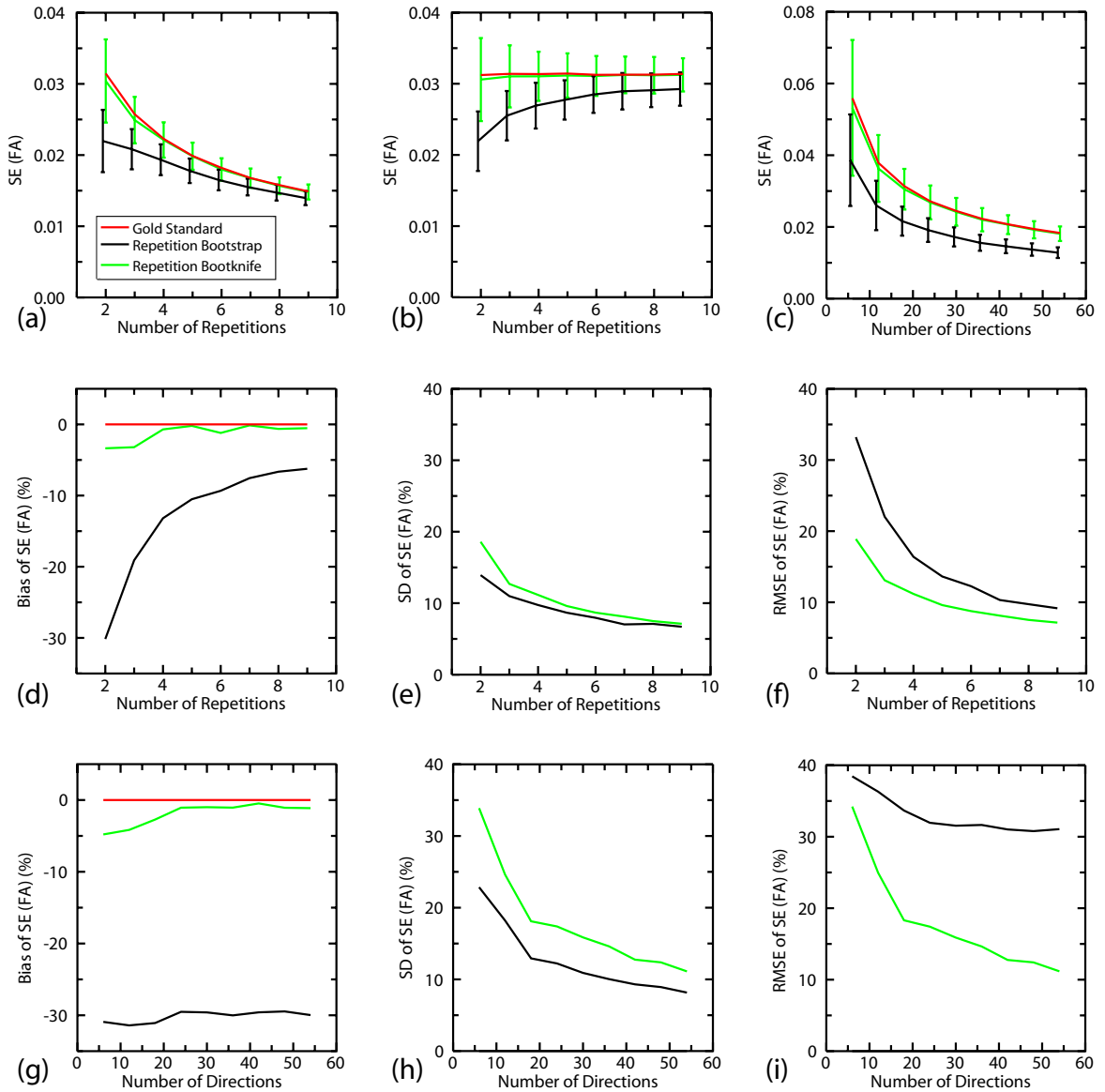


Figure 3-1: **(a)** Standard error (SE) of FA estimated by repetition bootstrap (black line) and repetition bootknife (green line) plotted as mean and standard deviation (SD) in vertical bars with varying number of repetitions while the number of directions is fixed to 18. The gold standard SE is shown in red. Noisy diffusion signals were created by adding noise corresponding to SNR=25 to noise-free signals from a prolate tensor with FA=0.5, $D_{av}=0.7 \times 10^{-3} \text{ mm}^2/\text{s}$. **(b)** Same as (a) except that SNR for individual acquisition is adjusted to keep the effective SNR after combining repetitions to be constant. SNRs used per acquisition in the repetitions from 2 to 9 are approximately 25.0, 20.4, 17.7, 15.8, 14.4, 13.3, 12.5, and 11.8. **(c)** Same as (a) except that number of directions is varied while the number of repetition is fixed to 2. Noise were added in the same way as in (a). **(d-f)** Bias, SD, and square root of MSE (RMSE) of bootstrap estimates of SE

in % of gold standard SE for the data displayed in (a). % Bias, SD, and RMSE for the data displayed in (b) is almost identical to (d-f). **(g-i)** Bias, SD, and RMSE of bootstrap estimates of SE in % of gold standard SE for the data displayed in (c).

3.4.2 Bootstrap methods in the diffusion signals from single tensor model

Figure 3-2 shows the performance of the four bootstrap approaches for estimating the 95% confidence interval (CI) of the angle of primary eigenvector with the number of repetitions between 1 and 6 and the number of directions fixed to 18. Since model-based resampling methods such as residual and wild bootstrap do not depend on repeated acquisitions, bootstrap performance can be shown even in the case of only one acquisition unlike the repetition methods for which results are displayed with number of repetitions starting at two. Figure 3-2a-c shows the bias, SD, and RMSE when the noise-free modeled diffusion tensor has a moderate anisotropy of $FA=0.5$. The repetition methods have a very similar pattern of bias, SD and RMSE to that found for FA in Figure 3-1d-f. Repetition bootstrap is substantially downward biased and the overall error is smaller with repetition bootknife. The residual bootstrap and wild bootstrap methods are shown to be nearly unbiased, have small SD, and have RMSE smaller than the repetition methods. The residual bootstrap has slightly lower bias and SD than the wild bootstrap. For each value of the number of repetitions, residual bootstrap has the smallest RMSE, followed by the wild bootstrap, then repetition bootknife, and lastly repetition bootstrap. This bootstrap performance pattern is nearly identical in high anisotropy of $FA=0.8$ (result not shown), while in

low anisotropy of $FA=0.2$ (Figure 3-2d-f), all bootstrap methods suffer from worse performance. Figure 3-2d-f show that the residual and wild bootstrap methods overestimate the CI especially for small numbers of repetitions, and the estimates of all the bootstrap methods are more dispersed at low FA than the estimates in medium or high anisotropy (note the scale difference of y axis). This implies that not only the primary eigenvector direction is more uncertain in the low anisotropy but also the ability of bootstrap to estimate the increased uncertainty is worse. Interestingly, all four bootstrap methods show similar RMSE in low anisotropy.

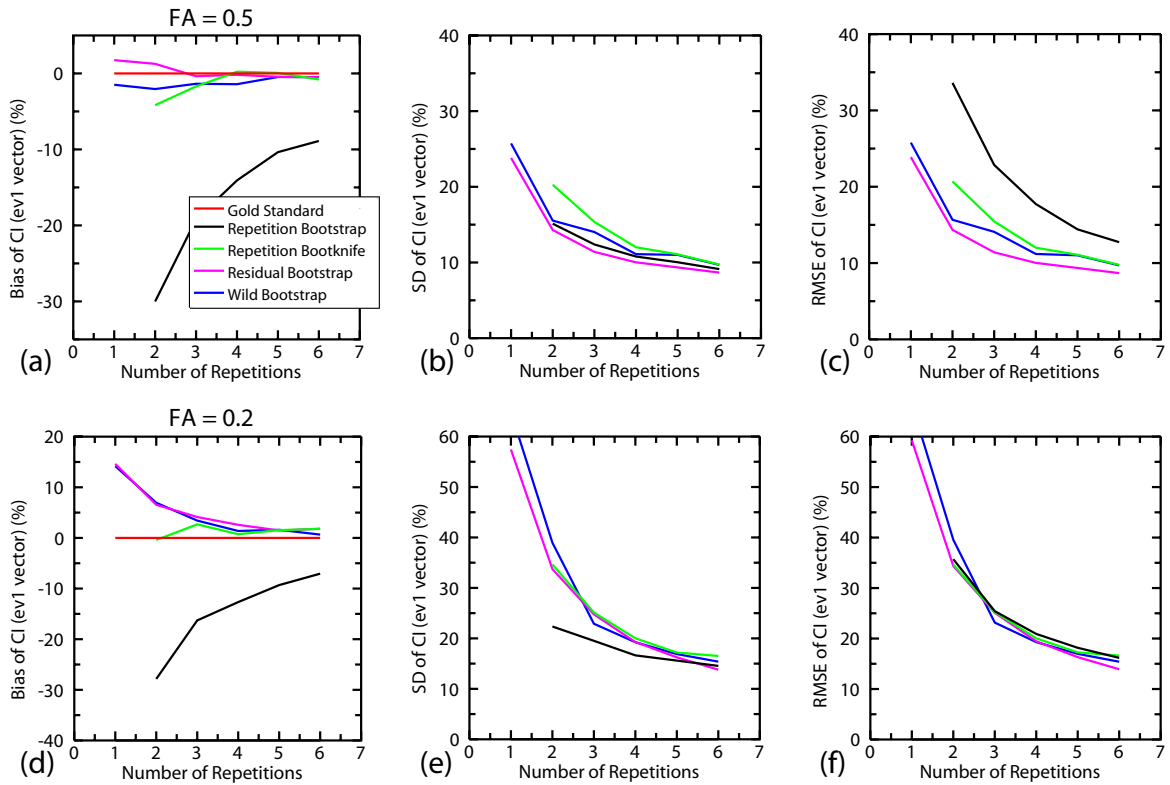


Figure 3-2: Bias, SD, and RMSE of the 95th percentile confidence interval of the the angle of primary eigenvector estimated by bootstrap methods with different number of repetitions while number of directions is fixed to 18. The DTI models were prolate tensors with FA of 0.5 (a-c), and 0.2 (d-f) and D_{av} of $0.7 \times 10^{-3} \text{ mm}^2/\text{s}$ for all. Data with FA of 0.8 have almost identical plots to (a-c).

The bootstrap performance for estimating the primary eigenvector angle CI when the number of diffusion gradient directions is increased while the number of acquisitions is fixed to two is demonstrated in Figure 3-3. As pointed out in Figure 3-1c and Figure 3-1g-i, repetition bootstrap bias remains relatively independent of the number of directions, leading to substantially larger RMSE for large numbers of directions. The other three methods show similar trends of bias, SD, and RMSE to Figure 3-2. The residual bootstrap is generally the least biased and variable followed by the wild bootstrap and repetition bootknife. Thus residual bootstrap seems to have better overall performance than the others. Just as in Figure 3-2, results for high anisotropy of $FA=0.8$ are almost identical to moderate anisotropy of $FA=0.5$, and for low anisotropy of $FA=0.2$, the model-based resampling shows some over-estimation at lower numbers of directions that rapidly disappears with more directions. All bootstrap methods except for the repetition bootstrap show relatively small differences in RMSE in low anisotropy as well. When the data is acquired only once (meaning that number of repetition is one), repetition bootstrap and repetition bootknife is no longer available, while model-based resampling can still be used. Results for the performance of residual and wild bootstrap without any repeated acquisition (not shown) indicate that the trend is very similar to Figure 3-3 though the bias, SD, and RMSE are larger than that of Figure 3-3, since the sample size is only half of Figure 3-3. Overall, residual bootstrap is less biased, less variable and has smaller RMSE.

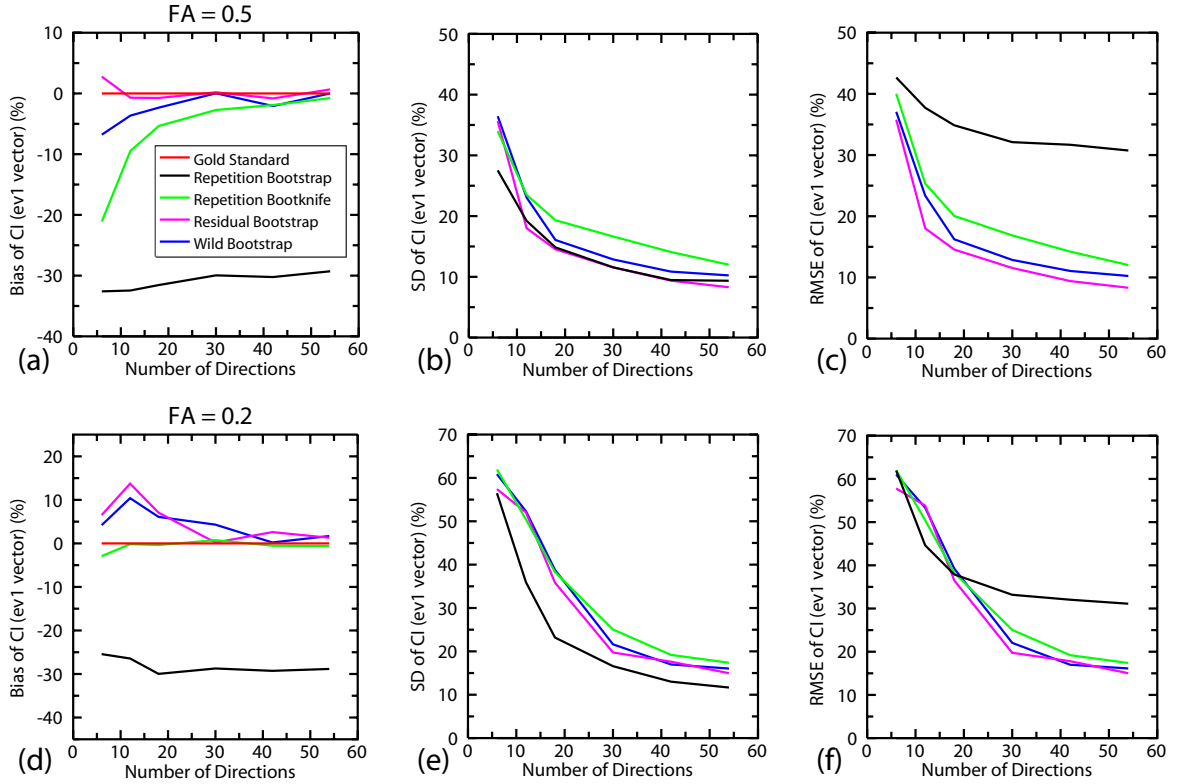


Figure 3-3: Bias, SD, and RMSE of the 95th percentile confidence interval of the angle of the primary eigenvector estimated by bootstrap methods with different numbers of directions while the number of repetitions is fixed to 2. The DTI models were prolate tensors with FA of 0.5 (a-c), and 0.2 (d-f) and D_{av} of $0.7 \times 10^{-3} \text{ mm}^2/\text{s}$ for all. Data with FA of 0.8 have almost identical plots to (a-c).

Figure 3-4 shows Figure 3-2c and Figure 3-3c plotted together with the common x-axis representing the number of samples (includes $b=0$) in order to clearly demonstrate the increase of sample size either by number of repetitions (solid lines) or directions (dashed lines). This shows that the residual bootstrap and wild bootstrap have a very similar trend of improvement of RMSE when either number of repetitions or directions is increased. As expected, the repetition bootstrap, on the other hand, does not benefit from increasing number of

directions as much as number of repetitions. Even the repetition bootknife has a slight tendency of better performance with larger repetitions rather than directions.

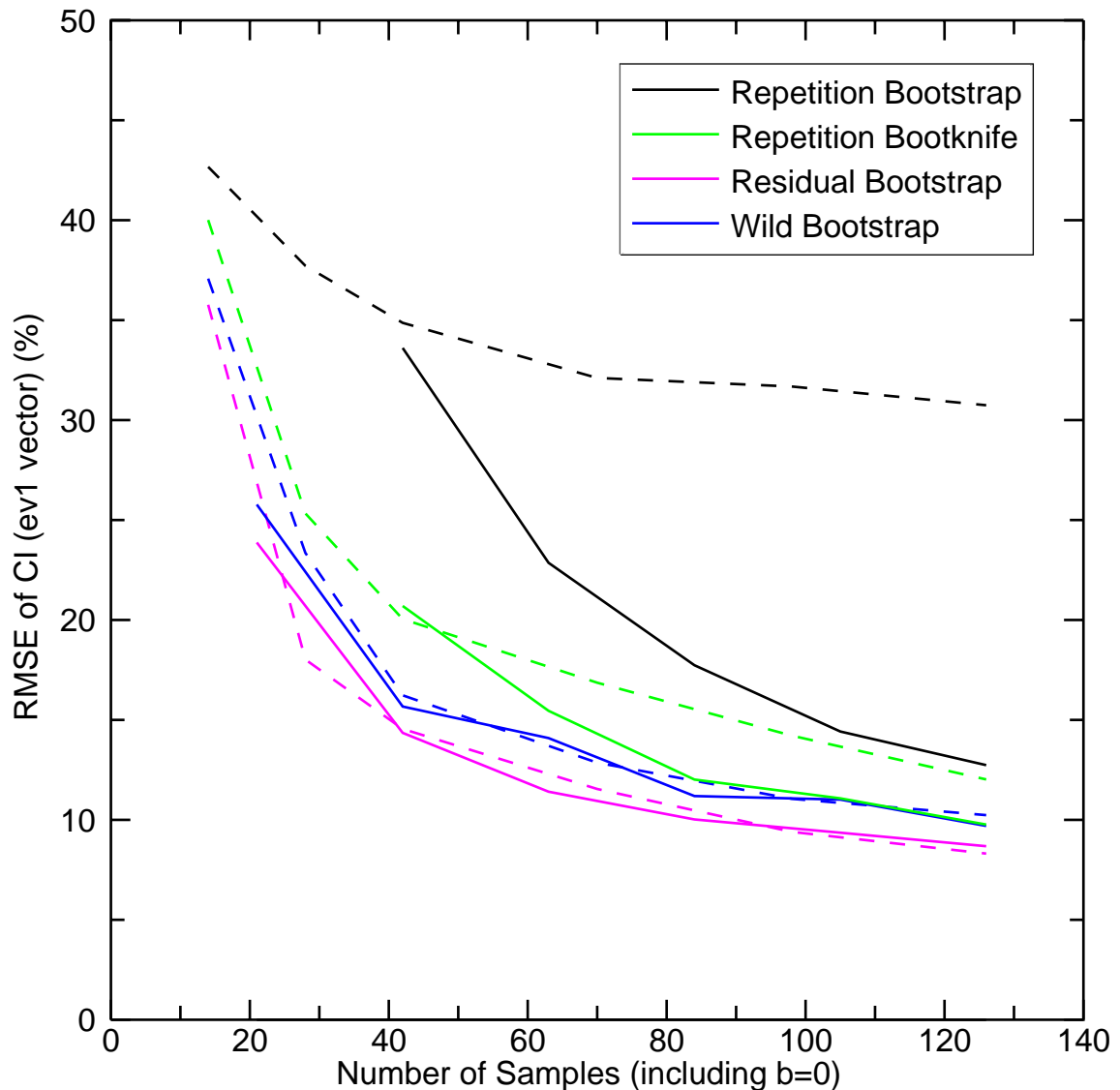


Figure 3-4: Comparison of the RMSE of the 95th percentile confidence interval of the angle of the primary eigenvector estimated by bootstrap methods for varying numbers of repetitions and directions. Solid lines are results with different number of repetitions (ranging from 1 to 6) while number of directions is fixed to 18. Dashed lines are results with different number of directions (ranging from 6 to 54) while number of repetitions is fixed to 2. Noisy diffusion signals are created by adding noise corresponding to SNR=25 to noise-free signals from a prolate tensor with FA=0.5, $D_{av}=0.7 \times 10^{-3} \text{ mm}^2/\text{s}$. That is, solid line results are same as figure 2c while dashed line results

are same as figure 3c. The number of samples include $b=0$ measurements (with the number 1/6 of different number of diffusion directions).

3.4.3 Bootstrap methods in the diffusion signals from tensor mixture model

Figure 3-5 shows the performance of the bootstrap methods when acquisitions are repeated one or two times with a relatively large number of diffusion directions of 54. This sampling scheme was chosen since the model (diffusion tensor) insufficiency is more likely to be an issue with the large number of directions, and clinical scans are not likely to be repeated more than one or two times with large number of directions. Typical diffusion weighting of $b=1000$ s/mm² was used, thus representing the scenario where resolving the PVE such as intravoxel crossing is not necessarily of primary interest. In order to separate the influence of PVE versus simply different diffusion tensor shape on the bootstrap, two separate results from different modeling is simultaneously displayed. Solid lines are results from modeling tensor mixture, while dashed lines are results from modeling single tensor equivalent to the fitted tensor to noise-free signals from tensor mixture. Of course, once the noisy diffusion signals are acquired in either way, then fitting a single tensor to the data is assumed, just as almost all the diffusion tensor analysis of real data is done (i.e. without applying HARDI or multiple tensor modeling of diffusion signals). In the case of two repetitions shown in Figure 3-5a-c, all bootstrap methods show relatively small differences between the solid and dashed lines implying that PVE is not a significant factor in the performance of any of bootstrap methods. This

trend is replicated in the case with only one repetition in Figure 3-5d-f, where the model-based resampling shows similar trends between solid and dashed lines.

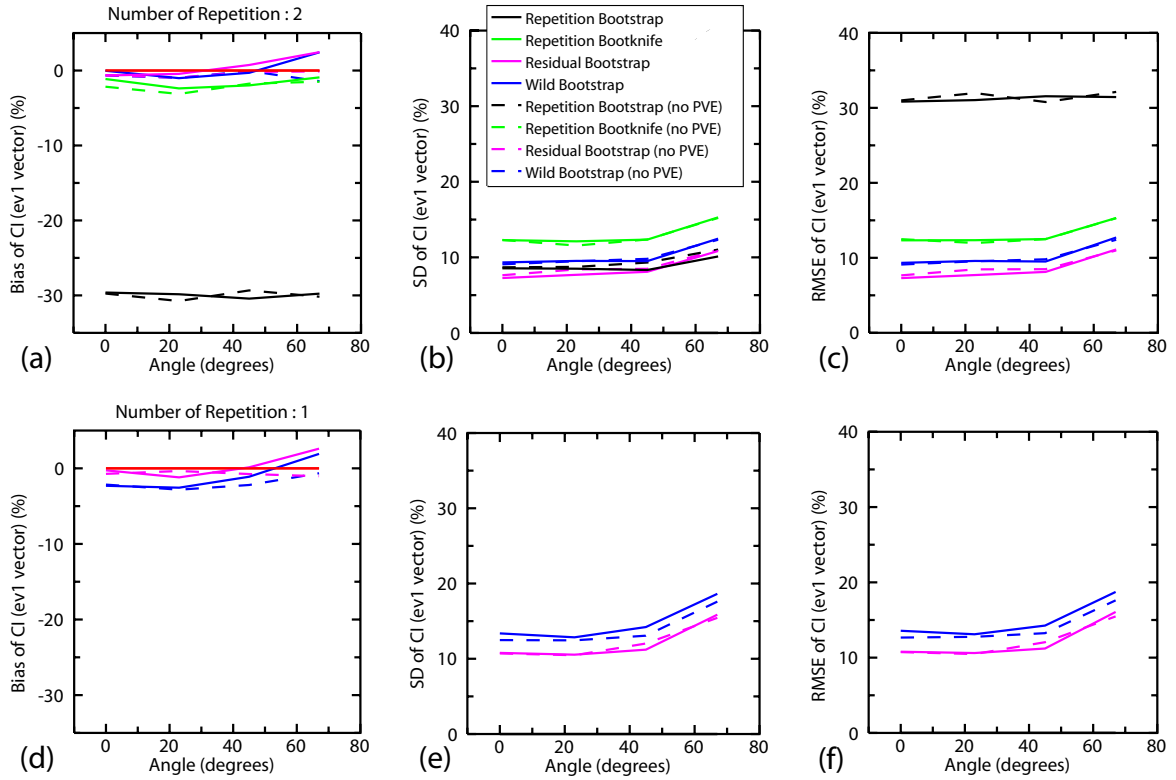


Figure 3-5: Bias, SD, and RMSE of the 95th percentile confidence interval of the angle of the primary eigenvector estimated by bootstrap methods with varying angles between two tensors within a voxel for two (a-c) and one (d-f) repetitions while number of directions is fixed to 54. The primary eigenvectors of the two prolate tensors, each with FA=0.7 and $D_{av}=0.7 \times 10^{-3} \text{ mm}^2/\text{s}$, were positioned at angles of 0, 23, 45, and 68 degrees. The solid lines are the usual single tensor fits to these modeled PVE. The dashed lines are single tensor fits to an equivalent single tensor (with the same FA, D_{av} , single tensor shape, etc) found from the single tensor fit to the noise-free partial volume model and then refitted with noise added.

Figure 3-6 shows the results when the b value is increased to 3000 s/mm² for more pronounced PVE [22]. When the angles between primary eigenvectors of the two modeled tensors are small, the difference between solid and dashed lines is minimal, but when the angle is large such as 68 degrees, solid lines of

residual and wild bootstrap show larger bias and SD leading to larger RMSE. This supports our theory that model-base resampling is more susceptible to PVE. On the other hand, the repetition bootstrap and repetition bootknife methods are insensitive to the presence of PVE. Even so, the residual bootstrap still have smaller RMSE than the other methods even with the presence of PVE, except for the tensor mixture with 68 degrees where the repetition bootknife seems to be slightly better than residual bootstrap.

