2016

Integrated Structural And Functional Biomarkers For Neurodegeneration

Benjamin Michael Kandel
University of Pennsylvania, bkandel@seas.upenn.edu

Follow this and additional works at: https://repository.upenn.edu/edissertations
Part of the Biomedical Commons, Computer Sciences Commons, and the Neuroscience and Neurobiology Commons

Recommended Citation
https://repository.upenn.edu/edissertations/2377

This paper is posted at ScholarlyCommons. https://repository.upenn.edu/edissertations/2377
For more information, please contact repository@pobox.upenn.edu.
Integrated Structural And Functional Biomarkers For Neurodegeneration

Abstract
Alzheimer’s Disease consists of a complex cascade of pathological processes, leading to the death of cortical neurons and development of dementia. Because it is impossible to regenerate neurons that have already died, a thorough understanding of the earlier stages of the disease, before significant neuronal death has occurred, is critical for developing disease-modifying therapies. The various components of Alzheimer’s Disease pathophysiology necessitate a variety of measurement techniques. Image-based measurements known as biomarkers can be used to assess cortical thinning and cerebral blood flow, but non-imaging characteristics such as performance on cognitive tests and age are also important determinants of risk of Alzheimer’s Disease. Incorporating the various imaging and non-imaging sources of information into a scientifically interpretable and statistically sound model is challenging. In this thesis, I present a method to include imaging data in standard regression analyses in a data-driven and anatomically interpretable manner. I also introduce a technique for disentangling the effect of cortical structure from blood flow, enabling a clearer picture of the signal carried by cerebral blood flow beyond the confounding effects of anatomical structure. In addition to these technical developments in multi-modal image analysis, I show the results of two clinically-oriented studies focusing on the relative importance of various biomarkers for predicting presence of Alzheimer’s Disease pathology in the earliest stages of disease. In the first, I present evidence that white matter hyperintensities, a marker of small vessel disease, are more highly associated with Alzheimer’s Disease pathology than current mainstream imaging biomarkers in elderly control patients. In the second, I show that once Alzheimer’s Disease has progressed to the point of noticeable cognitive decline, cognitive tests are as predictive of presence of Alzheimer’s pathology as standard imaging biomarkers. Taken together, these studies demonstrate that the relative importance of biomarkers and imaging modalities changes over the course of disease progression, and sophisticated data-driven methods for combining a variety of modalities is likely to lead to greater biological insight into the disease process than a single modality.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Bioengineering

First Advisor
James C. Gee

Second Advisor
Brian B. Avants

Keywords
Alzheimers, Biomarkers, Multivariate, Neuroimaging

This dissertation is available at ScholarlyCommons: https://repository.upenn.edu/edissertations/2377
Subject Categories
Biomedical | Computer Sciences | Neuroscience and Neurobiology

This dissertation is available at ScholarlyCommons: https://repository.upenn.edu/edissertations/2377
INTEGRATED STRUCTURAL AND FUNCTIONAL BIOMARKERS FOR NEURODEGENERATION

© COPYRIGHT

2017

Benjamin M. Kandel

This work is licensed under the Creative Commons Attribution NonCommercial-ShareAlike 3.0 License

To view a copy of this license, visit

http://creativecommons.org/licenses/by-nc-sa/3.0/
Alzheimer’s Disease consists of a complex cascade of pathological processes, leading to the death of cortical neurons and development of dementia. Because it is impossible to regenerate neurons that have already died, a thorough understanding of the earlier stages of the disease, before significant neuronal death has occurred, is critical for developing disease-modifying therapies. The various components of Alzheimer’s Disease pathophysiology necessitate a variety of measurement techniques. Image-based measurements known as biomarkers can be used to assess cortical thinning and cerebral blood flow, but non-imaging characteristics such as performance on cognitive tests and age are also important determinants of risk of Alzheimer’s Disease. Incorporating the various imaging and non-imaging sources of information into a scientifically interpretable and statistically sound model is challenging. In this thesis, I present a method to include imaging data in standard regression analyses in a data-driven and anatomically interpretable manner. I also introduce a technique for disentangling the effect of cortical structure from blood flow, enabling a clearer picture of the signal carried by cerebral blood flow beyond the confounding effects of anatomical structure. In addition to these technical developments in multi-modal image analysis, I show the results of two clinically-oriented studies focusing on the relative importance of various biomarkers for predicting presence of Alzheimer’s Disease pathology in the earliest stages of disease. In the first, I present evidence that white matter hyperintensities, a marker of small vessel disease, are more highly associated with Alzheimer’s Disease pathology than current mainstream imaging biomarkers in elderly control patients. In the second, I show that once Alzheimer’s Disease has progressed to
the point of noticeable cognitive decline, cognitive tests are as predictive of presence of Alzheimer’s pathology as standard imaging biomarkers. Taken together, these studies demonstrate that the relative importance of biomarkers and imaging modalities changes over the course of disease progression, and sophisticated data-driven methods for combining a variety of modalities is likely to lead to greater biological insight into the disease process than a single modality.
## Contents

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ABSTRACT</td>
<td>iv</td>
</tr>
<tr>
<td></td>
<td>LIST OF TABLES</td>
<td>vii</td>
</tr>
<tr>
<td></td>
<td>LIST OF FIGURES</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>CHAPTER 1 : Thesis Overview</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>CHAPTER 2 : Multimodal Sparse Regression</td>
<td>10</td>
</tr>
<tr>
<td>2.1</td>
<td>Introduction</td>
<td>10</td>
</tr>
<tr>
<td>2.2</td>
<td>Methods</td>
<td>13</td>
</tr>
<tr>
<td>2.3</td>
<td>Results</td>
<td>21</td>
</tr>
<tr>
<td>2.4</td>
<td>Discussion</td>
<td>28</td>
</tr>
<tr>
<td>2.5</td>
<td>Conclusion</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>CHAPTER 3 : Decomposing cerebral blood flow MRI into functional and structural components: A non-local approach based on prediction</td>
<td>36</td>
</tr>
<tr>
<td>3.1</td>
<td>Introduction</td>
<td>36</td>
</tr>
<tr>
<td>3.2</td>
<td>Methods</td>
<td>42</td>
</tr>
<tr>
<td>3.3</td>
<td>Results</td>
<td>52</td>
</tr>
<tr>
<td>3.4</td>
<td>Discussion</td>
<td>65</td>
</tr>
<tr>
<td>3.5</td>
<td>Conclusion</td>
<td>72</td>
</tr>
<tr>
<td>3.6</td>
<td>Supplementary Figures</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td>CHAPTER 4 : Neuropsychological Testing Predicts Cerebrospinal Fluid Aβ in Mild Cognitive Impairment (MCI)</td>
<td>75</td>
</tr>
<tr>
<td>4.1</td>
<td>Introduction</td>
<td>75</td>
</tr>
<tr>
<td>4.2</td>
<td>Methods</td>
<td>77</td>
</tr>
</tbody>
</table>
CHAPTER 5: White matter hyperintensities are more highly associated with pre-clinical Alzheimer’s disease than imaging and cognitive markers of neurodegeneration

5.1 Introduction ................................................. 97
5.2 Methods .................................................. 99
5.3 Results .................................................... 104
5.4 Conclusions .............................................. 114

CHAPTER 6: Conclusions and Future Directions .............................................. 115

6.1 Summary of Work ......................................... 115
6.2 Sparse Regression: Between Inference and Prediction ......................... 116
6.3 Perspectives on Different Sources of Imaging Information ................. 119
6.4 Future Directions in Predictions of Amyloid ................................ 121

APPENDIX ......................................................... 123

BIBLIOGRAPHY ................................................. 123
List of Tables

<table>
<thead>
<tr>
<th>TABLE 1</th>
<th>Regression analysis for prediction of age in training data using eigenanatomy, PCA, and ICA projections</th>
<th>34</th>
</tr>
</thead>
<tbody>
<tr>
<td>TABLE 2</td>
<td>Prediction performance (correlation, mean absolute error, and p-value for correlation) in testing data for predicting age from eigenanatomy, PCA, and ICA projections, and AAL/JHU regions. Predictions using multimodal data outperformed predictions using one modality in every instance, and eigenanatomy projections performed the best overall</td>
<td>35</td>
</tr>
<tr>
<td>TABLE 3</td>
<td>Prediction performance (correlation, mean absolute error, and p-value for correlation) in testing data, using random forests, for eigenanatomy, PCA, and ICA projections, and AAL regions. Predictions using all data were compared to predictions using the variance inflation factor (VIF) regression-selected predictors</td>
<td>35</td>
</tr>
<tr>
<td>TABLE 4</td>
<td>Statistics from linear models plotted in Figure 18. Although the raw and structurally predicted CBF values showed strong trends with age, the trends for the residual CBF was more variable. Residual CBF was strongly associated with age in the hippocampus and precuneus, but less so in superior occipital cortex and the precentral gyrus. This may suggest that there is a lesser degree of functional specialization in the precentral gyrus and occipital cortex than in hippocampus and precuneus throughout adolescence. Slope is given in units of CBF (ml/100g/min) per year</td>
<td>63</td>
</tr>
<tr>
<td>TABLE 5</td>
<td>Summary of demographics, psychometric scores, and imaging data for subjects</td>
<td>83</td>
</tr>
</tbody>
</table>
TABLE 6: Summary of univariate logistic regressions predicting Aβ status from each psychometric test and imaging biomarker. Age, gender, and education level (in years) were included as covariates. All data were scaled before regression to facilitate inspection of regression coefficients.

TABLE 7: Area under the curve (AUC), odds ratios, and positive and negative predictive values predicting Aβ status from biomarkers.

TABLE 8: Table of p-values of AUC's for each variable compared with every other variable (FDR corrected). p-values of less than 0.05 are color-coded to indicate which measure is better: Blue indicates that the test indicated in the row name is better, whereas green indicates that the test indicated in the column name is better.

TABLE 9: AUC values for prediction of Aβ status from cognitive tests when stratifying patients by ApoE ε4 status. Cognitive tests were overall more predictive of Aβ status in ε4 negative subjects than ε4 positive subjects.

TABLE 10: Summary of demographics, psychometric scores, and imaging data for ADNI subjects. Values are reported as mean±standard deviation except for white matter hyperintensities, which are reported as median and interquartile ranges.

TABLE 11: Summary of univariate logistic regressions predicting Aβ status from each psychometric test and imaging biomarker for ADNI subjects. Age, gender, and education level (in years) were included as covariates. All data were scaled before regression to facilitate inspection of regression coefficients. The only variables significant at the \( p=0.05 \) level were ApoE status and log white matter hyperintensity volume.

TABLE 12: Summary of PPMI cohort demographics.
**List of Figures**

| FIGURE 1 | Left: Distribution of ages among subjects. Right: Distribution of income among subjects. |
|          | ![Figure 1](image1.png) |
| FIGURE 2 | From top to bottom: Population-specific template; average cortical thickness image for all subjects; average perfusion image; average fractional anisotropy (FA) image for all subjects. |
| FIGURE 3 | A Sankey diagram showing the univariate relationships between image eigenvectors and age. Thicker links correspond to more highly significant correlations, and correlations with an FDR-corrected p-value of less than 0.05 are not shown. |
| FIGURE 4 | Correlation heatmap between projections on all eigenanatomy eigenvectors for cortical thickness, CBF, and FA. Left: Correlations of projections on eigenvectors before the eigenvector clustering step. Right: Correlations after clustering eigenvectors. Clustering eigenvectors significantly reduces correlation between eigenvectors of cortical thickness and FA, but less so for CBF. |
| FIGURE 5 | Sample eigenanatomy, PCA, and ICA eigenvectors. Positive components are in blue, with negative components in red. Eigenanatomy returns unsigned components, facilitating interpretation of regression coefficients. The spatial localization of eigenanatomy enables testing spatially specific neuroanatomical hypotheses. |
| FIGURE 6 | Age prediction on testing data using eigenanatomy, PCA, and ICA projections, and using AAL regions. Quantitative statistics can be found in Table 2. |
| FIGURE 7 | Eigenanatomy regions retained for prediction of age. Thickness predictors are blue; FA predictors are green. |
FIGURE 8: Age prediction on testing data using eigenanatomy, PCA, and ICA projections, and AAL regions as input to random forests. Random forests achieved lower error than linear models at the expense of model interpretability. Quantitative descriptions of the predictions are found in Table 3.

FIGURE 9: Schematic of predicting perfusion from structural MRI.

FIGURE 10: Graphical abstract of proposed method. Patches are sampled from image in modality 1 (here, T1) and SVD is used to learn optimal features (“eigenpatches”) to describe patches. Patches corresponding to each point in the image are then projected onto the “eigenpatches” to create a representation of the input image in feature space. We then use linear regression to predict the second image (here, perfusion image) from the feature-based description of the first image. This enables us to decompose the perfusion image into a component that is predicted from the structural image and the unique contribution of the perfusion image.

FIGURE 11: Histogram of ages of pediatric population.
FIGURE 12: Synthetic perfusion and anatomical data. Some aspects of the perfusion data, such as the higher activity at the intersection of the lines, can be deduced from the underlying anatomy (the intersection of the lines), but other aspects of the perfusion data, such as the increased activity on the upper right line, cannot be deduced from the anatomy. 12c, 12d: Decomposition of synthetic data using non-rotational invariant features. The constructed features include orientation, so the higher values in the horizontal line are correctly reconstructed. 12e, 12f: Reconstructed perfusion and residual perfusion decomposition of Figure 12b. Because orientation invariant features were used, the higher perfusion of the horizontal line is not predicted, but the intersection of the lines does indicate a greater predicted functional signal. Orientation invariance enables greater information sharing across regions, leading to lower variance in the reconstruction as compared to the reconstruction using non-rotational invariant features. 53

FIGURE 13: Negative log $p$-value map (FDR corrected) for raw perfusion (top), structurally predicted perfusion (middle), and residual perfusion (bottom). The structural, combined structural and functional, and functional areas all appear on the perfusion map. The structural and combined functional and structural blob appear on the structural $p$-value map, and the combined structural and functional blob and the purely functional blob appear on the functional $p$-value map. 56
FIGURE 14: Comparison of mean CBF image (top left), reconstruction from anatomy using RIPMMARC (top right), reconstruction from GM and WM probability images (bottom left), and residual perfusion image (bottom right). Mean CBF image is shown at ASL resolution (3.4mmx3.4mmx7.5mm); other images are shown at 2mm isotropic resolution.

FIGURE 15: Predictions of CBF within cortical GM using our structural predictors and probability maps. Our structural predictors account for much more variance than probability maps, which exhibit a strong ceiling effect. This figure indicates that even within the cortex, where additional tissue probability information does not predict CBF, the proposed structural predictors can find a meaningful relation between structure and CBF.

FIGURE 16: Correlation of CBF with: retest CBF; probability images; probability images and thickness; and our structural predictors. Our structural predictors are much better at predicting CBF than probability images, and account for roughly half the reproducible ASL signal. This result indicates that our structural predictors are more appropriate for structural correction of perfusion than using only tissue probability images and cortical thickness.

FIGURE 17: Reproducibility of mean CBF and derived CBF measures. Reproducibility is reported as the voxelwise correlation of the measure at two scans taken one hour apart. Although the reproducibility of the probability-derived CBF prediction is quite high, the low amount of variance explained by tissue probability values as opposed to RIPMMARC-derived predictors indicates that the reproducibility is most likely driven by whole-brain consistency in CBF values.
FIGURE 18: Raw, anatomically predicted, and residual CBF as a function of age. The raw CBF signal contains a mixture of the structurally predicted and residual CBF signals. The residual CBF shows a spatially heterogeneous longitudinal trajectory of functional specialization, with earlier-developing regions, such as the superior occipital lobe and precentral gyrus, showing less change over adolescence than the later-developing precuneus and hippocampus. This relative stability is not apparent in the raw or structurally predicted CBF signal.

FIGURE 19: Assessment of effect of resolution on CBF measurements. Raw CBF measurements are not affected at all by resolution. Both structural and residual CBF are minimally affected by processing resolution: The ROI-wise correlation between structural and residual CBF at 2mm and 3mm resolution are greater than 0.92.

FIGURE 20: Right-sided components of plots shown in 18.

FIGURE 21: Effect of parameter settings on correlation between predicted and true cerebral blood flow (CBF) for a sample subject. Algorithm performance was insensitive to reasonable choices of parameter settings, which included retaining enough eigenvectors to explain more than 50% of variance, using at least 3% of the cortex for training the regressors, and sampling at least 250 points for computing eigenvectors.

FIGURE 22: Boxplots comparing white matter hyperintensity volumes for the ADNI cohort, normalized to intracranial volume and log-transformed. Aβ+ subjects had significantly higher WMH volumes than Aβ- subjects.
FIGURE 23: Boxplots comparing log white matter hyperintensity volumes for the PPMI cohort, normalized to intracranial volume and log-transformed. As in the ADNI cohort, Aβ+ subjects had significantly higher WMH volumes than Aβ- subjects.
CHAPTER 1

Thesis Overview

1.0.1. Clinical Motivation

Alzheimer’s Disease (AD) is a progressive neurodegenerative disease that leads to dementia. As a chronic, irreversible disease, AD presents a significant load on caregivers and family members of affected individuals, as well as the general health care system. AD’s prevalence is estimated at 14% among Americans over the age of 70, and 24% among Americans between the ages of 80 and 89 (Plassman et al., 2007). According to the U.S. Department on Aging, the number of Americans over the age of 65 will increase from 17 million in 2010 to roughly 40 million by 2050, and the proportion of the population 65 and older will increase from 13% in 2010 to over 20% in 2050. The estimated cost of treating these patients in 2014 was $240 billion and will likely rise in the future.

Unfortunately, decades of research and multiple clinical trials (Doody et al., 2014; Doody, 2012; Salloway et al., 2009) have not produced a cure for AD after it has already caused noticeable cognitive decline. These negative results suggest that reversing pre-existing cognitive decline is most likely not a realistic possibility in the near future. Instead, new clinical trials are shifting their focus to preventing and arresting cognitive decline in subjects at risk of developing AD (Sperling et al., 2011b, 2014). Because of the key
role presymptomatic and early-stage dementia patients play in these clinical trials, finding accurate and robust biomarkers that can establish the stage of disease and likelihood of progressing to clinical AD has become a significant trend in AD research (Dubois et al., 2016; Teipel et al., 2015). The clinical interpretability of these biomarkers depends in turn on the extent to which they correspond with the etiology and pathophysiology of early-stage AD.

