Cover Letter

Please accept the attached grant “Statistical methods for longitudinal multivariate neuroimaging biomarkers” by Dr. Ciprian Crainiceanu (PI), Dr. Elizabeth Sweeney, Dr. Andrew Leroux, Dr. Daniel Reich, Dr. Peter Calabresi, Dr. Brian Caffo, and Dr. John Muschelli. This grant proposal is a competitive renewal application of the highly successful RO1 grant NS060910 ”Statistical methods for longitudinal multivariate neuroimaging biomarkers”. The grant proposes a natural and necessary research development, where previous measures obtained from neuroimaging are developed into potential biomarkers that are sensitive to longitudinal changes in brain pathology. The proposal deals with biostatistical methods for longitudinal, multivariate high dimensional biological signals obtained from neuroimaging studies.

Please assign the application to the following

SRG Biostatistical Methods and Research Design study section (**BMRD**).

Rationale. The focus of this grant is statistical methodology development for modern longitudinal data high dimensional data sets. The investigators are biostatisticians and the grant is written in a format for a biostatistical audience. We anticipate that the work will have an impact on biostatistical methodology research as well as the practice of data analysis in longitudinal neuroimaging studies in general, and in the study of the temporal dynamics of brain structure and MS lesions, in particular.

IC One of:

1. National Institute of Neurological Disorders and Stroke - NINDS
2. National Institute of Mental Health - NIMH

Rationale. Our application deals with clinical brain imaging, which is central to the scientific goals of these Institutes.

Sincerely,

Ciprian Crainiceanu
Elizabeth Sweeney
Andrew Leroux
Daniel Reich
Peter Calabresi
Brian Caffo
John Muschelli

Abstract. Multiple Sclerosis (MS) is an immune-mediated disease of the central nervous system (CNS) that affects an estimated $400,000$ people in the United States alone. MS is characterized by focal demyelinating lesions and causes physical and cognitive impairment. While imaging studies are widely used in clinical practice and research, the number of targeted and strongly predictive neuroimaging-based biomarkers is small. Thus, we focus on two promising imaging modalities that are likely to capture complementary information on MS disease severity and dynamics: (1) longitudinal changes in white matter integrity captured by Diffusion Tensor Imaging (DTI) across the corpus callosum; and (2) longitudinal changes in the voxel intensities of MS lesions in multi-sequence Magnetic Resonance Imaging (MRI). To address these emerging data structures we propose realistic biostatistical methods that can scale up and produce principled inference for longitudinal high dimensional data. The first Aim is focused on massively univariate generalized linear mixed effects models (MU-GLMMs) and proposes a simple inferential approach for dealing with the within- and between-study participant correlation. The second Aim is concerned with the joint modeling of dense longitudinal high dimensional data (e.g., lesion voxel intensities) and survival time (e.g., time to voxel recovery). The third Aim is designed to quantify the association between the longitudinal neuroimaging and established MS biomarkers. The fourth Aim is dedicated to implementation, software, and reproducibility.

Budget justification

Research team

Ciprian Crainiceanu, PhD, PI biostatistician (2.4 calendar months) is the co-founder and co-leader of the highly successful SMART research group and is a co-author of the book on Measurement Error in Nonlinear Models with Carroll, Ruppert and Stefanski. He has served as a biostatistician on multiple NIH funded research projects, including studies of sleep (PI: Dr. Naresh Punjabi), heavy metals exposure (PI: Dr. Eliseo Guallar), and cardiovascular diseases (PI: Josef Coresh). Dr. Crainiceanu is an expert in functional data analysis, nonparametric regression, Bayesian inference, brain imaging and wearable computing. He has over 150 publications on a variety of statistical and scientific topics and has successfully led the three precursor NINDS grants 1R01NS060910 "Statistical Methods for Multilevel Multivariate Functional Studies." Dr. Crainiceanu will coordinate all of the work for this grant and will ensure that all Aims will be addressed on time. Dr. Crainiceanu will organize a highly collaborative environment between the co-Investigators on the grant, their students and post-doctoral fellows, as well as the wider scientific community. This will build up on his extensive experience of building research groups and it is very likely to be successful as the members of the team have been working together for a long time (between 3 and 17 years). He will mentor the graduate students as well as lead and facilitate all of the research in the grant.

Daniel Reich, MD, PhD, neuroradiologist, directs the Translational Neuroradiology Section within the Intramural Research Program of the National Institute of Neurological Disorders and Stroke. Because he is an Intramural NIH researcher, Dr. Reich will not draw support from this grant, but will assist the PI and the other co-Investigators to conduct the Aims of the current proposal. Research in Dr. Reich's lab focuses on the use of advanced magnetic resonance imaging techniques to understand the sources of disability in multiple sclerosis and on ways of adapting those techniques for use in research trials and routine patient care. Dr. Reich is actively involved in training junior scientists and physicians. Trainees in his lab include postdoctoral and clinical fellows, graduate and medical students, and high school summer students. Dr. Reich is trained in diagnostic radiology and neurology, has a Certificate of Added Qualification in neuroradiology, and holds an active medical license in the state of Maryland. Dr. Reich and Dr. Crainiceanu have a successful collaboration that has resulted in more than 30 jointly published papers to date. Dr. Reich will work with the research group on all Aims of the grant. Dr. Reich has already hosted four students from Dr. Crainiceanu's research lab (Drs. Sweeney, Taki Shinohara, Jonathan Gellar, and James Pringle). Dr. Reich will continue the direct collaboration with Dr. Crainiceanu and the SMART research group, will provide access to his lab and data sources, and will assist Dr. Crainiceanu with defining and identifying the best avenues for biomarker discovery using longitudinal MS data.

Peter Calabresi, MD, PhD, neurologist (0.6 calendar months) is a Professor of Neurology at Johns Hopkins University, where he is the Director of the Division of Neuroimmunology as well as the Director of the Multiple Sclerosis Center, focusing on the diagnosis and management of MS. He is the principal investigator on several clinical trials and oversees research projects seeking to create new anti-inflammatory and neuroprotective therapies for MS. Dr. Calabresi is also directing a longitudinal magnetic resonance imaging (MRI) study using new ways of assessing nerve fiber and myelin integrity in the brain and spinal cord. Dr. Calabresi and Dr. Crainiceanu have a long term collaboration that resulted in multiple joint papers (15+) and regular meetings. Dr. Calabresi will provide scientific guidance and will work closely with the PI to ensure a high level of scientific focus and training for the PhD students and post-doctoral fellows involved in research. Drs. Crainiceanu and Calabresi will continue to develop their pilot program on pairing Biostatistical students and MS research fellows to

accelerate the research schedule. Dr. Calabresi will provide access to his lab and data sources, and will assist Dr. Crainiceanu with defining and identifying the best avenues for biomarker discovery using longitudinal MS data and will help co-advise the PhD and post-doctoral students working on the grant. Brian Caffo, PhD, biostatistician (1.8 calendar months) is a key co-investigator on this grant. Dr. Caffo is co-founder of the SMART research group, long term collaborator of the PI, and one of the most recognized Biostatisticians of his generation. His primary expertise is statistical modeling, computing algorithms and medical image analysis. He has over 100 publications on a variety of statistical topics such as the fMRI, MRI, SPECT, the Monte Carlo EM algorithm, Markov chain Monte Carlo algorithms, simulated annealing and generalized linear mixed models. Dr. Caffo has received a NIH K25 training grant where he has trained in the medical image instrumentation and analysis and won the (USA) Presidential Early Career Award for Scientists and Engineers (PECASE) award. Dr. Crainiceanu and Dr. Caffo have an already well established collaboration with a large number of co-authored papers (30+) and co-advised students (10+) and post-doctoral fellows (5+). Dr. Caffo will help coordinate all of the work for this grant and will assist in mentoring the graduate students and post-doctoral researchers supported by the grant as well as facilitate the research aims of the grant.

John Muschelli, PhD, biostatistician (3 calendar months) is a key member of the SMART research group and an expert on the statistical analysis of neuroimaging data. Dr. Muschelli has worked in neuroimaging for almost 10 years and developed over 25 packages in the R programming language, ranging from full neuroimaging pipelines to backend packages which access application programming interfaces (APIs) for Scopus and NIH Federal RePORTER. He has been the leader in developing a platform for neuroimaging analysis in R called Neuroconductor. As an Associate Scientist at Johns Hopkins University, he is analyzing large datasets of high-dimensional neuroimaging to automatically harmonize data sets and estimate clinically-relevant biomarkers with specific focus on Stroke and Multiple Sclerosis. In addition to research in neuroimaging, he, with Drs. Crainiceanu and Sweeney have also co-developed and taught a Massive Online Open Course (MOOC) on Coursera (Introduction to Neurohacking In R) on neuroimaging analysis that is free and open to the public. Dr. Muschelli will help coordinate the work on this grant and will assist in mentoring the graduate students supported by the grant as well as facilitate the research aims of the grant and lead Aim 4.

Adrian Gherman, MS, computer scientist (3 calendar months) is a PHP web developer/database administrator with 12+ years of experience and excellent design and programming skills. He has developed the SMART website, the SMART tools website and integrated the research ideas and code developed in a web-based, state-of-the-art research tool for MS lesion segmentation. Mr. Gherman is deeply integrated in the SMART group research, maintains a server dedicated to web applications for the group, collaborates with students and investigators, and helps design professional-grade software that is fast, easy to use, and intuitive. Mr. Gherman has proven himself to be an invaluable addition to the team. He served as lead developer for almost all of the projects he worked on, and therefore was involved in every stage of the project from the initial scoping, to delivery and client handover. Mr. Gherman will continue to be a crucial person for deploying state-of-the-art software, as described in the fourth Aim of the grant.

Two Ph.D. students (one at 100% and one at 50% support) will be supervised by Dr. Crainiceanu and the co-Investigators to execute the goals of this grant. Students will spend the necessary time (between 2 weeks and 3 months a year) in the laboratories of Drs. Reich and Calabresi to familiarize themselves with neuroimaging preprocessing and MS research. They will maintain continuous interactions between the members of the team and will be responsible for scientific and methodological developments. They will work on methods development, programming, writing, literature reviews and data processing, as needed. Dr. Crainiceanu will consult with the other members of the team and will

assign the most appropriate team of mentors for each individual student. Mentors will train the student on neuroimaging data processing, as needed, and the students will attend regular biostatistics and MS scientific research group meetings. Students will immerse themselves into the science as well as in biostatistics methods development.

Fringe benefits
Fringe benefits are calculated at 34% for faculty and staff effective July 1, 2021. Students do not receive fringe benefits. The rates are based upon Johns Hopkins University recent DHHS rate agreement dated June 9, 2020.