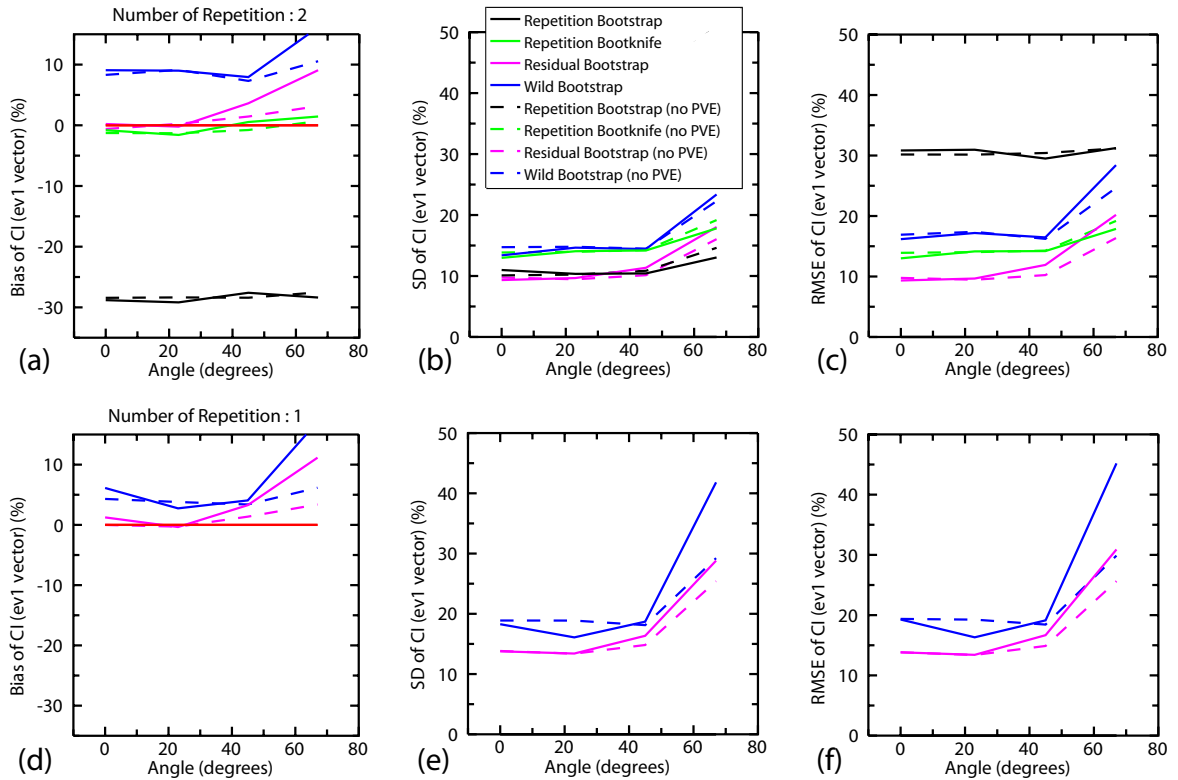


Figure 3-6: Same as figure 5 except that b value used is 3000 s/mm^2 .

3.5 Discussion

Bootstrap is a powerful method of estimating the uncertainties in DTI derived parameters and it has been successfully implemented and shown to be useful in

diverse applications such as probabilistic fiber tracking and quality assessment of DTI acquisitions. It also has the potential to be used for statistical tests such as voxel-based (or ROI-based) analysis of longitudinal (acquired at multiple time points) and cross-sectional DTI data. So far, only one particular implementation of bootstrap (repetition bootstrap) has been used in applications, but it is important to point out that bootstrap is not defined in a unique way, but rather a group of diverse algorithms sharing the basic concept. In this paper we implemented four DTI bootstrap approaches, including two previously unreported, and tested them by Monte Carlo simulation under diverse conditions in search of the optimal method that can calculate the uncertainty reliably.

We showed that repetition bootstrap is substantially downward biased and introduced the repetition bootknife that successfully reduced the bias and mean squared error. We also introduced the residual bootstrap as another model-based resampling technique, and compared the four bootstrap methods (repetition bootstrap, repetition bootknife, residual bootstrap, and wild bootstrap) for their performance in terms of bias, variance and mean squared error. Our simulations demonstrated that in the cases where DTI was acquired multiple times permitting all four bootstrap methods, model-base resampling outperforms repetition-based resampling if the model is true, suggesting that even if multiple acquisitions exist, model-based resampling might be the better choice. When data is acquired only once with possibly many different diffusion encoding directions, repetition-based bootstrap is not feasible but model-based resampling

can still be used. This allows greater freedom for researchers and/or clinicians in choosing a diffusion gradient sampling scheme when they are considering implementing bootstrap for DTI data. Within model-based resampling, residual bootstrap was consistently better than wild bootstrap, especially when the single tensor model was not sufficient. For repetition-based resampling, the modified version introduced here proved to be better than original version.

Another important result of this study is that with model-based bootstrap techniques, one can benefit from increasing number of directions just as much as increasing number of repetitions. Pajevic and Basser [4] postulated that with many distinct non-collinear directions fewer repetitions would be required to achieve the same reliability of bootstrap, but their data indicated that the relationship between the number of directions, repetitions, and the quality of bootstrap estimates were somewhat complicated. For instance, their data shows that CV of SE (RA) (similar to SD of SE (FA) (%) in our data) can actually increase with very large number of directions with fixed number of repetitions, and that CV of SE (Trace) actually gradually worsens with more directions with fixed repetitions. We believe that this is related to fixing number of $b=0$ images to one instead of increasing it to keep the ratio of $b=0$ to $b>0$ constant (such as 1:6 in this study). When we fixed the number of $b=0$ images to one regardless of the number of directions, we observed very similar trend that increasing the number of directions does not consistently improve the bootstrap performance (result not shown here). This effect is probably due to the strong leverage that the $b=0$ data

has on the least squares fit when only one $b=0$ data is acquired. The influence of a data point on the fit depends on the leverage and variance of the data point compared to the others. When many encoding directions are used, the influence for a single $b=0$ data point can be large due to the large leverage. Our data with fixed ratio of $b=0$ to $b>0$ indicate that model-based resampling with either large number of directions or large number of repetitions have very similar performance, which was clearly demonstrated in Figure 3-4. As long as model-based resampling is used and $b=0$ images are increased accordingly with more diffusion directions, total sample pool size alone determines the bootstrap performance. For repetition bootstrap, using more repetitions is always better than increasing the number of directions (even with increased number of $b=0$ images) since increasing the number of directions does not directly increase the resample pool size.

It is important to emphasize that for bootstrap to be reliable, the sample pool size should be large enough, though with model-based techniques repetition is not a requirement anymore. Residual or wild bootstrap can generate estimates of SE of DTI parameters in a single acquisition, but unless the number of directions is large, bootstrap estimates will be highly variable. It is difficult to generalize how large the total sample pool should be, because bootstrap performance depends on tensor anisotropy and shape, the DTI parameter of interest, and the definition of good performance depends on the sensitivity needed in the application. However, given an effect size, the bootstrap data will enable power calculations.

This study shows how different bootstrap methods can perform under a few selected demonstrative conditions, but more studies are needed in order to have a more complete picture of bootstrap performance issue.

The results shown in this study were focused on the performance of bootstrap in estimating the uncertainty of primary eigenvector and FA because incorporating bootstrap to fiber tracking is of a great interest, but similar results were obtained for the eigenvalues and D_{av} . However, it is not clear how the bootstrap methods performs in estimating the entire probability density function (*pdf*) of DTI parameters except for some evidence of repetition bootstrap properly capturing the characteristics of the *pdf* [4]. Objective Bayesian analysis [24] is another approach that has been used to compute the *pdf* of DTI parameters, though this can be computationally much more demanding. Now that bootstrap can be performed even without repeated acquisitions, it will be interesting to compare bootstrap and Bayesian approaches in certain situations such as probabilistic fiber tracking using the same dataset.

Inadequacy of the single tensor DTI model to describe the data (as assumed by the design matrix used in the WLS fit to obtain the tensor) has been shown here to increase the errors of the bootstrap estimates. The model-based wild and residual bootstrap methods are particularly sensitive to this effect. The wild bootstrap method as implemented here is based on the symmetry of the *pdf* of the residuals. This assumption is violated by both low SNR data (due to log

Rician noise) and single tensor assumptions. The residual bootstrap method is based on the similarity of the probability distribution functions between data points (and not the symmetry of the *pdfs*). However, depending on the alignment of the gradients relative to the tensor principal direction, the *pdfs* may vary among the data points due to log Rician noise and multi-tensor effects. Although also affected by the low SNR, the residual bootstrap method is less affected since it is not sensitive to the asymmetric *pdfs* due to log Rician noise, but only the differential effects on the *pdfs* among the gradient directions. Despite these effects, our results suggest that the model-based approaches generally perform better than the repetition-based methods. Model-based effects are not an issue when only 6 directions are used, since this effectively reduces the diffusion ODF (orientation distribution function) [25] to an exact effective single tensor.

Rician noise and PVE causing inadequacy of the single tensor model are important sources of uncertainty in the DTI derived parameters but there are other sources including cardiac pulsation, head motion, artifacts, eddy currents, magnetic susceptibility effects, etc [26]. The repetition-based methods are likely to better characterize the uncertainty from non-ideal variance caused by these sources than the model-based methods since repetition-based methods make less assumptions than model-based methods, though whether this holds true needs to be evaluated in some way. In this study, only ideal noise and PVE were considered because it is relatively straightforward to simulate these effects, but more studies are needed to evaluate how bootstrap methods perform with other

sources of variance as well, either by simulating some aspects of these sources or by using real data.

Weighted linear least squares estimation was used for calculating the diffusion tensor from the original as well as bootstrap samples, but alternative ways to estimate the diffusion tensor exist. Nonlinear least squares estimator was shown to be more robust at high b values or low SNR [27], and was less likely to produce unphysical negative eigenvalues [28] than linear least squares. Also, a robust estimator was shown to be effective against artifacts producing outliers [29, 30]. Bootstrap can be combined with these estimators as well, though the additional computation time required by these processing methods instead of the computationally efficient linear least squares solution can be a limitation. For instance, computation time for nonlinear least squares estimation can be up to 60 times more than that of linear estimation [29]. Considering the fact that bootstrap requires the tensor estimation to be iterated hundreds to thousands of times, whether it is beneficial to bootstrap with these more sophisticated tensor estimation and how bootstrap can be implemented more efficiently will be a subject of future study.

3.6 Conclusion

We have shown that a bias is present in the currently used repetition bootstrap method and have presented an alternate method (repetition bootknife) that corrects for this bias and therefore, better estimates the standard errors of DTI

parameters. We have also evaluated the model-based wild bootstrap which performs better than the repetition methods but is susceptible to model failures. We also present another model-based method (residual bootstrap) that generally performs better than all the other methods, but is also sensitive to failures of the tensor model to describe the data. These results can be used to design DTI experiments in terms of choosing number of averages and number of diffusion sensitizing gradient directions to achieve the standard errors that permit observation of a particular effect sizes. Also, importantly, the model-based methods enable probabilistic fiber tracking and hypothesis testing in longitudinal voxel-wise analysis with a single acquisition, which allows maximization of the number of diffusion sensitizing directions in a clinically feasible scan time.

3.7 Reference

1. Bassler, P.J., J. Mattiello, and D. LeBihan, *Estimation of the effective self-diffusion tensor from the NMR spin echo*. J Magn Reson B, 1994. **103**(3): p. 247-54.
2. Bassler, P.J., J. Mattiello, and D. LeBihan, *MR diffusion tensor spectroscopy and imaging*. Biophys J, 1994. **66**(1): p. 259-67.
3. Efron, B., *1977 Rietz Lecture - Bootstrap Methods - Another Look at the Jackknife*. Annals of Statistics, 1979. **7**(1): p. 1-26.
4. Pajevic, S. and P.J. Bassler, *Parametric and non-parametric statistical analysis of DT-MRI data*. J Magn Reson, 2003. **161**(1): p. 1-14.
5. Jones, D.K., *Determining and visualizing uncertainty in estimates of fiber orientation from diffusion tensor MRI*. Magn Reson Med, 2003. **49**(1): p. 7-12.
6. Jones, D.K. and C. Pierpaoli, *Confidence mapping in diffusion tensor magnetic resonance imaging tractography using a bootstrap approach*. Magn Reson Med, 2005. **53**(5): p. 1143-9.
7. Lazar, M. and A.L. Alexander, *Bootstrap white matter tractography (BOOT-TRAC)*. Neuroimage, 2005. **24**(2): p. 524-32.
8. Heim, S., et al., *Assessing DTI data quality using bootstrap analysis*. Magn Reson Med, 2004. **52**(3): p. 582-9.

9. Hasan, K.M., A.L. Alexander, and P.A. Narayana, *Does fractional anisotropy have better noise immunity characteristics than relative anisotropy in diffusion tensor MRI? An analytical approach*. Magn Reson Med, 2004. **51**(2): p. 413-7.
10. O'Gorman, R.L. and D.K. Jones. *How many bootstraps make a buckle?* in *13th Annual Meeting of ISMRM*. 2005. Miami Beach: ISMRM.
11. Whitcher, B., D.S. Tuch, and L. Wang. *The wild bootstrap to quantify variability in diffusion tensor MRI*. in *13th Annual Meeting of ISMRM*. 2005. Miami Beach: ISMRM.
12. Davison, A.C. and D.V. Hinkley, *Bootstrap methods and their application*. Reprinted with corrections. ed. Cambridge series on statistical and probabilistic mathematics. 2003, Cambridge, UK ; New York, NY: Cambridge University Press. x, 582.
13. Liu, R.Y., *Bootstrap Procedures under Some Non-IID Models*. Annals of Statistics, 1988. **16**(4): p. 1696-1708.
14. Hesterberg, T.C. *Unbiasing the Bootstrap-Bootknife Sampling vs. Smoothing*. in *Section on Statistics and the Environment*. 2004: American Statistical Association.
15. Efron, B. and R. Tibshirani, *An Introduction to the bootstrap*. 1993, New York: Chapman & Hall. xvi, 436.
16. Rao, J.N.K. and C.F.J. Wu, *Resampling Inference with Complex Survey Data*. Journal of the American Statistical Association, 1988. **83**(401): p. 231-241.
17. Shao, J., *Resampling methods in sample surveys*. Statistics, 1996. **27**(3-4): p. 203-254.
18. Shao, J., *Impact of the bootstrap on sample surveys*. Statistical Science, 2003. **18**(2): p. 191-198.
19. Davidson, R. and E. Flachaire, *The Wild Bootstrap, Tamed at Last*, in *Queen's University, Department of Economics, Working Papers*. 2001.
20. Jones, D.K., *The effect of gradient sampling schemes on measures derived from diffusion tensor MRI: a Monte Carlo study*. Magn Reson Med, 2004. **51**(4): p. 807-15.
21. Pierpaoli, C. and P.J. Basser, *Toward a quantitative assessment of diffusion anisotropy*. Magn Reson Med, 1996. **36**(6): p. 893-906.
22. Alexander, A.L., et al., *Analysis of partial volume effects in diffusion-tensor MRI*. Magn Reson Med, 2001. **45**(5): p. 770-80.
23. Schwartzman, A., R.F. Dougherty, and J.E. Taylor, *Cross-subject comparison of principal diffusion direction maps*. Magn Reson Med, 2005. **53**(6): p. 1423-31.
24. Behrens, T.E., et al., *Characterization and propagation of uncertainty in diffusion-weighted MR imaging*. Magn Reson Med, 2003. **50**(5): p. 1077-88.
25. Tuch, D.S., *Q-ball imaging*. Magn Reson Med, 2004. **52**(6): p. 1358-72.
26. Basser, P.J. and D.K. Jones, *Diffusion-tensor MRI: theory, experimental design and data analysis - a technical review*. NMR Biomed, 2002. **15**(7-8): p. 456-67.
27. Jones, D.K. and P.J. Basser, *"Squashing peanuts and smashing pumpkins": how noise distorts diffusion-weighted MR data*. Magn Reson Med, 2004. **52**(5): p. 979-93.
28. Koay, C.G., et al., *Investigation of anomalous estimates of tensor-derived quantities in diffusion tensor imaging*. Magn Reson Med, 2006. **55**(4): p. 930-6.

29. Chang, L.C., D.K. Jones, and C. Pierpaoli, *RESTORE: robust estimation of tensors by outlier rejection*. Magn Reson Med, 2005. **53**(5): p. 1088-95.
30. Mangin, J.F., et al., *Distortion correction and robust tensor estimation for MR diffusion imaging*. Med Image Anal, 2002. **6**(3): p. 191-8.

Chapter 4: Bootstrap Quantification of Cardiac Pulsation Artifact in DTI

4.1 Abstract

While several studies have shown the benefit of cardiac gating in the diffusion MRI with single-shot EPI acquisition, cardiac gating is still not commonly used, probably because it requires additional time and also many investigators may not be convinced that cardiac gating is worth the extra effort. We tested a clinically feasible protocol with minimal increase in the scan time, and quantified the effect of cardiac gating under partial or full Fourier acquisitions. Eight volunteers were scanned on a 3T scanner with a SENSE 8-channel head coil. DTI was performed with the use of a single-shot spin-echo EPI sequence. Diffusion weighted images were acquired along 32 directions, with or without cardiac gating and with partial or full Fourier acquisition. Vectorcardiography (VCG) was used to trigger at a minimum delay (30ms). Residual bootstrap was performed to estimate the uncertainties of DTI derived parameters. With partial Fourier, cardiac gating reduced the uncertainties and better efficiency in reducing the DTI parameter variability was also achieved even considering the increase in acquisition scan time. For full Fourier acquisition, gating slightly decreased the uncertainties but the efficiency was worse. Even with a minimum trigger delay which might not be the optimal scheme to avoid the majority of systole but allows clinically acceptable scan times, we have demonstrated that cardiac gating, especially of

partial Fourier acquisitions that are more susceptible to cardiac artifact, can reduce the uncertainties of DTI derived parameters in a time-efficient manner in DTI studies.

4.2 Introduction

Diffusion-weighted MRI (DWI), including diffusion tensor MRI (DTI), is designed to capture microscopic Brownian motion of water molecules, but these sequences are also sensitive to macroscopic movements as well, such as bulk subject motion and cardiac triggered brain pulsation. In order to minimize these sources of errors, single-shot EPI sequences are widely used for the acquisition of DWI. Single-shot EPI effectively freezes the bulk subject motion with its fast imaging time of usually around 100 ms or less, but cardiac pulsation artifact can still be found occasionally by visual inspection.

Several studies have shown the benefit of cardiac gating with single-shot EPI [1-5]. In practice, though, cardiac gating is not commonly used in clinical or research protocols, probably due to the following two reasons. First, cardiac gating calls for additional time and effort in the acquisition. It takes more time to acquire the same amount of data with gating, the amount of time depending on the gating scheme details (such as trigger delay used and number of slices acquired per cycle) and the subject heart rate. The additional step of attaching the triggering device to the patient (ECG or pulse oximeter devices) and the uncertainty introduced in not knowing the total scan time may also contribute.

Second, many investigators may not be convinced that cardiac gating is worth the extra effort, even given the cited studies advocating cardiac gating. This may partially stem from the way these studies were performed. In order to clearly demonstrate the cardiac pulsation artifact, quantification was focused on diffusion weighted images acquired in only one direction where the artifact was assumed to be strongest [3, 5], or DTI derived parameters were compared from datasets acquired only at specific points in the cardiac cycle [2, 4]. These conditions do not reflect the realistic effect of cardiac pulsation artifact on DTI since DTI data are acquired at multiple, diverse directions and from an array of points in the cardiac cycle. These are then combined by means of solving diffusion tensor elements, which are then used to calculate DTI derived parameters. For a researcher, the principal intent is the effect of cardiac pulsation in the DTI derived parameters.

In this work, we re-examined the benefit of cardiac gating in DTI with single-shot EPI acquisition by addressing the two issues above. First, cardiac gating was performed using a clinically feasible scheme which minimizes the increase in the scan time with some compromises in avoiding the systole. Second, the effect of cardiac gating was investigated by measuring the uncertainties in DTI derived parameters rather than the uncertainties in certain directions of the raw diffusion weighted images. Parameter uncertainties were estimated by bootstrap, a non-parametric, statistical technique based on data resampling. Finally, since it has

been suggested that partial Fourier acquisition, which is frequently used to save acquisition time, worsens the cardiac artifact [6] we examined the effect of full or partial Fourier acquisition.

4.3 Methods

4.3.1 Subjects / Acquisition

4.3.1.1 Group A

Five healthy, male volunteers (ages ranging from 24 to 60) were scanned on a Philips Intera 3T scanner (Philips, Best, The Netherlands) with a SENSitivity Encoding (SENSE) 8-channel head coil and gradients with a maximum strength of 33 mT/m. DTI was performed with the use of a single-shot spin-echo EPI sequence. Images were acquired with diffusion weighting ($b=1000\text{s/mm}^2$) along 32 non-collinear directions in addition to one image with minimal diffusion weighting. Acquisition parameters were as follows: field of view 240 x 240 mm; matrix size 96 x 96; 55 slices with thickness of 2.5 mm (no gap) covering the whole brain; measured and reconstructed voxel size 2.5 x 2.5 x 2.5 mm isotropic; TE 104 ms, SENSE reduction factor 2. These scans were acquired with or without cardiac gating and with full or partial (60%) Fourier acquisition. Each condition was repeated twice resulting in eight DTI acquisitions per subject.

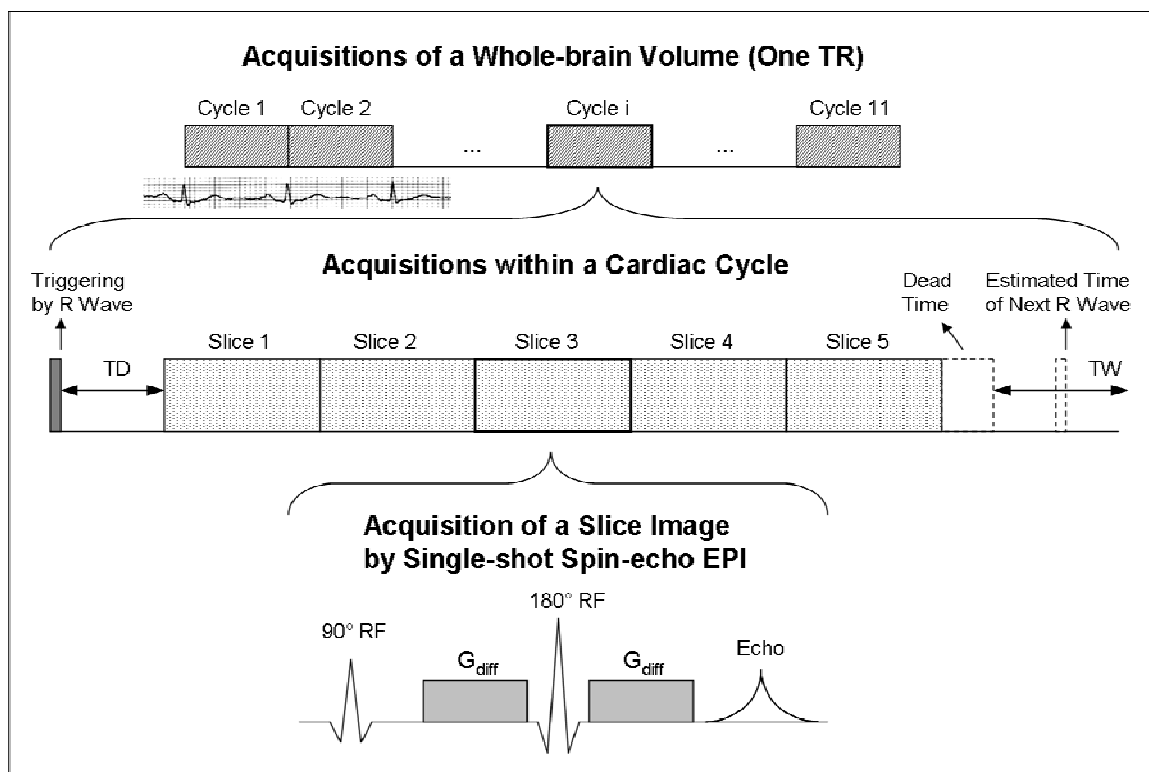


Figure 4-1. Schematics of cardiac gated DTI acquisition. TD: (trigger) delay time. TW: (trigger) window time.

Cardiac gating was performed in a way that can be easily implemented in clinical or research studies with a minimal increase in scan time, though it might not be ideal in terms of avoiding systole. We were interested in this approach since it was recognized that the cardiac gating schemes used in previous studies were often too lengthy and / or problematic to implement (such as using a dedicated pulse sequence). These approaches would discourage wider practice of cardiac gating even if it were indeed beneficial. It might be more realistic to use a compromised approach in gating as long as it proves be beneficial. The specific cardiac gating scheme implemented in our study was the following (Figure 4-1).

Triggering system was based on vectorcardiography (VCG) [7, 8]. Just like conventional ECG triggering system, VCG triggering utilizes the ECG signals from electrodes attached to the chest of the subject. By using spatial information provided by VCG, this triggering system can detect R waves more accurately. This not only makes triggering more reliable, but also reduces the patient setup time by eliminating the trial and error iterations of electrode placements for suitable ECG signals, often required with conventional ECG triggering system [7]. Time of trigger delay (TD) was set to the minimum value (30 ms) allowed by the scanner, and the maximum number of slices were acquired after triggering in each cardiac cycle. Since our system allows acquisition of one image (slice) every 145 ms, we could acquire around five slices per cycle, though the exact number could vary based on the heart rate. We used single-shot spin-echo EPI sequence available in most MR systems, as opposed to special sequence optimized for cardiac gating such as re-ordering slice acquisitions [3].

TR was set to a minimum value of 8 sec (≈ 145 ms per slice \times 55 slices) without cardiac gating. With gating, the nominal TR was in the range of 9.4 to 11 sec, based on the heart rate. In reality, the actual TR was slightly longer, though, since heart rate fluctuates during the scan and no data was acquired if heart rate is out of allowed range (-20 to 10% variation of heart rate). The nominal acquisition time for each DTI dataset was 4 min 48 sec without gating, and 6 min 7 sec with gating, 27% longer than without gating (see results for more details on the acquisition times).

4.3.1.2 Group B

Additional data were acquired from three subjects in order to further examine the effect of increased TD on the degree of cardiac artifact. Since large TD (such as 250 ms which would be sufficient to avoid the majority of cardiac pulsation induced brain motion; see discussions for details) increases the scan time to an unacceptable level, 50-100% in our system, we only considered maximal TD (TD_{MAX}) that does not increase the scan time compared to minimal TD (TD_{MIN}). This is possible since with TD_{MIN} there is some remaining time per cardiac cycle that are not utilized; single slice acquisition should be completed within a cardiac cycle and if not enough time remains at the end of the cycle for another slice acquisition, this time will simply be wasted. TD_{MAX} was determined by gradually increasing TD from minimal value until the acquisition time was increased.