Although the etiology of AD is not known, progression from normal aging to AD is associated with multiple quantifiable abnormalities, commonly referred to as “biomarkers.” Biomarkers have been identified that are sensitive to multiple disease domains: MR-derived measurements of cortical thickness and cerebrospinal fluid (CSF) tau and hyperphosphorylated tau are markers of neurodegeneration; CSF measurements and PET scans can show evidence of abnormal levels of cerebral amyloid; FDG-PET and arterial spin labeling MRI identify regions of diminished neuronal metabolism and perfusion (Jack et al., 2013); white matter hyperintensities show evidence of small vessel disease (Kester et al., 2014); and cognitive tests can identify specific areas of diminished cognitive function. The temporal ordering of these various manifestations of AD is complex and still not a totally settled matter, but it is clear that the relative importance of the various modalities changes over the course of AD. For example, some studies have demonstrated that Aβ accumulation predates AD by five to ten years, but that once threshold values are reached, cerebral Aβ does not increase further later on in the disease process (Buchhave et al., 2012). Additionally, cerebrovascular disease may play a role early in the disease process (Barnes et al., 2013; Kester et al., 2014). As the disease progresses from risk factors to actual brain damage, neurodegenerative markers such as τ (Buchhave et al., 2012) and cortical thickness (Davatzikos et al., 2011) become more evident, as do measurable impairments in cognitive function. Because of the heterogeneity of disease manifestation over time, studies encompassing a variety of imaging modalities and clinical measurements over time are necessary to provide a complete picture of the longitudinal progression of the disease.
Although some of these tests, such as CSF measurements, provide absolute scalar measurements, others, such as structural MR images, are extremely high-dimensional and require specialized processing before they can be combined with other biomarkers. Because these various biomarkers are direct or indirect markers of a neurodegenerative process, there is significant overlap in the information they provide. Assessing the relative contributions of the various imaging biomarkers is technically challenging, as the high-dimensional, spatially structured nature of image data makes many traditional statistical techniques not viable. In addition, mainstream AD research has focused on neurodegenerative and amyloid biomarkers, sometimes at the expense of cognitive testing and vascular markers. This thesis both proposes technical methods for multivariate analysis of imaging data and also demonstrates that cognitive testing and white matter hyperintensities, which have not been as thoroughly investigated as other biomarkers, are at least as highly associated with AD pathology as more conventional imaging biomarkers.

1.0.2. Gaps in Current Knowledge and Techniques

The plethora of biomarkers available for tracking the progression of AD raises several important methodological issues. Because of the potential for overlapping information between distinct biomarkers, even those that probe different pathological cascades, evaluation of any one biomarker must take place in the context of the various other biomarkers that are also available. Evaluating the amount of information shared between high-dimensional imaging and low-dimensional clinical measurements is difficult. The level of complexity involved in combining imaging information with scalar or low-dimensional clinical measurements depends on what technique is used to analyze the images.
1.0.2.1. Statistical Frameworks for Multi-Modal Medical Image Interpretation

There are three main analytical frameworks for using medical images to gain biological insight. The first reduces the information present in an image to one or two values by using methods specified by a priori knowledge of the disease or organ of interest. In neuroimaging, this most often takes the form of averaging a biologically relevant quantity, such as cortical thickness (Lerch et al., 2008) or brain metabolism (Landau et al., 2010), over a pre-specified region of interest. Examples from other clinical domains include measurement of maximal jet velocity for assessing the severity of aortic stenosis (Saikrishnan et al., 2014). These methods have two principal advantages. First, the measurements are directly clinically interpretable and correspond to known biological phenomena. Second, because the measurements are scalars, they can easily be combined with other clinical information in an algorithmic decision-making process. At the same time, reducing neuroimaging data to one number makes the measurement sensitive to exactly how the extraction of data was done. For example, if cortical thickness is averaged over a predefined region, the resulting measurement will depend on how exactly the region is defined. Precise delineations of the boundaries of many brain regions are not settled (Devlin and Poldrack, 2007; Bohland et al., 2009; Yushkevich et al., 2015). In addition, the mapping of anatomical labels from a manually labeled template brain to a subject brain is sensitive to subtle issues of bias in the selection of templates and normalization to the templates (Avants et al., 2010; Yushkevich et al., 2010).

A second method for drawing scientific inferences from medical imaging is univariate statistical testing. There are several varieties of this form of testing, depending on what exactly is being compared, but one of the most popular techniques is “voxel-based morphometry”, or VBM (Ashburner and Friston, 2000). VBM works by registering many scans to a common stereotactic space, smoothing the resulting images, and then performing voxel-wise statistical tests comparing the segmented gray matter probability between control and test groups. From the outset, VBM was designed as a research methodology applied to
groups of subjects and not as a source of individual-specific diagnoses. As such, the primary question that VBM answers is “across many people, is this voxel correlated with a condition of interest?” This is distinct from the question that biomarkers seek to answer, which is “judging from this person’s test, is she likely to have a disease of interest?” Although it is possible to derive a region of interest from a VBM study and use that region in a separate test population (Hirata et al., 2005), such an approach will likely be less successful than multivariate approaches (Davatzikos, 2004) and is in any case not what VBM was designed to do (Friston and Ashburner, 2004).

The third major group of methods for analyzing brain images falls under the general category of multivariate approaches. This group covers a wide range of techniques, but shares the practice of combining multiple voxels or brain regions to predict a result. Brain function is increasingly recognized as organized into interconnected networks (Bullmore and Sporns, 2009), and multivariate techniques are better able to account for this level of complexity than univariate methods. The recent explosion in interest in machine learning has produced multiple techniques suitable for dealing with high-dimensional data like medical imaging, and many of them have been adapted for neuroimage analysis. However, these methods were originally intended to predict an outcome from an input data source, not to derive scientific insight into the input data. In addition, combining multiple sources of data and still obtaining meaningful inference about individual brain regions is inherently complicated. As such, even visualizing the results of these techniques is not straightforward (Golland et al., 2005; Sabuncu, 2014) and principled interpretation of these studies is the subject of debate (Davis et al., 2014; Haufe et al.; Rosa and Seymour, 2014).

Neuroimage analysis for AD, which causes changes detectable by multiple imaging modalities, presents special challenges. Region of interest (ROI)-based methods, by reducing the information present in an image to a few scalar measurements, make combining multiple modalities into a single analysis amenable to traditional statistical analysis. On the other hand, univariate voxel-based techniques do not accommodate multiple modalities, and
the difficulties inherent to multivariate techniques are compounded when more than one modality is considered. Therefore, a need still exists for a method for incorporating multiple modalities of images into one analysis in an interpretable framework.

1.0.2.2. Assessing the Degree of Independence Between Related Imaging Modalities

Another important question in the context of multimodal imaging in AD research is the extent to which different imaging modalities contain overlapping information. As a concrete example, arterial spin labeling (ASL) imaging (Roberts et al., 1994) measures brain perfusion. Although ASL has been investigated as a possible AD biomarker (Wolk and Detre, 2012), recent research has suggested that there is more information shared between structural and ASL imaging than had previously been recognized (Franklin et al., 2013; Salgado-Pineda et al., 2006). These studies showed that changes in measured cortical thickness have been observed after acute interventions. Because the number of neurons present, the principal biological meaning of cortical thickness, is unlikely to increase after an acute intervention, it is likely that changes in cerebral blood flow are reflected in the observed changes in cortical thickness. At the same time, highly influential AD researchers have recently shifted from positing that perfusion and metabolic changes precede anatomical changes (Jack Jr et al., 2010) to saying that perfusion and anatomical changes happen concurrently (Jack et al., 2013). The mixing of signal between the anatomical and perfusion modalities may at least partially explain the finding that anatomical and perfusion imaging show abnormalities at the same stage of disease. Finding the information that is unique to ASL imaging and unrelated to T1 may highlight pathological processes that are primarily detected as changes in perfusion, but not visible, or at least not yet visible, as changes in anatomy.

An additional reason to investigate the degree of information shared between ASL and T1 images is that each additional MR modality acquired in a clinical scan adds to the
time and cost associated with MR, so paring down the number of modalities to the minimum necessary is imperative. In addition, implementing new MR modalities in clinical settings nationwide or world-wide requires a significant investment in standardizing scanning protocols and establishing clear guidelines for interpreting the images. Therefore, thoroughly investigating the importance of information in ASL that is orthogonal to the more standard T1-weighted anatomical images is an important step in justifying its collection as an additional modality in clinical practice. In addition, the amount of blood flow not explained by observed anatomic variability may provide a unique insight into regulation of blood flow and blood flow supply in cortical tissue.

1.0.2.3. Associations Between Biomarkers and AD Pathology in Cognitively Normal and Mild Cognitive Impairment Subjects

Beyond these technical questions, significant questions remain regarding which biomarkers or clinical tests are most highly associated with AD pathology. Although whether or not amyloid-beta (AB) has a causal role in AD is not entirely clear (Joseph et al., 2001), presence of AB has been accepted as a necessary component of preclinical AD (Sperling et al., 2011a). The ability of various biomarkers to predict conversion to clinical AD has attracted the lion’s share of attention from AD researchers (Davatzikos et al., 2011; Gomar JJ et al., 2011). Although this goal is important, examining the degree of association with AB as distinct from ability to predict clinical AD is informative for a few reasons. First, the clinical definition of probable AD is determined in part by the results of cognitive tests. Therefore, predicting cognitive decline using the same cognitive tests as those used to establish the existence of cognitive decline is somewhat circular. Recent studies have suggested that cognitive tests are at least as sensitive as imaging biomarkers to predict conversion to clinical AD (Ewers et al., 2012). Because of the potential circularity of the study design, it is difficult to disentangle the strength of association between cognitive decline and pathological processes from the association between present and future cognitive decline. The association
between cognitive testing and molecular markers of AD pathology such as AB is not subject to the same degree of circularity. Therefore, directly examining the association between AB and cognitive testing may give a clearer indication of the relationship between AD pathology and cognitive decline than would be obtained by predicting future AD conversion from cognitive testing.

Another area in which evaluating the association between multiple modalities and AD pathology is important is in preclinical AD. The relationship between preclinical AD, as defined by presence of AB in the absence of cognitive decline (Sperling et al., 2011a), and traditional AD biomarkers such as cortical thickness appears to be more tenuous than in patients with clinical AD or mild cognitive impairment (MCI) and different studies have found conflicting results (Whitwell et al., 2013; Dickerson and Wolk, 2012). Small vessel disease, as quantified by the presence of white matter hyperintensities (WMH), has been reported to be correlated with AD pathology (Kester et al., 2014), and signs of vascular disease correlate with presence of AD, especially in the early stages of the disease (Attems and Jellinger, 2014). Nevertheless, the strength of association between WMH and AD pathology in cognitively normal subjects has not been compared to that of cortical thickness and other, more traditional AD biomarkers. WMH are an important complement to other AD biomarkers because they are a marker of cardiovascular health, as opposed to neuronal integrity. A strong correlation between WMH and AD pathology could indicate that at least at the earliest stages of disease, cardiovascular stress plays an important role in initiating the pathologic cascade.

1.0.3. Primary Contributions

In this thesis, I present both technical and clinical studies that address the challenges and open questions outlined above. Specifically, the major contributions of this thesis are:
1. A framework for sparse regression that integrates multi-modal imaging into an anatomically and scientifically interpretable model.

2. A novel method that demonstrates that significantly more signal is shared between arterial spin labeling (ASL) imaging and T1-weighted anatomical imaging than previously known, and that both the shared signal and the signal unique to ASL are tightly correlated with age in a pediatric cohort.

3. Evidence that in mild cognitive impairment (MCI), cognitive testing is as accurate as diagnosing AD pathology as multivariate imaging biomarkers are.

4. A study suggesting that white matter hyperintensities, a marker of small vessel disease, is a more sensitive marker of AD pathology in the preclinical stage than current standard AD biomarkers are.

Taken together, the clinical studies presented provide evidence that different modalities and sources of information are important for diagnosing AD at different stages in the disease process, and the technical innovations provide tools for conducting clinically interpretable multi-modal imaging studies.
CHAPTER 2

Multimodal Sparse Regression

2.1. Introduction

Modern imaging datasets are increasingly multimodal. Virtually all modern large-scale imaging studies, even those that concentrate on a given modality, such as resting state fMRI (Biswal et al., 2010), include a variety of imaging measures (Mueller et al., 2005; Satterthwaite et al.). Although some groups have reported improvements in classification accuracy in Alzheimer’s Disease when using multimodal data (Zhang and Shen, 2011), others have claimed that multimodal classification does not tend to outperform a single sensitive test (Landau et al., 2010). This trend towards multimodal data presents challenges in data processing, visualization, and statistical inference. In particular, the extremely high dimensionality of medical imaging data presents challenges to classical linear model-based statistical analyses, which assume that there are more subjects than measured variables \((n > p)\). Several approaches exist to deal with the high-dimensional nature of medical imaging datasets.
2.1.1. Mass-Univariate Approaches

One of the most widely used methods to perform statistical analyses on medical images is to use voxel-based morphometry (VBM) \cite{Ashburner2000}. VBM performs a statistical test on each voxel in the image, producing a spatial map that describes how closely the values at a given voxel are correlated with an outcome measure. The massive number of multiple comparisons conducted when using VBM necessitate appropriate corrections \cite{Nichols2003}. In addition, because brain function is spread over regions larger than a single voxel \cite{Haxby2001}, multivariate approaches are more naturally suited to leveraging the spatially distributed information contained in medical imaging data \cite{Davatzikos2004}.

When examining multimodal data, univariate approaches are further restricted because they do not provide insight into the relationships between the various modalities. One way of using univariate approaches to analyze multimodal data is to perform separate mass-univariate analyses on each modality and examine the degree of spatial overlap between the resulting statistical maps \cite{Chen2011a, Tosun2012, Haier2009}. A drawback of this method is that spatial overlap alone does not give insight into the subject-wise interactions or correlations of the various modalities. To take a somewhat extreme example, if half the experimental population have increased cortical thickness as compared to controls and the other half have increased BOLD activation, a spatial map may show overlapping significant areas, even though no individual subject actually has increased cortical thickness and increased BOLD activation. To provide greater insight into the biological mechanisms underlying observed changes, several studies have begun investigating multivariate approaches to multimodal data \cite{Tosun2012, Anurova2014, Naylor2013}, looking at, for example, the correlation between cortical thickness and BOLD activation in a given region.
One challenge of integrating large multimodal datasets is the difficulty in visualizing and interpreting the results, especially when performing multivariate analyses of data. Interpretation of multivariate data is often made easier by sparse methods, which ensure that only a small part of the data set is used for predicting an outcome variable. Sparse methods have enjoyed a resurgence in popularity in recent years, with several groups proposing sparse methods tuned for neuroimaging data (Michel et al., 2012; Grosenick et al., 2008, 2013, 2009; Kandel et al., 2013; Varoquaux et al., 2012; Sabuncu and Van Leemput, 2012; Batmanghelich et al., 2009, 2011, 2012). Applying sparse techniques to multi-modal data enables specific and biologically interpretable statements to be made about data; for example, “Decreased cortical thickness in the left parietal lobe is correlated with decreased perfusion in the left and right parietal lobes, and this network together predicts a decrease in verbal ability.”

2.1.2. Data-Driven Dimensionality Reduction

Many clinical studies using multimodal imaging data average image values over predefined regions of interest (ROI’s) to reduce the dimensionality of the data so that it will be more amenable to standard statistical analyses. Although this approach may be ideal if the ROI’s are already known and have anatomically meaningful boundaries, this is not ideal for exploratory analyses which have minimal prior knowledge. Traditionally, linear regression from a high-dimensional dataset is performed after a dimensionality reduction step, such as principal component analysis (PCA) (Hotelling, 1957). However, PCA-derived eigenvectors have global support and therefore do not provide anatomical specificity. Sparse techniques can provide more local specificity. In particular, a recently introduced sparse dimensionality reduction technique, “eigenanatomy,” has proven to provide greater power to detect group differences than either voxel-based morphometry (VBM) (Avants et al., 2012a) or pre-defined ROI’s (McMillan et al., 2014) while maintaining anatomical interpretability. Here, we extend the eigenanatomy approach to a multi-modal setting. Although the
sparse eigenvectors are orthogonal in the image space, orthogonality is not enforced on the low-dimensional coefficients generated by projecting the imaging data onto the sparse eigenvectors. Therefore, care must be taken to prevent excessive collinearity among the predictor variables. We demonstrate that even with collinearity in the predictor variables, our method of extending eigenanatomy to multi-modal datasets produces a more accurate prediction of age in a pediatric population than principal component regression, independent component regression, or regression on average values within regions defined by the AAL atlas.

The eigenanatomy objective function is not new to this work. Here, we focus on the practical challenges, including validation, interpretation, and visualization of predictive models, involved in multimodal data analysis, and demonstrate the advantages of the eigenanatomy framework for multi-modal neuroimaging studies as compared to either classical dimensionality reduction techniques or predefined regions of interest (ROI’s). The release of all data and code used to generate the paper will facilitate the use of this technique as a template for future studies, as well as encourage reproduction of similar evaluations with different datasets.

2.2. Methods

2.2.1. Reproducibility

To facilitate the use of this study as a template for other multimodal population studies, we have attempted to make it as reproducible as possible. All the data is available from an open-access data repository. The paper itself is written using the R package knitr [Xie 2013], which facilitates on-the-fly production of figures from data, enhancing reproducibility and documenting all data processing steps. The full code for producing the paper, including raw
data and code for producing figures, is available from [https://bitbucket.org/bkandel/](https://bitbucket.org/bkandel/multimodaleanat)

### 2.2.2. Dimensionality Reduction Techniques

Dimensionality reduction is a technique to reduce the complexity of input data into a relatively small number of summary measures. Linear dimensionality techniques can be written as a matrix factorization problem. We assume the input data is given in an $n \times p$ input matrix $X$, where $n$ is the number of observations or subjects and $p$ is the number of variables associated with each observation. In the context of medical imaging, $n$ typically ranges from a few tens to a few hundred, and $p$ is on the order of $10^3 - 10^6$, depending on the size of images. Dimensionality reduction seeks to find a factorization of $X$ into an $n \times k$ coefficient matrix $U$ and a $p \times k$ loading or eigenvector matrix $V$ so that $X \approx UV^T$. The most well-established method for dimensionality reduction is Principal Component Analysis (PCA), which finds an orthogonal matrix (i.e. $V^TV = 1$) that projects the input matrix to a lower-dimensional subspace. More recently, Independent Component Analysis (ICA, e.g. [Hyvärinen, 1999]) has become widely used in the neuroimaging community. ICA seeks a decomposition of $X$ in which the components are independent, which is a stronger condition than orthogonality.

One drawback of standard PCA and ICA is that the eigenvectors often cover the entire input matrix, meaning that each entry in the coefficient matrix is a weighted average of all the voxels in the image. This makes interpretation of the output difficult for two related reasons. First, the lack of spatial specificity of the eigenvectors makes it difficult to use the coefficients to investigate anatomically-informed biological hypotheses. For example, it is impossible to use the coefficients from a PCA decomposition to look at the relation between left precuneal atrophy and age. In addition, because the eigenvectors have both positive and negative components, interpreting the weights in a linear regression model that relates the coefficients
to an outcome measure is not intuitive. This is for two reasons. First, the PCA eigenvectors can contain negative weights, even if the input data is strictly positive, as is the case in cortical thickness, perfusion, and fractional anisotropy (FA) images. If the weight of a given PCA coefficient in a linear model is positive but the corresponding entry in the eigenvector is negative, it follows that an *increase* in the input matrix corresponds to a *decrease* in the outcome variable. Second, this problem is compounded by the overlapping nature of the eigenvectors: Because a given brain region can contribute positively or negatively to each eigenvector, it is very difficult to go back from the coefficient weights to the biological meaning of the weight. Furthermore, interpreting the eigenvectors themselves without accounting for the weights is ill-advised (Biesmann et al., 2012), as the eigenvectors can be confounded by biases in the data in non-obvious ways (Haufe et al.). Therefore, although PCA may be used for predicting age in unseen data, it is not as useful for testing biological hypotheses.