Supplies
Research supplies are requested in the total amount of $9,000 in the first year and $3,000 per year in years 2-4 to cover costs for acquiring new computers and associated software and hardware.

Travel
Travel funds are requested in the amount of $5,000 per year for years 2-5 to cover the costs related to PI, Co-Investigators, postdoctoral fellow and graduate students to attend annual conferences and informational meetings.

Other expenses
Biostatistics Computing Support Service: The Biostatistics Computing Support Services (BCSS) Account covers the costs of computer support, software, maintenance and supplies essential to Biostatistics faculty successfully collaborating on sponsored research projects. The costs are not otherwise listed as direct costs on this or other grants, nor are they recovered through indirect costs. The BCSS charge is $\$78$ per annual percent unit of effort expended by a Biostatistics faculty member. We have budgeted for 20% effort by Dr. Crainiceanu, (20 * $\$78$ = $\$1,560$), 25% effort by Dr. Muscelli (25 * $\$78$ = $\$1,950$), and 15% effort for Dr. Caffo (15 * $\$78$ = $\$1,170$ each) for a total of $\$4,680$ per year.

Facilities and administration
MTDC, Organized Research On-Campus rate is 63.75% based on approved DHHS Facilities & Administration rate agreement dated June 9, 2020.

Consortium/contractual agreements

This proposal brings together a team that has an exceptional track record of collaborating on projects involving the development of novel statistical methods for large and complex brain imaging data. Drs. Crainiceanu, Caffo, and Muschelli are all part of the Hopkins Team. They have extensive expertise in the development of new statistical methods, statistical computing, data management, brain imaging analysis, and multiple sclerosis. Dr. Reich is a neuroradiologist who directs the Translational Neuroradiology Unit within the Neuroimmunology Branch of the National Institute of Neurological Disorders and Stroke. Because he is an NIH researcher, Dr. Reich will not draw support from this grant. Dr. Reich will collaborate closely with the PI to ensure data access and quality. Dr. Reich will continue to provide access to the NIH MS MRI data and will help train the PhD and post-doctoral students in MS neuroimaging research, as needed. Dr. Reich will collaborate closely with the PI and Dr. Calabresi to ensure data access and quality. Dr. Sweeney is an assistant professor in Division of Biostatistics in the department of Population Health Sciences at Weill Cornell Medicine. Her expertise is in the statistical analyses of magnetic resonance-based imaging data, including image segmentation, image normalization and harmonization, cross-sectional and longitudinal modeling, as well as software development. Dr. Leroux is an assistant professor in the Colorado School of Public Health, University of Colorado. Dr. Leroux will help coordinate the work on this grant and will assist in mentoring the graduate students and post-doctoral researchers supported by this grant as well as facilitate the research aims of this grant.

Dr. Crainiceanu (PI) will take the lead role in establishing the research groups, setting the schedule for accomplishing the sequence of tasks over the five years, and coordinating the groups on this schedule. Dr. Crainiceanu will be responsible for ensuring the success of the proposal.

Dr. Sweeney (co-I) will participate in research associated with all Aims with special emphasis on Aims 2 and 4. She will advise or co-advise students working on the specific aims of the grant.

Dr. Leroux (co-I) will participate in research associated with all Aims with special emphasis on Aims 1, 2, and 3. Dr. Leroux will advise or co-advise students working on the specific aims of the grant.

Dr. Reich (co-I) will not draw support from this grant, but will be involved in all Aims of the research with particular emphasis on providing guidance on biomarker development and interpretation. He will co-advise the PhD and post-doctoral students, especially in terms of neuroimaging training, access to data, and scientific guidance. Dr. Reich will help co-advise students and will host them in his lab, as needed. This collaboration builds on existent long term collaborations between the PI and Dr. Reich.

Dr. Calabresi (co-I) will participate in research associated with all Aims of the grant, will provide scientific support and guidance and will help co-advise students working on the specific aims of the grant. This collaboration builds on existent long term collaborations between the PI and Dr. Calabresi.

Dr. Caffo (co-I) will participate in all phases of research and will help co-advise students and coordinate the work on the grant.

Dr. Muschelli (co-I) will be involved in all Aims of the research and, especially, in Aim 4. Dr. Muschelli is the most prolific author of Neuroimaging packages in R and the co-leader of Neuroconductor. He will advise and co-advise PhD and post-doctoral students. The highly integrated collaboration between Dr. Muschelli and the PI has resulted in multiple funded grant applications and is a model of collaboration.

Project narrative.
The project provides statistical analysis methods for quantification of the evolution in the intensity of brain lesions on multi-sequence Magnetic Resonance Imaging (MRI). Methods are motivated by the need to develop new neuroimaging-based biomarkers for multiple sclerosis (MS), but can be applied to other types of brain diseases including stroke, Alzheimer disease, and cancer.

Introduction to Resubmission Application. We would like to thank the reviewers and the BMRD section for characterizing our application as "highly significant for multiple sclerosis research" and for highlighting the " significance of the ... sample size estimation methods for ... longitudinal multi-modal neuroimaging and software development". The previous sub aim D.1.1 is now published as a the paper "Fast Univariate Inference for Longitudinal Functional Models" in the Journal of Computational and Graphical Statistics. This highlights the high quality and relevance of this former sub aim. In the new version of the application we have added an entire new sub aim (D.1.2) to replace and complement the published one. We would also like to thank the reviewers for their insightful comments and recommendations. We have addressed them in detail, which substantially improved our proposal. This resulted in roughly two new pages of text, which is bracketed and in italics. The grant was methodologically strengthened by: (1) publishing on estimation and prediction of marginal GLMMs (former aim 1.1); (2) extending methods to non-Gaussian random effects, multivariate functional data, and data with nested and crossed designs (new aim 2.1); (3) comparing our methods with state of the art approaches using simulations (both in aims 1 and 2). Below we answer individual comments.

Resume and summary discussion. Unjustified efficiency and computational feasibility of employing bootstrap methodology; benchmarking against published methods. We have now published a paper [10] that shows that the bootstrap works extremely well in high dimensional applications. We have added extensive simulation studies indicating that the proposed methods are: (1) as accurate as state of the art approaches; (2) two orders of magnitude faster; and (3) applicable to high dimensional data. Conducting a comparison with benchmark approaches was actually an enormous undertaking, as the state of the art methods are extraordinarily slow and do not scale up. Indeed, it took more than 8.6 years of computation time (equivalent time on a personal laptop) to conduct simulations for the well established competing method (joint survival and longitudinal modeling) in Aim 2. **Not specifying a procedure for the landmark selection.** In response to this concern, we have added a procedure that uses an increasingly refined set of quantiles of the distribution of event times and a sensitivity analysis to how fine the quantile grid is. We also note that our methods are general and can work with any set of landmarks. In fact, our Aim 2.2 proposes the super-landmark functional Cox, which borrows strength across the landmarks. **Absence of theoretical support for the use of Bonferroni correction.** We apologize for the misunderstanding, which was likely due to our imperfect wording. We do not use the Bonferonni correction, which is not appropriate for functional data. We have corrected this in the current version of the proposal. **Limited sample size and number of longitudinal measurements.** These are the largest available longitudinal data sets in MS. The sizes of our data sets are comparable to longitudinal studies published in the statistical literature. Moreover, through our collaboration, the data sets will continue to grow in size and scope.

Additional comments/Reviewer 1. Lack of details on the models and evaluation. Our new paper [10], new simulation results, and software implementation details in Aim 2 address this problem in this version of the proposal. **Superiority of methods/Compare to benchmark procedures.** We now compare our methods with state of the art approaches using simulations; please see details in the response to the summary discussion. **Increase support for Muschelli and Gherman.** Done. **Additional comments/Reviewer 2.** Comments addressed; please see previous responses. **Additional comments/Reviewer 3. Minor extensions of existing methods.** We apologize for not explaining better the breakthrough contributions of our proposal. We now provide extensive evidence, both methodological and computational, about the superiority of our models; please see our answers to the summary discussion. **Insufficient attention to establishing statistical properties.** We have conducted simulation studies with a scope that far exceeds what is currently available in the literature (please note our discussion of simulation times for the competing joint survival and longitudinal models). We propose to expand these simulation studies to quantify the finite sample properties of our approaches. This is possible only because of the superior computational performance of our methods; indeed, for competing methods simulations are difficult or impossible to conduct in high dimensions.

A Specific Aims. The previous three funding cycles of this grant have anticipated the explosive rise in longitudinal functional and imaging data. Indeed, our research group has become the leader in the development of methods and software for the analysis of longitudinal high dimensional brain imaging data. Among other accomplishments, this research led to the development of: (1) intensity normalization techniques that have become part of multiple analytic pipelines; (2) automatic cross-sectional and longitudinal segmentation of multiple sclerosis (MS) lesions; (3) biostatistical methods for high-dimensional longitudinal studies of imaging; (4) a web-based imaging tools repository with $7,384$ unique users from 119 countries; and (5) Neuroconductor, the first reproducible, open-source platform for neuroimaging using R, with 115 packages and $18,495$ users from 135 countries.

While imaging studies are widely used in clinical practice and research, the number of neuroimaging-based biomarkers remains small. This lack of established neuroimaging-based biomarkers is, at least in part, due to the historical lack of a comprehensive, principled, reproducible neuroimaging analytic platform. The platform we have developed, Neuroconductor, provides the methodological and computational infrastructure for addressing this gap in the literature. We propose to focus on biomarker discovery and validation based on: (1) longitudinal changes in white matter integrity captured by Diffusion Tensor Imaging (DTI) across the corpus callosum; and (2) longitudinal changes in the voxel intensities in multi-sequence Magnetic Resonance Imaging (MRI) of lesions before and after lesions are identified. These two imaging modalities are likely to capture complementary, previously unavailable, information on MS disease dynamics. Both data structures require realistic biostatistical methods that can scale up and produce principled inference for longitudinal high dimensional data. To address this pressing need we propose the following framework.

Aim 1. Develop inferential methods for longitudinal brain imaging data. **1.1** Introduce fast inferential approaches for multiple spatially varying parameters and random effects; [**1.2** *Extend methods to non-Gaussian random effects and errors, multivariate functional data, and data with nested and crossed designs*;] **1.3** Develop methods for sample size estimation for complex longitudinal imaging studies.

Aim 2. Introduce a class of models for dynamic prediction of longitudinal brain imaging data. **2.1** Develop the landmark historical functional Cox model; **2.2** Develop the super-landmark historical functional Cox model; **2.3** Introduce methods that account for spatial dependence.