The acquisition parameters for group B were set to be the same as group A except for TE and TR. Stronger gradient system (80 mT/m) were available at the time of acquisitions for group B, reducing TE from 104 ms to 72 ms (both partial and full Fourier acquisitions) and TR from 8 sec to 6.3 sec (partial Fourier) or 7.2 sec (full Fourier). The nominal scan times for each DTI datasets were 3 min 45 sec without gating and partial Fourier acquisition, 4 min 19 sec without gating and full Fourier acquisition, 4 min 34 sec to 5 min 2 sec with gating and partial Fourier acquisition, and 5 min 45 sec to 6 min 32 sec with gating and full Fourier acquisition. Three gating schemes (no gating, gating with TD_{MIN} , and gating with

TD_{MAX}) and two Fourier acquisitions (partial or full) were tested without any repeated acquisitions, so each subject received six DTI acquisitions lasting approximately 30 minutes. Cardiac gating was done as described for group A, with TD_{MIN} of 30 ms and TD_{MAX} in the range of 63 to 198 ms (partial Fourier acquisition) or 292 to 329 ms (full Fourier acquisition).

4.3.2 Analysis

For all DTI datasets, subject head movement between different volume acquisitions was corrected by registration of diffusion weighted images to the minimally weighted ($b=0$) image using FLIRT (FMRIB's Linear Image Registration Tool, Oxford, UK) [9]. Transformation was modeled as linear equations with six degrees of freedom, and images were resampled by sinc interpolation (sinc full-width of seven voxels) in order to reduce interpolation errors and also to minimize modulation of uncertainties that can arise during the interpolation [10]. Then, images were slightly smoothed using a $3 \times 3 \times 3$ voxel Gaussian kernel with full width at half maximum (FWHM) of one voxel to suppress the high frequency noise. Diffusion tensors were estimated on each voxel by log-linear weighted least squares multiple regression, and then maps of diffusion tensor derived parameters such as fractional anisotropy (FA) were created [11].

To estimate the uncertainties of these DTI derived parameters on a voxel basis, residual bootstrap was performed [11]. Residual bootstrap is one of the model-based bootstrap resampling techniques where residuals, differences between

measured MR signals and the model-fitted signals, are resampled on the whole dataset to create multiple bootstrap resamples of the original sample (i.e. measured signals). The dispersion of DTI derived parameters calculated from these resamples reflects the uncertainty of the original parameter. Uncertainties of scalar parameters were expressed as standard errors, while the uncertainties in the direction of vectors such as primary eigenvectors were expressed as 95 percentile confidence intervals of minimum angle. Resampling from an original sample was iterated 200 times.

Initial DTI processing and bootstrap described above were carried out in the native space (i.e. without any registrations except for the correction for the bulk head motion), and bootstrap-estimated uncertainty maps were normalized to MNI space before the final analysis. Normalization was performed as follows. The transformation was initially approximated by linear transformation of b=0 images from each DTI datasets to the single subject T1 image in the MNI space using FLIRT. FA maps from each datasets were registered to the MNI space using the same transformation. Then, high-order non-linear registration was performed [12]. The transformation was parameterized using cubic B-spline functions, and empirically chosen node spacing of six voxels in x, y, and z directions was used. One FA map from a subject was arbitrarily chosen, and all the other FA maps, from different datasets of a subject and from different subjects, were all registered to the chosen FA map. Finally, all DTI uncertainty maps were

resampled to MNI space according to the transformation information acquired previously.

To visualize the effect of cardiac gating on the DTI uncertainty maps, for each subject, maps of gated to non-gated uncertainty ratios (R_U) were created for either partial or full Fourier space acquisitions. For group A, two DTI datasets were acquired under the same conditions, so two uncertainty maps were averaged before further analysis. Efficiency ratio (R_E) maps, which can demonstrate the brain regions where cardiac gating achieved smaller or larger degree of uncertainty per unit time, were also created. R_E is conceptually similar to SNR efficiency, which is a measure of achieved signal-to-noise ratio per unit time. The term SNR efficiency deemed narrowly defined in the context of MR signals and their noise expressed as the standard deviation. Definition of R_E accommodate DTI derived parameters and their uncertainties in more general forms. Nevertheless, the scheme of normalizing different acquisition time by dividing the SNR by square root of the time was equally applied to R_E , as shown in the equation below,

$$R_E = \frac{PUR_g / \sqrt{T_g}}{PUR_{ng} / \sqrt{T_{ng}}} = \frac{U_{ng}}{U_g} \times \sqrt{\frac{T_{ng}}{T_g}} = \frac{1}{R_U} \times \frac{1}{\sqrt{R_T}}$$

where PUR (parameter-to-uncertainty ratio; P/U) is DTI derived parameter divided by its degree of uncertainty, R_T (time ratio) is the ratio of gated to non-gated acquisition times, and g and ng refers to gating and non-gating each. It was assumed that the estimated DTI derived parameters with or without gating

were comparable enough so that PUR_g / PUR_{ng} could be reduced to the inverse of R_U . All these maps, either from one subject or from all subjects together, were filtered with 3x3x3 voxel median kernel to better demonstrate the general pattern.

In addition to ratio maps showing the spatially-dependent effects of cardiac artifact and gating, the effects were also quantified as a single value representing the whole brain. Since major white matter bundles, where researchers are usually more interested in, are highly anisotropic in general, ROI of voxels with $FA > 0.4$ was created in the FA map. The mean of ratios described above within this ROI was reported.

For group B, two different gating schemes exist, so the ratio maps were created for both 1) gating with TD_{MIN} versus no gating and 2) gating with TD_{MAX} versus no gating.

4.4 Results

4.4.1 Group A

The Figure 4-2 shows the acquisition times for each DTI datasets in all five subjects. It took 6 min 7 sec \pm 23 sec for the acquisitions with cardiac gating, and 4 min 48 sec without gating, making gated acquisitions 27% lengthier on average in scan time. The heart rates for these subjects were 65 \pm 4 beats per minutes.

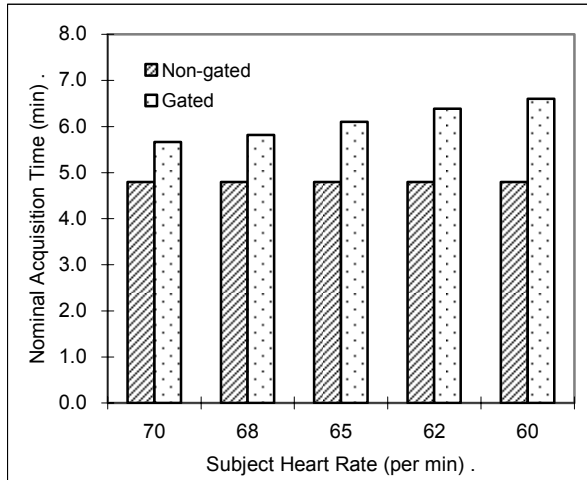


Figure 4-2. Nominal acquisition times for each subject in group A. Subjects are sorted by their heart rate during the acquisition in the x axis.

Typical diffusion weighted images acquired are shown in the Figure 4-3, with possible cardiac artifacts visible on the non-gated data. Artifacts were found to be in the diverse locations in the brain, and they were not confined to the images with diffusion weighting approximately in the z direction but could be found in other images as well. The severity of artifacts were also diverse, though artifacts as severe as what has been demonstrated previously [5] were rare, if any, except for those in the cerebellum.

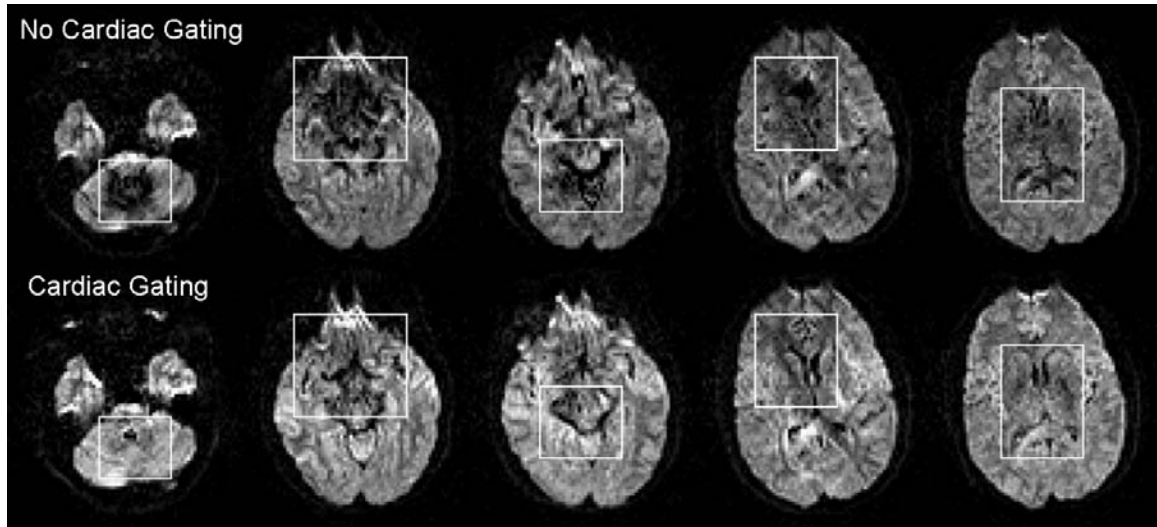


Figure 4-3. Typical diffusion weighted images without cardiac gating (top row) and with gating (bottom row). Images were acquired with partial Fourier acquisition, and top and bottom row images are from the same slice of the same subject. White boxes highlights possible cardiac pulsation artifacts without gating.

Figure 4-4 shows typical maps of DTI derived parameters as well as their bootstrap-estimated uncertainty maps from one non-gated partial Fourier dataset. Smaller confidence intervals (meaning high certainty in the estimation) of primary eigenvectors in the white matter can be appreciated. Increased FA standard errors are found in the central part of the brain, including deep gray matter and corpus callosum, possibly reflecting the cardiac artifacts, lower SNR of multiple channel coils in the center of the FOV, and the inherently low signals in the deep gray matter. From three diffusivity maps, $Ev1$ ($=\lambda_1$) had larger uncertainty compared to MD ($=(\lambda_1+\lambda_2+\lambda_3)/3$) or EvT ($=(\lambda_2+\lambda_3)/2$), since no averaging between different eigenvalues is done for $Ev1$, and the values of this parameter is also larger than MD and EvT values.

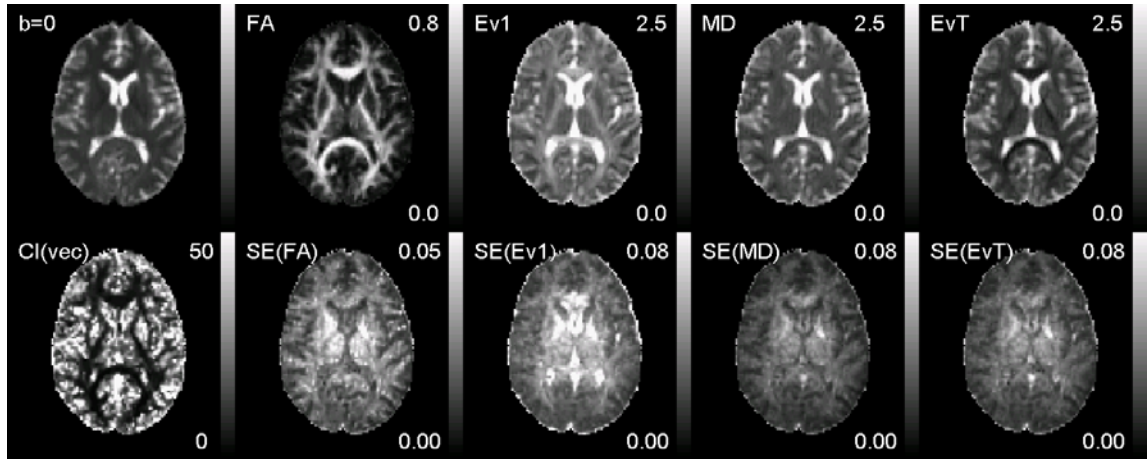


Figure 4-4. Typical maps of DTI derived parameters (top row) and their bootstrap-estimated uncertainty maps (bottom row). CI: confidence interval. SE: standard error. Ev1: primary eigenvalue = λ_1 , also known as axial diffusivity. MD: mean diffusivity = $(\lambda_1 + \lambda_2 + \lambda_3)/3$, also known as $\text{Tr}(\mathbf{D})/3$. EvT: transverse eigenvalue = $(\lambda_2 + \lambda_3)/2$, also known as radial diffusivity.

Figure 4-5 shows the maps of R_U and R_E with partial Fourier acquisition. In the middle row of the figure, virtually all voxels in the brain have R_U smaller than one (depicted in blue), meaning that cardiac gating reduces the uncertainties everywhere in the brain. In the bottom row, R_E values were larger than one (depicted in blue) in the voxels located mainly in the central part of the brain slice-wise, and at the level or below the corpus callosum. This suggests that cardiac gating not only reduces the uncertainties in these regions but that reduction can be big enough to compensate for the increased scan time in terms of time efficiency. In other words, if we were given the same time to scan, it can makes more sense to do cardiac gating and acquire less data (i.e. fewer averages or directions) than get more data without gating, assuming that we can still acquire sufficient data even with cardiac gating.

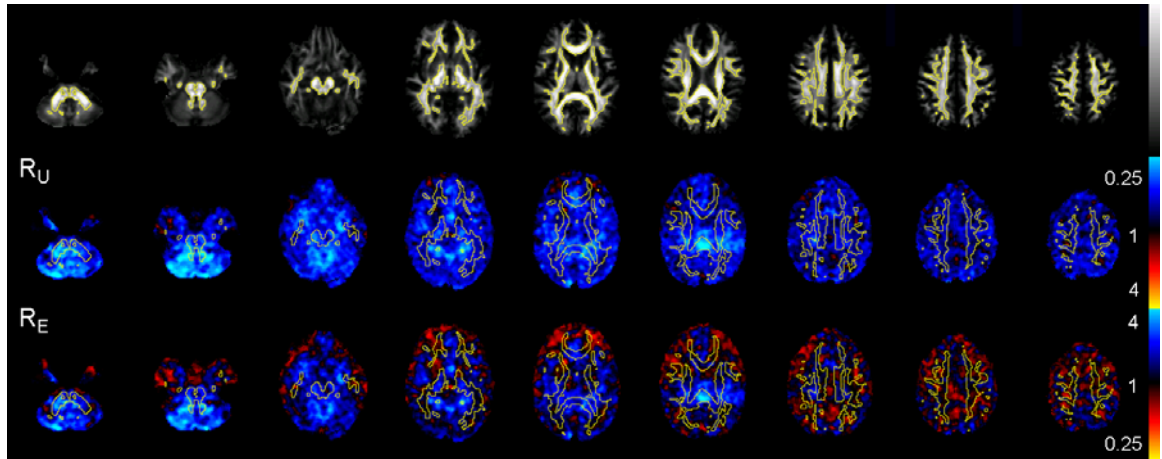


Figure 4-5. Maps of R_U (gated to non-gated uncertainty ratio) and R_E (gated to non-gated efficiency ratio) for the 95th percentile confidence intervals of the primary eigenvector directions, with partial Fourier acquisition. These maps show voxel-wise median values of ratios from all five subjects. Top row: FA maps from the average of five subjects with yellow contour depicting the high anisotropy region, which was used to report results in Figure 4-9 and Figure 4-10. Nine slices are shown, from the level of cerebellum to the top of the brain. The fifth slice in each row corresponds to the same slice shown in Figure 4-4. Middle / bottom row: R_U and R_E maps in the same slices as top row. The same yellow contours as top row are overlaid as well.

Figure 4-6 shows the same maps as Figure 4-5 but with full Fourier acquisition. In the middle row of the figure, R_U is no more completely dominated by values less than one, and especially in the periphery of the brain, numerous voxels with R_U larger than one (depicted in red) can be observed. In the bottom row, most of the voxels have R_E smaller than one (depicted in red). These maps suggest that cardiac gating is not time efficient anymore, though it is still effective in reducing the uncertainties.

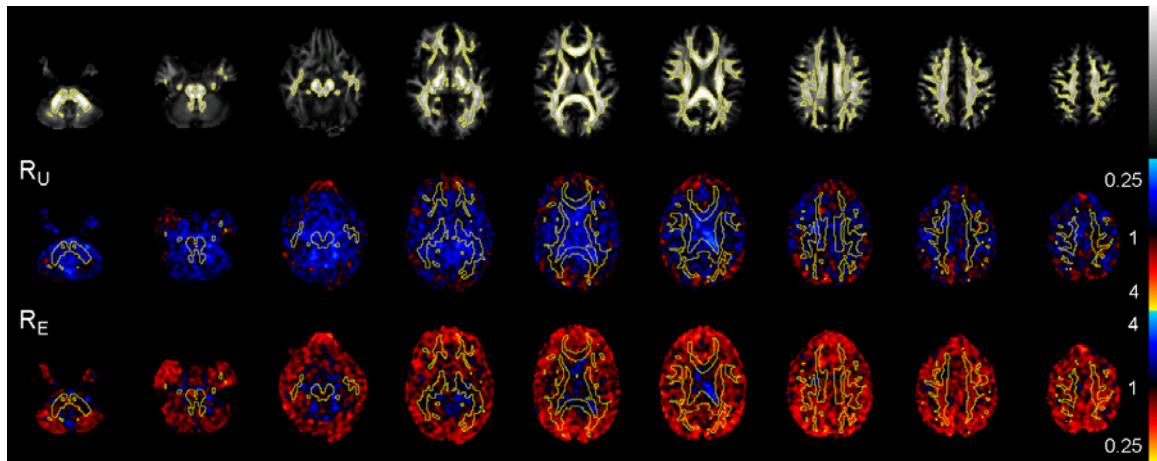


Figure 4-6. The same maps as Figure 4-5, except that data are from full Fourier acquisition.

These trends shown in the Figure 4-5 and Figure 4-6 were closely replicated with uncertainties of scalar DTI derived parameters. Figure 4-7 shows the maps of R_U and R_E with partial and full Fourier acquisitions in one slice, and uncertainties are standard errors of FA, Ev1, MD and EvT. Other than spatially less-varying, the maps show the same trend; with partial Fourier acquisition, cardiac gating globally reduced the uncertainties and was particularly time efficient generally in the central part of the brain; with full Fourier acquisition, while cardiac gating generally reduced the uncertainties, this was at the cost of reduced time efficiency compared to no gating.

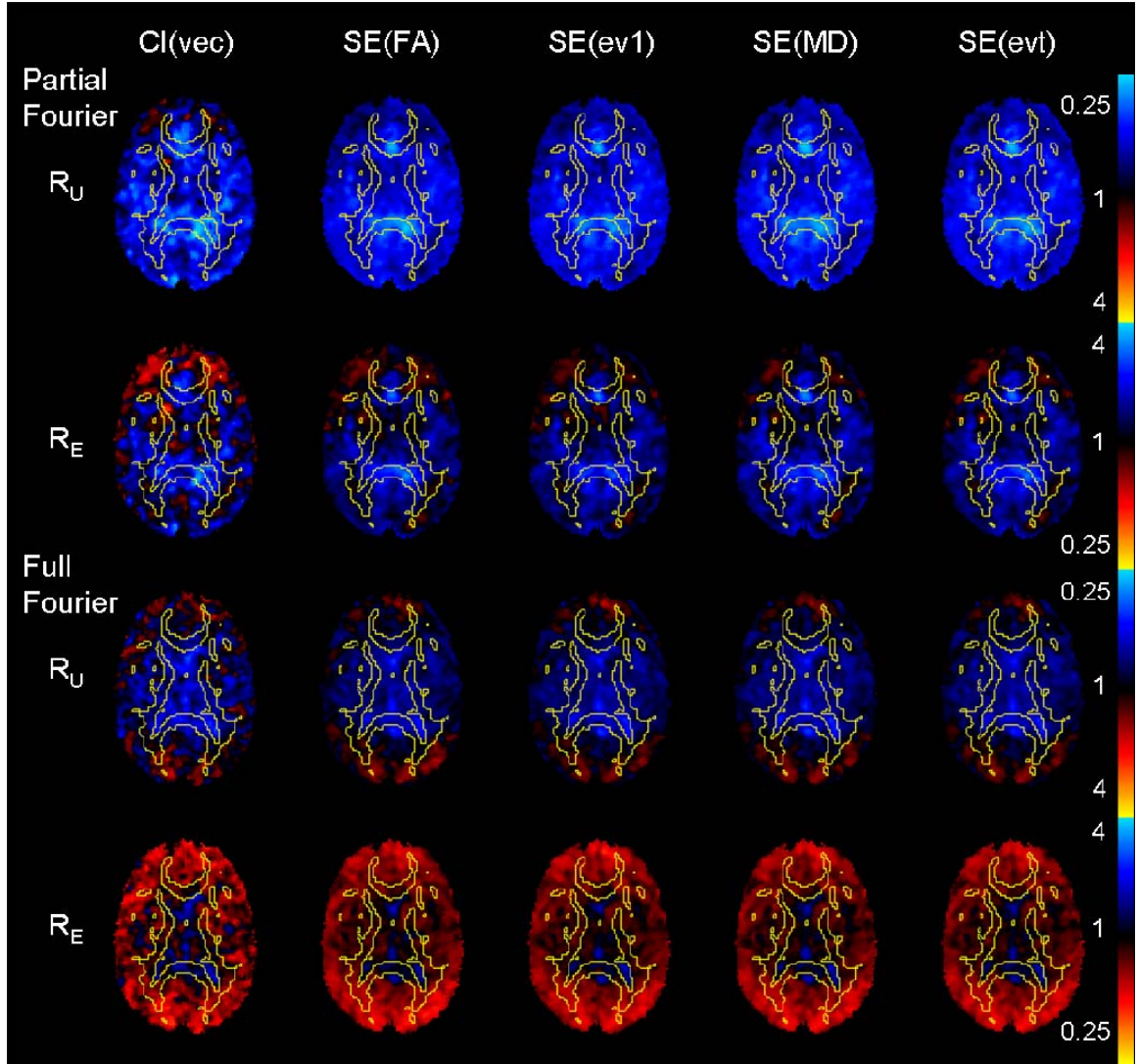


Figure 4-7. Maps of R_U and R_E with partial or full Fourier acquisitions in the same slice as Figure 4-4 (and the same slice as the fifth slice of Figure 4-5 and Figure 4-6). In addition to Cl(vec) which was already shown in Figure 4-5 and Figure 4-6, standard errors of scalar parameters (FA, Ev1, MD, EvT) are shown as well for comparison.

Figure 4-8 shows spatial intensity projections of R_U for partial and full Fourier acquisition. These images are presented to make easy comparison with previously reported results [3, 5]. For partial Fourier acquisition, our projections closely resembled the results presented in [5], with the uncertainty reduction effect of cardiac gating more pronounced in the lower half of the brain and in the

central region. Projections also make clear that with full Fourier acquisition the cardiac gating effect is much less. The distribution of the regions where cardiac gating was effective with full Fourier acquisition did not seem to correspond highly with the data for partial Fourier acquisition, other than the fact that they are both concentrated in the lower part of the brain.

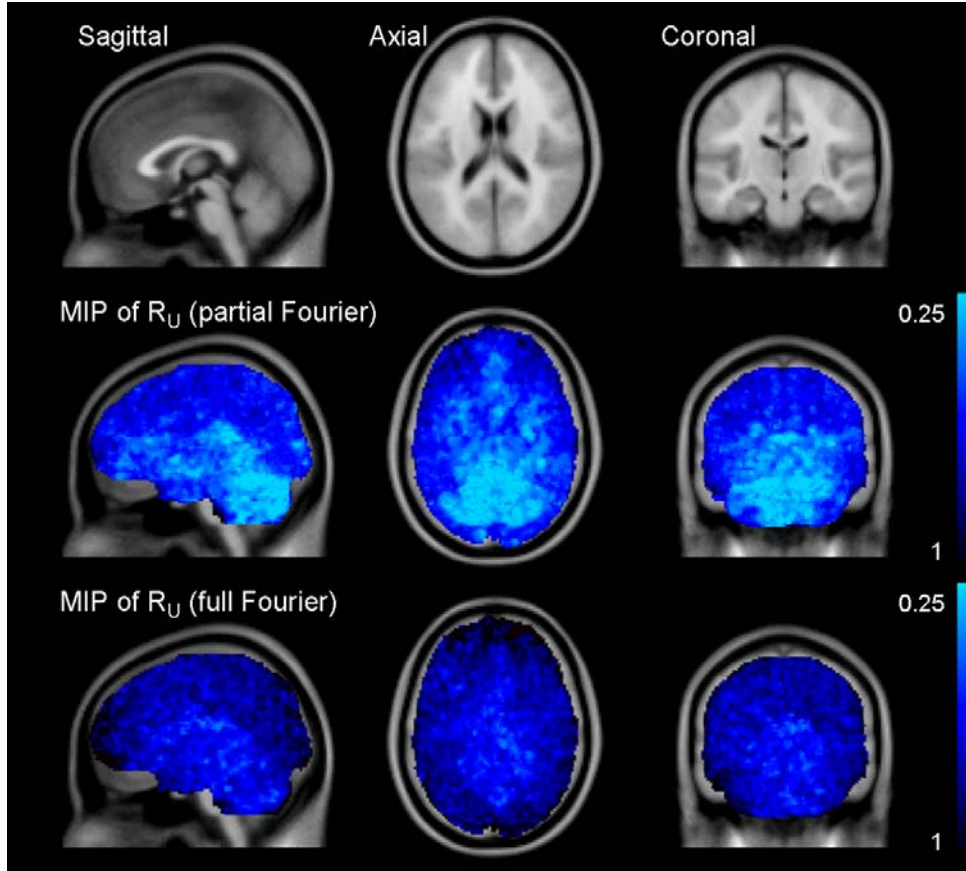


Figure 4-8. Spatial projections of R_U for partial and full Fourier acquisition (i.e. projections from the data shown in the middle row of Figure 4-5 and Figure 4-6) toward sagittal, axial and coronal planes. Since virtually all R_U values were below one and since lower values indicate stronger reduction of uncertainties by cardiac gating, the projections are actually minimum intensity projection, so that stronger reduction of uncertainties will stand out. R_U is calculated from $CI(\text{vec})$.