Sparse dimensionality reduction techniques (Zou et al., 2006; Witten et al., 2009) deal with the problems of global support of PCA eigenvectors by enforcing a sparsity constraint on the matrix decomposition. The sparseness forces the eigenvectors to have only a few non-zero entries, thus making the eigenvectors more amenable to anatomically-specific hypothesis testing. Eigenanatomy (Avants et al., 2012a; Dhillon et al., 2014) augments these constraints with smoothness and cluster thresholds, which ameliorate the inherent instability of sparse solutions (Xu et al., 2012). In addition, eigenanatomy requires the eigenvectors to be unsigned, thereby eliminating the difficulty of interpreting linear model weights: A positive weight in a linear model means that an increase in the predictor variable corresponds to an increase in the outcome variable and vice versa.
Mathematically, the eigenanatomy objective function is

\[
\text{arg min}_{\hat{U}, \hat{V}} \|X - \hat{U}\hat{V}^T\|^2_2 + \text{smooth}(G^*(V)) + \text{cluster}(V)
\]  

(2.1)

subject to \(\|V\|_0 < \gamma\),

\[V \succeq 0\]

where \(\gamma\) is the desired level of sparseness, the \(\ell_0\) “norm” operator \(\| \cdot \|_0\) is the number of non-zero entries in the argument, and \(G^*\) is the manifold in image space representing a geometric object \cite{AvantsGee2003}. Although optimizing over an \(\ell_0\) penalty is an NP-hard problem, thresholded gradient techniques have been proposed for sparsity-constrained optimization problems \cite{Donoho1995}, and we adopt the conjugate gradient version used in \cite{Avants2012a}. We use the open-source eigenanatomy decomposition function \texttt{sparseDecom} as implemented in the \texttt{ANTsR} package \url{github.com/stnava/antsr}.

One challenge with sparse dimensionality reduction techniques is that the projections on the sparse eigenvectors are not orthogonal \cite{ShenHuang2008}, which can lead to collinearity in the regressors. To deal with this problem, we first perform an eigendecomposition using very sparse eigenvectors (50 eigenvectors, each covering 2% of the image), and then use hierarchical graph clustering to cluster eigenvectors into fewer, but larger, eigenvectors. This strategy of oversegmenting into “supervoxels” and then clustering has proven fruitful in image segmentation \cite{Veksler2010}. Eigenvectors which, when the population of images is projected onto them, have correlated values are clustered together into communities. This clustering reduces the degree of collinearity between the eigenvectors. The clustering stops when the graph density of the original eigenvectors reaches a pre-set threshold. We have found that a graph density of 0.10 is a good value for maintaining anatomical specificity and predictive accuracy. We used the \texttt{igraph} package for graph-based community construction \cite{CsardiNepusz}. For PCA, we used the \texttt{princomp} function in R, and we used the \texttt{fastICA} package \cite{Hyvarinen1999} for ICA computations.
2.2.3. Multi-Modal Data Visualization and Integration

One key advantage of sparse dimensionality reduction techniques is that they facilitate incorporation of multiple modalities into a single interpretable multivariate model. In our experiments, we limit ourselves to linear and generalized linear models, as the coefficients and $p$-values of these models are well-established and can be interpreted using mature, standard statistical techniques. The predictor variables for the linear models are the coefficient scores for the various decomposition techniques, obtained by projecting the original image space data onto the eigenvectors to obtain an $n$ by $k$ matrix, where $n$ is the number of subjects and $k$ is the number of eigenvectors.

To choose which of the projections we include in the final linear model relating imaging data to an outcome, we use a model selection. The sparsity induced by model selection complements the sparsity of the individual eigenvectors, allowing multiple eigenvectors to enter the model if they each contribute to the outcome but still maintaining the anatomical interpretability afforded by the spatial sparsity of the eigenvectors. Because imaging data is generally highly correlated between regions, we use the VIF regression (Lin et al., 2011) model selection technique, which explicitly accounts for correlation among predictors. Because the focus of the study is on feature generation and not feature selection, we fix the model selection method and apply it to all decomposition techniques.

Visualization of multivariate regressions is challenging, especially when considering multiple input and multiple output variables. We use the Sankey diagrams from the d3 library for visualizing correlations between input and output variables. These diagrams permit quick and intuitive visual exploration of the correlation between each input and output variable.

Because multivariate models are more prone to overfitting than univariate models, a rigorous training-testing split of data is necessary to assess generalizability of results. Two steps are necessary to interpret an optimized multivariate linear model. First, the regression weights can be interpreted. A cautionary note is in order here: Interpretation of regression
coefficients in multivariate linear regression is not trivial and is subject to confounding factors and multicollinearity artifacts, even when using a model selection technique like VIF that attempts to minimize collinearity between predictors. Specifically, the regression coefficients from a multivariate linear model correspond to the effect of a given regressor keeping all other regressors constant. Second, statistical assessment of the significance of the model can be performed by applying the model to unseen testing data and evaluating the performance of the model. Note that because model selection is performed in the training data, naïvely looking at the $p$-value of the model in the training data is inappropriate and would constitute statistical “double-dipping” unless special care is taken [Berk et al., 2013]. The ability to evaluate a model in unseen testing data is a key advantage to predictive models as opposed to correlation-based approaches and is valuable even if predicting the outcome from neuroimaging data is not clinically useful.

2.2.4. Clinical Data

We use the Pediatric Template of Brain Perfusion dataset, which contains multimodal data for normally developing pediatric subjects. The full pediatric dataset is available at [http://figshare.com/articles/The_Pediatric_Template_of_Brain_Perfusion_PTBP_/923555](http://figshare.com/articles/The_Pediatric_Template_of_Brain_Perfusion_PTBP_/923555). Briefly, the dataset consists of 133 subjects, 82 females and 83
males, ages 7.07-17.99, mean 12.48 years. Cohort selection aimed to match the demographic
distribution of children in the United States, based on US census data, for race, ethnicity,
gender, and family income. Distributions of age and income with respect to sex are shown
in Figure 1.

2.2.5. Image Acquisition

All MRI experiments were performed on a Siemens 3T TIM Trio scanner. For
the T1-weighted acquisition, Magnetization-Prepared Rapid Acquisition Gradient Echo
(MPRAGE) images were acquired using a 3D inversion recovery sequence with TR/TE/TI =
2170/4.33/1100 ms. The resolution is 1x1x1mm$^3$ with a matrix size of 256x256x192. The flip
angle = 7° and total scan time was 8 minutes and 8 seconds. For perfusion imaging, pseudo
continuous arterial spin labeled (pCASL) images were acquired using gradient-echo echoplanar imaging (EPI) with TR/TE = 4000/12 ms. The resolution was 3.125x3.125x6mm
(5mm thickness with a 1mm gap) over a 64x64x24 matrix. 40 label/control pairs were
acquired. Generalized autocalibrating partially parallel acquisition (GRAPPA) was used
with an acceleration factor of 2. Labeling duration was 1.5s and the post-labeling delay
was 1.2s. Total imaging time was 5 minutes and 30 seconds. Diffusion weighted images
were acquired with single-shot spin-echo EPI with TR/TE = 9500/87ms. A single b=0
volume was acquired along diffusion weighted images for 30 directions with b-value=1000.
The resolution was 2x2x2mm with a matrix size of 128x128x75 voxels, with a flip angle of
90°. Total acquisition time was 11 minutes with two acquisitions.

2.2.6. Image Preprocessing

Our image processing pipeline is based on Advanced Normalization Tools (ANTs) and the
ANTsR package, which provides an interface from the R computing environment to the ANTs
image processing utilities. We used the ANTs template-building scripts to construct an unbiased template from the subject population (Avants et al., 2008). Although template-building methods have traditionally been used for single-modality data, we leverage recent advances in template-building techniques to construct a multi-modal atlas (Avants et al., 2011b; Jain et al., 2012). AAL (Tzourio-Mazoyer et al., 2002) and JHU (Wakana et al., 2004) labels and priors for brain extraction and tissue segmentation are warped from pre-existing template space to the custom template using a multi-atlas label fusion (MALF) approach (Wang et al., 2012). We generated a six-tissue segmentation: Cerebrospinal fluid (CSF), cortical gray matter (GM), white matter (WM), deep GM, brainstem, and cerebellum. Full details and code for constructing multi-modal templates, along with sample data, can be obtained from github.com/ntustison/TemplateBuildingExample.

For each subject’s T1 image, we performed bias correction (Tustison et al., 2010), brain extraction, normalization to the template (Avants et al., 2008), prior-based segmentation (Avants et al., 2011a), and cortical thickness calculation (Das et al., 2009). The T1-based cortical thickness measurement pipeline is based on the antsCorticalThickness.sh script (Tustison et al., 2013). Diffusion processing, including fractional anisotropy (FA) calculation, is performed using Camino (Cook et al.) as incorporated into ANTs (Tustison et al., 2012). pCASL images were processed using a robust regression approach (Maumet et al.; Avants et al., 2012b) that regressed tag from untagged images. The M0 image was obtained by averaging the non-tag images, and was used as a reference for motion correction. Motion correction was performed using the antsMotionCorr function in ANTs, and motion parameters (three for rotation and three for translation) were regressed out as nuisance variables. Physiological nuisance parameters were estimated using the CompCor approach, which estimates physiological noise by computing an eigendecomposition of the high-variance regions in an image (Behzadi et al., 2007). Full details are available in the open-source script at https://raw.github.com/stnava/ANTs/master/Scripts/antsASLProcessing.sh. The blood T1 value was adjusted for age and gender as $T1 =$
2.3. Results

2.3.1. Template and Average Images

The population-specific template and the average of all images warped to the template are shown in Figure 2. As expected, cortical thickness in the medial temporal and medial frontal
is relatively high, and cortical thickness in the occipital lobe and motor cortex is relatively low (Fischl and Dale, 2000). Perfusion in the cortex is higher than in white matter or CSF, and the higher perfusion in the deep gray matter structures is clearly visible, consistent with existing literature (Roberts et al., 1994). The average FA image shows higher FA in the corpus callosum, with lower FA in the gray matter.

2.3.2. Eigenanatomy Decompositions

We obtained eigenanatomy-derived eigenvectors for the populations of cortical thickness, CBF, and FA images. The correlation matrix of all the eigenvectors is shown in Figure 4. Cortical thickness tended to be more correlated with CBF than with FA, and each modality showed strong intra-modality correlations. To help alleviate the collinearity of the predictors, we clustered the eigenvectors to create fewer, but larger, eigenvectors. The correlations between these clustered eigenvectors are shown in Figure 4. Clustering significantly reduced the collinearity between the eigenvectors, especially for cortical thickness and FA. Because of the reduced collinearity, we used the clustered eigenvectors in our regressions.

A sample clustered eigenanatomy-derived eigenvector is shown in Figure 5. The eigenanatomy eigenvector covers primarily the precuneus. On the other hand, the PCA and ICA vectors have global support. The eigenanatomy vector also has only positive values, whereas the PCA and ICA vectors each have positive and negative components.

2.3.3. Univariate Data Exploration

Before looking at the multivariate regression, it is helpful to look at univariate correlations between the eigenanatomy regions and the outcome variable. We use a Sankey diagram
Figure 3: A Sankey diagram showing the univariate relationships between image eigenvectors and age. Thicker links correspond to more highly significant correlations, and correlations with an FDR-corrected p-value of less than 0.05 are not shown.

to show the links between age and the various eigenanatomy regions (Figure 3). Although many regions are correlated with age, not all survive FDR correction.

2.3.4. Regression Analysis in Training Data
Figure 4: Correlation heatmap between projections on all eigenanatomy eigenvectors for cortical thickness, CBF, and FA. Left: Correlations of projections on eigenvectors before the eigenvector clustering step. Right: Correlations after clustering eigenvectors. Clustering eigenvectors significantly reduces correlation between eigenvectors of cortical thickness and FA, but less so for CBF.
Figure 5: Sample eigenanatomy, PCA, and ICA eigenvectors. Positive components are in blue, with negative components in red. Eigenanatomy returns unsigned components, facilitating interpretation of regression coefficients. The spatial localization of eigenanatomy enables testing spatially specific neuroanatomical hypotheses.

We ran a VIF regression-based model selection strategy to select projections for prediction of age from neuroimage data. The model selection, when performed on the decomposition-based methods (Eigenanatomy, PCA, and ICA), retained only thickness and FA projections, but did not retain CBF projections; for the pre-defined ROI’s, thickness, CBF, and FA values were all retained. With all other regressors held constant, increases in thickness in bilateral anterior precuneus were weakly correlated with increased age, and decreases in thickness in bilateral posterior precuneus and cingulate were more strongly correlated with increases in age. Increases in FA in corpus callosum were correlated with increasing age, whereas with all other regressors held constant, decreases in FA in the inferior longitudinal fasciculus were correlated with increases in age. The eigenanatomy eigenvectors are rendered in Figure 7. Anatomically specific descriptions and coefficient interpretation are not possible with PCA and ICA eigenvectors.
Figure 6: Age prediction on testing data using eigenanatomy, PCA, and ICA projections, and using AAL regions. Quantitative statistics can be found in Table 2.
2.3.5. Evaluation on Testing Data

To verify the generalizability of the regression to unseen data, we predicted the age of half the data and evaluated the fit. The eigenanatomy predictors outperformed the PCA and ICA predictors, as well as the AAL/JHU regions (Figure 6). Although all three methods produced a highly significant fit, the eigenanatomy projections gave greater correlation between true and predicted age, lower root mean squared error, a lower $p$-value for the correlation (Table 2), and a slope relating predicted to true age closer to unity.

2.3.6. Comparison to Non-Linear Models

To examine the influence of regression method on prediction accuracy, we constructed a random forest predictor to relate eigenanatomy, PCA, and ICA projections, as well as AAL and JHU regions, to age. We compared the prediction using both the VIF regression-based model selection technique and using all of the projections. Random forests decreased the mean absolute error (MAE) and increased correlation for all methods, but eigenanatomy projections still outperformed PCA and ICA projections and predefined ROI’s. Giving all the data to the random forest resulted in more accurate predictions than giving only the
Figure 8: Age prediction on testing data using eigenanatomy, PCA, and ICA projections, and AAL regions as input to random forests. Random forests achieved lower error than linear models at the expense of model interpretability. Quantitative descriptions of the predictions are found in Table 3.

regions selected using the VIF model selection technique. Predicted vs. true age is plotted in Figure 8 with quantitative values in Table 3.

2.4. Discussion

We have presented here extensions to a sparse dimensionality reduction method that enable biologically interpretable predictions from high-dimensional multi-modal data that requires minimal prior knowledge. We demonstrated how incorporating a variety of techniques, including over-segmentation and then clustering of eigenvectors and variance inflation factor regression, minimize the impact of collinearity of regressors on prediction accuracy, making these methods suited for multi-modal image analysis problems.
2.4.1. Interpretation of Regression

One major advantage of sparse matrix decompositions for multi-modal image analysis is that the individual regressors in a linear model can be anatomically interpreted. Our results here for cortical thickness agree with standard results in developmental neuroscience that cortex generally thins throughout development (Gogtay et al., 2004). The observation that increased FA in the corpus callosum is strongly correlated with increasing age is also in consonant with previous research (Lebel et al., 2008). Similarly, although the FA in the inferior longitudinal fasciculus (ILF) increases with age, it increases at a lower rate than the corpus callosum (Lebel et al., 2008), and so when corrected for the corpus callosum, the regression coefficient becomes negative.

2.4.2. Collinearity of Sparse Eigenvectors

One confounding issue in sparse matrix decompositions is that projections onto eigenvectors are not generally constrained to be orthogonal (Shen and Huang, 2008). This lack of orthogonality can complicate regression from sparse eigenvectors, as the collinearity of the regressors makes the design matrix ill-conditioned and can confound the regression coefficients. Variance inflation factor (VIF) regression (Lin et al., 2011) ameliorates this problem somewhat by explicitly choosing a regressor set that minimizes covariance, but cannot change the pre-existing collinearity among the regressors. Clustering the sparse eigenvectors based on their correlation further helps to minimize this issue, but does not completely resolve it. In particular, the CBF values appeared to be highly collinear, even after clustering the eigenvectors. This collinearity is a result of the greater degree of collinearity in the data, and resolving this issue for CBF specifically is an important issue. Because CBF is generally so correlated across the brain, several groups have divided each subject’s CBF values by the mean of the whole brain CBF to create a relative CBF measure.
Aslan and Lu (2010). Although relative CBF may not be as predictive of age as raw CBF is, it may give better spatial discrimination of developmental trends and so may generate more meaningful spatial information.

### 2.4.3. Sparsity in Multi-Modal Image Analysis

Although sparsity can be a useful penalty in medical image analysis, it can also be overused and misinterpreted. For example, if a signal is not sparse and is instead distributed throughout the brain, enforcing sparsity on a solution can be misleading by implying that only certain parts of the brain contain signal. This is due to the fundamental nature of sparse solutions: By discarding redundant information, they generate a minimal set of information that can be used to predict the outcome—even if other parts of the brain also contain similar information. Put another way, there is currently no standard method for setting the sparsity parameter in a sparse decomposition. The dependence of the solution on the sparsity penalty has long been known (Tibshirani, 2011). Although we do not solve this issue here, clustering the sparse eigenvectors allows the eigenvectors to adapt naturally to a grouping that allows different levels of sparsity for different regions, thereby making the sparsity penalty less arbitrary.

With this caveat, though, sparsity is a useful tool for multi-modal image analysis problems. Because information is virtually always shared between different modalities, the reason to incorporate more than one modality into an analysis problem is to highlight the unique information given by one modality that is not given by another modality, ignoring the shared information between the modalities. This emphasis on extricating the unique information from the various modalities while ignoring shared information between the modalities naturally leads to sparse methods as a way to produce predictive models. In particular, while cortical thickness may measure the number of cells in the cortex, CBF measures cortical perfusion, while FA is a measure of white matter tract integrity. Understanding the unique
componets of these measures may give greater insight into how different neurobiological mechanisms develop and relate to each other.

2.4.4. Linear vs. Non-Linear Models

One important tradeoff when constructing predictive models is the expressive power of the model vs. the interpretability of the output of the model. In particular, non-linear predictive models such as random forests (Breiman, 2001), while achieving state-of-the-art classification accuracy, do not produce directly interpretable models. Although some attempts have been made at interpreting the prediction model, including a variable importance score (Breiman, 2001), the model still does not have the direct interpretation and statistical significance theory attached to linear regression coefficients. The method proposed for selecting a subset of eigenanatomy regions and then feeding those projections into a random forest attains a compromise solution: The predictors used by the random forest correspond to discrete anatomical regions, but the way those regions are combined does not have a simple interpretation. Still, this level of interpretability may be sufficient for some research questions, and gives more insight into the prediction mechanism than feeding all brain regions into a random forest.

2.4.5. Comparison to State-of-the-Art Results

Prediction of age from medical imaging data has drawn a significant amount of attention in recent years (Franke et al., 2010, 2012; Brown et al., 2012; Fjell et al., 2012; Dosenbach et al., 2010). Most of these methods use smooth, non-linear models to predict age from imaging data, so the results from the random forest predictor are most relevant to comparing against this work. Comparing directly against previous work is complicated by a variety of factors. The age range of the subjects can influence the accuracy of the prediction, as the prediction
is generally more accurate for younger subjects. In addition, different methods use different combinations of modalities. Furthermore, the studies that produced the highest accuracy used training datasets of several hundred to nearly a thousand subjects (Franke et al., 2012; Brown et al., 2012), whereas our database included a training set of only 82 subjects. Even with these caveats, though, the random forest achieved a mean absolute error (MAE) of 1.27, which is comparable to the MAE of 1.1 reported in other studies.