Aim 3. Quantify and validate the association with health outcomes. **3.1** Quantify the association between longitudinal measures of microstructure of intracranial white matter tracts and cognitive disability; **3.2** Quantify the dynamic association between the total lesion volume process, lesion voxel intensities and time to voxel recovery; **3.3** Validate methods using reader studies and subset analyses.

Aim 4. Develop software and research deliverables. **4.1** Maintain, develop, and enhance Neuroconductor; **4.2** Develop a curated, online, open access collection of analytic datasets; **4.3** Develop a massive online open course (MOOC) on Coursera dedicated to longitudinal neuroimaging analysis. **4.4** Deploy methods as web-based applications and executables.

Aim 1 provides novel solutions to a longstanding class of problems in statistics and quantifies longitudinal changes in measures of brain microstructure and structural connectivity. Aim 2 introduces novel solutions for quantifying the dynamic processes associated with MS lesions during the course of the disease. Aim 3 introduces a class of novel methodological approaches for validation of biomarkers identified in Aims 1 and 2. While our scientific problem is focused, the methods are general and can be applied to many longitudinal neuroimaging studies in many disease areas, including the ADNI, AIBL, HBC, and MISTIE, where our methods could be used to study changes in lesions or in white and gray matter image intensities.

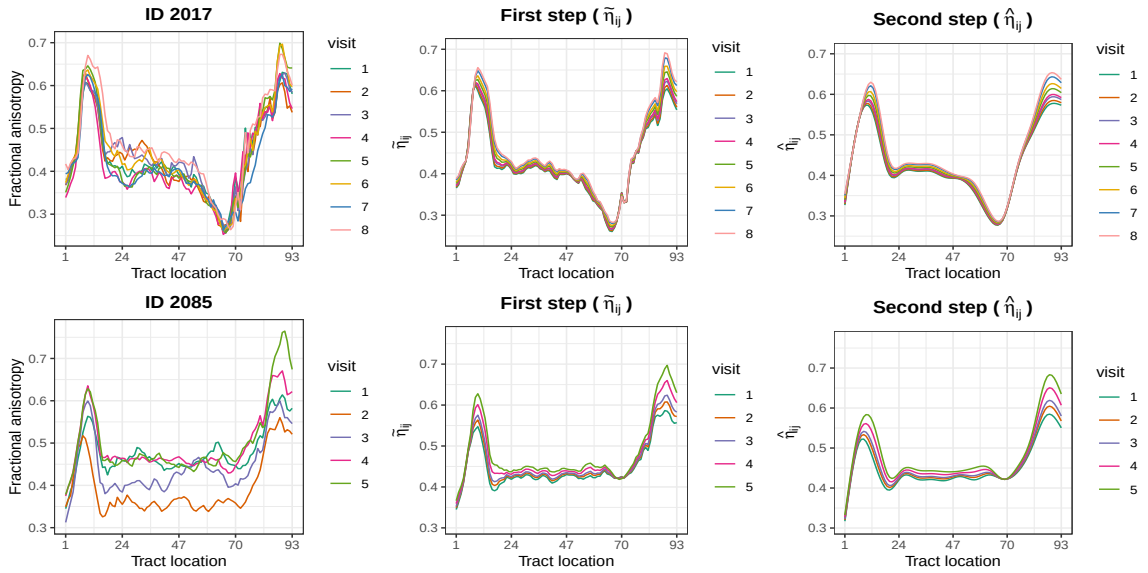

Figure 1: Fractional anisotropy (FA) shown for two study participants (displayed per row). The x-axis in every panel corresponds to numbered locations along the corpus callosum and the y-axis represents the FA. Each line (shown in a different color) corresponds to a different visit. Left panels: observed data. Middle panels: data smoothed using location-specific random intercept random slope models. Right panels: trajectories from the middle panels smoothed over locations.

B Significance and progress report. Multiple Sclerosis (MS) is an immune-mediated disease of the central nervous system (CNS) that affects an estimated $400,000$ people in the United States alone [33]. MS is characterized by diffuse white matter demyelination and focal lesions and causes physical and cognitive impairment [20]. While imaging studies are widely used in clinical practice and research, the number of neuroimaging-based biomarkers is small. Thus, we focus on two promising imaging modalities that are likely to capture complementary information on MS disease severity and dynamics: (1) longitudinal changes in white matter integrity captured by Diffusion Tensor Imaging (DTI) across the corpus callosum; and (2) longitudinal changes in the voxel intensities in multi-sequence Magnetic Resonance Imaging (MRI) of lesions.

MS is thought to affect white matter tracts, which are made up of myelinated axons. Axons are the long projections of neurons that transmit electrical signals and myelin is the primarily lipid sheath surrounding the axons, enabling action potentials to be propagated more quickly. DTI is a magnetic resonance imaging (MRI) modality that traces the diffusion of water in the brain. Because water diffuses with different degrees of anisotropy in different brain tissues, DTI indirectly measures the integrity and directionality of white matter tracts [3, 4, 11, 22]. Several voxel-level summaries of water diffusion can be obtained from processing DTI, including fractional anisotropy (FA) and mean diffusivity (MD). FA represents a measure of fiber integrity and directionality while MD is equal to the magnitude of diffusion [5]; these measures have been widely studied in MS [1, 19, 27]. Continuous summaries of white matter tracts, parameterized by distance along the tract can be derived from DTI.

By collecting longitudinal DTI data, researchers hope to better understand the relationship between MS and disability. Our first data set consists of 100 study participants, 66 women and 34 men, aged between 21 and 70 years at first visit. The number of visits per study participant ranged from 2 to 8, with a median of 3, and were approximately annual; a total of 340 visits were recorded. At each visit full DTI scans were obtained and used to create tract profiles, accompanied by several tests of cognitive and motor function with scalar outcomes. For example, the paced auditory serial addition test (PASAT) is a commonly used examination of cognitive function affected by MS with scores ranging

between 0 and 60. The two left panels in Figure 1 display the longitudinal FA data measured along the corpus callosum for two study participants, with a range of 0 to 1. The x-axis is the segment number along the corpus callosum, with segments being manually aligned to biological landmarks within- and between-study participants. Every line corresponds to a visit and every point on a line is the average FA in the corresponding corpus callosum segment. Visits are color coded (see legend). For every segment (point on the x-axis) we fit a separate random slope random intercept model without accounting for other covariates. The predictions from these models are shown in the middle of Figure 1. In both graphs the predictions are much closer together than the observed data, indicating that models interpret much of the differences between visits as noise. This is likely reasonable. Indeed, note that visit 2 for study participant 2 (Subject ID 2085, visit shown in green) indicates a very low value of FA in the middle of the function (around 0.35) while the FA values are closer 0.45 at visits 4 and 5, which is biologically implausible over a year. The right panels in Figure 1 display a smooth version of the results from the middle panels. Figure 1 indicates that post-fitting smoothing along locations in the brain can improve, at least visually, the overall fit. In this proposal we will show how to conduct inference for such models while accounting for the correlation across brain locations. To be specific, we will focus on the the baseline and longitudinal changes in measures of white matter integrity as a potential biomarker. We will explore which measure or combination of measures (e.g., FA, MD) and which location or combination of locations (e.g., ventral corpus callosum, upper corticospinal tract) are most associate with MS health outcomes. In Aim 1 we will address modeling of longitudinal DTI data, while in Aim 3 we will propose models for the longitudinal association between longitudinal DTI and cognitive and motor function tests.

In MS clinical trials the most common neuroimaging-based biomarkers are total lesion volume and the number of new and enhancing lesions. These biomarkers are essential, but do not capture the recovery process of lesions, which is thought to diminish in more severe, progressive disease. Moreover, brain lesion volume measured by MRI is not strongly correlated with disability [15, 16, 21] and patients often continue to have progressive symptoms while developing few new lesions. This phenomenon is known as the clinico-radiological paradox [2]. To address this paradox we propose to analyze the longitudinal dynamics of lesion intensities in multiple MRI contrasts and their association with covariates, disease-modifying treatment, and lesion volume processes.

In addition to slow changes in the brain structure over time, MS is characterized by the occurrence of lesions in the brain and spinal tract. These lesions have their own dynamics that could be associated with clinical outcomes and treatment. We have compiled a unique data set from a long term longitudinal brain imaging study conducted at NINDS. Thus, our second data set contains 48 study participants who developed 751 lesions with a total of $182,006$ voxels with an average number of visits of 24 and average time between visits of about a month. What makes this data set exceptional is the large number of MRI scans and the density of these scans, which allow the analysis of lesion dynamics, a process that changes substantially during a few months. This data set was created and analyzed by our research group [34]. The data set building process took us about 5 years as it started with developing methods for intensity normalization [30, 31], registration for clinical MS images [14], and cross-sectional and longitudinal segmentation of new lesions using multi sequence MRI [35, 36]. In collaboration with Dr. Reich (co-I on this grant) we have included some of these components in the current analytic pipeline used at NIH by Dr. Reich. These methodological developments would not have been possible without the previous funding cycles of this grant. In Aim 2 we will address joint modeling of lesion voxel intensity dynamics and time to event (voxel recovery), while in Aim 3 we will introduce methods for quantifying the association between these processes and the observed dynamics of established biomarkers of MS (e.g., lesion volume as a function of time).