Figure 4-9 shows the plots of R_U and R_E in each subject averaged in the high anisotropy white matter regions of the whole brain. The median value of R_U from

all subjects was 0.82, meaning that the uncertainties (95 percentile confidence interval of primary eigenvectors) are 18% smaller with cardiac gating and partial Fourier acquisition. With full Fourier acquisition, R_U values were more variable between subjects and the median value was 0.93, i.e. 7% smaller uncertainties with cardiac gating. The R_E values for all subjects were larger than one with partial Fourier acquisition, with the median value of 1.08. This suggests that cardiac gating is actually more time efficient by around 8%. With full Fourier acquisition, all but one subject had R_E lower than one with median of 0.95, meaning that for most subjects the time efficiency was reduced by cardiac gating.

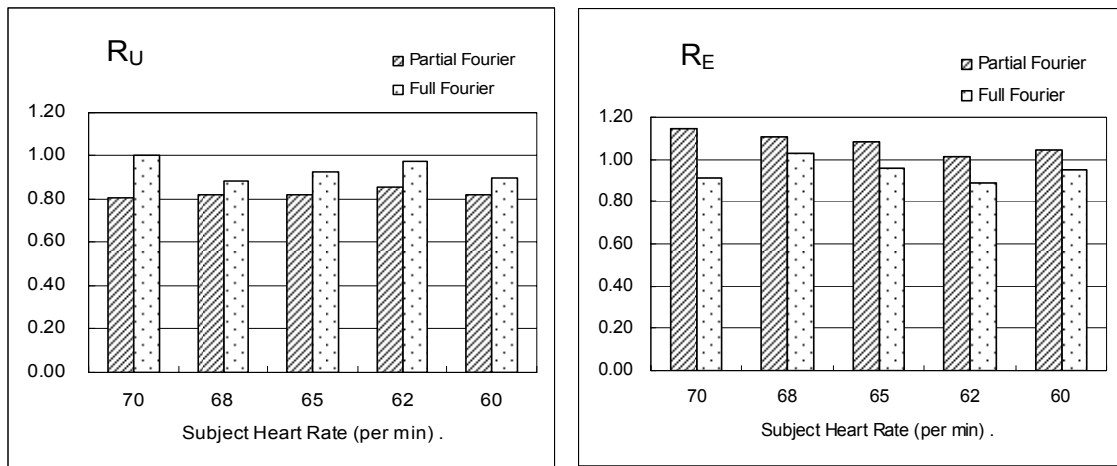


Figure 4-9. Plots of R_U and R_E in each subject averaged in the high anisotropy white matter regions of the whole brain.

4.4.2 Group B

Figure 4-10 shows the plots of R_U and R_E in each group B subject averaged in the high anisotropy white matter regions of the whole brain, similarly to Figure 4-9. With TD_{MIN} , the TD as group A, results were similar to group A. Cardiac gating was able to reduce the uncertainties by the median value of 15% (partial

Fourier acquisition) or 5% (full Fourier acquisition). Except for one case (subject with the heart rate of 55 per min, partial Fourier acquisition), TD_{MAX} always reduced R_U , meaning that the longer delay was more effective. Since TD_{MAX} does not increase scan time compared to TD_{MIN} , R_E was generally larger with TD_{MAX} , though the trend that partial Fourier acquisition benefits from cardiac gating but not full Fourier acquisition in terms of time efficiency still seems to hold with TD_{MAX} .

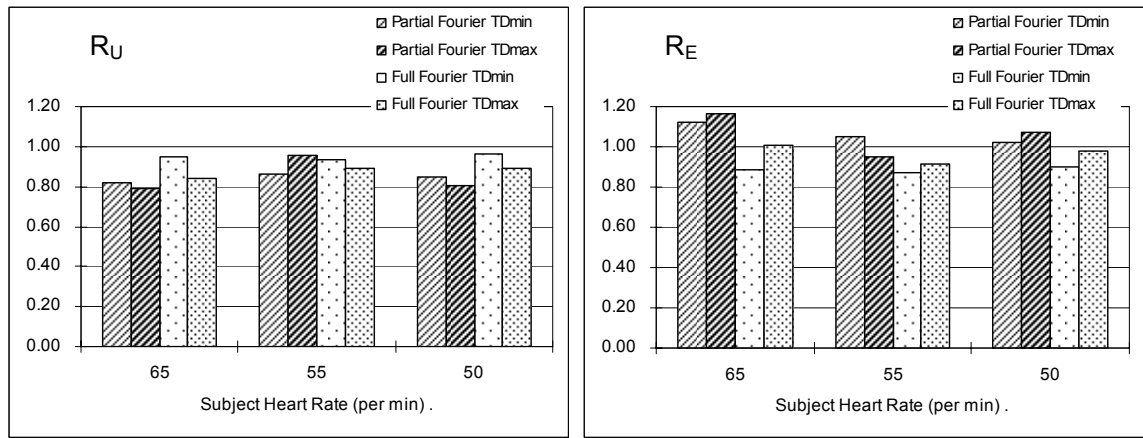


Figure 4-10. Plots of R_U and R_E in each group B subject averaged in the high anisotropy white matter regions of the whole brain. R_U and R_E calculated from cardiac gating with TD_{MAX} as well as TD_{MIN} are both shown side by side for easier comparison.

4.5 Discussion

Cardiac gating is still rarely used in brain DTI studies probably because investigators do not believe that it is not worth extra time and effort despite potential benefits. Indeed, according to the previous studies on cardiac gating [3, 5, 6], the total acquisition time with cardiac gating can be two to three times longer than without gating. Even in our system where the acquisition time per

slice is much shorter and cardiac gating is more efficient due to less dead time compared to typical DTI acquisitions including [3, 5, 6], the total acquisition time can be twice as long if TD is large enough as suggested by previous studies about brain pulsation (more details are below). Doubling or tripling the scan time is not a viable option for many, if not the majority of, clinical and even research studies due to study time constraints. An optimized scheme with only relatively minor increase in the scan time has been reported [3]. However, it is a customized sequence not widely available and it requires further modifications depending on the acquisition setting, and is an additional source of complexity in the cardiac gating.

In this study, we took an alternate approach, cardiac gating with minor increase of scan time and implementation effort at the expense of a less than ideal triggering scheme. Our results show that while our gating scheme of using minimum TD might not be as effective as a scheme with a longer TD, it clearly reduces cardiac artifact to the level where even time efficiency (or scanning efficiency, achieved degree of uncertainties under given time) is higher compared to no gating scheme. Higher time efficiency applies only to partial Fourier acquisition, but considering that most single-shot EPI studies these days are performed in partial Fourier scheme, it is likely that our results would generally apply to typical DTI studies.

We would like to point out that the mechanism for the effectiveness of cardiac gating with minimum TD could be effective is not well understood yet. Before discussing this issue further, it is worthwhile noting that triggering is typically achieved via ECG (and VCG) or pulse oximeter and that the time relationship between triggering and cardiac pulsation induced brain motion is quite different for these two triggering mechanisms. ECG detects R waves of the heart directly while pulse oximeter detects the pulse wave in the finger (thus called peripheral gating), and it has been reported that the ECG trigger precedes the peripheral trigger by around 250 ms fairly consistently [4]. This means that the ECG TD of 300 ms is equivalent to peripheral TD of 50 ms. Though exact details vary, there is an consensus that cardiac pulsation induced brain motion is greatest from 100 to 250 ms after the ECG detected R waves [13, 14]. This time period corresponds to the last period of 150 ms just before peripheral triggering, and this explains why minimum TD relative to the peripheral gate is almost as effective as longer TDs relative to the peripheral gate [5].

On the other hand, triggering in our study occurred by VCG (equivalent to ECG in the timing of triggering) and thus our minimum TD happens before the peak of brain motion. The acquisition of the first 2D image after the trigger occurs at the moment where the brain motion may be largest. Considering the fact that around five slices are acquired in each cardiac cycle and that first of those are constantly acquired at the time of systole, 20% of the images acquired by our gating scheme should be affected by cardiac pulsation artifact, actually similar to the

percentage reported without any gating at all [3, 5]. Still, both objective evidence of bootstrap estimated uncertainties and subjective visual inspection of the raw images suggest that our cardiac gating with minimum TD reduces the artifact effectively.

There is another study which observed this phenomenon. A study by Brockstedt et al. [1] compared three ECG triggering protocols with TDs of 100 ms, 400 ms, and no triggering. Their results showed that ECG gating with TD of 100 ms clearly reduces the uncertainties compared to no triggering (statistically significant) though not as much as TD of 400 ms (the differences in the degree of uncertainties between TD of 100 ms and 400 ms were not statistically significant). This is particularly interesting since they had chosen TD of 100 ms to represent diffusion encoding during the time period of maximum brain motion to be compared to TD of 400 ms that was expected to avoid most pulsation effects. Trigger delays of 100 ms was expected to yield worse cardiac artifacts than no gating at all but in reality this did not hold and the effectiveness of the short delay was almost comparable to TD of 400 ms.

Though the mechanism behind the effectiveness of ECG (or VCG) gating with minimum TD remain elusive, our results nevertheless suggest that with partial Fourier acquisition, this cardiac gating scheme not only reduces the artifact but is also more efficient in terms of achieved degree of uncertainties per unit acquisition time. This possibility was previously observed [5], but based on the

results 1) from only voxels with statistically significant cardiac artifacts, 2) only with diffusion weighting in the z direction (artifact expected to be the worst), 3) only for diffusion weighted signals (as opposed to DTI derived parameters), and 4) with TD of 500 ms, which will lead to clinically (or for many research protocols) unfeasible scan times (3.4 times of ungated). We drew our conclusion 1) from all the voxels of major white matter bundles, 2) with diffusion weighting in numerous directions which is typical in DTI, 3) based on the DTI derived parameters of interest which are the eventual metrics of interest, and 4) with TD of 30 ms, which only increases the total acquisition time by around 27%. We believe that our results provide convincing evidence that short TD provides better scanning efficiency and with realistic scan times.

Reducing the uncertainties in DTI derived parameters is not the only benefit of cardiac gating. As described previously, cardiac gating reduces the spatial varying characteristics of degree of uncertainties since cardiac artifacts are worse in the lower central part of the brain [3, 5] which was also confirmed by our study. Cardiac gating also reduces any potential bias in the estimated DTI derived parameters [2, 4], though their results may over-emphasize the bias effect since they compared datasets from different TDs rather than simply comparing results of gating versus no gating.

Another crucial finding of our study is that cardiac gating does not decrease the uncertainties in DTI derived parameters as much with full Fourier acquisition.

Though this might imply that cardiac gating is actually less effective, we believe that a more plausible explanation is that full Fourier acquisition is more immune to cardiac pulsation artifacts, leaving less room for improvement by cardiac gating. When we visually inspected raw diffusion weighted images, we were much less likely to find cardiac pulsation artifacts in the full Fourier data. Also, a study by Robson et al. [6], which motivated us to examine full Fourier acquisition data as well, suggested that Margosian reconstruction scheme [15] after partial Fourier acquisition is the major source of cardiac pulsation artifact. Typically, single-shot EPI DTI studies are acquired partially in the K-space (partial Fourier) and reconstructed in ways similar to the Margosian approach (e.g. homodyne method) by completion of K-space by the property of Hermitian conjugate symmetry and phase correction using symmetrically acquired data near the center of K-space. For this reason above, we therefore suggest that full Fourier acquisition would lead to less cardiac pulsation artifacts by avoiding this reconstruction approach.

The fact that full Fourier acquisition is more immune to cardiac pulsation artifacts provides us with the opportunity to reconsider the trend of always preferring partial over full Fourier acquisitions with EPI sequences for diffusion weighted images. Partial Fourier acquisitions have been preferred because they reduce the image blurring along the phase encoding direction due to less T_2 and T_2^* filter effect and TE can be shorter that leads to higher SNR and shorter scan times [16]. A recent study [17] concluded that the overall best SNR efficiency was

achieved with a SENSE reduction factor of 2.1 and partial Fourier encoding of 60%, with an eight-element head coil array and a b-value of 1000 s/mm^2 for diffusion weighting. Unfortunately, the optimizations studies for the fraction of K-space coverage with single-shot EPI sequences for diffusion weighted images so far have not considered effects of cardiac pulsation artifacts. Since the advancement in imaging gradient performance and parallel imaging techniques which make even full Fourier acquisition fairly easy without sacrificing spatial resolution, we have to wonder whether partial Fourier acquisition is still the better option. While even our practical cardiac gating scheme will increase the scan time by around 27%, full Fourier acquisition takes only marginally longer than partial acquisition and cardiac gating can be skipped saving lots of time. For the purpose of this discussion we should declare that we did observe lower standard errors and fewer artifacts with the ungated full Fourier compared to the partial Fourier data. However, due to difficulty in quantifying the effects of smoothing with full versus partial Fourier acquisitions, we did not present the direct comparison between the errors for these experiments and are wary of drawing conclusions from this result alone.

Only one particular cardiac gating scheme was tested in this study, but we believe that improvements can be made to make this gating scheme even more practical. First, currently our system allows fixed number of slices to be acquired in each cardiac cycle based on the heart rate entered just before the scan. Cardiac gating can be more efficient if the number of slices acquired per

heartbeat is adjusted during the acquisition, reflecting the trend of heart rate changes. For instance, if the heart rate decreases during the scan to a point where one more slice can be acquired, then one more slice can be acquired based on the re-adjusted heart rate based on the stability of the heart at some later time during the experiments. Secondly, while we used VCG gating because of its superior performance, peripheral gating might make more sense in cardiac gating for brain, not heart, imaging. As described above in the discussion, because of the inherent delay (≈ 250 ms) between the R wave and the peripheral pulse wave, most of the significant brain motions by cardiac pulsation have already occurred by the time of peripheral triggering. Contrary to ECG (or VCG) gating, peripheral gating with minimum TD makes more sense in terms of avoiding the brain motions. Alternatively, even with the cardiac R wave trigger, allowance for acquisition during the same period that a trigger is expected would enlarge the acquisition time window without risk of acquiring during the brain pulsation period. In fact, the known delay between cardiac R wave and the brain pulsations can be used to great advantage in this way since time normally lost while waiting for the next trigger would be fully utilized and only a period corresponding to worse brain pulsatile motion avoided. While the brain motions occur at the end of cardiac cycle with peripheral gating, acquisitions are not done at this time anyway in order to wait for the next trigger (called TW or window time). Using the same time to wait the next trigger and to avoid the brain motions will allow even more time for acquisitions. It will be a subject of future study whether peripheral gating, even with its shortcoming of less stable triggering, can

outperform the scheme presented in our study. Last, as gradient performance increases, cardiac gating will become more efficient since there will be less dead time, time that is not long enough for one slice acquisition that has to be wasted.

As an objective measure of uncertainties of DTI derived parameters, residual bootstrap was utilized. While residual bootstrap was shown to outperform other DTI bootstrap methods by a Monte Carlo simulation study [11], we have to take into consideration that the assumptions made by residual bootstrap might not be realized. We have analyzed the data with wild bootstrap [11, 18] and the results for the primary eigenvector direction were almost identical to the results with residual bootstrap (results not shown). This gave us confidence that the selection of bootstrap will not sway our results, though they are both model-based resampling techniques, and limitations with model-based resampling still remains. While our acquisition scheme of many different directions with only one or two repetition does not suit well with repetition-based resampling [11], it would be an interesting subject to quantify the cardiac pulsation artifacts with repetition-based resampling on DTI datasets with small number of directions and many repetitions.

4.6 Conclusion

We have demonstrated that VCG-triggered cardiac gating with minimum trigger delay could be a good compromise between the need for cardiac gating and maintaining practical scan time. Our gating scheme has higher time efficiency (or scanning efficiency) compared to no gating when partial Fourier acquisition was

performed. Full Fourier acquisition was relatively immune to cardiac pulsation artifacts decreasing the need for cardiac gating. When single-shot EPI DTI studies are done with partial Fourier acquisition, we recommend the practical cardiac gating scheme presented in this study. If full Fourier acquisition is viable then full Fourier acquisition without cardiac gating should be considered as well.

4.7 Reference

1. Brockstedt, S., et al., *Triggering in quantitative diffusion imaging with single-shot EPI*. Acta Radiol, 1999. **40**(3): p. 263-9.
2. Jones, D.K. and C. Pierpaoli. *Contribution of cardiac pulsation to variability of tractography results*. in *13th Annual Meeting of the ISMRM*. 2005. Miami Beach, FL.
3. Nunes, R.G., P. Jezzard, and S. Clare, *Investigations on the efficiency of cardiac-gated methods for the acquisition of diffusion-weighted images*. J Magn Reson, 2005. **177**(1): p. 102-10.
4. Pierpaoli, C., et al. *Analyzing the contribution of cardiac pulsation to the variability of quantities derived from the diffusion tensor*. in *11th Annual Meeting of the ISMRM*. 2003. Toronto, Ontario, Canada.
5. Skare, S. and J.L. Andersson, *On the effects of gating in diffusion imaging of the brain using single shot EPI*. Magn Reson Imaging, 2001. **19**(8): p. 1125-8.
6. Robson, M.D. and D.A. Porter, *Reconstruction as a source of artifact in non-gated single-shot diffusion-weighted EPI*. Magn Reson Imaging, 2005. **23**(9): p. 899-905.
7. Chia, J.M., et al., *Performance of QRS detection for cardiac magnetic resonance imaging with a novel vectorcardiographic triggering method*. J Magn Reson Imaging, 2000. **12**(5): p. 678-88.
8. Fischer, S.E., S.A. Wickline, and C.H. Lorenz, *Novel real-time R-wave detection algorithm based on the vectorcardiogram for accurate gated magnetic resonance acquisitions*. Magn Reson Med, 1999. **42**(2): p. 361-70.
9. Jenkinson, M. and S. Smith, *A global optimisation method for robust affine registration of brain images*. Med Image Anal, 2001. **5**(2): p. 143-56.
10. Rohde, G.K., et al., *Estimating intensity variance due to noise in registered images: applications to diffusion tensor MRI*. Neuroimage, 2005. **26**(3): p. 673-84.
11. Chung, S., Y. Lu, and R.G. Henry, *Comparison of bootstrap approaches for estimation of uncertainties of DTI parameters*. Neuroimage, 2006. **33**(2): p. 531-41.

12. Sdika, M., *A fast nonrigid image registration with constraints on the Jacobian using large scale constrained optimization*. IEEE Transactions on Medical Imaging, 2008. **27**(2): p. 271-281.
13. Greitz, D., et al., *Pulsatile brain movement and associated hydrodynamics studied by magnetic resonance phase imaging. The Monro-Kellie doctrine revisited*. Neuroradiology, 1992. **34**(5): p. 370-80.
14. Poncelet, B.P., et al., *Brain parenchyma motion: measurement with cine echo-planar MR imaging*. Radiology, 1992. **185**(3): p. 645-51.
15. Margosian, P. *Faster MR imaging - imaging with half the data*. in *Society for Magnetic Resonance in Medicine*. 1985.
16. Hyde, J.S., B.B. Biswal, and A. Jesmanowicz, *High-resolution fMRI using multislice partial k-space GR-EPI with cubic voxels*. Magn Reson Med, 2001. **46**(1): p. 114-25.
17. Jaermann, T., et al., *Influence of SENSE on image properties in high-resolution single-shot echo-planar DTI*. Magn Reson Med, 2006. **55**(2): p. 335-42.
18. Whitcher, B., et al., *Using the wild bootstrap to quantify uncertainty in diffusion tensor imaging*. Hum Brain Mapp, 2007.

Chapter 5: Bootstrap Voxel-wise Serial DTI Analysis

5.1 Abstract

Diffusion tensor MRI (DTI) is emerging as a powerful tool to overcome the limitations of conventional MRI and CT in the evaluation of traumatic brain injury (TBI), especially diffuse axonal injury (DAI). We developed a fully automated voxel-wise serial DTI analysis for robust and sensitive detection of progressive microstructural white matter changes in each TBI patient. This technique called Bootstrap-based Longitudinal Analysis of DTI Estimates (BLADE) identifies clusters of voxels with statistically significant FA changes over time, with the help of DTI residual bootstrap. High quality 3T DTI data with ASSET parallel imaging, 55 diffusion directions, and 1.8 mm isotropic voxels were acquired from six mild TBI patients and seven control subjects on three separate time points. BLADE was able to identify voxels with FA increase over time within lesions of TBI patients with cortical contusion or DAI. More interestingly, FA decrease over time in the normal appearing white matter adjacent to the lesions was also detected possibly indicating the presence of ongoing Wallerian degeneration from the lesions. BLADE is the first technique that detected voxel-wise FA changes in a single subject with high statistical confidence, which can be a powerful tool in analyzing neurological disease with high inter-subject variability in the time course and spatial distribution of pathologies.

5.2 Introduction

Traumatic brain injury (TBI) is a leading cause of mortality and morbidity in the US, especially in young people under age 45. Conventional MRI and CT are limited for the accurate evaluation of TBI due to their tendency to underestimate the extent of injury, especially diffuse axonal injury (DAI) [1], and absence of quantitative pathophysiological information. Diffusion tensor MRI (DTI) is emerging as a powerful tool to overcome these limitations. The extensive spatial heterogeneity of DTI fractional anisotropy (FA) presents a challenge for detecting changes in this parameter. Hence, manual ROI analysis has been the most commonly used method for quantitative DTI analysis despite being labor-intensive and operator-dependent. In this study, we developed a fully automated voxel-wise serial DTI analysis (BLADE) for robust and sensitive detection of progressive microstructural white matter changes in TBI.

5.3 Methods

5.3.1 Subjects / Acquisitions

Six patients with mild TBI (Glasgow Coma Scale 13-15 at presentation in the Emergency Dept.) and seven patients who suffered leg injuries but not head injuries were enrolled in the study. On conventional imaging studies, two TBI patients were suspected of cortical contusion, and the remaining four were suspected of having DAI. Patients with leg injuries were used to serve as control

subjects in testing the BLADE technique and in determining the cluster thresholding (see next section for details).

MR exams including conventional MRI and DTI were performed at three separate time points on a 3T GE Signa EXCITE scanner with an 8-channel phased array head coil, using ASSET parallel imaging with an acceleration factor of two. The three time points were within one week, after one month and after one year of the brain injury (simply called 1 week, 1 month and 1 year later in this chapter). Whole-brain DTI was performed with a single-shot multi-slice axial spin-echo EPI (TR / TE = 14 s / 64 ms, NEX=1) at 1.8-mm isotropic voxel resolution using 55 diffusion-encoding directions (distributed evenly by electrostatic repulsion scheme [2]) at $b=1000 \text{ s/mm}^2$, as well as one $b=10 \text{ s/mm}^2$ acquisition.

5.3.2 Analysis

For the detection of localized FA changes over time, we developed an analysis technique called Bootstrap-based Longitudinal Analysis of DTI Estimates (BLADE), briefly introduced elsewhere [3]. This technique can be summarized as creating an FA subtraction map scaled by standard errors such that thresholding and clustering can be used to select groups of voxels with statistically significant FA changes over time. The residual bootstrap [4] was used because it only requires a single DTI dataset to estimate the standard errors of FA. Utilizing data redundancy from the 55 distinct diffusion measurements at each voxel, we can perform voxel-wise statistical testing in each subject separately, without multiple

DTI acquisitions, pooling data from multiple subjects, or making assumptions about the underlying noise properties.

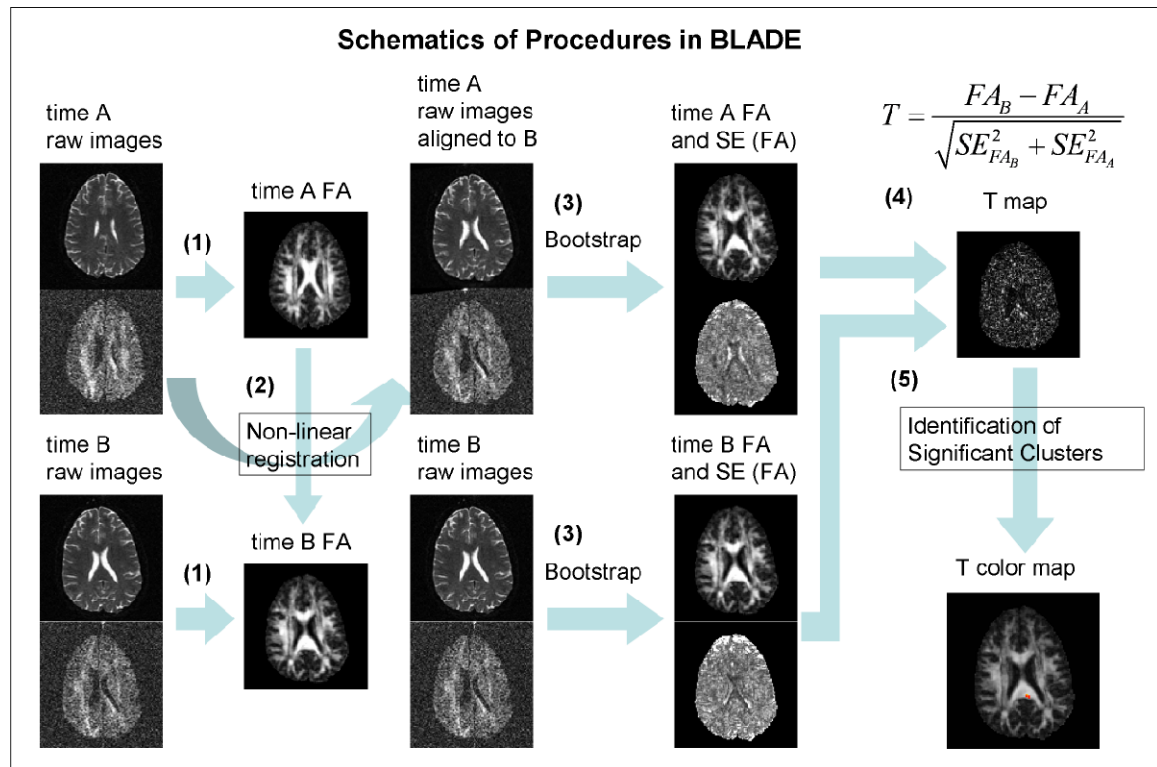


Figure 5-1. Schematics of procedures in BLADE.