### 2.4.6. Drawbacks of the Framework

Although the framework for multi-modal image analysis shows promise as compared to standard dimensionality reduction techniques, it still has some drawbacks. Even with well-behaved data, interpreting coefficients of multivariate linear regressions is not as straightforward as interpreting coefficients of univariate regressions. In particular, because the coefficients of a multivariate regression are corrected for the effects of all the other covariates, careful attention must be paid to the whole model, and each individual regressor cannot be viewed in isolation.

Another drawback of sparse decomposition techniques is the instability of the solutions, as demonstrated by the different regions selected by the AAL/JHU regions and the eigenanatomy decomposition (Table 1). Instability of sparse learning techniques is a fundamental characteristic of sparsity in general (Xu et al., 2012). Although bootstrapping-based techniques have been used to estimate the stability of sparse regression methods (Varoquaux et al., 2012), extending this technique to achieve stable multi-component sparse dimensionality reduction is a significant open problem in the field.
2.5. Conclusion

We have proposed a method for incorporating eigenanatomy decompositions into a multi-modal data analysis pipeline. We have proposed methods for dealing with collinearity of sparse predictors. In addition, we have demonstrated that the sparse eigendecomposition method outperforms traditional PCA and ICA-based dimensionality reduction methods for predicting age from imaging data, while maintaining anatomical interpretability of the regression. This method is suited to extracting the relevant and unique information from multi-modal imaging datasets.
| Method           | Projection      | Estimate | Std. Error | t value | Pr(>|t|) | Main Positions                              |
|------------------|----------------|----------|------------|---------|---------|---------------------------------------------|
| Eigenanatomy     | (Intercept)    | 10.508   | 4.288E+00  | 2.45    | 1.657E-02 |                                             |
|                  | ThickEanatProj01 | 0.547    | 6.618E-01  | 0.827   | 4.110E-01 | Bilateral Anterior Precuneus                 |
|                  | ThickEanatProj03 | -2.729   | 6.399E-01  | -4.26   | 5.711E-05 | Bilateral Posterior Precuneus               |
|                  | ThickEanatProj05 | -1.890   | 8.070E-01  | -2.34   | 2.180E-02 | Cingulate                                   |
|                  | FAEanatProj01   | 48.323   | 6.869E+00  | 7.03    | 7.559E-10 | Corpus Callosum                             |
|                  | FAEanatProj05   | -14.994  | 4.105E+00  | -3.65   | 4.742E-04 | Inferior Longitudinal Fasciculus            |
| PCA              | (Intercept)1    | 12.659   | 2.439E-01  | 51.9    | 1.126E-61 |                                             |
|                  | ThickPCAProj03  | -0.023   | 5.759E-03  | -3.93   | 1.838E-04 |                                             |
|                  | ThickPCAProj04  | -0.022   | 5.962E-03  | -3.74   | 3.495E-04 |                                             |
|                  | FAPCAProj02     | 0.286    | 6.414E-02  | 4.46    | 2.799E-05 |                                             |
|                  | FAPCAProj05     | -0.275   | 7.749E-02  | -3.55   | 6.690E-04 |                                             |
| ICA              | (Intercept)2    | 1.753    | 1.531E+00  | 1.14    | 2.557E-01 |                                             |
|                  | ThickICAProj10  | -0.036   | 8.731E-03  | -4.11   | 9.447E-05 |                                             |
|                  | FAICAProj08     | -0.290   | 7.515E-02  | -3.86   | 2.286E-04 |                                             |
| AAL/JHU labels   | (Intercept)3    | 11.715   | 4.685E+00  | 2.5     | 1.452E-02 |                                             |
|                  | Thickness.AALLabel11 | -2.295   | 8.341E-01  | -2.75   | 7.398E-03 |                                             |
|                  | CBF.AALLabel10   | 0.005    | 1.549E-02  | 0.354   | 7.242E-01 |                                             |
|                  | CBF.AALLabel28   | -0.083   | 2.298E-02  | -3.63   | 5.118E-04 |                                             |
|                  | FA.JHULabel17    | 31.026   | 6.622E+00  | 4.69    | 1.181E-05 |                                             |

Table 1: Regression analysis for prediction of age in training data using eigenanatomy, PCA, and ICA projections.
<table>
<thead>
<tr>
<th>Method</th>
<th>Correlation</th>
<th>MAE</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1   Eigenanatomy</td>
<td>0.772</td>
<td>1.59</td>
<td>1.40e-17</td>
</tr>
<tr>
<td>2   PCA</td>
<td>0.650</td>
<td>1.89</td>
<td>2.98e-11</td>
</tr>
<tr>
<td>3   ICA</td>
<td>0.496</td>
<td>2.16</td>
<td>1.84e-06</td>
</tr>
<tr>
<td>4   AAL/JHU Labels</td>
<td>0.691</td>
<td>1.74</td>
<td>5.07e-06</td>
</tr>
<tr>
<td>5   Eigenanatomy, Thickness</td>
<td>0.598</td>
<td>1.88</td>
<td>2.29e-09</td>
</tr>
<tr>
<td>6   PCA, Thickness</td>
<td>0.585</td>
<td>1.96</td>
<td>6.17e-09</td>
</tr>
<tr>
<td>7   ICA, Thickness</td>
<td>0.406</td>
<td>2.29</td>
<td>1.40e-04</td>
</tr>
<tr>
<td>8   AAL Labels, Thickness</td>
<td>0.221</td>
<td>2.50</td>
<td>4.50e-02</td>
</tr>
<tr>
<td>9   Eigenanatomy, CBF</td>
<td>0.529</td>
<td>2.08</td>
<td>2.66e-07</td>
</tr>
<tr>
<td>10  PCA, CBF</td>
<td>0.585</td>
<td>1.96</td>
<td>6.17e-09</td>
</tr>
<tr>
<td>11  ICA, CBF</td>
<td>0.178</td>
<td>2.52</td>
<td>1.08e-01</td>
</tr>
<tr>
<td>12  AAL Labels, CBF</td>
<td>0.494</td>
<td>2.04</td>
<td>2.09e-06</td>
</tr>
<tr>
<td>13  Eigenanatomy, FA</td>
<td>0.611</td>
<td>2.03</td>
<td>8.32e-10</td>
</tr>
<tr>
<td>14  PCA, FA</td>
<td>0.669</td>
<td>1.97</td>
<td>4.84e-12</td>
</tr>
<tr>
<td>15  ICA, FA</td>
<td>0.565</td>
<td>2.04</td>
<td>2.60e-08</td>
</tr>
<tr>
<td>16  JHU Labels, FA</td>
<td>0.617</td>
<td>1.84</td>
<td>5.16e-10</td>
</tr>
</tbody>
</table>

Table 2: Prediction performance (correlation, mean absolute error, and p-value for correlation) in testing data for predicting age from eigenanatomy, PCA, and ICA projections, and AAL/JHU regions. Predictions using multimodal data outperformed predictions using one modality in every instance, and eigenanatomy projections performed the best overall.

<table>
<thead>
<tr>
<th>Method</th>
<th>Correlation</th>
<th>MAE</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1   Eigenanatomy, VIF-selected predictors</td>
<td>0.849</td>
<td>1.36</td>
<td>&lt;1e-16</td>
</tr>
<tr>
<td>2   Eigenanatomy, all data</td>
<td>0.864</td>
<td>1.27</td>
<td>&lt;1e-16</td>
</tr>
<tr>
<td>3   PCA, VIF-selected predictors</td>
<td>0.693</td>
<td>1.62</td>
<td>4.1e-13</td>
</tr>
<tr>
<td>4   PCA, all data</td>
<td>0.785</td>
<td>1.44</td>
<td>&lt;1e-16</td>
</tr>
<tr>
<td>5   ICA, VIF-selected predictors</td>
<td>0.661</td>
<td>1.81</td>
<td>1e-11</td>
</tr>
<tr>
<td>6   ICA, all data</td>
<td>0.835</td>
<td>1.36</td>
<td>&lt;1e-16</td>
</tr>
<tr>
<td>7   AAL/JHU regions, VIF-selected predictors</td>
<td>0.722</td>
<td>1.57</td>
<td>1.4e-14</td>
</tr>
<tr>
<td>8   AAL/JHU regions, all data</td>
<td>0.851</td>
<td>1.26</td>
<td>&lt;1e-16</td>
</tr>
</tbody>
</table>

Table 3: Prediction performance (correlation, mean absolute error, and p-value for correlation) in testing data, using random forests, for eigenanatomy, PCA, and ICA projections, and AAL regions. Predictions using all data were compared to predictions using the variance inflation factor (VIF) regression-selected predictors.
CHAPTER 3

Decomposing cerebral blood flow MRI into functional and structural components: A non-local approach based on prediction

3.1. Introduction

A fundamental challenge in the interpretation of functional images of the brain is the extent to which the observed function is driven by underlying structure, since the goal of most functional imaging is to provide insight into physiological and pathophysiological processes that may not be manifested in structural changes. In particular, a body of prior work establishes that perfusion and structural signal is shared across modalities. Franklin et al. recently showed that acute baclofen-induced perfusion decreases induce changes in T1-derived gray matter (GM) density (Franklin et al., 2013). A prior study showed increases in observed GM density following acute administration of levodopa (Salgado-Pineda et al., 2006). Chronically, decreased perfusion may result in decreased cortical thickness (Fierstra
et al., 2010). This connection between brain perfusion and structure may confound efforts to correlate disease processes with either perfusion or structure (Villain et al., 2008; Chtelat et al., 2008; Chen et al., 2011a; Tosun et al., 2012, 2010; Jrnum et al., 2011). In brief, structural modalities are not purely structural and may inform and even directly predict functional signal.

To improve interpretability of effects that are correlated across modalities, it is common to apply a correction to emphasize the information unique to a given modality. For example, many perfusion image processing protocols correct the perfusion image for partial volume effects due to variations in gray matter/white matter ratios (Muller-Gartner et al., 1992), since gray matter and white matter have markedly different perfusion values (Roberts et al., 1994). In addition to partial volume and other technical challenges, though, perfusion in a given voxel may be at least partially determined by the underlying brain anatomy. Therefore, we seek to reframe this relation between brain anatomy and perfusion more broadly: Given a perfusion image and a structural anatomical image, how much information is unique to the perfusion image, and how much of the perfusion image can be reconstructed given the structural image? A schematic of this approach is shown in Figure 9.

As a motivating example problem, we consider perfusion measurements of typically developing adolescents. Perfusion studies of typically developing children have shown changes over development (Chiron et al., 1992; Wintermark et al., 2004; Biagi et al., 2007; Jain et al., 2012; Satterthwaite et al., 2013; Wang et al., 2003; Wang and Licht, 2006). In parallel, many studies have focused on structural brain changes over development, including such metrics as cortical thickness (Shaw et al., 2008) and white matter structure (Tamnes et al., 2010). Some of the changes in perfusion are likely due to development of the underlying anatomical substrate, including such developments as cortical thickness, gyrification indices (Blanton et al., 2001; Su et al., 2013), and possibly other, more subtle anatomical changes. On the other hand, it is possible that some of the changes in perfusion are due only to changes in the perfusion of specific cortical areas that are not explained
Figure 9: Schematic of predicting perfusion from structural MRI.

by structural changes. We seek to improve the interpretability of perfusion imaging by separating the component of cortical perfusion that can be explained by structural features from the component of cortical perfusion that is due to biological processes not driven by the underlying anatomy. This separation will help evaluate what unique information is gained by using perfusion imaging as compared to anatomical imaging modalities, thus enabling more principled and informative integration of perfusion imaging into multimodal neuroimaging population studies. The residual perfusion signal represents localized processes that are not explained by the global anatomy-perfusion relationship, signifying development of functional specialization.

Several image processing strategies incorporate knowledge of one modality to improve the interpretability of a second modality, especially where the two modalities offer complementary sources of information. One of the most commonly encountered variants of this problem occurs in positron emission tomography (PET) image processing. PET images have low spatial resolution, leading to significant partial volume effects (PVE) (Hoffman et al., 1979). A widespread method for correcting these partial volume effects is to divide the PET image by gray and white matter probability images (e.g., Muller-Gartner et al.)
By assuming that PET activity within white matter is known, it is then possible to reconstruct the amount of signal that would have resulted from a purely gray matter voxel. Similar strategies have been pursued for arterial spin labeling (ASL) perfusion (Williams et al., 1992) partial volume correction. Many ASL partial volume correction methods assume that white matter has perfusion that is 40% of a comparable unit of gray matter (Johnson et al., 2005), based on quantitative in vivo measures of ASL perfusion (Roberts et al., 1994), even though this ratio is almost certainly dependent on image resolution. More sophisticated models include partial volume correction based on locally determined gray matter activation (Asllani et al., 2008, 2009), a kinetic equation for multiple inversion time ASL (Chappell et al., 2011), and specially designed pulse sequences (Petr et al., 2012). In addition, some studies have incorporated the presence of brain lesions for partial volume correction of ASL images (Schuff et al., 2009).

Fundamentally, partial volume correction (PVC) aims to reconstruct the ideal image that the scanner would have seen had technical impediments, such as scanner resolution and point spread function, not interfered. Although this correction is an important consideration when interpreting perfusion images, it does not attempt to account for true effects of underlying brain structure. Besides technical difficulties with obtaining accurate perfusion measurements, there may be genuine interactions between the underlying anatomy and the observed perfusion that go beyond white and gray matter probabilities. In this work, we address a different problem from PVC: How much brain perfusion can be related to the underlying structure, and how much cannot be predicted from the underlying structure?

Moreover, generating a feature vector for each voxel that contains all the necessary information to reconstruct perfusion from anatomy is not straightforward. Gray matter and white matter probabilities are nearly always used when predicting perfusion from anatomical imaging, even though they provide only a limited model of the structure-perfusion relationship. Cortical thickness may also be correlated to perfusion. Here, we present RIPMMARC (Rotation Invariant Patch-based Multi-Modality Analysis aRChitecture), an
Step 1: Construct dictionary from random samples

Step 2: Project all $n$ voxels in image onto dictionary

Step 3: Regress raw CBF on training subset of structural data

Outcome: Prediction for entire image

Figure 10: Graphical abstract of proposed method. Patches are sampled from image in modality 1 (here, T1) and SVD is used to learn optimal features ("eigenpatches") to describe patches. Patches corresponding to each point in the image are then projected onto the "eigenpatches" to create a representation of the input image in feature space. We then use linear regression to predict the second image (here, perfusion image) from the feature-based description of the first image. This enables us to decompose the perfusion image into a component that is predicted from the structural image and the unique contribution of the perfusion image.
alternative data-driven strategy of deriving structure-perfusion relationships implicitly. RIPMMARC provides a way to encode more detailed local structural information about a given voxel in an image than a scalar intensity value, and this information can be used to predict the perfusion at that point. From concurrently acquired structural and perfusion images, we learn a dictionary of anatomical patch features that can be used to predict perfusion, with the atoms, or elements, in the dictionary corresponding to paradigmatic textural and anatomical features. Mean-centering each input patch ensures that the dictionary contains gradient information invariant to raw intensity value, with intensity represented in corresponding tissue probability values. In contrast to traditional dictionary learning approaches, we construct rotation-invariant dictionaries to enable more complete sharing between similar anatomical structures across the brain. This rotation invariance allows, for example, sharing of information between right and left sides of the brain, which would not be possible when using traditional dictionary learning techniques. Rotation invariance is particularly important in 3D images, as the number of possible orientations increases with the number of dimensions. Projecting patches focused at every voxel in the image onto the rotation invariant dictionary produces a locally varying feature weight image for each atom in the dictionary. We combine the structural feature weights with the probabilistic segmentation images in a linear model to predict perfusion from the structurally derived measures. This linear model then produces a “structurally predicted” perfusion image, corresponding to the predicted perfusion given the structural features, and a residual perfusion image, corresponding to the perfusion that cannot be explained by structural information. A graphical abstract of our method is shown in Figure 10.

RIPMMARC is inspired by feature learning methods (Ranzato et al., 2007; Aharon et al., 2006; Mairal et al., 2008); rotation-invariant feature transforms (Lowe, 1999; Ke and Sukthankar, 2004; Bay et al., 2006; Toews and Wells III, 2013) and dictionary learning methods (Chen et al., 2012; Barthelemy et al., 2012); and modality synthesis algorithms (Hertzmann et al., 2001; Wang et al., 2006; Rueda et al., 2013; Rousseau, 2010). To the best of our knowledge, this work is the first to use rotation invariance for image synthesis.
In addition, our work uses a much more expressive and accurate model for predicting CBF from structural information than prior work.

In sum, we make the following contributions: 1) We propose a novel rotation-invariant dictionary learning method for modality synthesis; 2) We show that these learned dictionaries are significantly better at predicting perfusion than segmentation probability or cortical thickness maps; 3) We demonstrate that this method produces consistent perfusion maps across session scans within a single subject; 4) We show that this method decomposes the raw CBF signal into structurally predicted and residual CBF signals, and all three signals are linked to age in a pediatric population; and 5) The residual perfusion values display a weaker correlation with age in the occipital cortex and precentral motor cortex and a stronger correlation with age in precuneus and hippocampus, suggesting regionally heterogeneous trajectories of functional specialization that are distinct from trajectories of cortical structural development.

3.2. Methods

3.2.1. Representations of Structure

Given an image $\mathcal{I}$, we denote the segmentation probability for white matter (WM) and gray matter (GM) at a voxel $x \in \mathcal{I}$ as $p_{GM,WM}(x)$. We additionally denote the observed cerebral blood flow (CBF) value as $c_{obs}(x)$, and the corrected CBF value as $c_{corr}(x)$. Standard ASL partial volume correction (Johnson et al., 2005) takes the form

$$c_{corr}(x) = \frac{c_{obs}(x)}{p_{GM}(x) + 0.4 \cdot p_{WM}(x)}.$$  (3.1)
This specific formulation derives from a more general assumption of a linear relationship between the voxelwise white matter and gray matter densities. Denoting the true GM and WM CBF levels at voxel $x$ as $c_{GM,WM}(x)$, we have

$$c_{GM}(x) \cdot p_{GM}(x) + c_{WM}(x) \cdot p_{WM}(x) = c_{obs}(x),$$  

(3.2)

where assuming that $c_{WM}(x) = 0.4 \cdot c_{GM}(x)$, as in the earliest work on PVC correction, leads to Equation 3.1. Alternatively, it is possible to learn the relation between GM and WM activity from the CBF image directly, either by sampling over lobes (Johnson et al., 2005) or a local kernel centered on the voxel of interest (Asllani et al., 2008). Both approaches directly analyze the gray matter and white matter probability images as they relate to perfusion.