Figure 2 displays the FLAIR voxel intensities in one slice of a lesion. Panels A and B display the intensity trajectories (shown in gray) as a function of time (x-axis) for each voxel, where time 0 is the time when the lesion was first identified. Each panel emphasizes a different voxel (A and B,

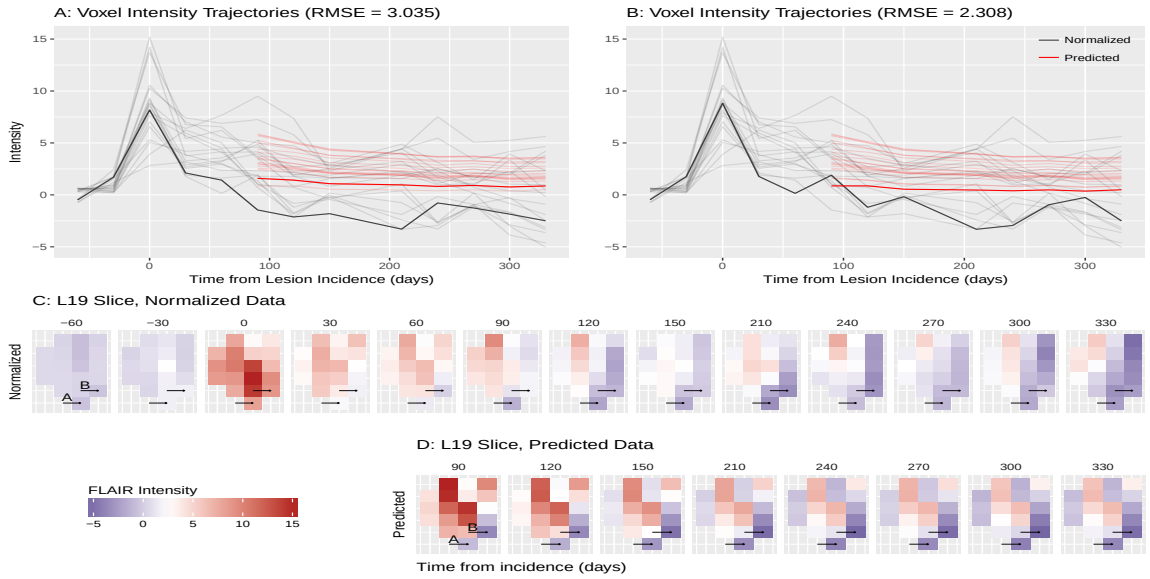

Figure 2: Normalized FLAIR voxel intensities with emphasis on two different voxels (Panels A and B). Intensity trajectories (shown in gray) as a function of time (x-axis) for each voxel. Red lines correspond to predicted trajectories using information up to day 90. Panel C displays the within-slice location of each voxel with voxels A and B identified by small arrows. Panel D provides the predicted intensities.

shown in black). Red lines correspond to predicted trajectories using linear regression models on the past history of the voxel up to day 90. Panel C displays the within-slice location of each voxel with voxels A and B identified by small arrows. Blue and red colors correspond to lower and higher voxel intensities respectively. Panel D provides the predicted intensities based on history up to day 90. Based on these dynamics, we are interested in multiple potential biomarkers: (1) the within-lesion intensity change from baseline to the first time the lesion forms (flare-up amplitude); (2) the rate and shape of the decrease in intensities from the maximum (recovery period); (3) the time from lesion formation to recovery; and (4) number and location of voxels recovered after a given time (e.g., 6 months). All of these biomarkers are calculated at the voxel-level and can be aggregated at the lesion and person level. There are biologically plausible hypotheses that all these biomarkers are associated with the disease severity and treatment. In Aim 2 we will address modeling the dynamics of lesion voxel intensity and their time to recovery, while in Aim 3 we will propose models for the longitudinal association between these processes and established MS biomarkers, including total lesion volume and number of lesions.

Both data examples considered have a longitudinal structure, as data are sampled at multiple times and are high-dimensional. The DTI example is a generalization of traditional longitudinal data analysis, with the distinction that the measurement at a visit is not univariate, but high dimensional (e.g., FA along the corpus callosum). The lesion example is a generalization of the joint longitudinal and time to event data structure, where the object measured is high dimensional (e.g., lesion intensity at every voxel). In both examples, strong, unknown, spatial correlation exists that cannot typically be reduced to parametric forms. The goal is to build a principled biostatistical infrastructure designed to discover and validate imaging-based biomarkers for MS using such novel data structures. Our close collaboration with clinical investigators ensures that this proposal will have a direct impact on clinical trials and practice. We are now in the position to make this translational breakthrough due to the continuous support of the NINDS grant R01NS060910, which has funded methodological research necessary for the analysis of multilevel and multivariate functional data.

Our proposal is significant because it provides an ideal combination of focused scientific

problems and development of appropriate statistical methods. Indeed, we aim to: (1) design a principled analytic platform for analyzing the slow longitudinal changes in brain structure and connectivity; (2) propose computationally feasible inferential methods for the joint analysis of longitudinal high dimensional brain imaging data and time to event; (3) introduce methods for quantification of the longitudinal association between high dimensional brain imaging and health outcomes; and (4) build the analytic and computational infrastructure to support these goals. These new measures promise to provide complementary precision medicine tools that will aid in the management of patients and play a key role in the assessment of new MS therapies.

The success of the current proposal is ensured by the fact that the NINDS grant R01NS060910 has helped us build an unique infrastructure in biostatistics that is well integrated with state-of-the-art neuroimaging research. The support of this grant has also led to several important successes.

First, the grant has supported the creation of one of the most successful and recognized research groups on statistical methods, the "Statistical Methods for Applications and Research in Technology" (SMART) working group. The most important achievement of the group was to become a center for excellence in neuroscience collaboration. Second, our research led to the publication or acceptance of at least 120 papers (40 in the last funding cycle), with an additional 20 papers in various stages of development. The focus of these papers and our research is on the development and application of best statistical methods to the best studies containing population-level neuroimaging data. To make our methodological developments reproducible and usable we have developed Neuroconductor, an open-source platform for rapid testing and dissemination of reproducible computational imaging software. The goals of the project are to: (a) provide a centralized repository of R software dedicated to image analysis; (b) disseminate quickly software updates; (c) educate a large, diverse community of scientists using detailed tutorials and short courses; (d) ensure quality via automatic and manual quality controls; and (e) promote reproducibility of image data analysis. This platform currently contains 115 R packages, with more than 40 developed by our group, analytic pipelines and vignettes. Currently this is the best curated resource for Neuroimaging with R and has more than $18,000$ users from 135 countries. Third, our research group has been at the forefront of development and deployment of Massive Open Online Courses (MOOCs). Dr. Caffo is leading our successful MOOC initiatives, which have so far enrolled millions of students. Drs. Crainiceanu, Sweeney and Muschelli have started the highly successful Coursera course Neurohacking in R (which, to date, has more than $19,549$ learners and $1,089$ course completers). Thus, the SMART working group will continue to be one of the centers of excellence in Neuroimaging education.

Relevance will be pursued by and entrusted to a team with a unique track record of cross-disciplinary translational research. Five biostatisticians (Crainiceanu, Sweeney, Leroux, Caffo, Muschelli) and two Multiple Sclerosis researchers (Reich and Calabresi) will pursue the aims of this grant. Please see letters of support from Drs. Reich and Calabresi. Dr. Crainiceanu will coordinate the entire process, ensure that all Aims will be addressed on time, and foster the collaborative environment. Research will be supported by a computer scientist (Gherman) with more than 12 years of experience and PhD students. The members of the team have a combined total of more than 60 years of research experience, over 300 peer reviewed publications, and more than 10 successful (internal or external) grant applications as PIs. The team members have successfully mentored more than 10 post-doctoral students and over 40 PhD and master students, and have more than 50 papers with at least 2 authors from the team. The team has an optimal combination of statistical and computational skills, neuroimaging and MS scientific experience, and state-of-the art computing to pursue the aims of this grant. [*The exceptional track record of the team is the best argument for the relevance of this proposal.*]

C Innovation. The novelty of the current proposal is defined by the close connection between the scientific questions and statistical methods. Indeed, the biostatistical methods proposed here are in direct response to the need to quantify the slow changes (over many years) in brain structure and connectivity and faster dynamics (over months) of brain lesions and study their association with clinical

outcomes. The scientific innovation of the current proposal is very high. Indeed, quantifying longitudinal dynamics of white matter integrity over time, identifying patterns of change in brain MRI intensities, characterizing the association of these processes with covariates and clinical outcomes, and building a reproducible statistically principled analytic framework are crucial steps towards discovery of MS biomarkers. The integration between the scientific and biostatistical team will help test, exclude, or add potential biomarkers. From a pure statistical methods perspective, the level of innovation is also very high. Indeed, building on our leadership in applied functional data analysis methods and software, we propose novel methods to deal with multivariate, longitudinal spatially correlated functional data. We provide special attention to feasible and scalable solutions that work for very large data sets. This proposal is the first to propose inferential methods for massively univariate generalized mixed effects models and joint high dimensional longitudinal and survival data analyses. All methods are "read and use", meaning they can be used immediately after reading, which will substantially improve their impact both in the statistical literature and in scientific applications. As the novelty of statistical methods is substantially enhanced by high-quality software, we dedicate an entire Aim with detailed milestones for software development and propose to integrate the software developers early on into the research process. These changes make a huge difference in practice and partially account for the success of the SMART research group. While methods are developed and tailored to brain imaging, they can be used in any applications that collect high dimensional data measured at multiple visits.

D Approach. This section describes the basic approach and timeline for the four aims.

D.1 Develop inferential methods for longitudinal brain imaging data. This Aim introduces new methods for the analysis of longitudinal functional data, as described in Figures 1 and 2. These data are longitudinal (because there are multiple visits), multivariate (because we are investigating multiple MRI sequences simultaneously), and high dimensional (because the measurement is the entire brain or a subregion), and sampled at a few or many unequally spaced times (according to the design of the experiment). The structure of the data is $[\{Y_{ijm}(v) : v \in V\}, t_{ij}, \boldsymbol{X}_{ij}]$, where $Y_{ijm}(v)$ could be, for example, the normalized intensity of the $m = 1, \ldots, M = 4$ MRI contrasts of the voxel v of the $i = 1, \ldots, I$ study participant at the $j = 1, \ldots, J_i$ visit recorded at time t_{ij}. The vector $\boldsymbol{X}_{ij}$ contains time-invariant and time varying covariates, including sex, which will be used as a biological variable in all analyses. About twice as many women have MS as men, which is reflected in our data sets.

Given this structure of the data, we are interested in modeling and inference for high dimensional brain imaging data measured at multiple visits. Many advanced modeling approaches for longitudinal high dimensional data have focused on joint modeling, which is extremely computationally expensive, especially when data are not Gaussian (e.g., binary or Poisson). [*To solve this problem, the previous proposal contained sub Aim 1.1, which intoduced the idea of combining massive univariate Generalized Linear Mixed Effects models (MU-GLMMs) with the bootstrap of study participants. This idea is now published in the Journal of Computational and Graphical Statistics [10]. The new method proved to be orders of magnitude faster and as accurate as state-of-the-art approaches [17, 29, 28]. Moreover, the new method is easily scalable to data sets where no other method, including these competing methods, is known to work. We will briefly overview this approach and then introduce the new sub aims, which will naturally build on these novel ideas.*]

[*We provide a short summary of the previous Aim 1.1, which is now published [10]. For*] presentation simplicity we drop the index m. [*Consider the*] MU-GLMM for $Y_{ij}(v)$, where $\eta_{ij}(v) = h[E\{Y_{ij}(v)\}] = \boldsymbol{X}_{ij}\boldsymbol{\beta}(v) + \boldsymbol{Z}_{ij}\boldsymbol{b}_i(v)$, $h(\cdot)$ is a link function, $\boldsymbol{Z}_{ij}$ is the design vector for the random effects, $\boldsymbol{b}_i(v) \sim D_v(b)$ and $D_v(b)$ is a voxel-specific distribution of the random effects. Denote by $\widehat{\boldsymbol{\beta}}(v)$ the estimators of the fixed effects, $\boldsymbol{\beta}(v)$, and by $\widehat{\boldsymbol{b}}_i(v)$ the predictors of the random effects, $\boldsymbol{b}_i(v)$. We focus on one spatially varying parameter, $\beta_k(v)$, but drop the bold font notation and the index k. The fixed effects estimators can be further smoothed, using, for example, $\widetilde{\beta}(v) = S\widehat{\beta}(v)$, where S is any $V \times V$ matrix.