The framework of BLADE (also illustrated as a graphical flow chart in Figure 5-1) consists of four steps. First, the raw DTI images at each time point were corrected for the subject motion during the scan by rigid-body transformation (FLIRT, <http://www.fmrib.ox.ac.uk/fsl/flirt/>) [5]. Then, FA maps were generated by diffusion tensor analysis (described in chapters 2 and 3), including fitting of diffusion tensor by weighted least squares, diagonalization of the tensor to yield eigenvectors / eigenvalues, and calculating FA from eigenvalues.

Second, these FA maps were used for the registration of DTI datasets between different time points. FA maps are preferred over other images since FA maps are better suited for registration of brain white matter due to the rich contrast within white matter (this rich contrast in fact makes visual inspection of FA maps tricky). An initial registration between FA maps by affine transformation (FLIRT) was followed by non-linear registration modeled as 3D 5th order polynomials (AIR 5.2.5, <http://bishopw.loni.ucla.edu/AIR5/>) [6, 7] in order to correct for any time-point dependent non-linear distortion of the images as well as any possible atrophy in specific white matter bundles. The transformation parameters were applied to all the raw DTI images such that data from all time points were registered to a common space.

Third, residual bootstrap with 200 iterations was used to create the standard error map of FA as described in [4]. The number of bootstrap iterations of 200 was chosen since this is a good compromise between reliable estimation of uncertainties and reducing computation time, according to my preliminary Monte Carlo simulation studies (results not shown) as well as a study by another group [8].

Fourth, FA maps and their standard error maps were combined for each pair of time points to calculate a T statistic map, using the equation below.

$$T = \frac{FA_B - FA_A}{\sqrt{SE(FA_B)^2 + SE(FA_A)^2}} \quad \text{Equation 5-1}$$

In reality, the distribution of calculated T statistics when there is no difference (i.e. null distribution) does not converge to the T distribution for a number of reasons (see the discussion section for details), and thus strictly speaking, *pseudo* T statistic might be a better terminology.

Finally, clusters of voxels with relatively large differences between time points were defined. All voxels in each cluster were $\text{abs}(T) > 2.6$ (roughly corresponding to $p < 0.01$), and only voxels connected by edges or faces in 3D were qualified as a cluster. It is difficult to assign a p-value to each cluster within our framework, so the distribution of clusters in the healthy subjects served as a guide to empirically choose statistically significant clusters. The clusters marked as statistically significant were overlaid in color on the average FA map. By using clusters of voxels, rather than the voxels themselves, the power to detect changes in the regions larger than the threshold of cluster size increases at the expense of decreased power of detecting changes in the smaller regions.

5.4 Results

5.4.1 Control Subjects

The MRI findings were normal in the seven control subjects. In these subjects, the size of largest cluster found in the whole brain white matter region after thresholding the T statistic map is shown in Figure 5-2. Clusters larger than 25

voxels were rare. Thus, for the analysis of TBI patients, the cluster size threshold of 30 voxels was empirically chosen as a criterion for defining significant clusters.

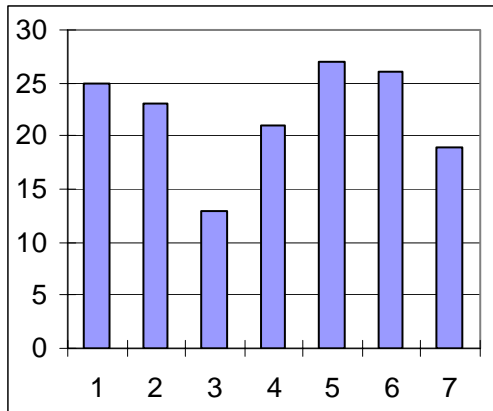


Figure 5-2. The plot showing the largest cluster with $\text{abs}(T) > 2.6$ in the whole brain of each healthy subject. X axis: subjects. Y axis: number of voxels in the largest cluster.

FA maps and T statistic maps in the corresponding slices in one of these control subjects are shown in Figure 5-3. The T statistic maps generally show random patterns with few noticeable structures. This suggests that the original FA differences are properly scaled by the bootstrap estimated standard errors to make the final T statistic values random and spatially independent. Note that voxels not included in the BLADE analysis (voxels outside white matter) are masked out in the images shown in Figure 5-3.

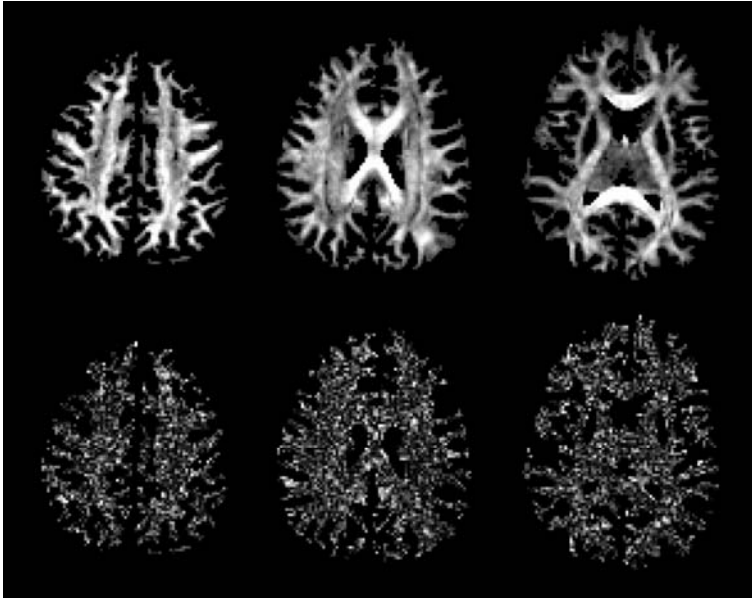


Figure 5-3. Top row: FA maps in three different slices. Bottom row: T statistics maps in the same slices as the top row.

5.4.2 Patients with Cortical Contusion

In both of the two TBI patients with cortical contusions, BLADE showed increased FA values in the interval between the 1-week and 1-month time points in the white matter region showing FLAIR (fluid attenuated inversion recovery) abnormality due to the contusions.

Results for one of the subjects are shown in Figure 5-4. While FA is increased at the site of contusion coinciding with the subsided hyperintensity in FLAIR images after 1 month (upper row, arrow), clusters of FA decrease is found in normal appearing voxels not contiguous to the FLAIR contusion region but right above the contusion site (lower row, arrowhead).

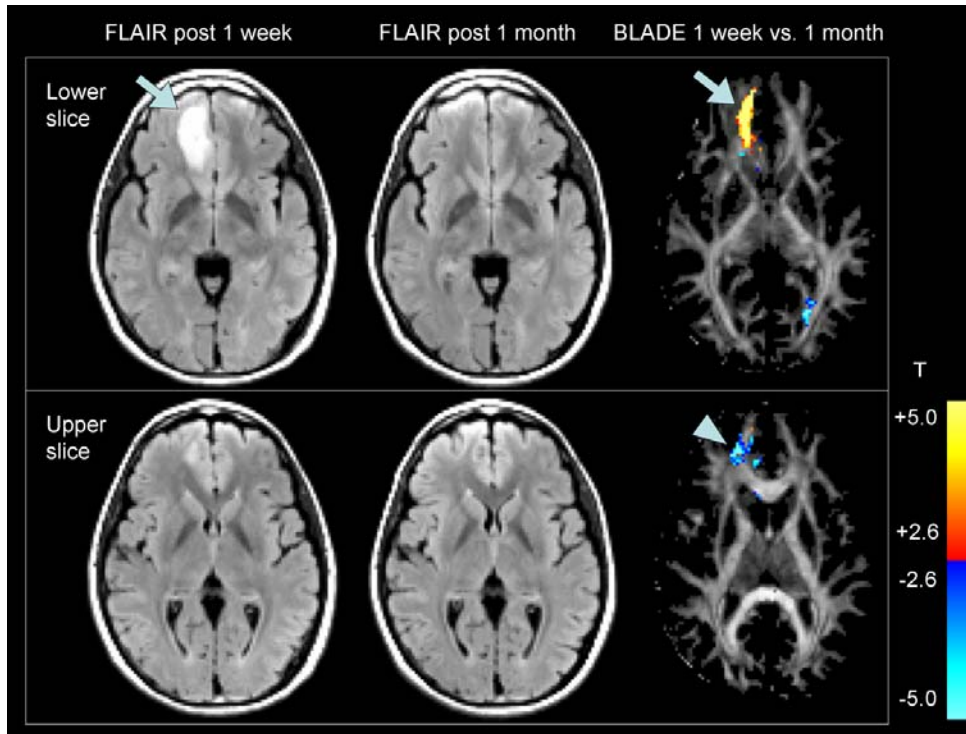


Figure 5-4. BLADE analysis between 1 week and 1 month post injury in a focal cortical contusion patient. FLAIR images in each time point are also presented. Warm colors indicate the positive T statistic values, i.e. larger FA in 1 month compared to 1 week post injury. Smaller FA in 1 month are displayed as cool colors. Images in each row are from the same anatomic location.

5.4.3 Patients with Diffuse Axonal Injury

Of the four mild TBI patients suspected of having DAI on conventional MRI, the patient with the most severe DAI (Figure 5-5 and Figure 5-6) showed increased FA values within the lesions (arrow) and the decreased FA in non-contiguous adjacent normal-appearing white matter (NAWM) (arrowheads). The largest component of this progressive FA decrease in NAWM occurred between the 1-week and 1-month time points, but further decreases were seen at 1-year following trauma.

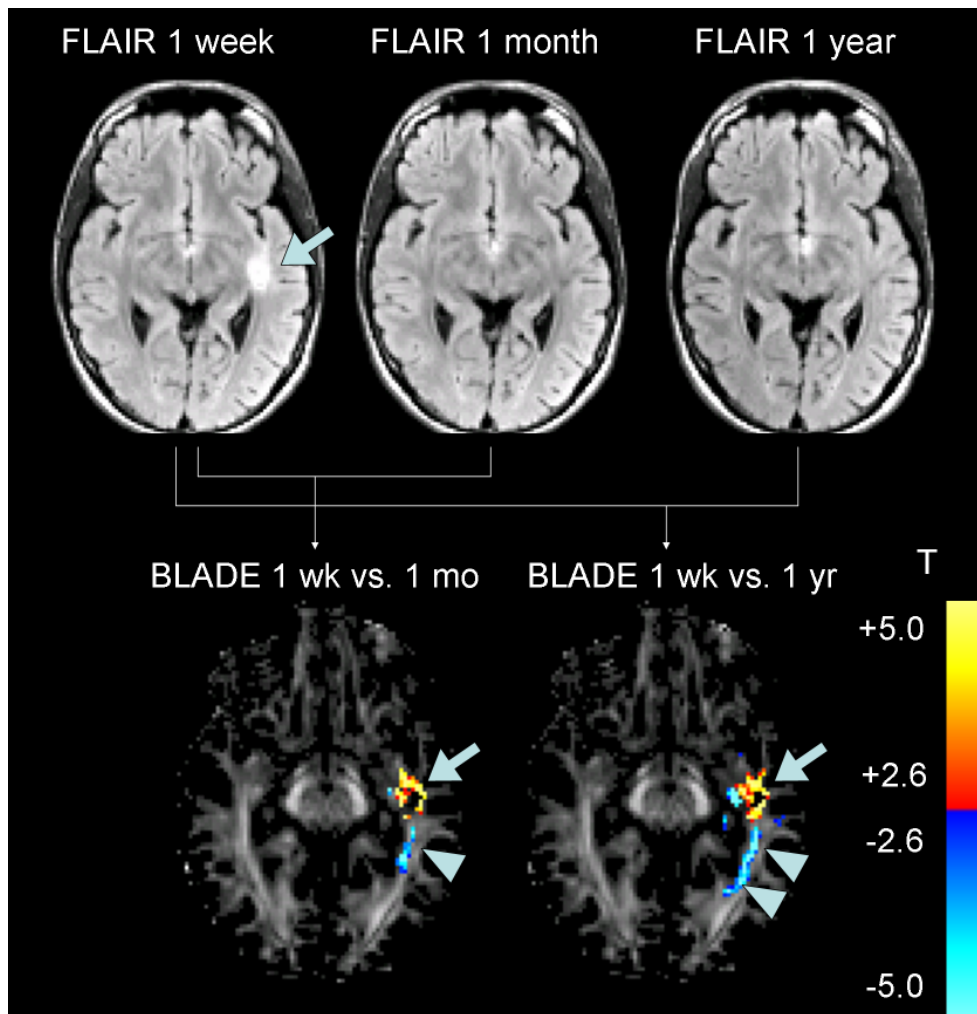


Figure 5-5. BLADE analysis between 1 week and 1 month, and 1 week and 1 year in a diffuse axonal injury patient.

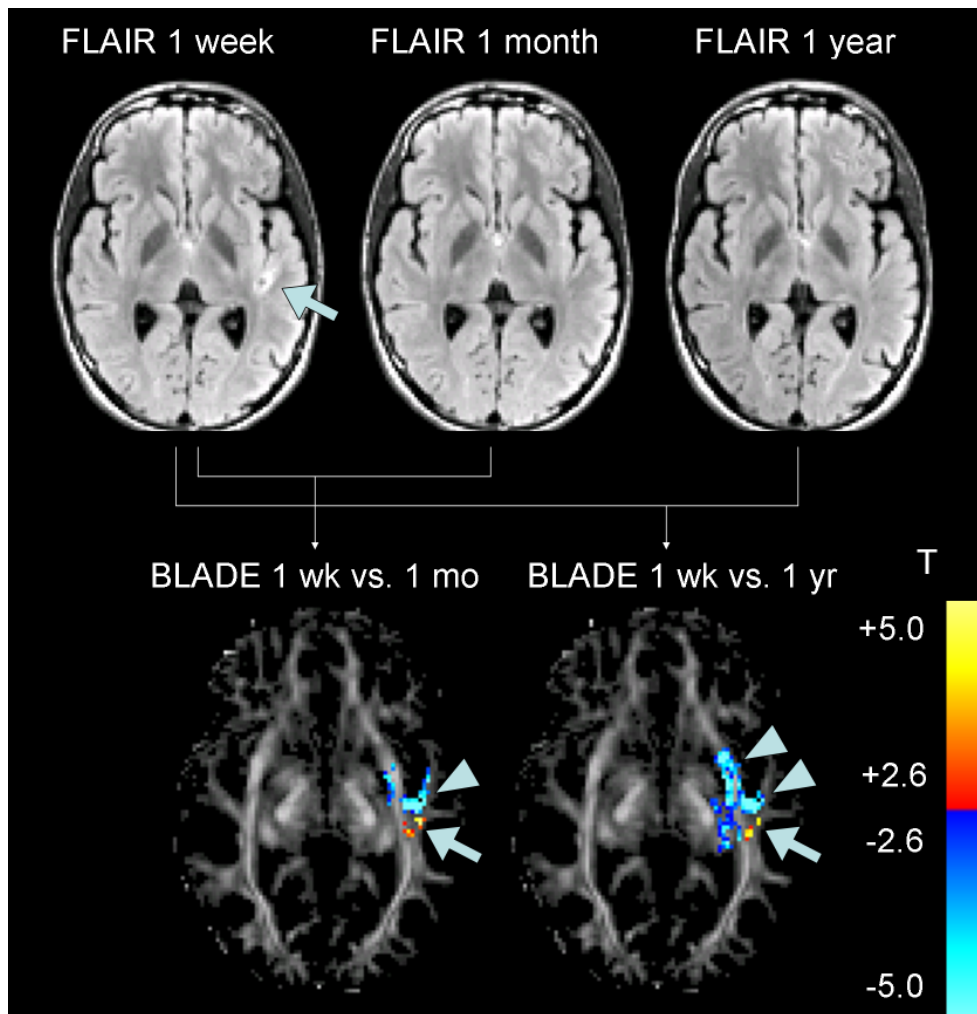


Figure 5-6. BLADE analysis in the different slice of the same patient shown in Figure 5-5.

The other three DAI patients did not show particular findings by the BLADE analyses, except for FA increases in the lesion visible by either FLAIR or T_2^* weighted images. The result for one of these patients is shown in Figure 5-7. The arrow indicates the lesion visible as a hyperintense region in the FLAIR image and a cluster of increased FA determined by the BLADE analysis.

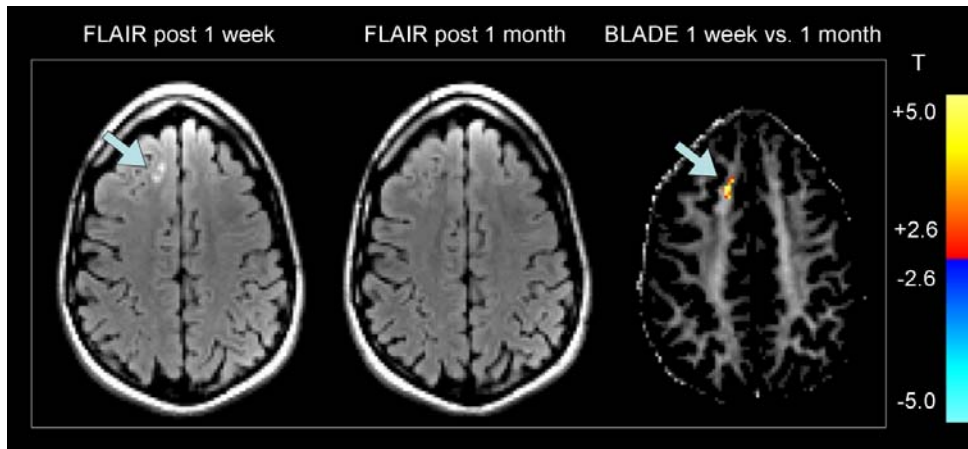


Figure 5-7. BLADE analysis between 1 week and 1 month in a diffuse axonal injury patient (different from the patient shown in Figure 5-5 and Figure 5-6).

5.5 Discussion

This is the first study demonstrating automated, subject-specific, voxel-wise serial analysis of whole-brain DTI, localizing progressive microstructural changes. While this study involves a relatively small number of subjects, the results are nevertheless as strong as studies with large number of subjects, due to the nature of BLADE to perform an independent statistical inference for each subject. The main observations made from the BLADE analysis is that FA values increased over time in the lesions visible by conventional MRI and decreased in the NAWM near the lesions. It is likely that the increase of FA in the lesions represent resolution of vasogenic edema and other local inflammatory changes. Increased water content in the lesions not only makes them visible to conventional MRI but they are known to decrease the FA values. Thus, it is not surprising that the increase in the FA values in the lesions over time is accompanied by less pronounced hyperintensity in the FLAIR images. The

progressive decrease of FA in NAWM adjacent to the lesions is particularly interesting since conventional MRI fails detect any abnormalities, and may represent Wallerian degeneration along white matter pathways.

It is extremely useful to detect changes but it is important to note that ambiguities can exist in the biological interpretation of these changes. FA values are generally known to be reduced in a wide variety of pathologies, but there is the possibility of paradoxical increase of FA when only selective populations of fibers degenerate in the regions of crossing fibers (although the frequently cited study by Pierpaoli et al. [9] actually does not report paradoxical increase in FA values). Normal (i.e. pathology-free) FA depends strongly on the subject and brain region, thus what is normal is not known unless measured in the same subject and same region prior to development of any pathology.

Attempts have been made to adapt the FA values in the anatomically equivalent region on the contralateral side of the same subject as normal (often quoted as 'internal reference') [10]. This approach works only with unilateral pathology and it is limited by asymmetry of anatomy and asymmetry of the FA values themselves that would make the estimate of normal FA values less certain. Alternatively, distributions of FA values from healthy subjects has been used to determine the deviation of FA values in individual subjects voxel-wise [11]. This approach is probably even less optimal in defining normal FA values, since inter-subject variability is likely to be larger than the variability between right and left

sides in the same subject. As a result, currently no effective method exist to determine the normal FA values except for measuring FA values before the injuries, and interpretation about FA changes should be made cautiously.

On a technical side, the statistics in BLADE is somewhat limited, leaving room for improvement. As commented in section 5.3.2, the T statistics calculated by BLADE are actually pseudo T statistics. Typically the T statistic is calculated as below.

$$T_{df} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{SE(\bar{X}_1)^2 + SE(\bar{X}_2)^2}} \quad \text{Equation 5-2}$$

\bar{X}_1 and \bar{X}_2 are n-sized sample means of i.i.d. random variables X_1 and X_2 , and standard errors (SE) of these sample means in the denominator are estimated by $SE(\bar{X}_1) = s_1 / \sqrt{n}$ and $SE(\bar{X}_2) = s_2 / \sqrt{n}$, where s_1 and s_2 are standard deviations of the same samples from X_1 and X_2 . The whole denominator is actually the SE of the random variable $\bar{X}_1 - \bar{X}_2$ and conceptually the T statistic above can be described as the difference in sample means divided by the SE of the difference.

$$T_{df} = \frac{\Delta \bar{X}}{SE(\Delta \bar{X})} \quad \text{Equation 5-3}$$

Unlike the Z statistic with similar expression $Z = (X - \mu) / \sigma$ where σ is known, the SE in Equation 5-3 has to be estimated from the samples as shown in Equation 5-2. Thus as the sample size increases the estimated SE becomes more accurate and the T statistic distribution approaches the standard normal

distribution. The degrees of freedom (df) dictates the shape of the T distribution at certain sample sizes, and df is crucial in calculating p-values.

The difference between testing ΔFA and testing $\Delta \bar{X}$ is that 1) ΔFA is not normally distributed while $\Delta \bar{X}$ is normally distributed and 2) $SE(\Delta FA)$ is estimated by bootstrap while $SE(\Delta \bar{X})$ is estimated by equations described above, and the df is not known for $SE(\Delta FA)$. Thus, even under the null hypothesis that there is no difference in FA, the null distribution of $T = \Delta FA / SE(\Delta FA)$ is unknown.

The problem that the null distribution of our pseudo T statistic is unknown can be circumvented if we use the pseudo T statistic just to define clusters of relatively strong changes rather than calculating voxel-wise p-values directly, a strategy we have used for BLADE. The greater challenge remains, though, regarding the multiple comparison problem. The null distribution of cluster size is currently unknown within the BLADE framework, and p-values cannot be assigned to each cluster based on the data from control subjects. These difficulties prompted me to investigate the feasibility of permutation testing in serial DTI analysis as introduced in the next chapter. Even with these limitations, BLADE is a significant step forward in that it is the first technique ever to analyze DTI data from a single subject quantitatively, opening new possibilities in research and clinical applications of DTI.

5.6 Conclusion

Bootstrap-based voxel-wise statistical testing was able to detect local FA changes over time in TBI patients, including within normal-appearing white matter on conventional 3T MRI, suggesting a role for DTI in the analysis of traumatic white matter injury.

5.7 Reference

1. Le, T.H. and A.D. Gean, *Imaging of head trauma*. Semin Roentgenol, 2006. **41**(3): p. 177-89.
2. Jones, D.K., M.A. Horsfield, and A. Simmons, *Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging*. Magn Reson Med, 1999. **42**(3): p. 515-25.
3. Chung, S., et al. *Bootstrap-based longitudinal analysis of DTI estimates (BLADE): applications in clinically isolated syndrome patients*. in *14th Annual Meeting of ISMRM*. 2006. Seattle, WA: ISMRM.
4. Chung, S., Y. Lu, and R.G. Henry, *Comparison of bootstrap approaches for estimation of uncertainties of DTI parameters*. Neuroimage, 2006. **33**(2): p. 531-41.
5. Jenkinson, M. and S. Smith, *A global optimisation method for robust affine registration of brain images*. Med Image Anal, 2001. **5**(2): p. 143-56.
6. Woods, R.P., et al., *Automated image registration: I. General methods and intrasubject, intramodality validation*. J Comput Assist Tomogr, 1998. **22**(1): p. 139-52.
7. Woods, R.P., et al., *Automated image registration: II. Intersubject validation of linear and nonlinear models*. J Comput Assist Tomogr, 1998. **22**(1): p. 153-65.
8. O'Gorman, R.L. and D.K. Jones, *Just how much data need to be collected for reliable bootstrap DT-MRI?* Magn Reson Med, 2006. **56**(4): p. 884-90.
9. Pierpaoli, C., et al., *Water diffusion changes in Wallerian degeneration and their dependence on white matter architecture*. Neuroimage, 2001. **13**(6 Pt 1): p. 1174-85.
10. Glenn, O.A., et al., *Diffusion Tensor MR Imaging Tractography of the Pyramidal Tracts Correlates with Clinical Motor Function in Children with Congenital Hemiparesis*. AJNR Am J Neuroradiol, 2007.
11. Le, T.H., et al. *Automated Detection of Traumatic White Matter Injury Using Voxel-Based Morphometry of Diffusion Tensor Images: A 3T Study with Parallel Imaging*. in *13th Annual Meeting of ISMRM*. 2005. Miami, FL: ISMRM.

Chapter 6: Permutation Voxel-wise Serial DTI Analysis

6.1 Abstract

Diffusion tensor MRI (DTI) has been widely used to investigate brain microstructural changes in pathological conditions as well as for normal development and aging. In particular, longitudinal changes are vital to the understanding of progression but these studies are typically designed for specific regions of interest. To analyze changes in these regions traditional statistical methods are often employed to elucidate group differences which are measured against the variability found in a control cohort. However, in some cases, rather than collecting multiple subjects into two groups, it is necessary and more informative to analyze the data for individual subjects. There is also a need for understanding changes in a single subject without prior information regarding the spatial distribution of the pathology, but no formal statistical framework exists for these voxel-wise analyses of DTI. In this study, we present PERVADE (PERmutation Voxel-wise Analysis of Diffusion Estimates), a whole brain analysis method for detecting localized FA changes between two separate points in time of any given subject, without any prior hypothesis about where changes might occur. Exploiting the nature of DTI that it is calculated from multiple diffusion-weighted images of each region, permutation testing, a non-parametric hypothesis testing technique, was modified for the analysis of serial DTI data and implemented for voxel-wise hypothesis tests of diffusion metric changes, as well

as for suprathreshold cluster analysis to correct for multiple comparisons. We describe PERVADE in detail and present results from Monte Carlo simulation supporting the validity of the technique as well as illustrative examples from a healthy subject and patients in the early stages of multiple sclerosis.