As explained in the introduction, we take a decidedly different approach to incorporating anatomy into CBF analysis. Instead of attempting to infer the unobservable true GM and WM perfusion in a voxelwise manner, we use all available anatomical information to create a “best guess” at what the observed perfusion would be given the anatomy at voxel $x$. Formulated as a prediction problem, we have

$$c_{obs}(x) = p_{GM}(x) \beta_{GM} + p_{WM}(x) \beta_{WM} + \text{residual}(x),$$  

(3.3)

where we have replaced $c_{GM,WM}(x)$ with $\beta_{GM,WM}$ to emphasize that they are learned values that are constant across the image. The “residual(x)” term accounts for the observed perfusion that cannot be accounted for by the other predictors. In addition to the tissue membership probability values, we incorporate a structural feature vector that describes the anatomy surrounding the voxel of interest. Denoting the value of the $n$’th feature of voxel $x$ as $s_n(x), n \in \{1, \ldots, k\}$, we obtain

$$c_{obs}(x) = p_{GM}(x) \beta_{GM} + p_{WM}(x) \beta_{WM} + s_1(x) \beta_1 + \ldots + s_k(x) \beta_k + \text{residual}(x),$$  

(3.4)
where $\beta_n$ is the weight for the $n$'th feature. As before, the $\beta_n$ weights are learned over the entire image. Concatenating the anatomically derived predictors for voxel $x$ on the right hand side of Equation 3.4 as $X_x = [p_{GM}(x), p_{WM}(x), s_1(x), \ldots, s_k(x)]$ and the weights as $\beta = [\beta_{GM}, \beta_{WM}, \beta_1, \ldots, \beta_k]^T$ allows us to reformulate Equation 3.4 as a standard linear regression:

$$c_{\text{obs}}(x) = X_x \beta + \epsilon,$$

where the $\epsilon$ term corresponds to the residual($x$) term in Equation 3.4. Unlike in standard linear regression, the $\epsilon$ term here is not i.i.d. Gaussian noise; it corresponds to the component of perfusion imaging that cannot be predicted from anatomical information. Although the presence of structured residuals may motivate the use of nonlinear prediction techniques, we have found that linear regression works well for this problem and does not suffer from overfitting, even when training on a small proportion of the data. Further concatenating the observed CBF value across the image as $C_{\text{obs}} = [c_{\text{obs}}(1), \ldots, c_{\text{obs}}(m)]$, where there are $m$ voxels in the image, and $X = [X_1; \ldots; X_m]$, where $[; ; ]$ indicates row-wise concatenation, we obtain

$$C_{\text{obs}} = X \beta + \epsilon.$$

The $X \beta$ term corresponds to the component of perfusion that can be predicted from anatomical features, and the $\epsilon$ term corresponds to the component of perfusion that cannot be predicted from anatomical features. A greater correlation between $C_{\text{obs}}$ and $X \beta$ indicates a more accurate reconstruction of observed perfusion from anatomical features.

### 3.2.2. Dictionary Construction

We consider an image $I$ with $N$ scalar-valued voxels at locations $x_n \in I$, $n = 1, \ldots, N$. We seek a function $s : I(x_n) \mapsto \mathbb{R}^k$ that produces a vector descriptor of the structure around each voxel. Denoting the neighborhood of voxel $x_n$ as $\mathcal{N}(x_n) = \{ x_j \mid \|x_j - x_n\|_2^2 \leq r \}$, where $r$ is the radius of the neighborhood, we generate for each voxel $x_n$ in the image the
vectorized patch $P(x_n) = \text{vectorize} \left( \{ I(x_j) \mid x_j \in \mathcal{N}(x_n) \} \right) \in \mathbb{R}^p$, where each neighborhood consists of $p$ voxels and the “vectorize” operation returns the vector representation of elements in a set. In addition, we denote the mean-centered patch as $P_c(x_n) = P(x_n) - \text{mean} \left( P(x_n) \right)$. Mean-centering each patch serves to minimize the effect of intensity inhomogeneity and concurrently emphasize the gradient and texture information. We denote the sets of voxels in cortical gray matter and white matter as $x_{GM,WM}$ respectively. Because our studies focus on cortical perfusion, we only work with cortical voxels.

To generate the $s$ feature descriptor function, we begin by constructing a rotation-invariant dictionary. Creating a rotation-invariant dictionary requires three steps: Determining a reference direction to reorient all patches to; reorienting all patches to that direction; and constructing a dictionary from the reoriented patches. To find the reference direction, we concatenate row-wise the vectorized patches $P_c(x_n)$ of 1000 voxels sampled randomly from around the cortex into a matrix $S \in \mathbb{R}^{1000 \times p}$. The number of sample voxels to take is limited only by computational power, but we did not observe any improvement in performance when sampling more than 1000 voxels. Next, we perform an SVD of $S$ to obtain its eigenvectors and consider the first eigenvector of $S$ the canonical reference frame.

We used a closed-form solution to align the image patches to the reference frame. The problem of aligning the orientation of two vectors is known as Wahba’s problem ([Wahba, 1965]), and the analytical solution is known as the Kabsch algorithm ([Kabsch, 1976]). Aligning two images corresponds to aligning the orientations of the first eigenvector (or two eigenvectors for a 3D image) of the covariance matrix of the gradient of the image. We denote the gradient operator $g : \mathcal{N}(x_n) \rightarrow \mathbb{R}^D$, where $D$ is the number of dimensions in the image. The covariance matrix $C(\mathcal{N}(x_n))$ of the gradient of the neighborhood $\mathcal{N}(x_n)$ is then given by

$$C(\mathcal{N}(x_n)) = \sum_{x_i \in \mathcal{N}(x_n)} g(I(x_i)) g(I(x_i))^T \in \mathbb{R}^{D \times D}.$$ (3.7)

To align the patches of two voxels $x_i$ and $x_j$, we denote the $k$’th eigenvector of $C(\mathcal{N}(x_i))$ as $w_k$ and the $k$’th eigenvector of $C(\mathcal{N}(x_j))$ as $v_k$ and calculate the rotation matrix $Q$ that
best aligns them:

$$\arg \min_Q \sum_k \|w_k - Qv_k\|^2 \quad (3.8)$$

Denoting $B = w_kv_k^T$, we compute the singular value decomposition (SVD) of $B$: $B = U\Sigma V^T$. Then the analytical solution to Equation 3.8 is given by $Q = UMV^T$, where $M = \text{diag}[1 \ 1 \ \det(U) \ \det(V)]$. We then rotate the voxel coordinates $x_i$ by $Q$ and use a linear interpolator to regenerate the neighborhood image after the rotation. A more computationally expensive alternative is to use the Radon transform to estimate orientation (Jafari-Khouzani and Soltanian-Zadeh, 2005). Using the Kabsch-based approach, we reorient all patches in $S$ to the principal eigenvector of $S$. Then, we perform a second SVD of the reoriented patches. These eigenvectors make up the rotation-invariant dictionary. We retain enough eigenvectors to account for 95% variance explained. Pseudocode for this algorithm can be found in Algorithm 1.

Once we have the rotation-invariant dictionary, we construct the feature vector for a given voxel $x_n \in x_{GM}$ by constructing the vectorized patch $P_c(x_n)$, reorienting the patch to the reference frame, and then multiplying the reoriented patch by each atom in the rotation-invariant dictionary. This procedure generates a $k$-vector for each voxel, with the $i$'th entry corresponding to the response of the $i$'th eigenvector to the voxel neighborhood. Pseudocode for this algorithm can be found in Algorithm 2.

### 3.2.3. Feature Learning

Once we have the rotation-invariant dictionary, we project the reoriented patches corresponding to each voxel in the image onto each rotation-invariant eigenpatch. This gives us an $n \times k$ feature matrix, where $n$ is the number of voxels in the image and $k$ is the number of eigenpatches. The columns of this feature matrix correspond to the response of each eigenpatch to the patch centered on each voxel. In addition to the structural feature matrix, we use the GM and WM probabilities for each voxel in the image. The GM and WM
probabilities are usually the two strongest predictors of blood flow in a given voxel, and we have found that they significantly increase the accuracy of CBF prediction. Thus, although modern PVC approaches (Asllani et al., 2008) move beyond this simplistic relationship between tissue type and perfusion, even this primitive method of incorporating tissue probabilities into the prediction results in a feature that is highly predictive of perfusion, as it is a reasonable first-order approximation to the true relationship between tissue type and perfusion. The GM and WM probabilities also model each voxel’s intensity value.

Once we have the final structural predictor matrix, we run a linear model relating CBF to our predictor matrix:

\[
\text{CBF signal} \approx \text{GM probability} \beta_{\text{GM}} + \text{WM probability} \beta_{\text{WM}} + \text{Structural predictors} \beta_{\text{structure}}.
\]  

(3.9)

To avoid overfitting, we train the model on 5% of the cortical voxels, sampled randomly, and then predict on the remaining 95% of the cortex. The use of a linear model and the fact that there are several orders of magnitude more training samples than predictors further minimize the risk of overfitting, and we did not observe a tendency to overfit in our data. We typically found a drop in variance explained of roughly 2% when going from training to testing data (Figure 21b). We note that in the current study, we learned the relationship between brain structure and perfusion on a per-subject basis. A graphical outline of the method is in Figure 10 and a more formal description of the algorithm is in Algorithm 2. An open-source ITK-based implementation can be found at https://github.com/bkandel/PatchAnalysis.

### 3.2.4. Parameter Settings

RIPMMARC has four free parameters: How many voxels to sample when constructing the dictionary; the ratio of testing to training data for the linear model; how many eigenvectors
Algorithm 1 Algorithm for generating reference frame and rotation-invariant dictionary.

**Input:** patch neighborhood operator $N(x_n)$, number of patches to sample $m$, input image $I$, target variance explained $v$. $\triangleright N(x_n)$ defines the points in the neighborhood of voxel $x_n$.

$p \leftarrow$ number of voxels in $N(x_n)$.

Initialize $S \leftarrow [ ]$. $\triangleright m \times P$ sample patch matrix.

for $i = 0, \ldots, m - 1$ do

$x_i \leftarrow$ random voxel in $I$.

t $\leftarrow$ vector representation of $\{x_j : x_j \in N(x_i)\}$

t $\leftarrow t - \text{mean}(t)$.

S $\leftarrow [ S \ t ]$. $\triangleright$ Mean-center patch.

$\triangleright$ Concatenate patches.

end for

Compute eigenvector matrix $V$ of $S$.

Retain $k_1$ eigenvectors to account for $v$ variance explained. $\triangleright V \in \mathbb{R}^{p \times k_1}$.

for $i = 0, \ldots, m - 1$ do

Reorient $S_i$ to $V_1$. $\triangleright$ Align each vectorized patch to reference frame.

end for

Compute eigenvector matrix $W$ of $S$. $\triangleright$ Construct rotation-invariant dictionary.

Retain $k_2$ eigenvectors to account for $v$ variance explained. $\triangleright$ Now, $W \in \mathbb{R}^{p \times k_2}$.

**Output:** $W$.

Algorithm 2 Algorithm for generating rotation-invariant patch-based description of image.

**Input:** patch neighborhood operator $N(x_n)$, input image $I$, rotation-invariant dictionary $W$ from Algorithm 1.

$N \leftarrow$ number of voxels in $I$.

$p \leftarrow$ number of voxels in $N(x_n)$.

Initialize $P \leftarrow [ ]$. $\triangleright N \times p$ patch matrix for every pixel in image.

for $i = 0, \ldots, N - 1$ do

$t \leftarrow$ vector representation of $\{x_j : x_j \in N(x_i)\}$.

t $\leftarrow t - \text{mean}(t)$.

Reorient $t$ to $W_1$.

P $\leftarrow [ P \ t ]$. $\triangleright$ Concatenate patches.

end for

$F \leftarrow PW$ $\triangleright$ Project patches of input image onto eigenvectors.

**Output:** $F$. $\triangleright$ Matrix with response of each image voxel to each eigenpatch.
to retain; and the size of the patches. We have found that algorithm performance is insensitive to reasonable settings of the first three parameters, and the final parameter should be chosen in a principled way (Figure 21). The number of voxels to sample when constructing the dictionary is limited by computational power, but we have not observed improvements in prediction accuracy or dictionary stability when using more than 1000 voxels (Figure 21c). Similarly, the prediction accuracy does not improve when trained on more than 5% of the cortex (Figure 21b). Choosing how many eigenvectors to retain is an issue that does not have a clear resolution, but we have found that retaining enough eigenvectors to account for 95% variance explained is a good rule of thumb and works well in our experience. Beyond 95% variance explained, no improvement is seen (Figure 21a). The only parameter that has a significant effect on algorithm performance is the patch size. However, the patch size can be chosen to emphasize the spatial scale of features of interest. Because we are interested in features such as position along a sulcus, we chose a radius of 1.4 cm, but it is likely that this parameter will need to change for different applications.

3.2.5. Clinical Data

3.2.5.1. Test-Retest Data:

The cohort consists of 12 healthy young adult participants (mean age 25.5±4.5 years, 7 female, 5 male). For each subject, data was acquired at two time points in the same day. For each time point, high resolution T1-weighted anatomic images were obtained using 3D MPRAGE imaging sequence and the following acquisition parameters: TR = 1620 ms, TI = 950 ms, TE = 3 ms, flip angle = 15°, 160 contiguous slices of 1.0 mm thickness, FOV = 192 × 256 mm², matrix = 192×256, 1 NEX with a scan time of 6 min. The resulting voxel size was 1 mm. Additionally, pseudo-continuous ASL (pCASL) images were aquired with 80 alternating tag/control images all with 14 contiguous slices of 7.5mm thickness, FOV = 220 × 220mm², matrix = 64 × 64; TR = 4000ms, tagging duration 1500ms, and postlabeling
3.2.5.2. Pediatric Data:

Our pediatric data consists of 88 subjects, with mean age 11.72, range 7.07-17.46 years (Figure 11). Magnetization-Prepared Rapid Acquisition Gradient Echo (MPRAGE) images were acquired using a 3D inversion recovery sequence with TR/TE/TI = 2170/4.33/1100 ms. The resolution was 1x1x1mm$^2$ with a matrix size of 256x256x192. Flip angle = 7° and total scan time was 8:08 minutes. Pseudo continuous arterial spin labeled (pCASL) images were acquired using TR/TE = 4000/22 ms, with resolution of 3.125x3.125x6mm$^3$ over a 64x64x24 matrix. The M0 image was estimated by averaging the control (non-tagged) images. 40 label/control pairs were acquired. Generalized autocalibrating partially parallel acquisition (GRAPPA) was done using an acceleration factor of 2. Labeling duration was 1.5s and the post-labeling delay was 1.2s. Total imaging time was 5:30 minutes.
3.2.5.3. Image Preprocessing:

The set of T1 images from the first session was used from each subject to construct a template using ANTs (Avants et al., 2011b). Additionally, a three-tissue segmentation of the template (Avants et al., 2011a) allowed the labels to be partially masked so only cortex and deep gray structures were labeled. For each time point, the T1 image was registered to the template image using SyN (Avants et al., 2008). The subject’s T1 image was also registered to the estimated M0 image as a reference for the pCASL using the antsIntrasubjectIntermodality.sh script in ANTs. These transforms were composed to map the cortical labels into ASL native space for each time point. All anatomical data was downsampled to 2mm isotropic resolution for analysis. For pCASL images, the M0 image served as a reference for motion-correction of all time-point volumes. Nuisance parameters, including motion and physiological confounds, were included as regressors, along with the tag-control binary label, in a robust regression scheme for CBF calculation (Avants et al., 2012b). Cerebral blood flow in physiological units was calculated from the difference between control and tagged images as

\[ f = \frac{\lambda \cdot \Delta M}{2\alpha \cdot M_0 \cdot T_{1b} \cdot (e^{-w/T_{1b}} - e^{-(\tau+w)/T_{1b}})}. \]  

(3.10)

where \( f \) is the perfusion in physiological units (mL/100g/min); \( \lambda \) is the blood-tissue water partition coefficient (0.9 g/mL); \( \Delta M \) is the mean difference between control and tagged images; \( \alpha \) is the tagging efficiency (0.85); \( M_0 \) is the equilibrium brain tissue magnetization, approximated by the mean of the control (non-tagged) images; \( T_{1b} \) is the blood T1 value, modified for each subject based on gender and age, as below; \( w \) is the postlabeling delay (1 second); and \( \tau \) is labeling duration (1.5 seconds). Full details are available in the open-source script at https://raw.githubusercontent.com/stnava/ANTs/master/Scripts/antsASLProcessing.sh. For the pediatric data, the blood T1 value was adjusted for age and gender as \( T_1 = (2115.6 - 21.5 \times \text{age} - 73.3 \times \text{sex}) \) ms, where female sex was set to 0 and male was set to 1, as suggested in (Wu et al., 2010). One subject was eliminated because of extreme
non-physiological CBF values, and two subjects were eliminated because of poor image quality with little differentiation between gray matter and white matter. Prior to dictionary construction, the pCASL and T1 images were resampled to 2mm isotropic resolution. To obtain region of interest (ROI) values for each subject, we warped the MNI template to the population-specific template generated with ANTs. This warp was concatenated with the template-to-subject warp to propagate the AAL label set to the subject space.

3.3. Results

Before analyzing real neuroimaging data, we first present two synthetic data analyses to provide a greater understanding of the motivation and mechanics of our method. We demonstrate the operation of the perfusion-anatomy decomposition on simple synthetic images to highlight the effect of orientation invariance when predicting perfusion. We then perform a simulated population experiment showing how observed changes in perfusion can in fact be due either to the underlying anatomy or changes in perfusion that are not explained by anatomical features. Following the synthetic experiments, we show that our anatomical features are much better than tissue probability maps or cortical thickness at predicting perfusion, and that both the anatomically predicted and residual functional images are highly reproducible within subjects. Finally, we demonstrate that the anatomically predicted and residual CBF signals in a pediatric population are tightly correlated with age in a region-specific manner, and that in certain instances have opposing trends.
Figure 12: Synthetic perfusion and anatomical data. Some aspects of the perfusion data, such as the higher activity at the intersection of the lines, can be deduced from the underlying anatomy (the intersection of the lines), but other aspects of the perfusion data, such as the increased activity on the upper right line, cannot be deduced from the anatomy. 12c,12d: Decomposition of synthetic data using non-rotational invariant features. The constructed features include orientation, so the higher values in the horizontal line are correctly reconstructed. 12e,12f: Reconstructed perfusion and residual perfusion decomposition of Figure 12b. Because orientation invariant features were used, the higher perfusion of the horizontal line is not predicted, but the intersection of the lines does indicate a greater predicted functional signal. Orientation invariance enables greater information sharing across regions, leading to lower variance in the reconstruction as compared to the reconstruction using non-rotationally invariant features (12c).
3.3.1. Synthetic Image Decomposition

We generated synthetic data to demonstrate how the proposed method decomposes simulated functional images into its purely functional component and to the component that can be inferred from structure. Figure 12 shows the “anatomical” and “perfusion” components of the data. Some aspects of the perfusion data, such as the increased activity at the intersections of the lines, can be inferred from the structure of the image (when trained on an appropriate reference functional image). Other aspects of the functional data, such as the increased activity on the upper right-hand line, cannot be inferred from the structural data: Given a patch-based descriptor of a given voxel in the structural image, it is impossible to tell whether the corresponding perfusion voxel has a high or low value. In addition, certain functional values can only be inferred from the orientation of the structure. For example, the horizontal central line has a higher functional value than the vertical lines. Given only an orientation-invariant feature description of the central line, it is impossible to tell what the functional value is. Figure 12 shows the result of the decomposition. As expected, both decompositions do not predict the increased activity in the upper right-hand line from the structural data, but do reconstruct the increased activity at the intersections of the lines. Only the non-rotation invariant decomposition reconstructs the increased activity on the horizontal line. On the other hand, constructing orientation-invariant features enables sharing more data across regions, leading to a lower-variance reconstruction (Figure 12e).