To account for spatial dependence and construct joint confidence intervals, the approach uses a nonparametric bootstrap of study participants. For every bootstrap sample $b = 1, \ldots, B$ we obtain the

estimator $\widehat{\beta}_b(v)$ and calculate $\bar{\beta}(v) = \frac{1}{B}\sum_{b=1}^{B}\widehat{\beta}_b(v)$ and $\widehat{\boldsymbol{S}}^2_\beta = \frac{1}{B}\sum_{b=1}^{B}\{\widehat{\beta}_b(v) - \bar{\beta}(v)\}\{\widehat{\beta}_b(v) - \bar{\beta}(v)\}^t$. The next step is to model $\{\widehat{\beta}_b(v) - \bar{\beta}(v)\}/\sqrt{\mathrm{trace}(\widehat{\boldsymbol{S}}^2_\beta)} = \sum_{q=1}^{Q}\xi_{bq}\phi_q(v) + \epsilon_{bv}$, where the square root and division operations on the left side of the equation are pointwise, $\xi_{iq} \sim N(0, \lambda_q)$, $\widehat{\lambda}_q$ and $\widehat{\phi}_q(\cdot)$ are the eigenvalues and eigenvectors of the correlation matrix of $\widehat{\beta}_b(v)$, and $\epsilon_{bv} \sim N(0, \sigma^2_\epsilon)$ are independent random errors. This can be implemented by functional PCA on the matrix with every row containing the parameters estimated from a bootstrap sample. If $w_{1-\alpha}$ is the $1-\alpha$ quantile of the distribution of $|\sum_{q=1}^{Q}\xi_{bq}\phi_q(v)+\epsilon_{bv}|$, where $|\cdot|$ is the absolute value, then the joint confidence intervals for $\beta(v)$ is $\bar{\beta}(v)\pm w_{1-\alpha}\sqrt{\mathrm{trace}(\widehat{\boldsymbol{S}}^2_\beta)}$. The quantile is obtained using simulations from the distribution of $|\sum_{q=1}^{Q}\xi_{bq}\phi_q(v) + \epsilon_{bv}|$, which can be done in seconds by simulating independently $\xi_{bq} \sim N(0, \lambda_q)$ and $\epsilon_{bv} \sim N(0, \sigma^2_\epsilon)$. [*This approach accounts for the correlation along the functional domain. The two limit cases, Bonferonni correction when $\widehat{\beta}_b(v)$ are independent and pointwise confidence intervals when $\widehat{\beta}_b(v)$ are perfectly correlated, would not be appropriate in most applications*.

[*While simple to implement, the approach is novel and addresses inference in longitudinal imaging analysis, a new problem in biostatistics. Random field theory [37, 38] and permutation testing [32] have been proposed for combining voxel-wise regression analyses for cross-sectional, not longitudinal, high dimensional data. However, recent papers [12, 13] have discredited the finite sample and asymptotic performance of random field theory. Moreover*], outcome permutation is inappropriate when accounting for more than just treatment effect while residual permutation is inappropriate for exponential family data. Our proposal provides a simple, easy to implement, generalizable alternative both for cross-sectional and longitudinal imaging data. Validation is conducted via realistic simulations.

D.1.1 Extend inferential approaches to multiple spatially varying parameters and random effects. In most applications there are multiple fixed effects for which one may want to explore the spatial varying effects. For example, one may want to fit a model at every voxel where the age and treatment effect depend on the particular location in the brain. For simplicity, consider the case when there are two spatially varying covariate effects, $\beta_1(v), \beta_2(v)$ and denote by $\boldsymbol{\beta}(v) = \{\beta^t_1(v), \beta^t_2(v)\}^t$, the $2V \times 1$ dimensional vector of parameters obtained by binding the spatially varying parameters. Currently there are no approaches that account for the correlation between the estimators of $\beta_1(v)$ and $\beta_2(v)$ either in cross-sectional or longitudinal studies. We propose the first solution using the analytic infrastructure introduced in [10] for single spatially varying parameters: bootstrap, smooth, decompose, and simulate.

We start with a nonparametric bootstrap of study participants. For every bootstrap sample $b = 1, \ldots, B$ we obtain the estimator $\widehat{\boldsymbol{\beta}}_b(v)$ and calculate $\bar{\boldsymbol{\beta}}(v) = \sum_{b=1}^{B}\widehat{\boldsymbol{\beta}}_b(v)/B$ and $\widehat{\boldsymbol{S}}^2_\beta = \sum_{b=1}^{B}\{\widehat{\boldsymbol{\beta}}_b(v) - \bar{\boldsymbol{\beta}}(v)\}\{\widehat{\boldsymbol{\beta}}_b(v) - \bar{\boldsymbol{\beta}}(v)\}^t/B$. The difference from [10] is that the correlation operator is $2V \times 2V$ dimensional and incorporates the correlation between the two spatially varying parameters. As the number of spatially dependent parameters increases, the dimension of the covariance operator increases, which leads to substantial increase in computational burden. Our solution circumvents this problem by taking advantage of the low rank of the correlation matrix . More precisely, $\{\widehat{\boldsymbol{\beta}}_b(v) - \bar{\boldsymbol{\beta}}(v)\}/\sqrt{\mathrm{trace}(\widehat{\boldsymbol{S}}^2_\beta)}$ is expressed as $\sum_{q=1}^{Q}\xi_{bq}\boldsymbol{\phi}_q(v)+\epsilon_{bv}$. Conducting inference separately for each spatially varying covariate as in [10] would be equivalent to assuming that $\beta_1(v)$ and $\beta_2(v)$ are not correlated. Here the assumption is relaxed and the joint confidence intervals account for cross-correlations.

We now focus on the spatially varying random effects, $b_i(v)$, and present the approach for the case of one random effect. The generalization to multiple random effects follows the same logic with the one for multiple fixed effects. [*The conditional distribution* $[b_i(v)|\mathrm{Data}]$ *obtained from univariate fits provides the pointwise prediction intervals for the random effects. However, we propose to solve a new problem: constructing the joint prediction for the high dimensional, spatially varying, study participant-specific random effect,* $[b_i(v)]$*. The procedure is similar to the one introduced for fixed effects: (1) bootstrap*

study participants; (2) obtain the quantiles of the maximum absolute values of the absolute values of the z-scores, $\max_v |\{b_i(v) - \overline{b}_i(v)\}|/s_{i,b}(v)$, where $s_{i,b}(v)$ is the standard deviation of the $[b_i(v)|\mathrm{Data}]$ distribution. One problem with this approach is that in some bootstrap samples the study participant i may not be selected and the distribution $[b_i(v)|\mathrm{Data}]$ cannot be computed.] To overcome this problem [*we propose the novel leave-one-in bootstrap*], which keeps study participant i and uses a nonparametric bootstrap of all the other study participants. The novelty of this proposal is multi-faceted. [*First, we have introduced two new problems: how to conduct joint inference on multiple spatially varying parameters and random effects? Second, we have proposed the first approaches for addressing these problems. And third, our methods are computationally feasible, scalable, and intuitive. Indeed, running pointwise mixed effects models is now routine either in a frequentist or Bayesian context. Such models are also easy to interpret and have been widely adopted in the scientific literature. In some sense, we are just saying: if one is comfortable with one mixed effects model, then one should be comfortable with many univariate mixed effects models. A crucial innovation of our proposal is to provide methods for combining the results of these many univariate models using statistical principles without the price of running computationally expensive, possibly unstable, high dimensional functional mixed effects models.*] Here we propose a method that is computationally feasible for high and ultra high-dimensional spatially varying random effects. Moreover, the approach can provide both pointwise and joint prediction intervals for the random effects.

D.1.2 Extend methods to non-Gaussian random effects and errors, multivariate functional data, and data with nested and crossed designs. [*It is standard practice to model the random effects $b_i(v)$ as Gaussian, which is appropriate in many situations. However, in many applications functional data can exhibit substantial skewness or heavy tail behavior. This is the case in our DTI-MRI application in MS. The advantage of using univariate mixed effects models is that we can allow random effects to have non-Gaussian distributions. We propose to explore the family of skew-normal/independent (SNI) distributions, which contains the skew-t, skew-slash, and skew-contaminated normal distributions [18]. While these approaches can be viewed as more exotic, they can be fit relatively easily using Bayesian posterior simulations. The team has extensive experience with Bayesian nonparametric mixed effects models; see, for example, [7, 9]. More importantly, the approaches can capture salient characteristics of the data, such as an individual curve or part of the curve that is unusually far from the population average.*]

[*So far, we have discussed univariate functional data, that is the case when one function is measured at every visit. Alas, most data are multivariate. For example, in our MS application we have both multiple measures of water diffusivity along the tract (multivariate functional data with the same domain) and along different tracts (multivariate functional data along multiple white matter tracts). We propose to fit separate univariate mixed effects models to each modality, and obtain the random effects $b_i^m(\cdot)$ for each modality $m = 1, \ldots, M$, where M is the number of multivariate functional measurements. We propose to describe the correlation between modalities by modeling the joint vector $\{b_i^1(\cdot), \ldots, b_i^M(\cdot)\}$, where each component, $b_i^m(\cdot)$ is a row vector of length equal to the domain of the specific random effect. The simplest approach is to conduct a PCA decomposition of the covariance matrix of these vectors. The most novel component of this approach is that it is the first to conduct mixed effects inference for multivariate structured functional data. While disagreement about the best methods for addressing the problem may exist, it is unarguable that the data structures described in this proposal are here to stay. Indeed, longitudinal, multivariate complex data are now ubiquitous.*]

[*While we have focused on longitudinal functional data, the methods extend easily to functional data with nested and crossed designs. Indeed, the only change required is to use the appropriate mixed effects model. This contribution is novel because it substantially expands the types of applications that can be handled by the new methods.*]

D.1.3 Develop methods for sample size estimation for complex longitudinal imaging studies. So far we have focused on inference for datasets of the same size as the original dataset. In many

applications we may be interested in datasets with the same structure but with with either fewer or more observations. For example, in a study where the null hypothesis of no association, $H_0 : \beta(v) = 0$, cannot be rejected at any spatial location, s, one may be interested in: (1) at what locations s would the null hypothesis be rejected with high probability if the dataset were twice as large? (2) at what sample sizes would the null hypothesis be rejected at all locations in a particular region with high probability? or (3) at what sample size would the global test of no association in a region be rejected with high probability? These are tough questions that do not currently have a solution.

