Research Statement
John Muschelli

My research philosophy is to solve relevant problems. Many times, those problems require applying existing or developing new statistical methods. Other times, porting methods developed for a certain area of research to another can open new avenues of analysis. My role has been primarily to perform or develop analytic methods, develop the tools (i.e. software) to enact those methods, and convey the results of those methods in a clear, succinct manner. Moreover, I believe the analysis should be reproducible so that others can use the tools or results for future work.

When choosing projects, I find that those that can directly improve quality of life, patient care, or outcomes to be the most fulfilling. Hence, I have focused on projects rooted in a well-defined clinical problem. My interests include brain imaging, complex data, impact of data preprocessing, and software development.

Brain Imaging   I believe there exists a plethora of clinical imaging that is underutilized. When a researcher is viewing an individual patient's image, they cannot adequately describe population-level metrics about the imaging done. Adding the population-level information about diseases, such as stroke, give additional important to researchers and clinicians. There is less methodological work in clinical imaging compared to research imaging. Most importantly, there is a need for easily-used software and interpretable outputs for clinical researchers to use.

Hemorrhagic Stroke   My clinical work has been primarily been involving a Phase II (MISTIE) and Phase III (CLEAR) clinical trial of hemorrhagic stroke. My first duties involved creating reports for the data safety monitoring board (DSMB). By implementing an automated reporting framework, I reduced the turnaround time for the report from 2 weeks to being able to generate daily reports with new data. This automated lead to an iterative process of revision, resulting in a higher-quality report.

While analyzing the endpoint outcomes, we observed the well-known relationship between stroke volume and likelihood of positive functional outcome. I observed the measurements for stroke volume either were fast and coarse or time-consuming and accurate. The accurate method relied on human tracings on slices of a computed tomography (CT) scan to determine stroke location and size.

Using these manual segmentations in "Quantitative Intracerebral Hemorrhage Localization", we developed a population-level map describing where strokes occurred in the trial population, which had been previously described only with coarse measurements and not spatially. We also determined anatomic regions of the brain that relate to stroke severity scores; having a stroke in these regions indicates worse stroke severity.

To make these analyses possible on a larger scale, we developed a statistical model to automatically estimate the probability a given voxel was stroke. I implemented the entire process: image conversion, image pre-processing, image analysis, and presentation of results. One crucial step was to separate the brain from the rest of the image (such as bone, face, etc.). In "Validated automatic brain extraction of head CT images", we adapted a previously published method for brain extraction used for magnetic resonance imaging (MRI) for CT scans, which was suggested previously but never fully validated as we did.

Multiple Sclerosis (Current Work)   As we used simple statistical methods to segment a hemorrhagic area from a CT scan, we used a T1-weighted and T2-weighted MRI to segment areas of gadolinium-enhancing lesions in patients with multiple sclerosis (MS). These specific lesions are markers for increased inflammation of a lesion and are clinically relevant. I compared

2 intensity-based normalization schemes and developed a series of features to use in modeling. The end goal is to use this segmentation in clinical trials as endpoints and disease markers.

Software Development I have been programming in R for the last 8 years. As such, I realize that methods are rarely used unless a well-documented code or a package has been created. I have created a series of R packages and contributed to many others so that others can employ and compare methods our group has developed. My most significant contributions have been the fslr (presented in "fslr: Connecting the FSL Software with R"), brainR ("brainR: Interactive 3 and 4d Images of High Resolution Neuroimage Data") and extrantsr (`https://github.com/muschellij2/extrantsr`) packages.

The fslr package allows R users to call FSL, a widely used and tested neuroimaging software, directly from R. This allows users to process their imaging within the same framework they did analysis. Users also do not need to learn the FSL-dependent syntax of each command.

The extrantsr package create convenient wrapper functions for users to preprocess structural neuroimaging data within R, calling heavily upon the ANTsR package (`https://github.com/stnava/ANTsR`). The ANTsR package is also a re-implementation of a widely used and tested neuroimaging software, advanced normalization tools (ANTs). The extrantsr package allows users to preprocess their data with simple and standardized functions with a limited amount of code.

Future Research I am currently validating the regions that relate to higher stroke severity in an independent cohort of similar patients with stroke. We also are gaining access to thousands of CT scans from patients with stroke and will test our automated segmentation to that data.

My future work is determining a general procedure for preprocessing structural imaging data, normalizing image intensities, and extracting features for image segmentation. The goal would be to apply the same principles to different diseases and conditions, such as brain tumors, to determine if the framework is generalizable. Using this information and clinical information, potential biomarkers can be determined for different diseases.