We consider the structure of neuroimaging data to be “rotation-invariant” in the sense that a gyrus pointing superiorly is equivalent to a gyrus pointing inferiorly. This rotation invariance enables information to be shared across hemispheres of the brain and reduces the chances of overfitting to a specific region.
### 3.3.2. Simulated Population Study

To demonstrate the need for a structure-function decomposition that differentiates between changes in perfusion that are due to structural abnormalities and those that are unrelated to the underlying structural substrate, we constructed a simulated data set that includes structural and functional effects. Throughout the brain, we simulated an ASL perfusion image based on the gray and white matter probability maps, with added noise. Using the notation from Section 3.2.1 at voxel \( x \in I \):

\[
c_{\text{obs}}(x) = 100 \cdot p_{\text{GM}}(x) + 40 \cdot p_{\text{WM}}(x) + \text{noise} \tag{3.11}
\]

To the images in the experimental group, we added additional anatomical and perfusion blobs in the following manner (Figure 13). In one blob (the “anatomical” blob), we increased the probability of gray matter. This caused a corresponding increase in the perfusion images. In the second blob, we increased the perfusion without a corresponding increase in GM probability, creating a perfusion increase that does not have a corresponding structural abnormality. Denoting CBF that is not predicted from the underlying anatomy as \( c_r(x) \),

\[
c_{\text{obs}}(x) = 100 \cdot p_{\text{GM}}(x) + 40 \cdot p_{\text{WM}}(x) + c_r(x) + \text{noise} \tag{3.12}
\]

In the third blob, we increased the GM probability and also added additional perfusion above that predicted by the increased GM probability. This blob represents an area that has both a structural abnormality (increased GM probability) and a perfusion abnormality (increased perfusion above that predicted by GM content). To recover an anatomy-perfusion decomposition of the images, we regressed out the anatomical information (GM and WM probability maps) from the perfusion images following the method in Section 3.2.3. This regression gave us two images: The perfusion predicted from structure, and the residual functional activation that is not explained by structure, in addition to the original perfusion images.
We ran a voxelwise t-test comparing control vs. experimental groups on the three types of images. The results are shown in Figure 13. The voxelwise $p$-statistic maps on the raw perfusion images show all three blobs, because all three blobs indeed had increased perfusion in the experimental group (Figure 13, top). $p$-statistic maps on the residual functional images show both the residual perfusion blob and the combined anatomical and perfusion blob (Figure 13, middle). This image, however, ignores the potentially biologically important role of decreased perfusion caused by abnormal anatomy. The $p$-statistic map on the perfusion images as predicted by anatomy shows this missing information (Figure 13, bottom).

3.3.3. Sample subject

The raw perfusion image, the perfusion that can be predicted from structure, and the residual perfusion images for a sample subject are shown in Figure 14. For reference,
the perfusion that can be predicted from probability maps is also shown. Our structural predictors are better at predicting CBF than the probability maps, and in particular predicts higher perfusion in sulcal pits. A quantitative depiction of the correlation between predicted and actual CBF is given in Figure 15.

3.3.4. Variance explained

The structural features we compute are significantly better at predicting perfusion data than gray and white matter probability masks and than cortical thickness maps. Figure 16 compares predicted vs. actual perfusion values using the proposed method, segmentation probability maps, and cortical thickness for the test-retest cohort. The correlation is computed voxel-wise across the gray matter, and each sample corresponds to one subject. The higher correlation of our structural predictors with CBF as compared to the controls indicates that our predictors are more effective at explaining observed perfusion than the control predictors.

3.3.5. Reproducibility

A key measure of the reliability of a clinical measurement is its test-retest reproducibility within a given subject. We evaluated the test-retest reproducibility of our anatomically-predicted and residual perfusion images and compared them to the reproducibility of the raw CBF signal and reproducibility of perfusion as predicted by tissue probability maps and cortical thickness (Figure 17). We evaluated reproducibility by voxel-wise correlation between the images at two time points for a given subject. The most reproducible measure was the CBF predicted by the probability maps, as this value is dependent only the CBF value averaged across an entire tissue compartment and is therefore highly reproducible. The voxel-wise reproducibility of CBF measurement was found to be 0.71±0.09, and
Figure 14: Comparison of mean CBF image (top left), reconstruction from anatomy using RIPMMARC (top right), reconstruction from GM and WM probability images (bottom left), and residual perfusion image (bottom right). Mean CBF image is shown at ASL resolution (3.4mmx3.4mmx7.5mm); other images are shown at 2mm isotropic resolution.
Figure 15: Predictions of CBF within cortical GM using our structural predictors and probability maps. Our structural predictors account for much more variance than probability maps, which exhibit a strong ceiling effect. This figure indicates that even within the cortex, where additional tissue probability information does not predict CBF, the proposed structural predictors can find a meaningful relation between structure and CBF.
Figure 16: Correlation of CBF with: retest CBF; probability images; probability images and thickness; and our structural predictors. Our structural predictors are much better at predicting CBF than probability images, and account for roughly half the reproducible ASL signal. This result indicates that our structural predictors are more appropriate for structural correction of perfusion than using only tissue probability images and cortical thickness.

Figure 17: Reproducibility of mean CBF and derived CBF measures. Reproducibility is reported as the voxelwise correlation of the measure at two scans taken one hour apart. Although the reproducibility of the probability-derived CBF prediction is quite high, the low amount of variance explained by tissue probability values as opposed to RIPMMARC-derived predictors indicates that the reproducibility is most likely driven by whole-brain consistency in CBF values.
this value serves as the upper bound on the reproducibility of predictions from spatially varying anatomical predictors. Predictions from probability maps and thickness on the one hand and our structural predictors have similar reproducibility to the raw CBF images. The residual CBF image was less reproducible as compared to raw CBF reproducibility ($p$-value=$6.69 \times 10^{-6}$), but still displayed relatively high reproducibility across subjects ($0.52\pm0.07$). Although the high reproducibility of the structurally predicted CBF was expected, the high reproducibility of the residual CBF indicates that it is not simply random noise and varies in a consistent way across subjects.

3.3.6. Pediatric Population Study
Figure 18: Raw, anatomically predicted, and residual CBF as a function of age. The raw CBF signal contains a mixture of the structurally predicted and residual CBF signals. The residual CBF shows a spatially heterogeneous longitudinal trajectory of functional specialization, with earlier-developing regions, such as the superior occipital lobe and precentral gyrus, showing less change over adolescence than the later-developing precuneus and hippocampus. This relative stability is not apparent in the raw or structurally predicted CBF signal.
<table>
<thead>
<tr>
<th></th>
<th>Raw CBF</th>
<th>Structural CBF</th>
<th>Residual CBF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Slope</td>
<td>p-value</td>
<td>Slope</td>
</tr>
<tr>
<td>Left Hippocampus</td>
<td>-1.62±0.32</td>
<td>1.12×10⁻⁶</td>
<td>-2.87±0.42</td>
</tr>
<tr>
<td>Right Hippocampus</td>
<td>-1.28±0.34</td>
<td>2.3×10⁻⁴</td>
<td>-2.89±0.42</td>
</tr>
<tr>
<td>Left Precuneus</td>
<td>-4.28±0.55</td>
<td>7.27×10⁻¹³</td>
<td>-2.8±0.41</td>
</tr>
<tr>
<td>Right Precuneus</td>
<td>-4.09±0.54</td>
<td>2.04×10⁻¹²</td>
<td>-2.84±0.4</td>
</tr>
<tr>
<td>Left Precentral</td>
<td>-2.37±0.56</td>
<td>4.37×10⁻⁵</td>
<td>-2.56±0.4</td>
</tr>
<tr>
<td>Right Precentral</td>
<td>-2.2±0.55</td>
<td>10⁻⁴</td>
<td>-2.55±0.4</td>
</tr>
<tr>
<td>Left Occipital</td>
<td>-2.1±0.47</td>
<td>1.73×10⁻⁵</td>
<td>-2.68±0.4</td>
</tr>
<tr>
<td>Right Occipital</td>
<td>-2.64±0.47</td>
<td>9.79×10⁻⁸</td>
<td>-2.65±0.4</td>
</tr>
<tr>
<td>DMN</td>
<td>-3.42±0.47</td>
<td>1.12×10⁻¹¹</td>
<td>-2.69±0.41</td>
</tr>
</tbody>
</table>

Table 4: Statistics from linear models plotted in Figure 18. Although the raw and structurally predicted CBF values showed strong trends with age, the trends for the residual CBF was more variable. Residual CBF was strongly associated with age in the hippocampus and precuneus, but less so in superior occipital cortex and the precentral gyrus. This may suggest that there is a lesser degree of functional specialization in the precentral gyrus and occipital cortex than in hippocampus and precuneus throughout adolescence. Slope is given in units of CBF (ml/100g/min) per year.
To return to the motivating problem of this work, we examined whether observed perfusion changes throughout adolescent development are predicted by a global model relating brain structure to perfusion. We examined trends from a variety of areas representing distinct functional domains and developmental characteristics. The hippocampus and precuneus represent higher-order memory and cognitive functions [Cavanna and Trimble, 2006], and the occipital cortex and precentral gyrus represent sensorimotor regions that are presumed to mature relatively early in development [Gogtay et al., 2004; Rueckriegel et al., 2008]. The default mode network (DMN), a collection of regions that are most active when subjects are not specifically engaged in any externally directed task [Buckner et al., 2008], continues to undergo maturation during adolescence [Uddin et al., 2011; Supekar et al., 2010]. We therefore also examined the CBF trends for the most consistent and conservative definition of the DMN, consisting of left and right precuneus, medial orbitofrontal cortex, and angular gyrus [Buckner et al., 2008].

CBF trends are plotted in Figure 18 with quantitative results in Table 4. We found that although both the raw perfusion values and structurally predicted perfusion changed throughout adolescence in all regions examined, the functional specialization of different regions, as measured by the residual CBF values, followed a regionally varying trajectory. Hippocampal and precuneal residual CBF values showed a strong correlation with age, whereas the residual CBF values were not as strongly associated with age in the superior occipital cortex and the precentral gyrus. These trends were bilateral (see plots for the right hemisphere in Supplementary Material, section 3.6). With the exception of the hippocampus, the structurally predicted CBF had lower variance than the raw CBF, and in all areas the residual CBF had lower variance than the raw CBF. For simplicity and to minimize overfitting, we used linear regression and did not include an interaction between age and gender, but it is possible that this analysis masks nonlinear effects.
3.4. Discussion

We have presented here a method to separate the anatomically predicted from the residual components of perfusion images as measured by ASL MRI. Our method to generate structural predictors gives much better prediction accuracy for predicting CBF than either probability maps or cortical thickness. The test-rest reproducibility of both the structurally predicted and residual CBF is close to that of the raw CBF, implying that both the structurally predicted and residual CBF maps contain stable signals. In addition, we found that although the anatomically predicted and raw CBF were closely related to age, the residual CBF showed a regionally heterogeneous pattern, suggesting that different brain regions undergo different amounts of functional specialization during development.

3.4.1. Interpretation of Structurally Predicted and Residual CBF

RIPMMAR takes CBF and structural images as input, and produces as output a structurally predicted CBF image and a residual CBF image. At first glance, the interpretation of these two outputs may be somewhat obscure, but we believe that when properly understood, each image has an intuitively clear interpretation that can be directly incorporated into clinical characterization of a subject. By way of analogy, we imagine an experiment tracking subject performance on a test of verbal ability in a group of children. A researcher may regress out “nuisance variables,” such as subject age and familial income, before examining the results. At the group level, the effect of these nuisance variables may in fact be of interest, but looking at an individual’s score without accounting for these nuisance variables would be misleading. In our method, we consider the “group effects” to be structural effects shared across the brain, whereas the “subject-level” measurements are the perfusion values at a given voxel. The group effects of underlying brain structure, similarly to age and familial income in our imagined verbal ability study, may be of independent
interest, and we may be interested in looking at regional variations in perfusion as predicted by structural measures. When looking at a given voxel, though, we may also be interested in the amount of perfusion that is not predicted by the underlying neural architecture, just as one may look at a verbal ability result for a given subject when corrected for age and family income. For both the structurally predicted and residual CBF measurements, the units are in the same units of blood flow as the original mean CBF image. Negative values for the residual CBF image correspond to areas with less-than-expected perfusion as compared to structurally homologous regions elsewhere in the brain. Finding regions of the brain that consistently have lower CBF than other structurally similar regions may help clarify which specific anatomical or microstructural characteristics drive regional perfusion variations.

**Biological Interpretation of CBF Measurements**

The reproducibility of the residual CBF, as well as its correlation with age, indicate that it is not only the result of measurement noise, but these results do not provide a true validation that the residual CBF results from a discrete biological process. Although this study establishes an empirical link between local cortical structure and ASL-measured perfusion, it does not conclusively demonstrate a specific biological mechanism for this link. One plausible biological mechanism for a link between cortical structure and perfusion is astrocyte-mediated vasodilation. Astrocyte morphology and distribution is known to vary across the cortex (Mittelbronn et al., 2001), and recent work has demonstrated that astrocytes are capable of modulating arteriole vasoconstriction (Howarth, 2014). A careful evaluation of possible correlations between cortical structure and cytoarchitectonic and vascular modulation of perfusion is necessary to establish a clear causal link between cortical structure and perfusion.
3.4.2. Results from Population Study

We examined how structurally predicted and residual CBF vary across age in a pediatric population. We found that although the raw and structurally predicted CBF decreased across all regions throughout adolescence, the trends for residual CBF exhibited a spatially heterogeneous pattern. In the precuneus, the residual CBF decreased with age, whereas in the hippocampus, the residual CBF increased with age. In both regions, the residual CBF showed a strong bilateral correlation with age. In contrast, the residual CBF in the precentral gyrus and superior occipital cortex showed a much weaker correlation with age. These findings suggest that the functional specialization in some areas follows the cortical structural development, but in other areas displays a distinct trajectory. For example, the precentral gyrus and the occipital lobe are known to reach their mature cortical thickness relatively early in development (Gogtay et al., 2004), and we found that the residual CBF of these areas did not show a strong correlation with age. On the other hand, the hippocampus has also been found to reach structural maturity relatively early in adolescence (Gogtay et al., 2004), but we found a strong correlation between residual CBF and age here. The precuneus, in contrast, displays significant structural changes throughout adolescence (Tamnes et al., 2013), and the precuneal residual CBF was also found to correlate strongly with age. As a whole, these findings indicate that functional specialization may follow a trajectory that is distinct from that of cortical structural development.

In all regions examined, the trend throughout adolescence was for the residual CBF to move towards zero, implying that in older adolescents, a global model relating brain structure to perfusion is more accurate than in younger adolescents.
3.4.3. Comparison to Partial Volume Correction Techniques

Although the method proposed here falls into the general category of atrophy and structure correction techniques, it has a fundamentally different purpose from standard partial volume correction (PVC) techniques [Meltzer et al., 1990; Muller-Gartner et al., 1992; Thomas et al., 2011]. Our method directly addresses the question of the relationship between brain structure and perfusion, which is not the purpose of PVC techniques. PVC aims to recover what the scanner would have seen had technical impediments, such as partial volume effects, not interfered with the imaging. This correction is crucial to appropriately interpreting observed perfusion values, but does not aim to discover what proportion of perfusion can be accounted for by underlying brain structure. In contrast, we aim to recover both the effect of anatomy on the perfusion image and the perfusion that is independent of anatomy. The separation of structural from non-structural perfusion effects is distinct from PVC-based approaches, which incorporate the structural information directly into the output image. We did not explicitly investigate the relationship between the eigenpatch-derived predictions and partial volume effects. It is possible that the eigenpatch predictions are affected by partial volume artifacts. However, the gradient information included in the eigenpatch descriptors should also indicate how close a given voxel is to the edge of the cortex, thus implicitly accounting for partial volume effects.

3.4.4. Consideration of Resolution

The different resolutions of arterial spin labeling MRI as compared to T1 MRI present significant challenges when attempting to analyze the relationship between the two modalities. Because the T1 image is at a much higher resolution than the ASL image, it is difficult to disentangle the effects of scanner characteristics on observed perfusion from true perfusion results. As opposed to PET imaging, quantitative analysis of ASL scanners
Figure 19: Assessment of effect of resolution on CBF measurements. Raw CBF measurements are not affected at all by resolution. Both structural and residual CBF are minimally affected by processing resolution: The ROI-wise correlation between structural and residual CBF at 2mm and 3mm resolution are greater than 0.92.

using physical or computational phantoms is not widespread, although some initial efforts have been reported (Noguchi et al., 2007). The lack of quantitative tools for analyzing scanner properties complicates the effort to work across resolutions. To examine the effect of anatomical variation on observed perfusion, we resampled both the ASL images and the T1 images to 2mm isotropic resolution. This resolution was observed to minimize interpolation artifacts from the ASL native space while still providing adequate anatomic detail. In addition, we explicitly examined the dependence of the results on resolution. We found that choosing the resolution to be 2mm or 3mm had a minimal effect on either structurally predicted or residual CBF (Figure 19).

3.4.5. Limitations

Although this work demonstrates that the proposed method has promise, it does leave some unanswered questions that require further study. First, although the results in the pediatric population imply that the signal present in the residual CBF has biological significance, more study is necessary to validate this finding in a variety of populations to further elucidate its utility in broader applications. Second, we have not rigorously examined here
how the dictionaries and coefficients vary across patients. Using only the predicted value from the dictionary learning approach without examining how the predictions are made may in fact throw away useful data, as the relationship between structure and perfusion itself may contain biologically significant information. To compare the structure-CBF relationship across subjects, though, it would be necessary to learn a consistent dictionary and apply it to all subjects. Carefully examining the variability of learned dictionaries across subjects and across different populations is necessary to establish appropriate techniques for constructing population-wide dictionaries. Third, the residual CBF retains significant amounts of the noise present in the raw CBF images, such as transit effect artifacts from large blood vessels as evident in Figure[14]. The residual CBF image will in general be noisier than the structural CBF image, as the residual image will contain both biologically significant signal and noise, whereas the structurally predicted CBF image will not be affected by ASL noise. For ASL sequences that are noisier than pCASL, such as PASL-derived sequences, the additional noise may interfere with detection of the residual CBF signal. Finally, this method assumes a good registration between the ASL and T1 images. We have found that using an affine registration coupled with a small deformable registration is provides a reliable and accurate way to align cortical perfusion images. In applications with dynamic structural images, such as cardiac or muscular imaging, finding a good correspondence between structural and perfusion images may be more difficult.

3.4.6. Future Work

3.4.6.1. Variations of the Technique

In this work, we learned the relationship between brain structure and perfusion on a per-subject basis. The motivation for this is that although there may be global variations in the function that relates brain structure and perfusion, the function is a global signal over the entire brain, whereas the use of imaging is intended to highlight regionally varying measures
of perfusion. This correction for global signal changes is similar in spirit to the use of relative CBF \cite{Aslan and Lu 2010}, where correcting for global perfusion has been found to increase the ability to find regional differences in blood flow. For application to patient populations, though, it may be more appropriate to learn the structure-perfusion relationship in an age-matched control cohort and apply the structure correction to the patient population. Alternatively, it may be ideal to learn eigenpatches from an independent population and project all subjects in the test population to that basis.

A related question that this study raises is how the structure-perfusion relationship changes across the brain. It may be more appropriate to learn the structure-perfusion relationship across individual lobes, rather than over the entire brain. RIPMARC can be easily modified to perform such an analysis by sampling the patches and training only over lobes, as opposed to over the whole brain.