We propose the upstrap [8], which samples with replacement either more or fewer samples than the original sample size. Using this approach we can obtain the joint distribution of spatially varying fixed and random effects for any sample size based on the methods described in Aims 1.1 and 1.2. Once this is done, the sample size can be varied to answer the difficult questions described in the previous paragraph. The simplicity of the approach should not be mistaken for lack of innovation or mathematical complexity. Indeed, sample size calculations cannot currently be done even for standard GLMMs without restrictive assumptions or extensive, often unrealistic, simulations that are limited by number of covariates, mutual dependencies, and distribution of random effects. The upstrap simply avoids all of these methodological traps, by up- or down-sampling the observed dataset. In high dimensional longitudinal brain imaging studies, there are no other methods that can even come close to the range of problems that the upstrap can address.

D.2 Introduce a class of models for dynamic prediction of longitudinal brain imaging data. Figure 2 illustrated the complexity of the problem where we observe hundreds of thousands of voxels within hundreds of lesions over tens of visits. Figure 3 provides the details of the conceptual problem at the voxel level. The two panels contain data for two different voxels, where the x-axis is time from lesion incidence and the y-axis is the normalized voxel intensity. Each dot is a voxel-specific intensity expressed as standard deviation units from the baseline mean. The green shaded area is one baseline standard deviation away from the baseline mean. When the voxel intensity falls within the green area, the voxel is considered to have recovered (to its pre-lesion occurrence intensity). While this problem was motivated by our work on MS lesion recovery, it is actually a well-known methodological problem in biostatistics. Two common analytic approaches for this type of data are joint modeling of the longitudinal and survival processes and landmarking. However, there has been little work dedicated to densely measured time-varying covariates using either approach. Moreover, the software for joint modeling is slow, especially for large datasets, and rather limited for landmarking. Indeed, current implementation of state-of-the-art joint modeling takes about 100 times longer to fit (in simulations) than the methods we propose here for small sample sizes and has serious convergence issues (models do not converge for more than 10% of simulated datasets). For moderate and large sample sizes (e.g., our brain lesion prediction problem) joint modeling does not work. Thus, there is an urgent need to develop methods that easily scale up, can handle densely sampled biomarkers, are stable, and account for complex associations between the biomarker and survival time. Unfortunately, we do need to take a deeper methodological dive to describe the main components of our approach.

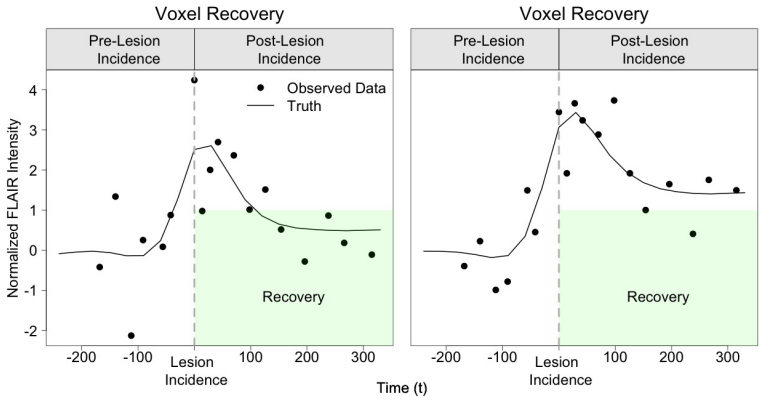

Figure 3: Conceptual problem for two voxels. Each dot is the normalized voxel intensity at one visit. Green area: one baseline standard deviation away from the baseline mean.

D.2.1 Develop the landmark historical functional Cox model. The statistical problem is to make predictions of survival for some time t given data up to time $s < t$. As a result, the time-dependent covariate, $Z_i(\cdot)$, is only partially observed. We propose to address this problem using a combination of landmarking, functional data analysis, and nonparametric smoothing. We start by using a set of landmark times $\mathrm{s} = \{s_1, \ldots, s_L\}$ and for each landmark, s_l, focus on the data for study participants who survived beyond s_l. Here voxels are treated as independent "study participants" and survival is "voxel has not yet recovered". Methods are exetnded to account for spatial correlation in Aims 2.3 and 3.2. We propose the landmark historical functional Cox model

$$\log \lambda_i(t|X_i, Z_i^{\mathcal{H}}(s_l), s_l) = \log \lambda_0(t|s_l) + X_i\beta(s_l) + \int_0^{s_l} Z_i(u)\gamma(s_l, u)du \ , \tag{1}$$

where $s_l \leq t \leq s_l + w_l$ are the landmark times and w_l are the prediction windows for $l = 1, \ldots, L$. The simplest landmark historical model is the "separate landmark model", where a separate regression model is fit to each landmark time. This approach results in L "separate" regression models. Often, it makes sense to borrow strength across models by allowing the parameters $\beta(s)$ and $\gamma(s, \cdot)$ to vary smoothly with s. These "super-landmark models", which combine the L "separate" models, are introduced in Aim 2.2. If we let $\eta_i(s_l) = X_i\beta(s_l) + \int_0^{s_l} Z_i(u)\gamma(s_l, u)du$, then the "separate" landmark model is fit by maximizing the penalized partial log-likelihood (ppl) separately for each landmark time, s_l

$$\text{ppl}(\beta, \gamma, s_l) \ = \ \textstyle\sum_{\{i:s_l<T_i\leq s_l+w_l\}} d_i\{\eta_i - \log(\sum_{T_j\geq T_i} e^{\eta_j})\} - P_{s_l}(\gamma) \ . \tag{2}$$

Equation (2) is the standard formulation of the log partial likelihood [6] minus a penalty term, $P_{s_l}(\gamma)$, on the functional parameter, $\gamma(s_l, \cdot)$. The summation over event times in the interval $(s_l, s_l + w_l]$ implies administrative censoring for event times beyond $s_l + w_l$. If $\gamma(s_l, \cdot) = \sum_{k=1}^{K} \xi_k\phi_k(s_l, \cdot)$ is expanded in a basis (e.g., splines) then the term $\int_0^{s_l} Z_i(u)\gamma(s_l, u)du = \sum_{k=1}^{K} \xi_k \int_0^{s_l} Z_i(u)\phi_k(s_l, u)du$. Thus, without penalization this is a standard Cox model with predictors $\int_0^{s_l} Z_i(u)\phi_k(s_l, u)du$ and associated parameters ξ_k, $k = 1, \ldots, K$. In practice $Z_i(\cdot)$ is not observed continuously and may be measured with noise. Because Z_i is densely measured, we numerically approximate the integral. The basis representation of $\gamma(s_l, \cdot)$ ensures smoothness while the penalty term balances model fit and complexity. While these models are complex, we will show how to implement them using existing software. Model fitting is performed by stratifying on landmark time and is illustrated below. The subtler component, and an important contribution of this proposal, is to use existing software to maximize exactly these likelihoods, even though it was not designed for this purpose. [*In the previous version of the proposal we had an extensive explanation of how the code is implemented. We were asked to reduce the description of the code component and replace it with comparisons with existing methods. We have done that, though we kept the R code while deleting the extensive explanations of individual components. Indeed, the simplicity and reproducibility of our software is a major novel contribution of our approach. At the end of the day, it matters less how complicated is the model if one cannot reproducibly implement it.*]

```
fit_separate <- c() # Create container for separate fits for each landmark
uS           <- unique(data_lm$svec) #Create vector of unique landmark times
for(s in seq_along(uS) {
    #Subset the data to landmark time corresponding to the current iteration
    data_lm_s <- subset(data_lm, svec == uS[s])
    #Fit the separate landmark Cox functional model
    fit_separate[[s]] <- mgcv::gam(formula=T ~ X + s(umat, k=8, by=zlmat),
                                   family=cox.ph, weights=d, data=data_lm_s)}
```

The code loops over landmark times, fitting a separate model to the corresponding subset of the landmark dataset, and can be parallelized. At each landmark time, a separate model is fit via a call to

the `mgcv::gam()` function. Within the call to `gam`, the `formula` argument specifies the outcome and the linear predictor using the syntax `outcome ~ linear predictor`. The outcome is event time `T` and the linear predictor is constructed using the syntax `X + s(umat, k=8, by=zlmat)`, where `umat` and `zlmat` are specific matrices obtained by transforming the survival data in appropriate format. The family argument specifies the likelihood to be maximized. Setting `family=cox.ph` maximizes the Cox partial likelihood corresponding to the outcome, linear predictor, and the event indicator (`weights=d`) using the current subset of the data (`data=data_lm_s`).

[*In response to requests from the reviewers of the first version of this proposal, we have conducted extensive simulation studies. In Aim D.2.2 we provide a summary of our findings.*]

D.2.2 Develop the super-landmark historical functional Cox model. A limitation of the separate landmark models is that some of the models may not be estimable due to lack of information, especially when landmark times are close to the baseline or far into the future. To address this limitation, "super-landmark models" borrow strength across the landmarks by maximizing the following pseudo penalized partial log likelihood (psppl):

$$\text{psppl}(\beta, \gamma, s) \;=\; \sum_{l=1}^{L} \sum_{\{i: s_l < T_i \leq s_l + w_l\}} d_i \{\eta_i - \ln(\textstyle\sum_{T_j \geq T_i} e^{\eta_j})\} - \{P(\gamma) + P(\beta)\}\,. \qquad (3)$$

This penalized log likelihood resembles equation (2), but with several important distinctions. First, the summation is done over all landmark time indexes, $l = 1, \ldots, L$, in addition to the summation over study participants that survived beyond each individual landmark time, s_l. Second, $\beta(s_l)$ is now a function of the landmark times and can be smoothed by adding the penalty $P(\beta)$. Third, the function $\gamma(\cdot, \cdot)$ is modeled as a bivariate smooth function with a bivariate smoothing penalty, $P(\gamma)$. The change from the separate landmark model in equation (2) to the super landmark model in equation (3) is implemented in R as:

```
fit_super <- mgcv::gam(formula = cbind(T, svec) ~  s(svec, k = 8, by = X) +
                         s(smat, umat, k = 30, by = zlmat),
                       family = cox.ph, weights = d, data = data_lm)
```

The code above builds on the ideas described in the fitting of the separate model, but uses a slightly different syntax. Fitting the landmark super model does not require looping over landmark times; instead, model fitting is done using the whole landmark data set. The outcome, now expressed as `cbind(T, svec)`, fits a Cox model stratified on landmark time; each landmark time is allowed to have its own baseline hazard. Syntactically, this implements the summation over landmark times in the log-likelihood presented in equation (3). The linear predictor has changed to allow both $\beta(s_l)$ and $\gamma(s_l, u)$ to vary smoothly across landmark times. First, `s(svec, k=8, by=X)` creates a spline basis expansion of `svec` (landmark time) multiplied by `X`, resulting in $X_i \sum_{k_x=1}^{8} \xi_{k_x} \phi_{k_x}(s_l) = X_i \beta(s_l)$ being added to the linear predictor. A second derivative penalty on $\beta(s_l)$ is automatically added to the log-likelihood. Second, consider the historical functional term, created using the code `s(smat, umat, k=30, by=zlmat)`.