6.2 Introduction

In recent years, diffusion tensor magnetic resonance imaging (DTI) [1, 2] has been widely used in studying normal development and aging, and diverse pathological conditions of the human brain due to its unique ability to identify microstructural abnormalities. Since DTI is a relatively new technique and is fundamentally different from other imaging techniques in that each voxel contains not a single value but a 3x3 matrix with 6 unique elements called the diffusion tensor, the statistical analysis of DTI is still under development. Multiple approaches exist for defining the target regions for statistical comparisons, including manually traced region of interest (ROI), fiber tracking defined ROI, whole brain histogram, and whole brain voxel-wise analysis by statistical parametric mapping SPM [3] or by tract-based spatial statistics [4]. There are also diverse DTI-derived parameters to compare, such as mean diffusivity (D_{av}), parallel / transverse diffusivity, fractional anisotropy (FA) and the primary eigenvector [5].

Most DTI studies with statistical testing performed in one or more of the aforementioned ways have one thing in common; they are all multiple-subject

group comparison studies, usually one group of experimental subjects compared with a matched group of healthy control subjects. Group comparison studies are possible when the effects of interest are located in stereotypical anatomic structures, such as the same white matter tracts, across subjects. In certain conditions where the effects are expected to be focal (or multi-focal) with spatial distributions that are highly specific for individual subjects, finding and grouping subjects that share effects in similar anatomic locations may be difficult. Furthermore, group analyses demand that significant effects be larger than group variability and therefore may suffer from decreased sensitivity. A statistical analysis that can be performed in individual subjects is thus required.

Due to the non-Gaussianity, DTI parameters such as anisotropy indices [6, 7] makes parametric testing less optimal. Non-parametric resampling techniques have great potential in the statistical testing of these DTI data since the recent trend of oversampling DTI data makes it suitable for resampling. One of these techniques called bootstrap has been shown to be very useful in DTI [6, 8-12]. These works have shown the ability of bootstrap to estimate the uncertainties of DTI parameters (descriptive statistics) but have not addressed the issue of statistical significance testing (inferential statistics). For statistical testing when comparing two groups, permutation testing [13-15], another non-parametric resampling technique, can be useful as well.

Permutation testing provides statistical significance testing of differences between groups, with the unique ability of directly estimating the null distribution of the statistic describing the difference, rather than assuming a null distribution of analytically known form (such as T-distribution). Multiple groups are required for resampling and permutation testing provides exact (or almost exact) p-values but does not estimate uncertainties of sample statistics (such as the standard error of mean). Bootstrap is mainly used to estimate the accuracy of sample statistics by resampling from one group (multiple groups are not required). Though it can be used not only in descriptive statistics (i.e. standard errors) but also in inferential statistics (i.e. significance testing), bootstrap testing is not able to estimate a data-driven null distribution and bootstrap-estimated p-value is only approximate.

We focus on permutation testing in this work. Permutation testing is well-established in the field of neuroimaging, especially functional MRI where numerous works advocate the strength of this approach in the last decade (for a review refer to [16]). Permutation based method 1) makes minimal assumptions and thus can be applied even in situations where the assumptions of parametric approaches are not met or cannot be verified, 2) is conceptually simple and provides intuitive solutions to the multiple comparison problem, and 3) is easily applicable to any test statistics allowing the researchers to freely choose the statistic best suited for their studies. Unlike bootstrap, permutation testing can elegantly incorporate the whole statistical procedure of voxel-wise comparison

and multiple-comparison correction in a completely non-parametric way due to the ability to estimate data-specific null distribution as described above, possibly with more accurate p-values.

In this study, we describe how permutation testing can be properly implemented in voxel-wise analysis of single-subject serial DTI studies. We will pay special attention to not violating exchangeability assumption for permutation testing that if groups are not different then any re-grouping of the samples (permutation) are equally likely as the original grouping (observation). Then, we present a novel statistical analysis framework called PERVADE (PERmutation Voxel-wise Analysis of Diffusion Estimates) that is designed to localize subtle and local microstructural changes over time in the whole brain of a single subject without any prior hypothesis. PERVADE includes 1) non-linear registration between two time-points to account for any morphological changes over time, 2) voxel-wise calculation of p-values by permutation testing, and 3) suprathreshold cluster analysis with permutation testing to deal with the multiple comparison problem. Our preliminary results of local microstructural changes detected outside as well as inside the focal lesions of patients in the earliest stage of multiple sclerosis show the potential of this technique to provide additional information about microstructural white matter injury.

6.3 Methods

6.3.1 DTI Permutation Testing

6.3.1.1 Permutation Testing in Group Comparison

Before presenting the proposed voxel-wise permutation testing in a single-subject serial DTI study, we first examine how it would be done in a multiple-subject group comparison study (i.e. comparison of group-averaged metric between two cohorts), a simple and typical scenario of permutation testing. To test whether the observed difference in group-averaged FA, i.e. $\hat{\theta} = \overline{FA_B} - \overline{FA_A}$, is statistically significant, permutation testing can be performed as follows. All subjects are randomly assigned to either group A or B while maintaining the original number of subjects (thus re-defining groups A, B to A^* , B^*), and one permutation replication of $\hat{\theta}$ defined as $\hat{\theta}_i^* = \overline{FA_{B^*}} - \overline{FA_{A^*}}$ is calculated. If the null hypothesis H_0 that the FA values of groups A and B are not different holds, then this re-grouping (or re-labeling to either group in each subject) would be equally likely to happen as the original grouping, meaning that any possible $\hat{\theta}_i^*$ are equally likely as $\hat{\theta}$. Thus, the random variable $\hat{\theta}^*$ has the null distribution of $\hat{\theta}$ if H_0 is true, and it is called permutation distribution.

The statistical significance (p-value) in the permutation testing is defined as,

$$p = \text{Prob}_{\text{perm}} \{ |\hat{\theta}^*| \geq |\hat{\theta}| \} = \# \{ |\hat{\theta}^*| \geq |\hat{\theta}| \} / N \quad \text{Equation 6-1}$$

where N is the total number of possible labelings. This simply quantifies the probability of observing group-averaged FA difference more extreme than $\hat{\theta}$ by random chance. If this probability is less than 5 percent ($p \leq 0.05$) then $\hat{\theta}$ is unlikely under H_0 , and we reject H_0 concluding that two groups are different in terms of FA. If N is too large to compute all possible labelings, the p-value can be approximated by using a randomly chosen subset of all possible labelings such that

$$p \approx \# \{ |\hat{\theta}_i^*| \geq |\hat{\theta}| \} / N' \quad (i=1,2,\dots,N') \quad \text{Equation 6-2}$$

where $N' < N$. Typically $N'=1000$ and $\hat{\theta}_1^* = \hat{\theta}$ (i.e. the first labeling of $\hat{\theta}^*$ is just the observed statistic $\hat{\theta}$).

6.3.1.2 Permutation Testing in Intra-subject Serial Comparison

On the other hand, if we want to test whether FA from two time points of a single subject is different, we can no longer permute the observed FA values since there are only two such values available, one from each time point. Instead, since the FA at one time point is calculated from the diffusion weighted images (DWIs, this refers to the raw images acquired in a single diffusion encoding direction for DTI in this study) at that time point (just like \overline{FA} is calculated using the FA values from the individual subjects of the group), the DWIs can be permuted. Assuming that same DTI acquisition protocol was used for two time

points, which is a typical scenario since it is always strongly encouraged to use the same protocol in the same scanner for follow-up MRI studies, the simplest approach satisfying exchangeability is to permute only the DWIs that have the same diffusion encoding (diffusion gradient direction and strength). This stratified permutation scheme shares similarity with permutation of scans only within an exchangeability block in functional neuroimaging [16], and with stratified (or repetition) DTI bootstrap [6, 8]. While multiple repeated acquisitions are mandated for stratified DTI bootstrap, scans for stratified DTI permutation do not need to be repeated at each time point since pooling two time points naturally guarantees at least two repetitions.

A caveat here, though, is that even with the same diffusion encodings, the DWIs may still not be theoretically exchangeable due to differences in the gain factors in the raw MR signal intensities and possible difference in head positioning (effectively changing the diffusion encoding directions relative to the brain tissue). Any drift in the gain factors of the MR system will result in changes of the DWI signal intensities. However, this difference can be easily remedied and exchangeability secured by simply estimating the ratio of the signal intensities from the images where we can assume that there are no biological changes, such as normal-appearing white matter (NAWM) in the non-diffusion-weighted images, simply called $b=0$ images. For the head positioning issue, if the relative angular locations of the brain to the MRI system are different between two time points, the brain will effectively be under different diffusion encodings even if the

diffusion gradients are applied exactly the same way. Thus, we propose to permute the diffusion encoding gradients (rotated in the same degree as any registration performed in order to correct for the different head positioning) in conjunction with the DWIs themselves. While DWIs themselves might not be exchangeable in a strict sense, the complete set of permuted DWIs paired with properly adjusted diffusion encoding directions could be exchangeable.

The only assumption required to justify this approach is that slightly different configurations of diffusion encoding directions will not affect the probability distributions of DTI metrics. Our experience tells us that in typical serial MRI studies, the rotational components by different head angulations are usually a few degrees (axis of rotation either x, y, or z). After the permutations above, each of diffusion encoding direction will be 'jittered' around a few degrees, but the diffusion gradient scheme will be similar overall. Notice that even if we are permuting the pairs of DWIs and diffusion directions that are slightly jittered, we still opt for the stratified permutation since drastic variations of the set of diffusion gradients after permutation can distort the permutation distribution due to gradient configurations effects on the DTI parameter estimations [17]. Limiting the permutation within the exchangeability block defined as the same physical diffusion encodings guarantees similar diffusion gradient sets. The permutation testing to calculate voxel-wise p-value is illustrated in Figure 6-1 steps (a) to (e) and Figure 6-2.

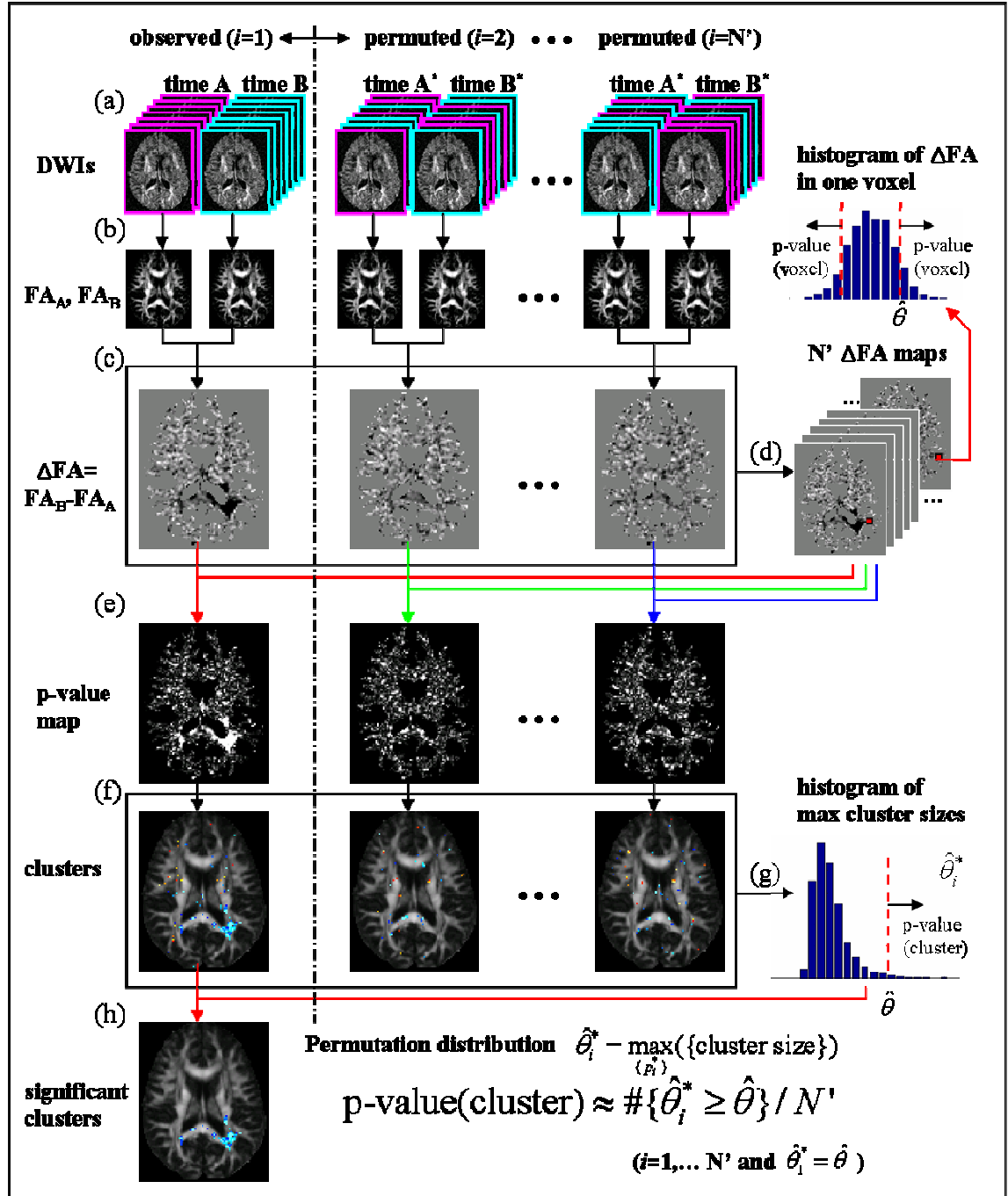


Figure 6-1: Overall schematics of permutation testing to calculate voxel-wise p-values and multiple-comparison-corrected cluster-wise p-values in a serial DTI study of a single subject. This corresponds to steps 4 and 5 in the PERVADE procedures (see human brain data analysis section on the methods). **(a)** DWIs (including $b=0$) are either kept as observed ($i=1$) or permuted ($i=2, \dots, N'$). For the simplicity of illustration, it is assumed that only seven images are acquired in each time point (A or B). Colored boundaries indicate the observed time point of each DWI. Permutation is done separately in each gradient direction (see figure 2). **(b)** Two FA maps are

calculated from (original or reassigned) DWIs by diffusion tensor processing. **(c)** Two FA maps are subtracted to create Δ FA map. **(d)** All Δ FA maps (total N') represent the map of voxel-wise permutation distribution of Δ FA. **(e)** Voxel-wise p-value maps are estimated by comparing Δ FA maps with the permutation distribution from (d). **(f)** Clusters are defined by thresholding p-value maps. **(g)** Sizes of the largest clusters in each cluster mask represent the permutation distribution of image-wise maximal cluster size. **(h)** All the clusters in the observed cluster mask are compared to the permutation distribution from (g) to estimate cluster-wise p-values, and only clusters $p < 0.05$ are declared significant.

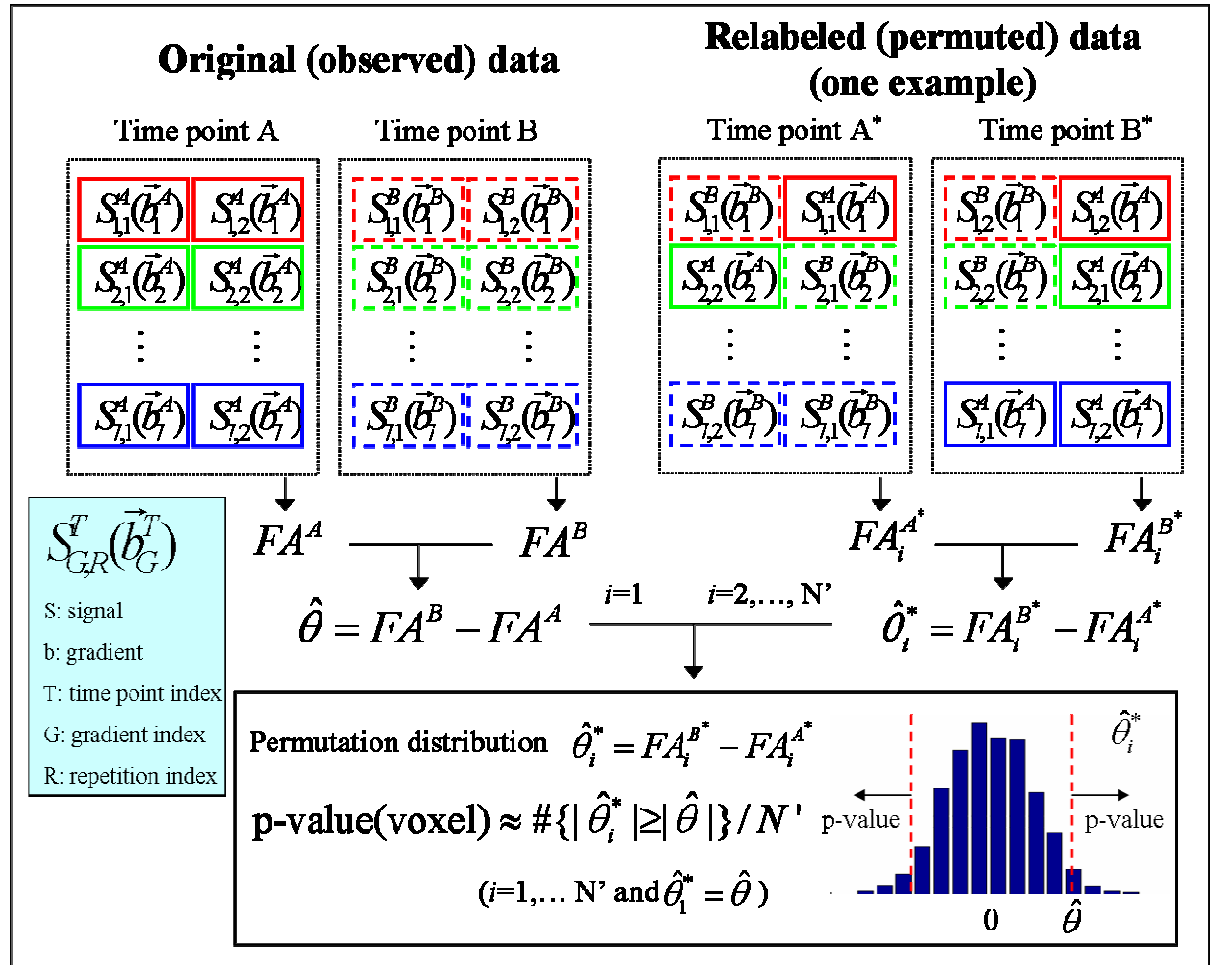


Figure 6-2: Detailed schematics of permutation testing to estimate voxel-wise p-values in a serial DTI study of a single subject. It is assumed that in each time point the DTI scan of seven DWIs (first DWI can be regarded as $b=0$) are repeated twice. In each DTI dataset (inside the black dotted box), diffusion-weighted signals $S(\vec{b})$ in the same row are acquired by applying the same diffusion gradients. \vec{b}^A or \vec{b}^B stresses that the effective gradients can be different in two time

points due to the head positioning. Boxes around the signals are colored based on the diffusion gradients in the observed data while solid and dashed linestyles indicate the original time point A and B. On the right side is one possible permutation of DTI dataset. Permutation is done separately for each gradient (thus signals in the same gradient forms the exchangeability block), easily identified by the colors. For each $N'-1$ permuted and one observed DTI data, ΔFA from two time points are calculated and voxel-wise p-value of observed FA difference are estimated as the probability of more (or equal) extreme difference than the observed $\hat{\theta}$ from the null distribution of N' differences, i.e. proportion of counts outside the red dashed lines to all N' counts in the histogram of right lower corner.

6.3.1.3 Permutation Testing for Multiple Comparisons

The permutation scheme can be used not only for voxel-wise p-values but also to deal with the multiple comparison problem arising from testing multiple voxels simultaneously. Specifically, we chose the suprathreshold cluster analysis (STCA) method that is known to provide a good balance between the statistical power and the ability to localize the changes [16, 18-21]. In STCA scheme, the statistic map is initially thresholded at some arbitrary single voxel level to define clusters (spatially contiguous voxels with contiguous defined as sharing a side). The multiple comparison corrected p-values for clusters are then calculated based on the distribution of maximum cluster sizes identified in each permutation originally done for the voxel-wise analysis; i.e., the same set of permutations are used for all voxels in the analysis domain. Equation 2 is valid here as well except that taking the absolute values of $\hat{\theta}_i^*$ and $\hat{\theta}$ is not necessary since cluster sizes are always positive (thus it is one-tailed testing, not two-tailed) and the equation is shown below,

$$p \approx \#\{\hat{\theta}_i^* \geq \hat{\theta}\} / N' \quad (i=1,2,\dots,N')$$
Equation 6-3

where $\hat{\theta} = \max_{\{p\}}(\{\text{cluster size}\})$ and $\hat{\theta}_i^* = \max_{\{p_i^*\}}(\{\text{cluster size}\})$. $\{p\}$ and $\{p_i^*\}$ denotes the observed p-value map and p-value map from i^{th} permutation each, in the domain of whole white matter.

The implemented permutation-based STCA is similar to that described previously [22-24], except that we use the p-value maps created from voxel-wise permutation tests to define clusters instead of some other statistic maps (like t-statistic). During the voxel-wise permutation testing, the whole image of DWIs are permuted together instead of permuting individual voxels independently. While not affecting the voxel-wise permutation testing in any way, this helps to create permutation replicas of the statistic maps that are equally likely in terms of spatial correlation structures as well as the distribution of p-values. Each cluster in the observed p-value map are assigned the cluster-wise p-value by comparing the cluster size with the maximal cluster permutation distribution $\hat{\theta}^*$, i.e. the collection of image-wise maximal cluster size from all the permutation replicas. The cluster-wise p-values estimated as such are by definition multiple-comparison corrected since the permutation distribution describes the probability of having any cluster(s) above certain size in the image by random chance. It has been shown that a strong control of family-wise error is achieved in this way, meaning that we can localize significantly changing regions by rejecting the null hypothesis for the clusters with adjusted p-values smaller than 0.05, rather than

simply knowing that changes are occurring somewhere in the brain [23]. The permutation testing to calculate multiple-comparison-corrected cluster-wise p-values is illustrated in Figure 6-1 steps (f) to (h).

Identification of significant clusters was performed with a multi-step iterative algorithm called the step-down test [24]. After rejecting the null hypothesis for clusters exceeding the 95th percentile of the permutation distribution

$\hat{\theta}^* = \max_{\{p^*\}}(\{\text{cluster size}\})$, a new white matter domain that does not include these

clusters was defined. $\hat{\theta}^*$ was then re-estimated from this new domain using the same p-value maps, and any previously undetected clusters were searched.

These steps were repeated until no more clusters were rejected. Then, $\hat{\theta}^*$ in the final step was considered the most appropriate null distribution least affected by any real changes, and all the clusters in the whole white matter domain were tested against this distribution. Actually, this approach of re-estimating $\hat{\theta}^*$ after excluding all rejected tests simultaneously rather than only the most significant test is called the jump-down variant of the step-down approach. Jump-down testing can substantially decrease the computation time due to fewer steps.

6.3.2 Non-linear Registration

As is the case for most voxel-wise analysis techniques, PERVADE depends on the accuracy of registration techniques. It is designed for detecting changes over time in the same subject, and it is easier to register data from the same subject

than across subjects since there is much less intra-individual than inter-individual biological variability. This makes PERVADE less vulnerable to mis-registration than group comparison methods. Still, any morphological changes over time such as atrophy should be accounted for, and, if not, they may be detected as significant changes. While acknowledging that detecting atrophy of specific white matter structures based on the unique ability of DTI to create rich contrast within white matter is also a subject of interest, the focus of PERVADE for now is to detect ‘microstructural’ changes over time, parametrized by DTI scalar indices such as FA, rather than ‘macrostructural’ volumetric changes associated with atrophy. Thus we attempted to eliminate volume changes by non-linear registration but remain sensitive to microstructural related intensity changes due to the atrophy process.

A high-order non-linear registration algorithm [25] was used for registration of the diffusion data between time points. The transformation was parametrized using cubic B-splines, making it intrinsically smooth, local, compact and fast to compute. The local support property of B-spline is particularly important as it ensures that a deformation in a given region of the brain has only a local effect on the transformation. An empirically chosen node spacing of 12 voxels in x, y, and z directions was used. To prevent folding in the transformation, the registration was constrained. These constraints made the Jacobian of the transformation greater than a small positive threshold on the pixels. Negative Jacobians between the pixels were penalized by making Jacobian derivatives

small on pixels. The constrained optimization problem was solved using a combination of the multipliers method and the limited memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) method [26] with a non-monotonic line search. The cost function was the squared difference between the reference and the deformed version of the floating image, suitable for intramodal registration of quantitative images such as FA.

6.3.3 Monte Carlo Simulation

Monte Carlo simulation studies were performed to investigate the effects of rotated diffusion encoding directions and different gain factors on the permutation exchangeability between different time points. DTI acquisition in each time point was assumed to be six diffusion directions (dual gradient scheme, $b=1000$ s/mm²) in addition to one $b=0$ s/mm² acquisition, repeated three times (total 21 measurements). The noise-free diffusion tensor \mathbf{D} was defined with FA=0.5, D_{av} ($=\text{Tr}(\mathbf{D})/3$) $=0.7 \times 10^{-3}$ mm²/s, and $\lambda_2=\lambda_3$.

The simulation of noisy diffusion weighted signals and DTI processing was done in a similar manner as described elsewhere [8]. From \mathbf{D} , noise-free diffusion weighted signals were calculated by $S(\mathbf{g}) = S_0 \exp(-b\mathbf{g}^T \mathbf{D} \mathbf{g})$ where S_0 is the signal without diffusion weighting, b is the diffusion weighting factor, and \mathbf{g} is 3x1 unit vector of the diffusion-encoding gradient direction. For time point B, in addition to using the same gradient set and scaling factors, we also tested scenarios where \mathbf{g} was replaced by $\mathbf{g}'=\mathbf{R}\mathbf{g}$ (\mathbf{R} is a product of three 3x3 matrices, each describing

rotations around x, y, and z axis each) or S_0 was replaced by S_0' in order to simulate the rotated diffusion directions or drift in gain factors. Next, the noise modeled as complex random number with real and imaginary parts following Gaussian distribution of zero mean and standard deviation σ ($=S_0/\text{SNR}$) was added to $S(\mathbf{g})$ and the magnitude of the noisy signal was calculated. The magnitude of the added noise (SNR=25 in S_0 or S_0') was constant for both time points, and for all the simulation studies.