The infrastructure for constructing a patch-based representation of imaging data has many other applications. It may be possible, for example, to use the patch descriptors to drive registration of images in cases where scalar intensity values are not sufficiently discriminatory. The patch-based descriptors would allow for a more expansive description of anatomy, similar to landmark-based registration techniques \cite{Thompson and Toga 1996}, while still enabling a dense representation of the images, as is common in voxel-based registration techniques \cite{Avants et al. 2008}.

### 3.4.6.2. Additional Applications

Although this study is limited to the connection between brain structure and perfusion, the method is fundamentally agnostic to imaging modality and can be applied across a wide range of imaging techniques. An obvious application of this work is atrophy correction for neurodegenerative populations \cite{Chen et al. 2011a}. Although several studies have shown that brain perfusion, as measured by ASL imaging, changes in Alzheimer’s Disease \cite{Wolk and Detre 2012}, the extent to which this decrease could be determined by atrophic and
other structural changes has not been addressed using methods similar to the proposed work. Some studies have shown hippocampal hyperperfusion in early Alzheimer’s Disease, notwithstanding hippocampal atrophy (Alsop et al., 2008). These contrasting trends of increased perfusion and atrophy highlight the need for rigorous structural correction of perfusion imaging.

RIPMMARC may also be useful for missing image imputation. Standard methods for data imputation rely on methods borrowed from matrix imputation (Xiang et al., 2013; Thung et al., 2014). Using an imputation method that incorporates image characteristics into the imputation may result in a more accurate imputation method. Traumatic brain injury represents a disease state where RIPMMARC may be particularly useful. In many cases, pre-injury perfusion scans of subjects are not available, confounding disease effects with natural inter-subject variation. By imputing perfusion based on structurally similar areas of the brain, RIPMMARC can provide a subject- and region-specific estimate of expected brain perfusion for the damaged region. Finally, RIPMMARC provides a scalable approach to novelty detection in multi-modal imaging studies.

3.5. Conclusion

The method presented here shows promise in decomposing CBF images into anatomically predicted and residual perfusion components. The algorithm proposed explains significantly more of the variance in CBF images than the segmentation probability maps commonly used for performing partial volume correction, and therefore may be more suitable for structural correction of perfusion images than tissue segmentation images. In addition, the method can be used to improve the interpretability of perfusion images by indicating how much of the observed changes in perfusion are caused by global structural trends and how much
Raw, Structural, and Residual CBF vs. Age, R. Hipp.

(a) Structurally predicted and residual CBF for right hippocampus.

Raw, Structural, and Residual CBF vs. Age, R. Precent.

(c) Structurally predicted and residual CBF for right precentral gyrus.

Raw, Structural, and Residual CBF vs. Age, R. Precun.

(b) Structurally predicted and residual CBF for right precuneus.

Raw, Structural, and Residual CBF vs. Age, R. Sup. Occ.

(d) Structurally predicted and residual CBF for right superior occipital cortex.

Figure 20: Right-sided components of plots shown in [18]

by localized processes. This separation of global from local effects can provide greater sensitivity for correlating spatially localized neuronal processes with perfusion images.

3.6. Supplementary Figures
(a) Correlation between predicted and true cerebral blood flow as a function of percent variance explained by eigenvectors of patch matrix. Correlation in testing data changed little as long as enough eigenvectors were retained to explain 50% of the variance, and no additional benefit was gained by retaining enough to explain more than 95% of the variance.

(b) Correlation between predicted and true cerebral blood flow as a function of percent of cortex used for training. Correlation was stable as long as 3% of the cortex was used for training the regressors.

(c) Correlation between predicted and true cerebral blood flow as a function of number of points sampled for estimating eigenvectors. The number of samples to take is limited only by computer memory, but the percent variance explained was insensitive to how many points were sampled.

Figure 21: Effect of parameter settings on correlation between predicted and true cerebral blood flow (CBF) for a sample subject. Algorithm performance was insensitive to reasonable choices of parameter settings, which included retaining enough eigenvectors to explain more than 50% of variance, using at least 3% of the cortex for training the regressors, and sampling at least 250 points for computing eigenvectors.
CHAPTER 4

Neuropsychological Testing Predicts Cerebrospinal Fluid $A\beta$ in Mild Cognitive Impairment (MCI)

4.1. Introduction

Recent guidelines for diagnosing mild cognitive impairment (MCI) due to Alzheimer’s Disease (AD) have emphasized the importance of psychometric testing for establishing the existence of MCI, and subsequently relying on biomarkers based on imaging and biofluids to assess the likelihood that the existing cognitive impairment is “due to AD” relative to a different cause (Albert et al., 2011). In particular, cognitive testing is a component of the core clinical criteria for MCI, which requires that impairment greater than expected for age must be present in at least one cognitive domain. Once clinical categorization of MCI is established, the guidelines suggest that the likelihood that the cognitive phenotype is due to AD should rely on various imaging and molecular biomarkers (each classified as either a biomarker of neurodegeneration or cerebral amyloid), without specifically taking into account the severity of the cognitive deficit within the MCI category.
Although imaging-derived biomarkers for diagnosis of AD and prediction of conversion from MCI to AD have been the subject of intensive research (Chtelat et al., 2003; Davatzikos et al., 2011; Killiany et al., 2000), how these biomarkers can be used most effectively in the presence of alternative sources of clinical information about a subject's status, such as cognitive testing, is still not settled. Several recent studies have examined the relative utility of cognitive testing, imaging, or molecular biomarkers for predicting conversion from MCI to AD (Eckerstrm et al., 2013; Ewers et al., 2014, 2012; Gomar JJ et al., 2011; Runtti et al., 2014). These studies have generally found that cognitive testing performs similarly to other biomarkers, but a potential criticism of these study designs is that using psychometric measurements to predict conversion to AD is circular, as the diagnosis of AD is itself determined in large part based on psychometric tests that are the same as or similar to those used to predict conversion.

To avoid this circularity, we sought to determine if cognitive testing with standard psychometric measures can predict the presence of cerebral amyloid based on a well-established CSF molecular biomarker, the detection of which is independent of cognitive scores, unlike clinical diagnosis of conversion to AD. Although post-mortem histology remains the gold standard for establishing AD pathology, measures of CSF Aβ1-42 and amyloid PET imaging are the closest currently available surrogate (Shaw et al., 2009; Clark CM et al., 2011). For the present study, we used CSF Aβ as a marker for AD pathology given its higher uniform availability in the studied cohort. We choose CSF Aβ in isolation, as opposed to τ/Aβ ratio, because we are specifically comparing the relationship between cognitive and neuroimaging neurodegenerative biomarkers and evidence of AD molecular pathology; thus, incorporating a molecular neurodegenerative marker like τ may confound the results. Moreover, we want to determine the relative and combined predictive value of psychometric testing with neuroimaging biomarkers of neuronal injury or neurodegeneration.
In particular, we examine several cognitive measures, including verbal memory, given their putative sensitivity to prodromal AD. We used diverse imaging-derived biomarkers to accurately represent both standard and developing measurement approaches. Further, we chose structural MRI and FDG PET measures given their emphasis in the MCI guidelines. For MRI data, we used an automated hippocampal volume measurement, several cortical-thickness measurements including a summary measure of several regions associated with AD-related tissue loss (Das et al., 2009; Tustison et al., 2014), and multivariate analysis of voxelwise measurements of cortical thickness (Avants et al., 2014; Dhillon et al., 2014). Hippocampal volume is considered to be one of the most established biomarkers of AD with numerous studies demonstrating its predictive value in MCI. We also used FDG-PET data from a set of regions (meta-ROI) previously determined to be sensitive to early AD and prediction of clinical conversion to AD in MCI cohorts (Landau et al., 2011). To obtain such a wide variety of clinical data in a sufficiently large population, we utilized the Alzheimers Disease Neuroimaging Initiative (ADNI) dataset. If cognitive measures perform similarly to both more standard and developing imaging biomarkers in prediction of AD pathology with MCI patients, they can provide a cost-effective and easily accessible method for assessing the likelihood of prodromal AD in patients with MCI.

### 4.2. Methods

#### 4.2.1. Clinical Data

Subjects: This study was a retrospective analysis of data obtained from the Alzheimers Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration
(FDA), private pharmaceutical companies and non-profit organizations, as a $60 million, 5-year public-private partnership. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer’s disease (AD). Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials. Data used in this article were downloaded from the ADNI website in January 2014. We included only MCI subjects with complete datasets for the current analysis, including CSF Aβ levels, all neuropsychological tests examined, and FDG-PET. Only those subjects with Freesurfer cortical and hippocampal segmentations of acceptable quality, as determined by the publicly available Freesurfer dataset available through ADNI, were included. In the ADNI study, MCI is split into two groups, early MCI (EMCI) and late MCI (LMCI). Diagnostic criteria for both EMCI and LMCI subjects were as follows: MMSE scores between 24-30 (inclusive), a subjective memory concern reported by subject, informant, or clinician, a CDR of 0.5, absence of significant levels of impairment in other cognitive domains, essentially preserved activities of daily living, and an absence of dementia. They also were required to have objective memory loss measured by education adjusted scores on delayed recall of one paragraph from Wechsler Memory Scale Logical Memory II, which further determined EMCI (16 years: 9-11; 8-15 years: 5-9; 0-7 years: 3-6) or LMCI (16 years: 8; 8-15 years: 4; 0-7 years: 2) status. Here, MCI refers to both EMCI and LMCI. The ADNI study includes a variety of collection sites around the United States and Canada, and a full list is available at [http://adni.loni.usc.edu/about/centers-cores/study-sites/](http://adni.loni.usc.edu/about/centers-cores/study-sites/). Recruitment for the ADNI study aimed to achieve a balance of normal controls, MCI, and AD subjects. For ADNI 1, a random subsample of subjects was selected for FDG imaging; in ADNI 2/GO, all subjects enrolled received FDG imaging. For up-to-date information on specific inclusion and exclusion criteria, please see [www.adni-info.org](http://www.adni-info.org).
4.2.2. Psychometric Testing

We aimed to include a battery of psychometric tests that would cover a broad range of cognitive domains, with special focus on memory due to its importance in AD. For memory, we included components of the Rey Auditory Verbal Learning Test (AVLT) (Rey, 1964) given its richness of measures for various aspects of mnemonic processing (e.g. immediate versus delayed recall versus delayed recognition); for assessment of cognitive speed, sequencing, and executive function, the Trail Making Test (Reitan, 1958) [Trails A and Trails B] was used; for language/semantics, category fluency (Butters et al., 1987) [Animals] and the Boston Naming Test (Kaplan et al., 1983) were examined; and as a measure of global cognition, the Mini-Mental State Examination was utilized (Folstein et al., 1975). We examined several of the AVLT measures, which depend on differential aspects of episodic and working memory (Wolk and Dickerson, 2011). The AVLT consists of five learning trials in which a list of 15 words is read and the subject is asked to immediately recall as many items as possible. After an interference list of 15 novel words is read and recalled, subjects are then asked to recall words from the initial list (5-min delayed recall). A 30-min delayed recall trial and recognition test follow. For the recognition test, subjects are presented with a list of the 15 studied words and 15 nonstudied foils and are asked to circle all words previously studied. To account for false alarms (FA) to nonstudied items, we calculated a measure of discriminability, d-prime (d’), in a standard fashion based on classic signal detection theory (Snodgrass and Corwin, 1988). Because d’ is undefined when either proportion is 0 or 1, we used standard formulas to convert these values:

\[
\text{Hits} = \frac{\text{Number of hits} + 0.5}{\text{Number of studied items} + 1} \quad (4.1)
\]

and

\[
\text{FA} = \frac{\text{Number of FA} + 0.5}{\text{Number of unstudied items} + 1} \quad (4.2)
\]
We analyzed performance on the fifth immediate memory trial (AVLT Trial 5 Recall), 5- and 30-minute delayed recall (AVLT 5-min Recall, AVLT 30-min Recall), and recognition memory discrimination (AVLT Recognition Discrimination). In addition, we calculated a retention score, which is the number of items remembered after a 30-minute delay (AVLT 30-min Recall) divided by the number of items remembered during the last immediate memory trial (AVLT Trial 5 Recall).

### 4.2.3. Determination of Amyloid and ApoE Status

Cerebrospinal fluid (CSF)-based molecular biomarkers were processed by the University of Pennsylvania/ADNI Biomarker Core Laboratory as previously described (Shaw et al., 2009, 2011). An Aβ_{1–42} value of less than or equal to 191 pg/ml was considered to be positive for the presence of amyloid pathology based on a prior autopsy-based study performed at the University of Pennsylvania (Shaw et al., 2009). For analyses involving ApoE status, subjects were dichotomized into ApoE ε4 positive and negative groups. ApoE ε4 positive status is defined as having at least one ApoE ε4 allele.

### 4.2.4. Neuroimaging Measures

Processing of neuroimaging data included both analyses made publicly available by ADNI and in-house image processing. The following analyses were based on preprocessed data downloaded from the ADNI website: FDG-PET scans were acquired and analyzed in accordance with a standard protocol (Landau et al., 2011). Mean FDG uptake was averaged over 5 ROIs that are sensitive to AD-related changes in metabolism, including right and left angular gyri, right and left inferior temporal regions, and bilateral posterior cingulate. These regions were averaged into a meta-ROI and normalized to an ROI focused on the pons and cerebellar vermis to give a summary FDG PET measure. Cortical thickness and
hippocampal measurement of the MRI scans were performed according to the standard ADNI Freesurfer [Dale et al., 1999] processing pipeline, and downloaded from the ADNI website. Only images that passed ADNI quality control for the temporal, occipital, temporal, and parietal lobe were included. Cortical thickness in the caudal portion of the middle frontal gyrus, medial portion of the orbital frontal cortex, inferior parietal lobule, lateral portion of the occipital cortex, inferior temporal gyrus, entorhinal cortex, temporal pole, and the isthmus of the cingulate cortex were averaged to form a meta-ROI thought sensitive to early AD related neurodegeneration, as previously suggested (Desikan et al., 2010).

4.2.5. Image Analysis

In addition to the image analysis performed by various ADNI investigators, we ran additional analyses of MR images to supplement standard approaches with a state of the art multivariate analysis technique. 1.5T and 3T non-accelerated T1-weighted MPRAGE and SPGR MRI scans of all MCI subjects from ADNI1 and ADNI2/GO were downloaded from adni.loni.usc.edu. We computed an alternative measure of cortical thickness using DiReCT [Das et al., 2009; Tustison et al., 2014], and used the AAL label set [Tzourio-Mazoyer et al., 2002] to define medial temporal and precuneal regions of interest (ROIs), as these areas are known to atrophy in early AD. We performed a singular value decomposition (SVD) analysis of the whole-brain cortical thickness data, as this analysis has proven useful in differentiating AD from frontotemporal dementia and predicting CSF-based biomarkers in this population (McMillan et al., 2013, 2014). The SVD was performed using the princomp function in R, and we retained the top 10 components. A grid search strategy using bootstrapping with 100 repetitions, with half the subjects left out for a validation cohort, was used to determine the optimal number of components to retain. Statistical Analysis All statistical analysis was performed using the R programming language, version 3.1.0. For predictive studies, we randomly split the subjects 5 times into training and testing cohorts, retaining
half the subjects for training and using the other half for testing in a 5x2 cross-validation scheme (Dietterich 1998). All area under the curve (AUC), odds ratios, and positive and negative predictive values are on the testing cohorts. Two-tail t-tests were used to compare AUC values between testing cohorts of different models to calculate a p-value for differences in mean AUC; false discovery rate (FDR) correction was applied to correct for multiple comparisons. For all analyses, patient age, gender, and education were used as additional predictors; for all MR-based imaging analyses, magnet field strength (1.5 or 3T) was included as a covariate. In addition to univariate predictions of Aβ status from psychometrics and imaging modalities, we performed principal component regression, using 3 principal components, on all the psychometric scores, as well as the psychometric and imaging values combined. Area under the curve (AUC) analysis was performed using the ROCR package in R (Sing et al. 2005).

4.3. Results

4.3.1. Subject Demographics

Subject data was collected between January 2006 and January 2013. A total of 622 MCI subjects with CSF-derived Aβ values were identified, and 407 of those were Aβ positive. Of these, 547 (350 Aβ positive) had FDG scans; 450 (286 Aβ positive) had complete Freesurfer segmentations without failures; 433 (273 Aβ positive) had intracranial volume available; and 408 subjects (257 Aβ positive) had complete psychometric scores available. There was a mean difference of 15 days between the psychometric tests and imaging studies, with 95% of subjects having the imaging and psychometric tests done within 55 days of each other. The maximum time difference was 138 days. A total of 62 adverse events were reported from the lumbar punctures, most of which were headaches (25 cases) or pain (23 cases),
<table>
<thead>
<tr>
<th></th>
<th>All subjects (mean ± standard deviation)</th>
<th>Aβ+</th>
<th>Aβ−</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of subjects</td>
<td>408</td>
<td>257</td>
<td>151</td>
</tr>
<tr>
<td>Number of males</td>
<td>232</td>
<td>151</td>
<td>81</td>
</tr>
<tr>
<td>Number of ApoE ε4+</td>
<td>207</td>
<td>178</td>
<td>29</td>
</tr>
<tr>
<td>Age</td>
<td>71.61±7.16</td>
<td>72.66±6.76</td>
<td>69.79±7.47</td>
</tr>
<tr>
<td>Education</td>
<td>16.24±2.71</td>
<td>16.14±2.79</td>
<td>16.41±2.59</td>
</tr>
<tr>
<td>Mini-Mental Status Examination</td>
<td>28.0±1.74</td>
<td>27.7±1.80</td>
<td>28.4±1.54</td>
</tr>
<tr>
<td>AVLT Trial 5 Recall</td>
<td>9.03±3.00</td>
<td>8.35±2.85</td>
<td>10.19±2.90</td>
</tr>
<tr>
<td>AVLT 5-min Recall</td>
<td>5.65±3.74</td>
<td>4.82±3.42</td>
<td>7.05±3.87</td>
</tr>
<tr>
<td>AVLT 30-min Recall</td>
<td>4.27±3.92</td>
<td>3.30±3.33</td>
<td>5.92±4.29</td>
</tr>
<tr>
<td>AVLT Recognition Discrimination</td>
<td>2.31±1.21</td>
<td>2.07±1.18</td>
<td>2.72±1.14</td>
</tr>
<tr>
<td>Retention</td>
<td>0.41±0.31</td>
<td>0.34±0.29</td>
<td>0.53±0.31</td>
</tr>
<tr>
<td>Trail Making Test</td>
<td>39.00±16.71</td>
<td>41.64±18.21</td>
<td>34.50±12.63</td>
</tr>
<tr>
<td>Trail Making Test</td>
<td>105.70±57.60</td>
<td>116.30±62.47</td>
<td>87.66±42.69</td>
</tr>
<tr>
<td>Boston Naming Test</td>
<td>26.92±3.28</td>
<td>26.73±3.20</td>
<td>27.26±3.39</td>
</tr>
<tr>
<td>Category fluency (animals)</td>
<td>18.05±4.93</td>
<td>17.44±4.88</td>
<td>19.08±4.84</td>
</tr>
<tr>
<td>Hippocampal volume</td>
<td>3497.62±577.07</td>
<td>3386.02±537.17</td>
<td>3687.56±537.17</td>
</tr>
<tr>
<td>Medial Temporal Thickness</td>
<td>3.83±0.60</td>
<td>3.78±0.61</td>
<td>3.93±0.57</td>
</tr>
<tr>
<td>Precuneus Thickness</td>
<td>1.54±0.39</td>
<td>1.52±0.39</td>
<td>1.58±0.37</td>
</tr>
<tr>
<td>Mean Cortical Thickness of AD Meta-ROI</td>
<td>2.64±0.17</td>
<td>2.61±0.17</td>
<td>2.68±0.16</td>
</tr>
<tr>
<td>Mean FDG-PET SUVR of AD Meta-ROI</td>
<td>1.26±0.14</td>
<td>1.23±0.15</td>
<td>1.31±0.11</td>
</tr>
</tbody>
</table>

Table 5: Summary of demographics, psychometric scores, and imaging data for subjects.
with 2 subjects reporting nausea and a few reporting a variety of other effects, including bruising, tenderness, and swelling. One adverse event, transient procedural anxiety, occurred during the imaging.