[*We compared our methods with state-of-the-art joint modeling [26]. Software for estimating joint models and performing dynamic prediction is available in*] `R` [23] via the `JM` [24] and `JMbayes` [25] packages. [*These methods were compared both with the landmark and super-landmark historical functional Cox models. We have considered* 18 *different simulation scenarios and compared methods in terms of discrimination accuracy, calibration, and computational efficiency. The most striking findings were that joint modeling: (1) is extraordinarily computationally intensive even for simple models and moderately sized data (hours to days for fitting one data set even for the simplest models); (2) does not work for data sets of the size and complexity found in our applications; (3) fails to converge in a large proportion of data sets (around* 10% *in our simulations); and (4) only works with simple parametric links between the survival and longitudinal process. In contrast, our landmark semiparametric models are orders of magnitude faster, perform similarly to joint modeling when the joint model is correctly specified and substantially out-performs it when it is not.*]

[*Answering this particular request required an enormous computational effort. Indeed, the total simulation time was* 8.4 *computation-years (time estimated to run on a standard laptop), over* 99% *of which was devoted to fitting and predicting the state-of-the-art joint models. This could not have been done without the exceptional computational resources (computing cluster) available at Johns Hopkins. It also emphasizes why the state-of-the-art approaches have not been evaluated in realistic simulation scenarios. Indeed, such studies are computationally onerous. Our newly proposed methods are roughly* 100 *times faster for cases when state-of-the-art models can be fit. They also scale up well to much larger data sets, where competing methods cannot be used. Thus, the newly proposed models are not just new, but also much more practical than current state-of-the-art methods.*]

[*The choice of landmarks is an open statistical problem. To address this problem, we propose to use an increasingly refined set of quantiles of the distribution of event times and a sensitivity analysis to how fine the quantile grid is. We also note that our methods are general and can work with any set of landmarks, while Aim 2.2 borrows strength across landmarks.*]

D.2.3 Introduce methods that account for spatial dependence. To incorporate spatial dependence let $N(v)$ be the spatial neighborhood of the voxel v. In addition to $Z_i(v)$ we now have $Z_i^n(v)$, where $n \in N(v)$ and the integral $\int Z_i(u)\gamma(s,u)$ is replaced with $\sum_n \int Z_i^n(u)\gamma_n(s,u)$. This functional regression with multiple predictors is implementable using methods described in Aims 2.1 and 2.2. We will also consider simpler versions, where we replace $Z_i^n(v)$ with a summary (e.g., mean, standard deviation) as well as by considering more complex models of the type $\sum_n \int F\{Z_i^n(t),t\}dt$.

D.3 Quantify and validate the association with health outcomes.

D.3.1 Quantify the association between longitudinal measures of microstructure of intracranial white matter tracts and cognitive disability. In addition to longitudinal DTI measurements, several cognitive and motor function tests were conducted at each visit. For example, the paced auditory serial addition test (PASAT) is a commonly used examination of cognitive function with scores ranging between 0 and 60. Thus, at every visit we have a health outcome, W_{ij}, in addition to $[\{Y_{ij}(v) : v \in V\}, t_{ij}, \boldsymbol{X}_{ij}]$, where $Y_{ij}(v)$ could be, for example, the FA at the voxel v of the $i = 1,\ldots,I$ study participant at the $j = 1,\ldots,J_i$ visit recorded at time t_{ij}. The vector $\boldsymbol{X}_{ij}$ contains time-invariant and time varying covariates. To quantify the association between longitudinal health outcomes and brain imaging we propose the MU-GLMM $h\{E(W_{ij})\} = X_{ij}\boldsymbol{\beta}(v) + Y_{ij}(v)\boldsymbol{\gamma}(v) + Z_{ij}\boldsymbol{b}_i$, where $h(\cdot)$ is a link function, $\boldsymbol{\gamma}(v)$ is the confounder-corrected longitudinal association between the brain image and health outcome, and all other components are the same as in Aim 1. For inference we propose the same structure as described in Aim 1 including voxel-wise GLMM fitting, potential smoothing across $v \in V$, bootstrap of study participants, and functional PCA decompositions across bootstrap samples. For inference on the random effects we propose the leave-one-in bootstrap approach and for sample size calculation we propose the upstrap. While the methods are similar, models and interpretation is fundamentally different. Indeed, the inferential solution is completely new in the literature. Quantifying the dynamic relationship between health outcomes (as measured by cognitive or physical function test) and white matter integrity (as measured by DTI-MRI) is a fundamental problem in MS in particular and in neurodegenerative diseases in general. Here we propose the first fully integrated, statistically principled, and computationally feasible approach for conducting this type of analysis. As for novelty, we are not aware of any other alternative that is feasible in this context.

D.3.2 Quantify the dynamic association between the total lesion volume process, lesion voxel intensities and time to voxel recovery. In Aim 2 we have focused on modeling time to recovery in a lesion voxel. We are now interested in connecting the voxel recovery process and a dynamic measure of disease burden or health outcomes. An example of such measure is the total lesion volume at every time, say $V_i(t)$. To study how the dynamics of the lesion volume affects the time to voxel recovery we can add the term $\int_0^{s_l} V_i(u)\delta(s_l,u)du$ to the landmark historical functional Cox model (1) or its supermodel version. Methods described in Aim 2 apply to this model completed with smoothing and inference. We use two approaches for incorporating improving prediction: (1) add the neighbors of the

voxel v in the models; (2) incorporate a spatial random effect; and (3) add lesion- and person-specific random effects. To study how lesion dynamics impact lesion volume, denote by $R_{ip}(t)$ $p = 1, \ldots, P$ a collection of P summaries, such as mean, percent recovered, or percentile over voxels in all the lesion of study participant i. We consider models of the type $E\{V_i(t)\} = \boldsymbol{X}_i\boldsymbol{\beta}(t) + \sum_{p=1}^{P}\int_0^t R_{ip}(u)\gamma_p(u)du$. Inferential approaches follow the same structure with what we described in Aims 1 and 2. Quantifying the dynamic relationship between health outcomes (as measured by cognitive or physical function test) and lesion recovery (as measured by white matter MRI intensity) is a fundamental problem in MS. Here we propose the first fully integrated, statistically principled, and computationally feasible approach for conducting this type of analysis. As for novelty, the only methodological alternative we are aware of is the joint longitudinal and survival modeling. Unfortunately, current implementations of this alternative approach are extraordinarily slow even for small sample sizes and completely fail for moderate and large sample sizes. Our proposed alternative is both statistically principled, orders of magnitude faster, and scalable to very large datasets.

D.3.3 Validate methods using reader studies and subset analyses. We propose to use internal validation via leave-one-out, k-fold, and upstrap validation. Upstrap would sample smaller subsets, fit the models and then apply the fitted model to the remainder of the data. Model prediction and fit measures will then be plotted as a function of the upstrap sample size. We will also use reader studies to check whether results are compatible with a human observer looking at the same data. For example, in the case of lesion recovery, we will present longitudinal images of lesions and ask the readers to rate how well lesion recovery predictions perform. We will not conduct reader studies for longitudinal DTI-MRI images because changes are too subtle to be observed by the human eye.

D.4 Develop software and research deliverables. Our group has extensive experience with open source, web-based and executable software development and dissemination. We are the only group that can develop the tools we propose below in the time and with the resources of one RO1 grant.

D.4.1 Maintain, develop, and enhance Neuroconductor. Our research group developed Neuroconductor, an open-source platform for rapid testing and dissemination of reproducible computational imaging software with the following goals: (1) provide a centralized repository of R software dedicated to image analysis; (2) disseminate quickly software updates; (3) educate a large, diverse community of scientists using detailed tutorials and short courses; (4) ensure quality via automatic and manual quality controls; and (5) promote reproducibility of image data analysis. As of October 2020, Neuroconductor hosts 115 packages with an average increase per year above 20% and has $17,872$ users from 135 countries. By developing R packages dedicated to longitudinal data analysis, we will provide Neuroconductor with an adavantage that no other platform has: a statistically principled way for dealing with high-dimensional longitudinal data using computationally feasible software.

D.4.2 Develop a curated, online, open access collection of analytic datasets. We propose to make available high quality curated longitudinal neuroimaging datasets. This will serve two purposes: (1) introduce a large number of biostatistical and computational experts to longitudinal neuroimaging research; and (2) make analytic tools and high quality datasets available to neuroimaging researchers.

D.4.3 Develop a massive online open course (MOOC) on Coursera dedicated to longitudinal neuroimaging analysis. Dr. Crainiceanu has already led one successful MOOC on Neurohacking in R with almost $19,549$ enrolled in the course and 1089 course completers as of October 2021. We propose to develop and deploy a MOOC to disseminate the analytic methods developed in this proposal.

D.4.4 Deploy methods as web-based applications and executables. We propose to deploy algorithms for the analysis of longitudinal neuroimaging data as web-based executables. Our web-based platform currently has $7,384$ users from 119 countries.

D.5 Timeline. Aim 1.1, 1.2: Years 1-2; Aim 1.3: Years 2-4; Aim 2.1, 2.2: Years 1-4; Aim 2.3: Years 3-5; Aim 3.1, 3.2: Years 2-5; Aim 3.3: Years 3-5; Aim 4.1: Years 4-5; Aim 4.2: Years 1-3; Aim 4.3: Years 4-5; Aim 4.4: Years 1-5.

Human subjects. This project will utilize data sets that have already been collected under the direction of Drs. Daniel Reich and Peter Calabresi, who are co-investigators on this project. Data collection, storage and analysis are protected under Drs. Reich's and Calabresi IRBs. The patient data for this project comes from retrospective reviews of archived longitudinal MRI data sets of patients with multiple sclerosis. The project involves working with highly processed MRI data: DTI along neuronal tracts and lesion intensities. Given the selection criteria, the samples that will be analyzed will have roughly the same proportions of women and minorities as the original samples (please see the attached Inclusion Enrollment Reports). We expect a larger proportion of women in the sample as MS affects disproportionately women, but the sample will include men, as well. Data will include minorities, roughly in the same proportion with the original study. Data for patients under 18 years of age are not included. The original brain MRIs are archived on NIH and Johns Hopkins Medical School computers and storage media as well as the medical records pertaining to patient outcome related to MS. We will use only processed MRI data and de-identified information. The analytic team will not use any personal identifier.s

Potential Risks. With any study involving the storage of patient records, there is the potential risk of patient information becoming available to unauthorized persons.