From a complete set of noisy signals, the noisy diffusion tensor, FA, and finally the statistic of interest $\hat{\theta} = FA_B - FA_A$ were calculated. Then, the stratified DTI permutation algorithm was performed on a pool of 42 signals (7x3 from each time point) 999 times, with and without the 'corrected' permutation scheme that ensures exchangeability by permuting \mathbf{g} (\mathbf{g}' for B) in conjunction with $S(\mathbf{g})$ (scaled in B to match A). Using the statistic from the original labeling $\hat{\theta}$ and the values from the 999 permuted labelings $\hat{\theta}_i^*$, the p-value of $\hat{\theta}$ was calculated by equation 2. This experiment of estimating a p-value from a noisy dataset was repeated 10,000 times to derive a distribution of p-values. We assumed that the distribution of FA at both time points was equivalent and therefore the distribution of p-values should have a uniform probability between 0 and 1. This fundamental statistical property was used to evaluate our methods. The gold standard (a uniform distribution) and derived p-value distributions were compared using quantile-quantile plots (Q-Q plots).

6.3.4 Human Brain Data

6.3.4.1 Subjects / Acquisitions

Clinically isolated syndrome (CIS) is a diagnosis given to patients who experience for the first time neurological symptoms indicative of central nervous system dysfunction (visual, brain/brainstem, or spinal cord) thought to be due to primary demyelinating disease. It is considered by clinicians to be the earliest stage of multiple sclerosis (MS), even before a definite diagnosis can be confirmed by subsequent clinical symptoms or new lesions seen on MRI scans. Three CIS subjects at risk of developing MS and one healthy volunteer were included in this study. Patients were part of an ongoing follow-up study of CIS patients, with scans at baseline (within 3 month of initial clinical presentation) and every 3 to 6 months afterward for two years. A 27-year-old healthy male subject was scanned four times within two months in order to validate the technique when no biological changes over time can be assumed. All subjects gave informed consent and the study was approved by our institution's committee on human research.

All images were acquired on a 1.5T GE Signa system (General Electric, Milwaukee, WI) equipped with 4G/cm gradients and a standard quadrature head coil. The DTI scans were performed with a single-shot multislice axial spin echo EPI sequence (TR/TE = 7s/105ms, 9 NEX) at 1.7x1.7x2.1 mm voxel resolution using six diffusion-encoding directions (dual gradient scheme) at $b=2000 \text{ s/mm}^2$

in addition to one $b=0$ s/mm² image. The brain was covered from the level of midbrain to the vertex with 40 slices. Data averaging in the scanner was done up to 3 NEX, which was acquired three times to come up with 3 repetitions of 3 NEX data.

6.3.4.2 Analysis

The overall procedure of PERVADE can be split into 5 steps, and the details of each step are given below.

Step 1: Initial DTI processing at each time point

After motion correction by linear image registration (FLIRT) [27], the DWIs were slightly smoothed using a 3x3x3 voxel Gaussian kernel (FWHM 1.5 voxel) to increase the SNR and were processed to calculate FA, as described elsewhere [8]. Brain masks were created from the $b=0$ images using the brain extraction tool (BET) [28] in order to mask out the artificially high FA regions outside the brain so that these will not affect the inter-time-point registration of FA maps in the next step.

Step 2: Inter-time-point registration

Non-linear registration was done after the initial linear registrations by FLIRT. For both linear and non-linear registrations, FA maps from all the time points were matched to a reference FA map that was created as follows. First, a linear transformation between FA maps from two arbitrarily chosen time points was

calculated. Then, the FA map from one of these time points was half-transformed to the other time point such that the resulting FA map will be spatially in the middle in between these two FA maps. This procedure ensured that none of the time points benefited from having no registration / interpolation error that can arise in other time points. Also, sinc interpolation was used for all the registrations in this study. The DWIs were resampled based on the transformation between the FA maps.

Step 3: Masking Out Unnecessary Regions

We masked out brain regions where subsequent permutation tests were not to be performed. Not only does this save computation time, but this also increases the power of detecting statistically significant clusters by reducing the number of multiple comparisons. By default, white matter from the whole brain was defined as the domain for permutation testing. The whole brain white matter was defined as voxels within the brain that had FA values larger than 0.2 from the average of the FA maps from the two time points, and that were not contaminated by CSF (defined by segmenting $b=0$ image using an automated segmentation tool (FAST) [29]).

Step 4: Permutation testing for voxel-wise p-values

The DWIs that were resampled to the common space by step 2 were smoothed using the same smoothing kernel of step 1 before voxel-wise permutation testing. Voxels outside the white matter were not included in the smoothing kernel. The

first time point (baseline) was compared to subsequent time points. For each time point, the directions of the diffusion gradients were rotated to the same degree as determined from the linear transformation in step 2. In order to be exact, the degree of rotation should include the additional components of the non-linear transformation, but this was deemed minor in comparison to rotation described by linear transformation and thus was neglected. The scaling factor correcting for the drift in the gain factors over time was estimated from the median of ratio of signal intensities from NAWM (lesion masked out from white matter defined in step 3) and then multiplied back to DWIs. Permutation testing was performed with the observed labeling plus a randomly chosen set of 999 labelings, and the voxel-wise p-values were estimated as described above.

Step 5: Permutation testing for suprathreshold cluster analysis (STCA)

In step 4, the permutation of the DWIs were done at the image-level, thus each permutation could generate a null p-value map maintaining the inter-voxel spatial correlation structures. For each of the 1000 p-value maps, the maximum cluster size found after thresholding the voxel-wise p-values at $p=0.01$ was entered as a data point in the cluster null distribution. Actual definition of clusters was not only based on the voxel-wise p-values but also the sign of the changes; in order to be a cluster, all the voxels in the cluster have to be changing in the same direction (increased or decreased FA values). This eliminated the chance of defining clusters of mixed changes (which is difficult to interpret biologically) and increased the power of detecting clusters with unidirectional changes. Significant

clusters identified by the jump-down method were overlaid on the averaged FA maps for better visualization. In some cases, fiber tracking was performed to reveal the spatial connectivity between a focal lesion and significant clusters in the NAWM. The tracking algorithm was based on FACT [30], with termination criteria of $FA < 0.2$ and angles of primary eigenvectors between connecting voxels $> 50^\circ$.

6.4 Results

6.4.1 Monte Carlo Simulation

Comparison of the distribution of observed p-values from the simulation of permutation testing and the expected p-values are shown (Figure 6-3). Figure 6-3a shows that when the diffusion directions were rotated by 20 degrees each around x, y and z axes for one time point, and if the permutation was carried out ignoring the diffusion gradient directions, the observed p-values were deviated from the ideal distribution, with the trend of over-estimating p-values. This is not surprising since increased variability due to the mismatched diffusion gradient directions will lead to more dispersed permutation null distributions that will eventually inflate the estimated p-values.

Permuting rotated gradients in pair with the signals led to observed p-values that closely followed the expected distribution, demonstrating the effectiveness of the modified permutation strategy. Note that the p-values hit the floor of 0.001 since

using 1000 labelings does not allow the p-values to be lower than 0.001. This is not an important issue for this study since the voxel-wise p-values are to be thresholded at 0.01 for the cluster analysis. However, if estimating smaller p-values are desired, more permutation labelings would be necessary at the cost of increased computation time.

Results for smaller degrees of rotation of 10 degrees show that the deviation from the expected distribution without permuting diffusion gradients was much less. Since the degrees of rotation for real follow-up studies were typically within the range of 5 degrees (see below), we expect that the problem of over-estimating p-values will not be a serious issue even if diffusion gradients were not permuted, and any small effects will be eliminated by our proposed correction algorithm.

The differences in signal gain factors can bias the p-values in the similar way as diffusion gradient rotation (Figure 6-3c,d), and this effect was eliminated by scaling the intensities as described above.

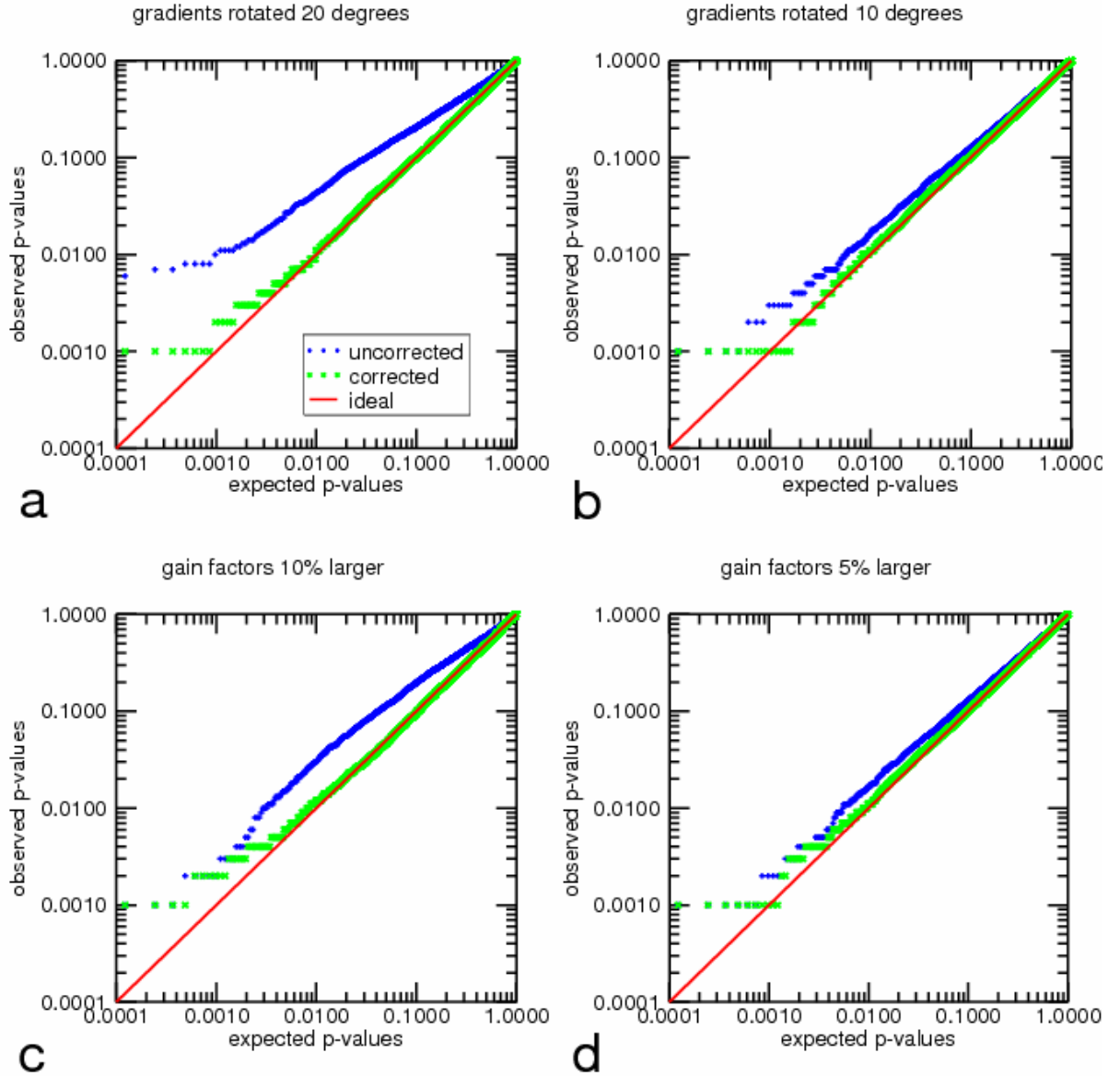


Figure 6-3: Qunatile-quantile plots for expected p-values following the uniform distribution between 0 and 1 versus observed p-values by permutation testing under diverse simulation conditions. Blue and green symbols indicate uncorrected and corrected (permuting the rotated gradients paired with the diffusion-weighted signals and equalizing the gain factors) permutation scheme while red correspond to perfect match between the expected and the observed. (a) Directions of diffusion encoding gradients set rotated by 20 degrees each in x, y and z in one of two time points only. (b) Gradients rotated by 10 degrees. (c) MR signal gain factor larger by 10% in one of two time points. (d) Gain factor larger by 5%.

6.4.2 Healthy Volunteer

Figure 6-4 shows voxel-wise p-value maps from the observed Δ Fa estimated by permutation testing from three pairs (0 month versus 0 month (head rotated), 1 month, and 2 months) for a healthy subject. We can appreciate the smoothed random field pattern with few noticeable non-random structures. Further evidence of the validity of PERVADE in human data is demonstrated by the Q-Q plot showing excellent agreement of single-voxel p-values from the whole brain white matter with the expected distribution. The permutation distribution $\hat{\theta}^* = \max_{\{p^*\}}(\{\text{cluster size}\})$ indicates that the probability of having any clusters larger than (or equal to) 11 voxels is 0.05, thus the critical value for significant clusters (for 0 versus 2 months) could be set as 11. Overall, only one significant cluster (slightly larger than the critical value) was identified in the three pairs of image-wise cluster analysis (data not shown).

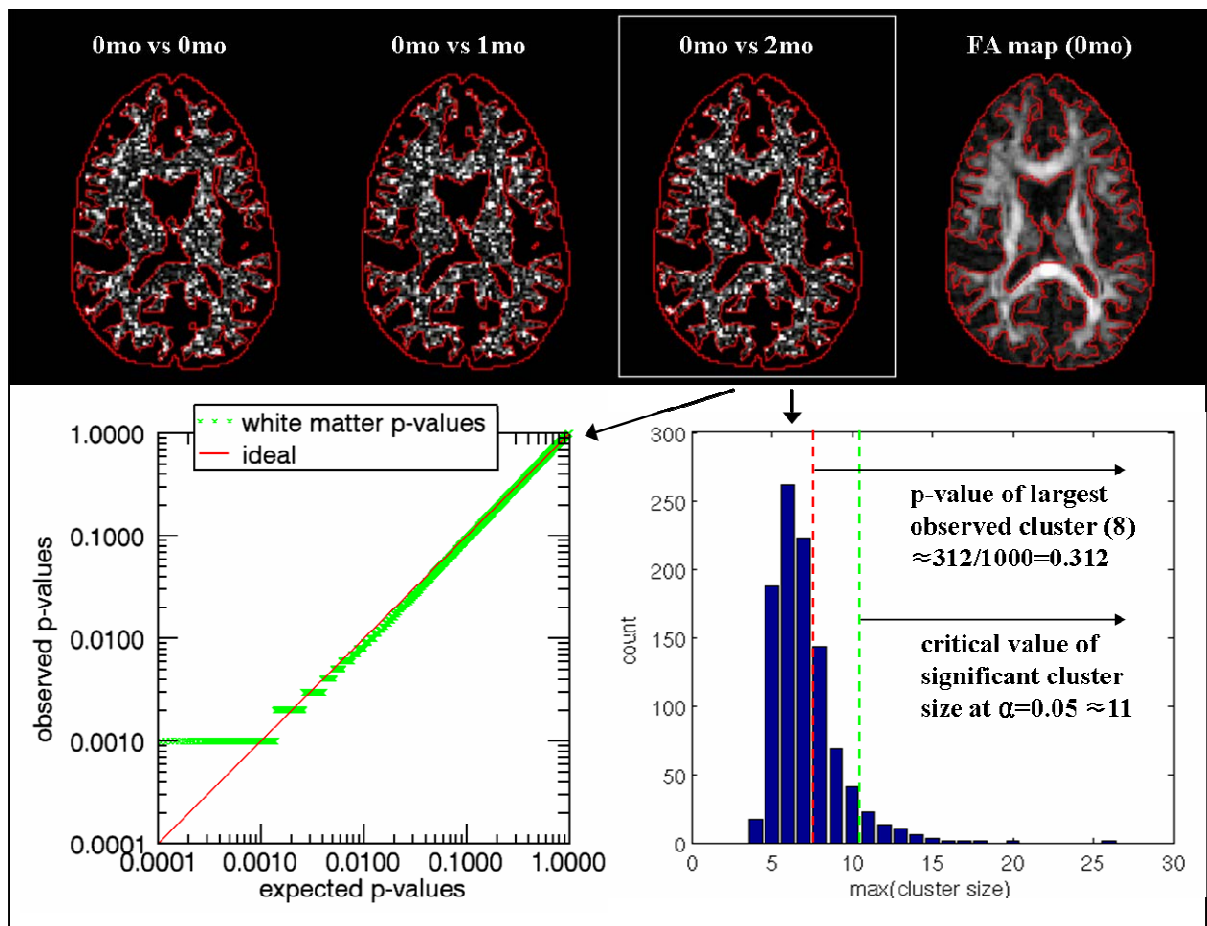


Figure 6-4: Healthy volunteer. [Top row] p-value maps from the baseline (0 month) versus 0 month (same day, subject was moved in between), 1 month, and 2 month as well as FA map from the baseline. P-values are rescaled by log transformation, and the maximal intensity displayed is $p=0.03$. White matter regions (as well as the whole brain) used for masking before permutation procedures are shown in red contours. [Bottom left] Q-Q plot of estimated voxel-wise p-values in the whole brain white matter of baseline versus 2 month. [Bottom right] Histogram of baseline versus 2 month permutation distribution of image-wise maximal cluster size from $N'=1000$ permutation trials. Critical value was 11 (i.e. probability of maximal cluster size ≥ 11 voxels was around 0.05). The largest cluster in the observed p-value map was 8 ($p=0.312$) thus no clusters were declared significant.

6.4.3 Patient #1

This patient had a significant thinning of the splenium of the corpus callosum over time that could be easily identified visually. For this patient, we compared

the result from linear and non-linear registrations to investigate the effects of atrophy (tissue loss) with or without the non-linear registration. The baseline FA map registered to the FA map 10 months later, either by linear alone and by linear + non-linear transformation, are shown (Figure 6-5). The corpus callosum in the registered baseline FA map was ‘thicker’ by about a half voxel than the reference image with linear registration only, while addition of the non-linear registration brought a better match in a qualitative sense. With the linear registration, significant clusters were identified at the border of the corpus callosum by PERVADE and those were not detected after non-linear registration. Even after non-linear registration, p-value maps of baseline versus 16 months showed the trend of general FA decrease in the corpus callosum (most voxels are in ‘blue’ color) though they were not detected by PERVADE as clusters.

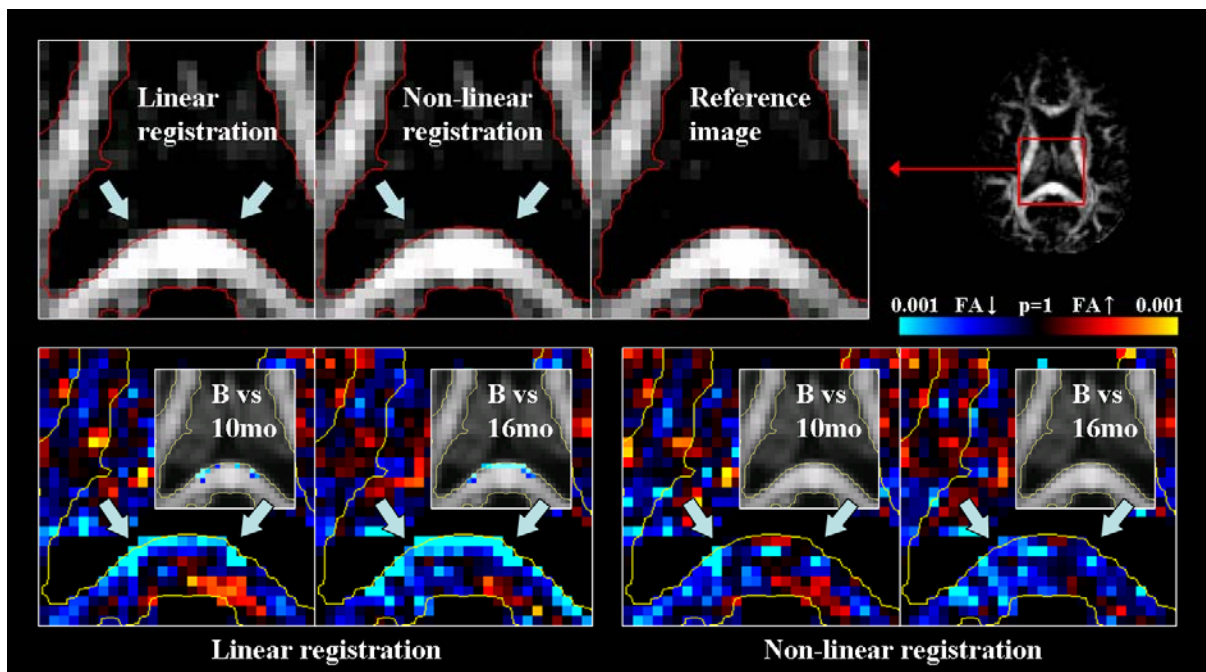


Figure 6-5: Patient #1. [Top row] Registration of the baseline FA map to the reference FA map (10 months) by linear registration only and non-linear registration (after linear registration). Images are magnified to splenium of corpus callosum and nearby structures. Red contour

denotes the border defined in the reference image. [Bottom row] P-value maps for the same anatomic structure either by linear or non-linear registration for baseline versus 10 months or 16 months. Significant clusters detected by PERVADE are displayed in the smaller embedded boxes.

6.4.4 Patient #2

This patient developed three new lesions in and around the splenium of the corpus callosum at 9 months after the baseline scan. While a few clusters (most of them attributable to lesions visible in conventional MRI as well) were detected in the analysis of baseline versus follow-up scans before 9 months (data not shown), the analysis of baseline versus 9 months revealed numerous clusters with $p < 0.05$, not only within lesions but also in the NAWM (Figure 6-6). In the later time points, there was a tendency for the FA clusters to remain longer than T2 hyperintensities in some lesions. Using DTI fiber tracking, part of the forceps major (a subset of corpus callosum fibers connecting occipital lobes on both sides) was delineated that intersects one of the new lesions at 9 months (the lesion within the splenium of corpus callosum close to the midline). Significant NAWM clusters were located within or near the tracks, suggesting that the detected NAWM changes might be related to the new lesion.

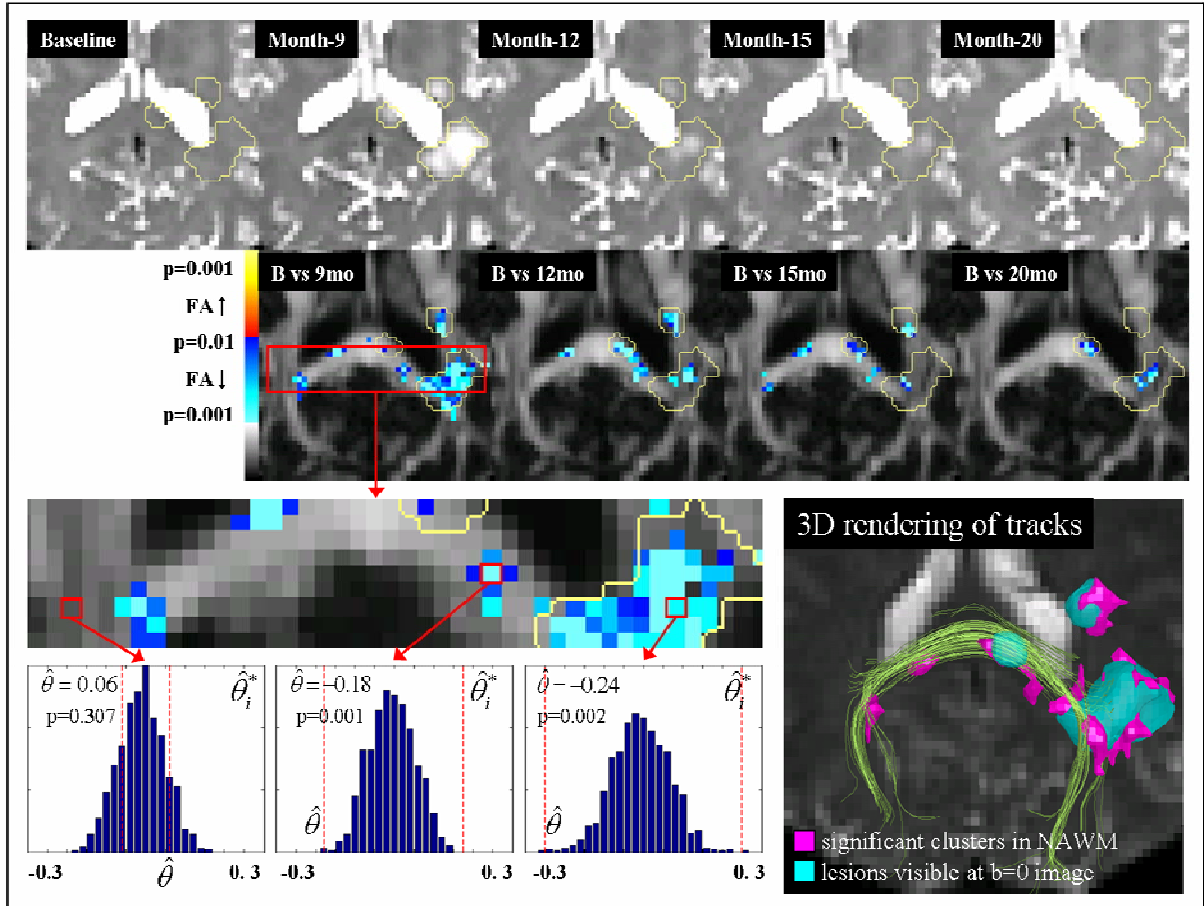


Figure 6-6: Patient #2. [Top row] $b=0$ images at different time points showing lesions around the left posterior horn of the lateral ventricle emerging and fading over time. Images are smoothed to the same degree as PERVADE and contrast / level were adjusted for the best visibility of lesions. Yellow contours demarcate the regions of the new lesions at the time of the appearance. [Middle row] Detected significant clusters for baseline versus subsequent time points. [Bottom row left] Permutation distribution of ΔFA as well as observed ΔFA for three voxels (NAWM and not within a significant cluster, NAWM and within a significant cluster, and core of a new lesion). Locations of these voxels are shown as well. [Bottom row right] 3D rendering of the tracks (light green), significant clusters (magenta) and lesions visible at $b=0$ images (cyan). Note that significant clusters are also present within the lesions though they are obscured by rendered lesions.