Adequacy of Protecting Against Risks.
a. Recruitment and Informed Consent: Patient scans are selected according to our inclusion criteria from the clinical data archived at NIH in Dr. Reich's lab and at JHMI in Dr. Calabresi's lab. Data were collected with informed consent that includes retrospective analyses of brain imaging studies.
b. Protection Against Risk: Patient data files are coded and the codes are kept in a locked office. Statisticians and students who will work with the coded MRI data files have completed IRB and HIPPA training. They cannot identify the patient from the code and do not have access to the code. Data will be provided to the PI of this grant without any identification information. The PI will not have access to the linkage information. All imaging data will be analyzed in NIfTI format with all header and face information removed. The identified data will remain on the NIH servers and will not be available to the analytic team. No identifiable information will be stored on the Biostatistics computer cluster. No member of the analytic team will require or receive access to identified data. De-identified data sets will be provided, as needed by the research teams of Drs. Reich and Calabresi.

Potential Benefits of the Proposed Research to the Subjects and Others. There is no direct benefit to the subjects whose data are reviewed and analyzed. The potential benefit to society as summarized in the abstract outweighs the risk to the subjects.

Importance of the Knowledge to be Gained. Our long-term objective is to improve the care MS subjects. To achieve this goal, we will use the data to understand the dynamics of structural brain connectivity as measured by DTI-MRI and of MS lesions over time and their potential associations with health outcomes and treatment. IRB Approval: Data sets are protected under the IRB for data collection for the studeis conducted by Drs. Reich and Calabresi.

Inclusion of women and minorities.
The patient data for this project comes from retrospective reviews of an archived longitudinal MRI data of patients with multiple sclerosis. The project involves working with MRI summarized data that meet the selection criteria. For the MS lesion data the selection criteria includes study participants whose longitudinal MRIs co-registration passes expert validation, who have at least one scan within 40 days of lesion incidence and at least one scan 200 days after lesion incidence. Selection criteria do not take into account gender or minority information. Thus, we expect the sample that will be analyzed to have roughly the same proportions of women and minorities as the original NIH sample (please see the attached Inclusion Enrollment Reports).

Inclusion of children.
Children were not included in the original MRI study data set. Thus, no children will be included in our data set and analysis.

References

[1] Yaniv Assaf and Ofer Pasternak. Diffusion tensor imaging (dti)-based white matter mapping in brain research: a review. Journal of molecular neuroscience, 34(1):51–61, 2008.

[2] F. Barkhof. The clinico-radiological paradox in multiple sclerosis revisited. Current Opinion in Neurology, 15(3):239–245, 2002.

[3] P. Basser, J. Mattiello, and D. LeBihan. Mr diffusion tensor spectroscopy and imaging. Biophysical Journal, Jan 1994.

[4] P. Basser and C. Pierpaoli. Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor mri. J Magn Reson B, 111:209–219, 1996.

[5] Peter J Basser. Inferring microstructural features and the physiological state of tissues from diffusion-weighted images. NMR in Biomedicine, 8(7):333–344, 1995.

[6] David R Cox. Regression models and life-tables. Journal of the Royal Statistical Society: Series B (Methodological), 34(2):187–202, 1972.

[7] C. Crainiceanu, D. Ruppert, and M.P. Wand. Bayesian analysis for penalized spline regression using WinBUGS. Journal of Statistical Software, 14(14):1–24, 2005.

[8] Ciprian M Crainiceanu and Adina Crainiceanu. The upstrap. Biostatistics, 21(2):e164–e166, 2020.

[9] C.M. Crainiceanu and A.J. Goldsmith. Bayesian functional data analysis using WinBUGS. Journal of Statistical Software, 32(11):1–33, 2010.

[10] E. Cui, A. Leroux, E. Smirnova, and C.M. Crainiceanu. Fast univariate inference for longitudinal functional models. Journal of Computational and Graphical Statistics, 0(0):1–12, 2021.

[11] Le Bihan. D., J.F. Mangin, C. Poupon, C.A. Clark, S. Pappata, N. Molko, and H. Chabriat. Diffusion tensor imaging: concepts and applications. Journal of Magnetic Resonance Imaging, 13(4):534–546, 2001.

[12] A. Eklund, T. Nichols, and H. Knutsson. Cluster failure: why fmri inferences for spatial extent have inflated false positive rates. PNAS, 113:7900–7905, 2016.

[13] A. Eklund, T. Nichols, and H. Knutsson. Cluster failure revisited: Impact of first level design and physiological noise on cluster false positive rates. Human Brain Mapping, 40:2017–2032, 2019.

[14] A. Eloyan, H. Shou, R.T. Shinohara, E.M. Sweeney, M.B. Nebel, J.L. Cuzzocreo, P.A. Calabresi PA, D.S. Reich, M.A. Lindquist, and C.M. Crainiceanu. Health effects of lesion localization in multiple sclerosis: spatial registration and confounding adjustment. PLoS One, 9, 2014.

[15] M. Filippi, D.W. Paty, L. Kappos, F. Barkhof F, D.A. Compston, and A.J. Thompson. Correlations between changes in disability and T2-weighted brain MRI activity in multiple sclerosis: a follow-up study. Neurology, 45(2):255–260, 1995.

[16] J. Furby, T. Hayton, D. Altmann, R. Brenner, J. Chataway, and K.J. Smith. A longitudinal study of MRI-detected atrophy in secondary progressive multiple sclerosis. Journal of Neurology, 257:1508–1516, 2010.

[17] J. Goldsmith, V. Zipunnikov, and J. Schrack. Generalized multilevel function-on-scalar regression and principal component analysis. Biometrics, 71(2):344–353, 2015.

[18] Dagne G. Huang Y. Bayesian semiparametric nonlinear mixed-effects joint models for data with skewness, missing responses, and measurement errors in covariates. Biometrics, 68(3):943–953, 2012.

[19] M Inglese and Maxim Bester. Diffusion imaging in multiple sclerosis: research and clinical implications. NMR in Biomedicine, 23(7):865–872, 2010.

[20] B.C. Jones, G. Nair, C.D. Shea, C.M. Crainiceanu, I.C. Cortese, and D.S. Reich. Quantification of multiple-sclerosis-related brain atrophy in two heterogeneous MRI datasets using mixed-effects modeling. Neuroimage Clinical, 3:171–179, 2013.

[21] L. Kappos, D. Moeri, E.W. Radue, A. Schoetzau, K. Schweikert, and F. Barkhof. Predictive value of gadolinium-enhanced magnetic resonance imaging for relapse rate and changes in disability or impairment in multiple sclerosis: a meta-analysis. Lancet, 353:964–969, 1999.

[22] S. Mori and P.B. van Barker. Diffusion magnetic resonance imaging: its principle and applications. Anatomical record, 257(3):102–109, 1999.

[23] R Development Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2009. ISBN 3-900051-07-0.

[24] Dimitris Rizopoulos. JM: An R package for the joint modelling of longitudinal and time-to-event data. Journal of Statistical Software, 35(9):1–33, 2010.

[25] Dimitris Rizopoulos. The R package JMbayes for fitting joint models for longitudinal and time-to-event data using mcmc. Journal of Statistical Software, 72(7):1–45, 2016.

[26] Dimitris Rizopoulos, Geert Molenberghs, and Emmanuel M.E.H. Lesaffre. Dynamic predictions with time-dependent covariates in survival analysis using joint modeling and landmarking. Biometrical Journal, 59(6):1261–1276, 2017.

[27] Emilia Sbardella, Francesca Tona, Nikolaos Petsas, and Patrizia Pantano. Dti measurements in multiple sclerosis: evaluation of brain damage and clinical implications. Multiple sclerosis international, 2013, 2013.

[28] Fabian Scheipl, Jan Gertheiss, Sonja Greven, et al. Generalized functional additive mixed models. Electronic Journal of Statistics, 10(1):1455–1492, 2016.

[29] Fabian Scheipl, Ana-Maria Staicu, and Sonja Greven. Functional additive mixed models. Journal of Computational and Graphical Statistics, 24(2):477–501, 2015.

[30] R.T. Shinohara, C.M. Crainiceanu, B. Caffo, M.I. Gaitán, and D.S. Reich. Population-wide principal component-based quantification of blood-brain-barrier dynamics in multiple sclerosis. Neuroimage, 57:1430–1446, 2011.

[31] R.T. Shinohara, E.M. Sweeney, A.J. Goldsmith, N. Shiee, F.J. Mateen, P.A. Calabresi, S. Jarso, D.L. Pham, D.S. Reich, and C.M. Crainiceanu. Statistical normalization techniques for magnetic resonance imaging. Neuroimage Clinical, 6:9–19, 2014.

[32] S. Smith and T. Nichols. Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. NeuroImage, 44:83–98, 2009.

[33] National Multiple Sclerosis Society. Estimating the prevalence of MS, 2002.

[34] E.M. Sweeney, R.T. Shinohara, B.E. Dewey, M.K. Schindler, J. Muschelli, D.S. Reich, C.M. Crainiceanu, and A. Eloyan. Relating multi-sequence longitudinal intensity profiles and clinical covariates in incident multiple sclerosis lesions. Neuroimage Clinical, 10:1–17, 2015.

[35] E.M. Sweeney, R.T. Shinohara, C.D. Shea, D.S. Reich, and C.M. Crainiceanu. Automatic lesion incidence estimation and detection in multiple sclerosis using multisequence longitudinal MRI. American Journal of Neuroradiology, 34:68–73, 2013.

[36] E.M. Sweeney, R.T. Shinohara, N. Shiee, F.J. Mateen, A.A. Chudgar, J.L. Cuzzocreo, P.A. Calabresi, D.L. Pham, D.S. Reich, and C.M. Crainiceanu. OASIS is Automated Statistical Inference for Segmentation, with applications to multiple sclerosis lesion segmentation in MRI. Neuroimage Clinical, 2:402–413, 2013.

[37] Keith J Worsley, Alan C Evans, Sean Marrett, and P Neelin. A three-dimensional statistical analysis for CBF activation studies in human brain. Journal of Cerebral Blood Flow & Metabolism, 12(6):900–918, 1992.

[38] Keith J Worsley, Sean Marrett, Peter Neelin, Alain C Vandal, Karl J Friston, and Alan C Evans. A unified statistical approach for determining significant signals in images of cerebral activation. Human brain mapping, 4(1):58–73, 1996.