6.4.5 Patient #3

This patient had a lesion within the right pyramidal tract at the superior level of the internal capsule and inferior aspect of the corona radiata. The degree of T2

hyperintensity within the lesion lessened after 3 month compared to the baseline scan, possibly indicating edema resolution. Results from the serial DTI permutation tests highlighted two significant clusters, one in the lesion with increasing FA over time, and another in the NAWM at the superior aspect of the corona radiata with decreasing FA over time (Figure 6-7). Tracks (part of pyramidal tract) seeded from the lesion showed that the distal cluster of decreased FA was connected to the lesion, suggesting that they are possibly related.

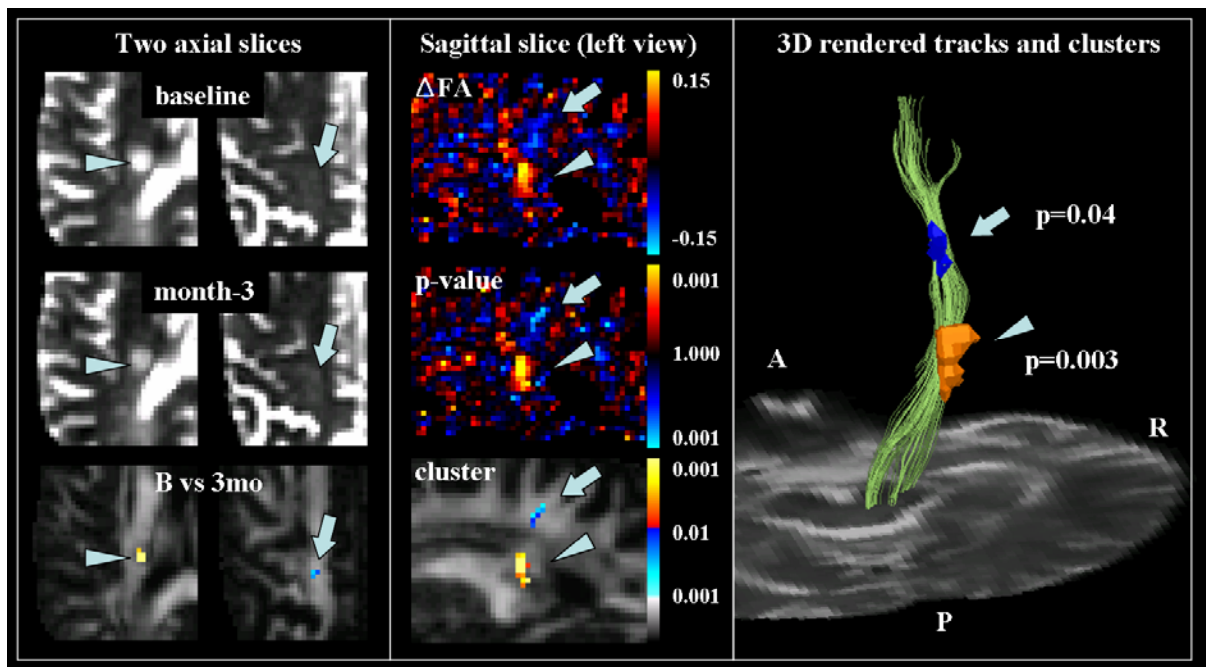


Figure 6-7: Patient #3. [Left column] $b=0$ images of baseline versus 3 months as well as detected clusters. Arrowheads and arrows indicate the lesion fading over time and detected NAWM cluster each. [Middle column] Δ FA, p-values and significant clusters in a sagittal slice (viewed from left) including these clusters. [Right column] 3D rendering of the tracks (light green), cluster within the lesion (orange) and cluster in the NAWM (blue). Cluster-wise p-values are shown as well.

6.4.6 Gradient tables rotation and gain factors

For all the human data analyzed by PERVADE, the average difference of angulations between the two time points were 2.52, 1.55 and 1.30 degrees around x-, y- and z-axes, respectively. The average difference in the gain factors was 5.6%.

6.5 Discussion

In this paper, we have proposed a novel non-parametric statistical framework for detecting subtle and local diffusion MRI changes over time in a single subject. To our knowledge, this is the first study to analyze single-subject serial DTI data at the voxel / cluster level. This was possible with the ability of DTI permutation testing to calculate voxel-wise statistic p-values and to correct for multiple comparisons. We have demonstrated that we can ensure exchangeability of the DWIs which justified our implementation of permutations for testing two DTI datasets. We have also demonstrated that voxel-wise p-values from the DTI permutation tests follow the expected null distribution in a control subject, and that we can detect statistically significant clusters in lesions and normal-appearing white matter of patients at the earliest stage of MS.

Diffusion MRI, including three-directional diffusion-weighted MRI and DTI, has been widely used in the study of MS [31]. Numerous studies have shown that diffusion MRI is sensitive to subtle damage in the normal appearing brain tissue as well as lesions and that it might even provide insights about pathological

processes such as inflammation, demyelination and axonal injury [32]. Serial studies can be especially valuable in studying disease evolution with or without treatment that can lead to a better understanding of pathophysiology and drug response, although they are rarely reported, probably due to more involved recruitment of patients. Four serial studies used ROI analysis of apparent diffusion coefficient (ADC) on either lesion or NAWM [33-36] and one study used histogram analysis of D_{av} and FA [37]. PERVADE is a completely different approach of analyzing serial DTI that can complement the techniques described above.

Before further discussing the merits and challenges of the technique presented here, we would like to point out that there have been numerous efforts in the MRI literature to use voxel-wise techniques to detect single-subject structural changes over time. Since the demonstration of detecting subtle brain changes by subvoxel registration and subtraction of serially acquired conventional MRIs [38, 39], image subtraction techniques have been widely applied in detecting intensity changes (such as lesions in MS or tumor evolution), morphological changes (such as atrophy) or both. For localization and quantification of atrophy, image subtraction schemes have either focused on the structural boundary shift [40-42] or Jacobian matrix [43-46]. For multiple sclerosis, the major interest was in the analysis of lesion evolution [47-50].

The techniques listed above have used conventional MRIs, and usually do not provide the statistical significance for the difference detected since conventional MRIs yield a simple one-image datum that does not allow uncertainties to be calculated in each voxel. One study, though, has attempted statistical testing with conventional MRIs [51]. In DTI, a study utilized subtraction of FA maps over two time points to visualize the spatial distribution of FA changes [52], though no formal voxel-wise statistical tests were attempted. DTI is unique in that numerous images are acquired, and this allows sufficient number of labelings to be used for permutation. We took advantage of this property to create statistical procedures for detection, rather than simple subtraction.

The most important strength of PERVADE is that it can lead to an almost exact statistical testing with minimal assumptions, while parametric approaches will have to make numerous compromises to reach the same goal. For instance, if the classical statistical framework of SPM is attempted for this study, we would first need to estimate the standard error of FA by an analytical formula [53, 54] with approximations from matrix perturbations and assumptions of known SNR or non-parametrically [55]. Then, we would need to create t-statistic maps assuming that FA probabilistic distribution can be approximated by a Gaussian distribution. Finally, for multiple comparison corrections, Gaussian random field theory [56, 57] can be used to investigate the distribution of the maximal statistic (largest t or cluster size) within the image, with numerous assumptions [58] that might not be satisfied in our case (such as heavy smoothing of FWHM of at least 2-3 times the

voxel size). The feasibility of a parametric statistical framework for serial DTI voxel-wise analysis would be a subject of future study.

While we have presented detection of significant FA clusters, the flexibility of the DTI permutation in the test statistic θ to be tested ($\theta = FA_B - FA_A$ in this study) allows it to be easily applied to different types of problems, as long as θ can be formulated. One simple example would be the comparison of FA averaged within a ROI. A more complicated scenario would be to test whether the connectivity between two predefined ROIs was changing over time in a subject. Connectivity testing are neither whole-brain voxel-wise nor ROI based tests, and certainly not well-defined in terms of parametric statistics since these problems are very specific to DTI analysis. For permutation testing, however, tests can be carried out in exactly the same way once θ is defined. Let's assume that $\theta = N_B - N_A$ where N is the number of reconstructed tracks connecting two regions, at time point A or B. If the permutation distribution of θ is created by tracking and counting the number tracks in each of the relabeled DTI datasets, we can calculate the p-value of observing the difference in the track counts by random chance. Last, permutation testing may be used in testing statistics derived from high angular resolution diffusion-weighted MRI or q-space imaging such as difference in generalized fractional anisotropy (GFA) or peaks of diffusion orientation diffusion function (ODF) [59].

In order to properly implement PERVADE, there are a few requirements to be met. First, it is designed to be suitable for serial DTI with the same DTI protocol. If comparing data from different sets of diffusion gradients is desired, stratified permutation of DWIs might not be optimal and the permutation distribution would include the additional variability from the effect of differing gradient configuration. Permuting residuals from fitting DWIs to the diffusion tensor model, similar to the way that residual bootstrap was implemented [8], might be an alternative that can be investigated in the future. Next, the ability of the permutation test to estimate small p-values depends on how many different labelings are possible. Suppose we have seven images (minimum for DTI) from each time point. Then, $({}_2C_1)^7=128$ different labelings exist and the smallest possible p-value is $1/128=0.0078$ (or 0.0156 for two-tailed tests due to symmetry of labeling), which is not small enough if a p-value threshold of 0.01 is desired, for instance. Fortunately, most of the recent DTI studies acquire many more than seven images, and the number of possible labelings increase dramatically with more DWIs, such as $({}_2C_1)^{14}=16,384$ for 14 images with no repeated acquisitions, or $({}_6C_3)^7 \approx 1.28 \times 10^9$ for the acquisition scheme used in this study.

It is also important to understand the limitations of PERVADE. This technique will be susceptible to artifacts that dominate the random noise, such as incomplete fat saturation. In a group comparison, the effects of artifacts will tend to be diluted due to the nature of having multiple subjects, but in a single subject study with only two datasets, image artifacts will play a more important role. For all the

results presented here, we made sure that detected clusters were not the result of any possible unsuppressed fat signals at one of the time points. Another major artifact is due to cardiac pulsation, but this is believed to be less of an issue since the two time points will be affected to a similar degree. However, the cardiac artifacts could still have different effects at different time points. The fact that we do not tend to see these artifactual effects in our analyses may arise from the permutation method, which to some degree will increase the variability in the null distribution in response to time-wise random effects during the individual DTI acquisitions.

Perhaps a more challenging issue is the effect of atrophy, as described in the registration method section. Even with a perfect registration, a complete distinction of microscopic and macroscopic changes may not be easy because of the following reasons. Technically, it is conceivable that changes in the microstructure will lead to slightly different definition of the 'boundary' of specific white matter structures that guides the registration, since no imaging technique other than DTI itself provides such rich contrast within white matter for the registration. Biologically, any irreversible damage in the axon / myelin structures will likely be followed by atrophy, thus it is not surprising that the two conditions often co-exist. Detecting these effects are both of much interest, but at some point we may have to accept that in trying to detect one of these two, the other may often exist to some degree. This difficulty does not invalidate the technique

of detecting biologic changes, but rather adds some uncertainty to the interpretation in terms of pure microscopic versus macroscopic changes.

In general, the statistical power of detecting biologically-driven significant clusters is somewhat unclear at this moment. Not only do we have little prior information about the effect size we want to detect, but this is made more complicated by the fact that FA in the different ranges have different uncertainties even with the same SNR of the images. Also, it is not known how similar degrees of injuries will affect FA in white matter with different magnitudes of FA and whether more than one fiber population exists within a voxel. Thus the power to detect a similar degree of injury in different white matter structures can be variable. With more information available on how pathological changes affect DTI metrics, we will be able to estimate the requirements for the SNR of acquisition and more generally the reliability of this technique.

6.6 Conclusion

We presented a novel statistical framework for whole brain voxel / cluster-wise analysis of single-subject serial DTI based on non-parametric permutation testing. This completely automated technique called PERVADE can be useful in detecting subtle and local DTI changes over time in a specific subject, especially in disease processes such as MS or traumatic brain injury that have heterogeneous regional distributions across different patients.

6.7 Reference

1. Bassar, P.J., J. Mattiello, and D. LeBihan, *Estimation of the effective self-diffusion tensor from the NMR spin echo*. J Magn Reson B, 1994. **103**(3): p. 247-54.
2. Bassar, P.J., J. Mattiello, and D. LeBihan, *MR diffusion tensor spectroscopy and imaging*. Biophys J, 1994. **66**(1): p. 259-67.
3. Friston, K.J., et al., *Statistical parametric maps in functional imaging: a general linear approach*. Hum Brain Mapp, 1995. **2**: p. 189-210.
4. Smith, S.M., et al., *Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data*. Neuroimage, 2006. **31**(4): p. 1487-505.
5. Schwartzman, A., R.F. Dougherty, and J.E. Taylor, *Cross-subject comparison of principal diffusion direction maps*. Magn Reson Med, 2005. **53**(6): p. 1423-31.
6. Pajevic, S. and P.J. Bassar, *Parametric and non-parametric statistical analysis of DT-MRI data*. J Magn Reson, 2003. **161**(1): p. 1-14.
7. Skare, S., et al., *Noise considerations in the determination of diffusion tensor anisotropy*. Magn Reson Imaging, 2000. **18**(6): p. 659-69.
8. Chung, S., Y. Lu, and R.G. Henry, *Comparison of bootstrap approaches for estimation of uncertainties of DTI parameters*. Neuroimage, 2006. **33**(2): p. 531-41.
9. Heim, S., et al., *Assessing DTI data quality using bootstrap analysis*. Magn Reson Med, 2004. **52**(3): p. 582-9.
10. Jones, D.K., *Determining and visualizing uncertainty in estimates of fiber orientation from diffusion tensor MRI*. Magn Reson Med, 2003. **49**(1): p. 7-12.
11. Lazar, M. and A.L. Alexander, *Bootstrap white matter tractography (BOOT-TRAC)*. Neuroimage, 2005. **24**(2): p. 524-32.
12. Whitcher, B., et al., *Using the wild bootstrap to quantify uncertainty in diffusion tensor imaging*. Hum Brain Mapp, 2007.
13. Edgington, E.S., *Randomization tests*. 3rd ed. 1995, New York: Marcel Dekker. xxii, 409.
14. Good, P.I., *Permutation, parametric and bootstrap tests of hypotheses*. 3rd ed. Springer series in statistics. 2005, New York: Springer. xix, 315 p.
15. Manly, B.F.J., *Randomization, bootstrap, and Monte Carlo methods in biology*. 3rd ed. Texts in statistical science. 2007, Boca Raton, FL: Chapman & Hall/ CRC. 455 p.
16. Nichols, T.E. and A.P. Holmes, *Nonparametric permutation tests for functional neuroimaging: a primer with examples*. Hum Brain Mapp, 2002. **15**(1): p. 1-25.
17. Jones, D.K., *The effect of gradient sampling schemes on measures derived from diffusion tensor MRI: a Monte Carlo study*. Magn Reson Med, 2004. **51**(4): p. 807-15.
18. Roland, P.E., et al., *Three-dimensional analysis of clustered voxels in O-butanol brain activation images*. Human Brain Mapping, 1993. **1**(1): p. 3-19.
19. Bullmore, E.T., et al., *Global, voxel, and cluster tests, by theory and permutation, for a difference between two groups of structural MR images of the brain*. IEEE Trans Med Imaging, 1999. **18**(1): p. 32-42.
20. Friston, K.J., et al., *Detecting activations in PET and fMRI: levels of inference and power*. Neuroimage, 1996. **4**(3 Pt 1): p. 223-35.

21. Poline, J.B., et al., *Combining spatial extent and peak intensity to test for activations in functional imaging*. Neuroimage, 1997. **5**(2): p. 83-96.
22. Locascio, J.J., et al., *Time series analysis in the time domain and resampling methods for studies of functional magnetic resonance brain imaging*. Human Brain Mapping, 1997. **5**(3): p. 168-193.
23. Holmes, A.P., et al., *Nonparametric analysis of statistic images from functional mapping experiments*. J Cereb Blood Flow Metab, 1996. **16**(1): p. 7-22.
24. Holmes, A.P., *Statistical issues in functional brain mapping*. 1994, University of Glasgow.
25. Sdika, M., *A fast non rigid image registration with constraints on the Jacobian using large scale constrained optimization*. Submitted to IEEE on Medical Imaging, 2007.
26. Liu, D.C. and J. Nocedal, *On the Limited Memory Bfgs Method for Large-Scale Optimization*. Mathematical Programming, 1989. **45**(3): p. 503-528.
27. Jenkinson, M. and S. Smith, *A global optimisation method for robust affine registration of brain images*. Med Image Anal, 2001. **5**(2): p. 143-56.
28. Smith, S.M., *Fast robust automated brain extraction*. Hum Brain Mapp, 2002. **17**(3): p. 143-55.
29. Zhang, Y., M. Brady, and S. Smith, *Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm*. IEEE Trans Med Imaging, 2001. **20**(1): p. 45-57.
30. Mori, S., et al., *Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging*. Ann Neurol, 1999. **45**(2): p. 265-9.
31. Rovaris, M., et al., *Diffusion MRI in multiple sclerosis*. Neurology, 2005. **65**(10): p. 1526-32.
32. Henry, R.G., et al., *Directional diffusion in relapsing-remitting multiple sclerosis: a possible in vivo signature of Wallerian degeneration*. J Magn Reson Imaging, 2003. **18**(4): p. 420-6.
33. Caramia, F., et al., *A longitudinal study of MR diffusion changes in normal appearing white matter of patients with early multiple sclerosis*. Magn Reson Imaging, 2002. **20**(5): p. 383-8.
34. Castriota-Scanderbeg, A., et al., *Diffusion of water in large demyelinating lesions: a follow-up study*. Neuroradiology, 2002. **44**(9): p. 764-7.
35. Rocca, M.A., et al., *Weekly diffusion-weighted imaging of normal-appearing white matter in MS*. Neurology, 2000. **55**(6): p. 882-4.
36. Werring, D.J., et al., *The pathogenesis of lesions and normal-appearing white matter changes in multiple sclerosis: a serial diffusion MRI study*. Brain, 2000. **123** (Pt 8): p. 1667-76.
37. Oreja-Guevara, C., et al., *Progressive gray matter damage in patients with relapsing-remitting multiple sclerosis: a longitudinal diffusion tensor magnetic resonance imaging study*. Arch Neurol, 2005. **62**(4): p. 578-84.
38. Hajnal, J.V., et al., *Detection of Subtle Brain Changes Using Subvoxel Registration and Subtraction of Serial Mr-Images*. Journal of Computer Assisted Tomography, 1995. **19**(5): p. 677-691.

39. Hajnal, J.V., et al., *A Registration and Interpolation Procedure for Subvoxel Matching of Serially Acquired Mr-Images*. Journal of Computer Assisted Tomography, 1995. **19**(2): p. 289-296.
40. Freeborough, P.A. and N.C. Fox, *The boundary shift integral: an accurate and robust measure of cerebral volume changes from registered repeat MRI*. IEEE Trans Med Imaging, 1997. **16**(5): p. 623-9.
41. Smith, S.M., et al., *Normalized accurate measurement of longitudinal brain change*. J Comput Assist Tomogr, 2001. **25**(3): p. 466-75.
42. Wang, D. and D.M. Doddrell, *MR image-based measurement of rates of change in volumes of brain structures. Part I: method and validation*. Magn Reson Imaging, 2002. **20**(1): p. 27-40.
43. Crum, W.R., R.I. Scahill, and N.C. Fox, *Automated hippocampal segmentation by regional fluid registration of serial MRI: validation and application in Alzheimer's disease*. Neuroimage, 2001. **13**(5): p. 847-55.
44. Freeborough, P.A. and N.C. Fox, *Modeling brain deformations in Alzheimer disease by fluid registration of serial 3D MR images*. J Comput Assist Tomogr, 1998. **22**(5): p. 838-43.
45. Thompson, P.M., et al., *Dynamics of gray matter loss in Alzheimer's disease*. J Neurosci, 2003. **23**(3): p. 994-1005.
46. Studholme, C., et al., *An intensity consistent filtering approach to the analysis of deformation tensor derived maps of brain shape*. Neuroimage, 2003. **19**(4): p. 1638-49.
47. Gerig, G., et al., *Exploring the discrimination power of the time domain for segmentation and characterization of active lesions in serial MR data*. Med Image Anal, 2000. **4**(1): p. 31-42.
48. Lee, M.A., et al., *Defining multiple sclerosis disease activity using MRI T2-weighted difference imaging*. Brain, 1998. **121** (Pt 11): p. 2095-102.
49. Meier, D.S. and C.R. Guttmann, *Time-series analysis of MRI intensity patterns in multiple sclerosis*. Neuroimage, 2003. **20**(2): p. 1193-209.
50. Rey, D., et al., *Automatic detection and segmentation of evolving processes in 3D medical images: Application to multiple sclerosis*. Med Image Anal, 2002. **6**(2): p. 163-79.
51. Lemieux, L., et al., *The detection and significance of subtle changes in mixed-signal brain lesions by serial MRI scan matching and spatial normalization*. Med Image Anal, 1998. **2**(3): p. 227-42.
52. Qiu, D., et al., *Mapping radiation dose distribution on the fractional anisotropy map: applications in the assessment of treatment-induced white matter injury*. Neuroimage, 2006. **31**(1): p. 109-15.
53. Anderson, A.W., *Theoretical analysis of the effects of noise on diffusion tensor imaging*. Magn Reson Med, 2001. **46**(6): p. 1174-88.
54. Chang, L.C., et al., *Variance of estimated DTI-derived parameters via first-order perturbation methods*. Magn Reson Med, 2007. **57**(1): p. 141-9.
55. Chung, S., et al. *Bootstrap-based longitudinal analysis of DTI estimates (BLADE): applications in clinically isolated syndrome patients*. in *14th Annual Meeting of ISMRM*. 2006. Seattle, WA: ISMRM.

56. Worsley, K.J., et al., *A three-dimensional statistical analysis for CBF activation studies in human brain*. J Cereb Blood Flow Metab, 1992. **12**(6): p. 900-18.
57. Friston, K.J., et al., *Comparing functional (PET) images: the assessment of significant change*. J Cereb Blood Flow Metab, 1991. **11**(4): p. 690-9.
58. Petersson, K.M., et al., *Statistical limitations in functional neuroimaging. II. Signal detection and statistical inference*. Philos Trans R Soc Lond B Biol Sci, 1999. **354**(1387): p. 1261-81.
59. Tuch, D.S., *Q-ball imaging*. Magn Reson Med, 2004. **52**(6): p. 1358-72.

Chapter 7: Main Findings and Conclusions

In this dissertation, I have developed DTI-optimized resampling statistical tools. The bootstrap technique has emerged as a strong candidate to fulfill the need for the accurate uncertainty estimation of DTI derived parameters. It could help us to assess changes observed by serial DTI studies, which was the initial motivation behind of this dissertation project. Before this project, though, DTI bootstrap was still in its infancy, substantially biased and not applicable to data without repeated acquisitions. Thus, as the first part of the project, various bootstrap algorithms suitable for DTI were comprehensively investigated with Monte Carlo simulation. This study provided a strong foundation for the applications of bootstrap in the second part of the project, and greatly enhanced our understanding about bootstrap techniques in the DTI community.

In the second part of this dissertation project, residual bootstrap was used in tackling two different problems, cardiac artifact quantification and serial single-subject analysis. Prior to this project, the benefit of cardiac gating in DTI studies with single-shot EPI acquisitions was not clear. Bootstrap analysis on the study of cardiac gating offered convincing evidence that cardiac gating is necessary with partial Fourier acquisition. Full Fourier acquisition was shown to be less sensitive to cardiac pulsation artifacts. These results have a profound effect in the DTI community, since virtually every DTI acquisitions can be significantly improved. To better analyze serially acquired DTI data in an individual subject,

the other problem which motivated this project, a new statistical framework called BLADE was developed. BLADE was able to detect FA decrease in the normal appearing white matter of TBI patients, a very interesting finding of possible Wallerian degeneration that could not be detected otherwise.

Third and last part of this project was devoted to developing another statistical framework for serial DTI data analysis. This technique called PERVADE was designed to overcome limitations of BLADE such as null distribution estimation and multiple comparison correction. Instead of bootstrap, PERVADE is based on permutation testing, another resampling statistical technique, which has been optimized to accommodate the complex data structure of DTI. PERVADE was able to detect FA changes in the normal appearing white matter just like BLADE, with even higher confidence in the results due to the robust statistics used in PERVADE.

This project has shown that resampling statistics can enhance the capability of DTI with new analyses frameworks such as serial single-subject analysis. It is expected that the need for bootstrap and permutation testing will only increase in the future for diffusion MRI applications, but also for other MRI and imaging modalities with data oversampling. Currently, the next generation of modeling tissue water diffusion with diffusion MRI such as high angular resolution diffusion MR imaging (HARDI) or q-space imaging is under active development. Fiber tracking, already a product of complicated data processing, is getting even more

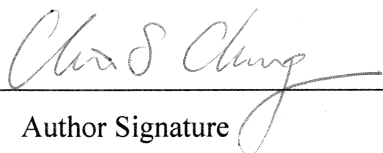
sophisticated, with the help of new diffusion models and new approaches to the problem of connectivity, such as the structural connectivity matrix. As the data processing gets more complex, parametric methods for uncertainty estimation such as the analytical approach will be increasingly difficult to implement. On the other hand, implementing resampling statistics is relatively straightforward since the same principles apply. It will be exciting to watch how resampling statistics evolve in the next generation diffusion MRI studies.

Publishing Agreement

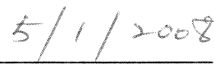
It is the policy of the University to encourage the distribution of all theses and dissertations. Copies of all UCSF theses and dissertations will be routed to the library via the Graduate Division. The library will make all theses and dissertations accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis or dissertation to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.



Author Signature



Date