<table>
<thead>
<tr>
<th></th>
<th>β Estimate</th>
<th>Std. Error</th>
<th>z-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mini-Mental State</td>
<td>-0.36</td>
<td>0.12</td>
<td>-3.11</td>
<td>1.9E-3</td>
</tr>
<tr>
<td>Examination</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AVLT Trial 5 Recall</td>
<td>-0.57</td>
<td>0.12</td>
<td>-4.95</td>
<td>7.6E-7</td>
</tr>
<tr>
<td>AVLT 5-min Recall</td>
<td>-0.55</td>
<td>0.11</td>
<td>-4.83</td>
<td>1.3E-6</td>
</tr>
<tr>
<td>AVLT 30-min Recall</td>
<td>-0.63</td>
<td>0.11</td>
<td>-5.47</td>
<td>4.4E-8</td>
</tr>
<tr>
<td>Trail Making Test A</td>
<td>0.44</td>
<td>0.14</td>
<td>3.18</td>
<td>1.5E-3</td>
</tr>
<tr>
<td>AVLT Recognition</td>
<td>-0.50</td>
<td>0.11</td>
<td>-4.45</td>
<td>8.7E-6</td>
</tr>
<tr>
<td>Discrimination</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retention</td>
<td>-0.59</td>
<td>0.11</td>
<td>-5.25</td>
<td>1.5E-7</td>
</tr>
<tr>
<td>Trail Making Test B</td>
<td>0.52</td>
<td>0.15</td>
<td>3.57</td>
<td>3.6E-4</td>
</tr>
<tr>
<td>Boston Naming Test</td>
<td>-0.08</td>
<td>0.11</td>
<td>-0.76</td>
<td>4.5E-1</td>
</tr>
<tr>
<td>Category fluency (animals)</td>
<td>-0.26</td>
<td>0.11</td>
<td>-2.39</td>
<td>1.7E-2</td>
</tr>
<tr>
<td>Hippocampal volume</td>
<td>-0.43</td>
<td>0.13</td>
<td>-3.44</td>
<td>5.9E-4</td>
</tr>
<tr>
<td>Medial Temporal Thickness</td>
<td>-0.12</td>
<td>0.11</td>
<td>-1.01</td>
<td>0.31</td>
</tr>
<tr>
<td>Precuneal Thickness</td>
<td>-0.01</td>
<td>0.12</td>
<td>-0.05</td>
<td>0.96</td>
</tr>
<tr>
<td>Mean Cortical Thickness of AD Meta-ROI</td>
<td>-0.23</td>
<td>0.12</td>
<td>-1.88</td>
<td>6.1E-2</td>
</tr>
<tr>
<td>Mean FDG-PET SUVR of AD Meta-ROI</td>
<td>-0.54</td>
<td>0.12</td>
<td>-4.57</td>
<td>4.9E-6</td>
</tr>
</tbody>
</table>

Table 6: Summary of univariate logistic regressions predicting Aβ status from each psychometric test and imaging biomarker. Age, gender, and education level (in years) were included as covariates. All data were scaled before regression to facilitate inspection of regression coefficients.
A summary of the demographics of the study population, including the psychometric and imaging information, is given in Table 5. We computed a logistic regression relating each psychometric test and modality with A status, while covarying for age, gender, and education (Table 6). The logistic regression results indicated that the psychometric tests and imaging modalities were predictive of Aβ status, even when included in a univariate model.

### 4.3.2. Predictive Models

The associations between the various psychometric scores and Aβ status were strong enough to predict Aβ status when the data used to train the model was separate from the data used for evaluation. While many of the psychometric measures displayed predictive value, varying in range of AUCs from 0.59 to 0.67, immediate and delayed recall measures performed particularly well, reaching an AUC of 0.65 and 0.67 respectively, corresponding to odds ratios of 3.0 and 2.5 (Figure 1, Table 7). The 30-minute delayed recall test was significantly better than both Trails tests, the Boston Naming Test, category fluency, and MMSE. The standard imaging modalities were similar to each other and the individual psychometric tests in prediction of A status with FDG-PET displaying the highest AUC at 0.67, followed by hippocampal volume at 0.64. Delayed recall performed significantly better than all of the cortical thickness-based measurements and trended better, but was not statistically significantly better, than hippocampal volume. Delayed recall performed similarly to FDG-PET. Despite the prior evidence of SVD analysis of the whole-brain cortical thickness data in prediction of CSF Aβ measures in a cohort of AD and FTD patients, this approach did not appear to enhance prediction (AUC=0.59) versus more standard structural MRI measures. Performing a PCA on the psychometric scores and using the resulting components boosted the AUC slightly to 0.68 with an odds ratio of 3.38; adding the imaging modalities to that model increased the AUC to 0.69, but the increase was not significant (Table 8). The multivariate analysis of the cognitive tests, however, was statistically significantly better.
than hippocampal volume, which was not true for any individual cognitive test. Repeating
the analysis using only subjects with 3T MR scans did not significantly change the results.

<table>
<thead>
<tr>
<th>Test</th>
<th>AUC ± 0.03</th>
<th>Odds Ratio ± 0.03</th>
<th>PPV ± 0.05</th>
<th>NPV ± 0.05</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mini-Mental Status Examination</td>
<td>0.61 ± 0.03</td>
<td>1.94 ± 0.60</td>
<td>0.71 ± 0.05</td>
<td>0.43 ± 0.05</td>
</tr>
<tr>
<td>AVLT Trial 5 Recall</td>
<td>0.65 ± 0.03</td>
<td>3.01 ± 0.36</td>
<td>0.75 ± 0.04</td>
<td>0.50 ± 0.05</td>
</tr>
<tr>
<td>AVLT 5-min Recall</td>
<td>0.65 ± 0.02</td>
<td>2.50 ± 0.44</td>
<td>0.73 ± 0.05</td>
<td>0.47 ± 0.04</td>
</tr>
<tr>
<td>AVLT 30-min Recall</td>
<td>0.67 ± 0.02</td>
<td>2.46 ± 0.52</td>
<td>0.73 ± 0.05</td>
<td>0.48 ± 0.06</td>
</tr>
<tr>
<td>AVLT Recognition Discrimination</td>
<td>0.64 ± 0.03</td>
<td>2.44 ± 0.55</td>
<td>0.73 ± 0.02</td>
<td>0.48 ± 0.07</td>
</tr>
<tr>
<td>Retention</td>
<td>0.67 ± 0.03</td>
<td>2.48 ± 0.48</td>
<td>0.73 ± 0.03</td>
<td>0.47 ± 0.06</td>
</tr>
<tr>
<td>Trail Making Test A</td>
<td>0.62 ± 0.02</td>
<td>2.13 ± 0.46</td>
<td>0.73 ± 0.04</td>
<td>0.44 ± 0.05</td>
</tr>
<tr>
<td>Trail Making Test B</td>
<td>0.63 ± 0.02</td>
<td>2.49 ± 0.48</td>
<td>0.75 ± 0.05</td>
<td>0.45 ± 0.05</td>
</tr>
<tr>
<td>Boston Naming Test</td>
<td>0.59 ± 0.02</td>
<td>1.66 ± 0.17</td>
<td>0.70 ± 0.03</td>
<td>0.42 ± 0.04</td>
</tr>
<tr>
<td>Category fluency (animals)</td>
<td>0.60 ± 0.02</td>
<td>1.88 ± 0.43</td>
<td>0.71 ± 0.05</td>
<td>0.42 ± 0.03</td>
</tr>
<tr>
<td>Hippocampal Volume</td>
<td>0.64 ± 0.02</td>
<td>2.41 ± 0.34</td>
<td>0.74 ± 0.04</td>
<td>0.46 ± 0.04</td>
</tr>
<tr>
<td>Medial Temporal Thickness</td>
<td>0.59 ± 0.01</td>
<td>1.67 ± 0.07</td>
<td>0.70 ± 0.04</td>
<td>0.42 ± 0.04</td>
</tr>
<tr>
<td>Precuneal Thickness</td>
<td>0.59 ± 0.02</td>
<td>1.83 ± 0.25</td>
<td>0.71 ± 0.03</td>
<td>0.43 ± 0.05</td>
</tr>
<tr>
<td>Mean Cortical Thickness of AD</td>
<td>0.61 ± 0.02</td>
<td>1.90 ± 0.31</td>
<td>0.71 ± 0.04</td>
<td>0.43 ± 0.04</td>
</tr>
<tr>
<td>Meta-ROI</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean FDG-PET SUVR of AD</td>
<td>0.67 ± 0.03</td>
<td>3.19 ± 1.22</td>
<td>0.76 ± 0.05</td>
<td>0.49 ± 0.08</td>
</tr>
<tr>
<td>Meta-ROI</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Principal component analysis of psychometric scores</td>
<td>0.68 ± 0.02</td>
<td>3.38 ± 1.16</td>
<td>0.71 ± 0.03</td>
<td>0.560.10</td>
</tr>
<tr>
<td>Principal component analysis of psychometric scores and imaging biomarkers</td>
<td>0.69 ± 0.02</td>
<td>3.18 ± 0.76</td>
<td>0.71 ± 0.03</td>
<td>0.55 ± 0.08</td>
</tr>
<tr>
<td>Principal component analysis of cortex-wide cortical thickness</td>
<td>0.59 ± 0.03</td>
<td>1.57 ± 0.21</td>
<td>0.67 ± 0.04</td>
<td>0.43 ± 0.02</td>
</tr>
</tbody>
</table>
Table 7: Area under the curve (AUC), odds ratios, and positive and negative predictive values predicting Aβ status from biomarkers.
Table 8: Table of p-values of AUC’s for each variable compared with every other variable (FDR corrected). p-values of less than 0.05 are color-coded to indicate which measure is better: Blue indicates that the test indicated in the row name is better, whereas green indicates that the test indicated in the column name is better.
4.3.3. Effect of ApoE Allele

Because of the tight link between ApoE ε4 and Aβ pathology, we sought to determine, as a secondary analysis, whether the observed effects are modulated by ε4 status. We divided the subjects into ε4 positive and ε4 negative groups and performed the analyses in the same way as before (Table 9). The results were broadly the same in that imaging did not significantly improve diagnostic accuracy over psychometric tests. Nearly all psychometric and neuroimaging biomarkers were more predictive of Aβ status in ε4 negative as compared to ε4 positive subjects. This trend was highly statistically significant (p<0.001 using a paired t-test).

4.4. Discussion

4.4.1. Impact

The results shown here indicate that a psychometric evaluation can be as useful as FDG-PET or quantitative MR imaging in predicting whether or not a given amnestic MCI patient likely has underlying AD pathology. The low cost and ready availability of psychometric batteries as compared to imaging studies makes them an attractive and useful alternative to specialized imaging techniques in clinical evaluation. Although the psychometric batteries do not approach perfect classification between Aβ-positive and Aβ-negative subjects, they can be useful in clinical practice to broadly estimate risk of prodromal AD and, perhaps, guide the process of obtaining additional studies, including molecular biomarkers. For situations in which obtaining an accurate measure of Aβ is paramount, such as evaluating appropriateness of a future anti-amyloid therapy, direct
<table>
<thead>
<tr>
<th>Test</th>
<th>AUC  ε4+</th>
<th>AUC  ε4-</th>
</tr>
</thead>
<tbody>
<tr>
<td>AVLT Trial 5 recall</td>
<td>0.72±0.03</td>
<td>0.7±0.04</td>
</tr>
<tr>
<td>AVLT 5-minute recall</td>
<td>0.70±0.04</td>
<td>0.72±0.04</td>
</tr>
<tr>
<td>AVLT 30-minute recall</td>
<td>0.70±0.03</td>
<td>0.74±0.03</td>
</tr>
<tr>
<td>Trails A</td>
<td>0.68±0.03</td>
<td>0.75±0.03</td>
</tr>
<tr>
<td>Trails B</td>
<td>0.67±0.02</td>
<td>0.76±0.03</td>
</tr>
<tr>
<td>Boston Naming Test</td>
<td>0.67±0.02</td>
<td>0.72±0.06</td>
</tr>
<tr>
<td>Category Fluency (animals)</td>
<td>0.70±0.03</td>
<td>0.72±0.05</td>
</tr>
<tr>
<td>MMSE</td>
<td>0.68±0.03</td>
<td>0.73±0.02</td>
</tr>
<tr>
<td>Discrimination</td>
<td>0.69±0.04</td>
<td>0.72±0.02</td>
</tr>
<tr>
<td>Retention</td>
<td>0.70±0.03</td>
<td>0.73±0.03</td>
</tr>
<tr>
<td>Medial Temporal Thickness</td>
<td>0.68±0.03</td>
<td>0.72±0.04</td>
</tr>
<tr>
<td>Precuneus Thickness</td>
<td>0.68±0.03</td>
<td>0.70±0.05</td>
</tr>
<tr>
<td>Mean FDG</td>
<td>0.70±0.02</td>
<td>0.75±0.03</td>
</tr>
<tr>
<td>Hippocampal Volume</td>
<td>0.70±0.04</td>
<td>0.74±0.04</td>
</tr>
<tr>
<td>Thickness of Meta-ROI</td>
<td>0.67±0.03</td>
<td>0.69±0.03</td>
</tr>
<tr>
<td>PCA of psychometrics</td>
<td>0.69±0.03</td>
<td>0.74±0.03</td>
</tr>
<tr>
<td>PCA of psychometrics and imaging</td>
<td>0.69±0.03</td>
<td>0.73±0.04</td>
</tr>
</tbody>
</table>

Table 9: AUC values for prediction of Aβ status from cognitive tests when stratifying patients by ApoE ε4 status. Cognitive tests were overall more predictive of Aβ status in ε4 negative subjects than ε4 positive subjects.
molecular imaging or CSF measurement of A is still necessary, perhaps after initial screening with psychometrics to enrich with amyloid positive patients.

One intriguing finding of this study is that multivariate analysis using principal components analysis (PCA) of the psychometric scores only marginally improved on the single best psychometric test, and the difference in AUC was not statistically significant at the $p < 0.05$ level. At the same time, the modest boost in AUC achieved by a multivariate analysis was sufficient to give a statistically significant improvement over hippocampal volume, but not over FDG-PET. These results suggest that improvements in diagnostic capability by using a multivariate cognitive profile as opposed to a single test offer only marginal improvements while at the same time suffering from less interpretability than a single test. Adding the imaging biomarkers to the multivariate analysis did not significantly improve the AUC, suggesting that imaging offers little added value over a cognitive profile when screening for underlying AD pathology.

Further, the fact that even the standard cognitive measures examined here displayed some success in determining the likelihood of AD pathology suggests that more research is warranted on designing and evaluating psychometric tests optimized for detection of early AD-related cognitive decline. In particular, measures guided by the cognitive neuroscience literature may be particularly useful in this regard [Rentz et al., 2013]. Finally, the results here indicate that the ability of psychometric scores to identify patients who will progress to AD is not due solely to the fact that those same scores are used to establish presence of probable AD. Instead, it appears that the predictive value of psychometric tests are due, at least in part, to their ability to separate MCI patients into sub-populations with higher and lower prevalence of AD pathology.
4.4.2. Limitations

Although this study does indicate that a psychometric battery should be an important component of the evaluation of MCI subjects beyond initial categorization to the MCI designation, there are several factors that may influence the relative ability of imaging to predict AD pathology. First, this study focused exclusively on cross-sectional imaging studies. Longitudinal imaging may provide a more reliable representation of disease progression. Nevertheless, longitudinal imaging may not be feasible for many care settings, so evaluating the diagnostic power of cross-sectional imaging is also important. It is worth noting that this study is meant to help guide providers caring for patients with MCI, not to detect AD pathology in presymptomatic patients. By the time cognitive scores become clearly abnormal, significant neurodegeneration has likely already occurred while this may be more variable in the preclinical phase. Thus, it is unclear whether the same relative predictive value of cognitive versus neuroimaging methods would hold in that context. The patient selection criteria also may limit the applicability of the findings presented here to a broader range of patients. This study focused on amnestic MCI subjects. It is possible that in a broader selection of MCI subjects, the memory tests proposed may provide even greater capability in prediction of amyloid status. On the other hand, in non-amnestic MCI populations, these tests may be less predictive due to differences in the loci of neurodegenerative change in amnestic versus non-amnestic prodromal AD. In addition, the ADNI study population is enriched in AD- or AD-like pathology. In a more general clinical setting, providers must also consider the possibility of other sources of cognitive impairment, such as depression or stroke. It is uncertain how this greater heterogeneity would impact the predictive value of both cognitive and neuroimaging measures. Another drawback to the current study is the sampling procedure. We excluded subjects who did not have all the biomarkers examined here, including those for whom the automated hippocampal segmentation failed. As such, the subset in this study would, if anything, overestimate the ability of hippocampal segmentation to track AD pathology; had we not excluded patients
with unreliable segmentations, the predictive ability of hippocampal volumes would likely be lower.

It is also possible that advances in image processing techniques may improve the diagnostic capability of neuroimaging data. Although it is impossible to rule out such advances, the variety of imaging modalities and image processing techniques used here make it less likely that new analytic approaches would improve the predictive power of imaging data enough to supplant psychometric measures as a key method for characterization of MCI patients. Indeed, the current work did use a promising analytic approach involving singular value decomposition across the entire cortical mantle, which had previously demonstrated good predictive value of CSF t-τ/ Aβ in patients with AD and frontotemporal dementia [McMillan et al., 2013]. Nonetheless, this approach did not display significant advantages over more traditional measures (e.g. hippocampal volume) or psychometric tests. In any case, psychometric tests are more accessible than sophisticated image processing techniques, especially to physicians who do not work in academic medical centers.

An obvious limitation of this study is the use of CSF-derived Aβ status as a gold standard in the prediction models, as CSF Aβ does not perfectly reflect brain AD pathology. While we took this approach to avoid the circularity of longitudinal studies of conversion, a better design would have autopsy-confirmed AD pathology for comparison with the other biomarkers. Nonetheless, CSF Aβ, along with amyloid PET, are the closest surrogates to histopathologic evaluation presently available and have displayed high sensitivity in autopsy studies [Shaw et al., 2009; Clark CM et al., 2011].

Finally, the limited accuracy for prediction of amyloid status of even the most accurate models indicates that caution should be exercised when using values from these models to guide clinical decision-making and, at most, they should be considered another piece in the overall assessment of risk in MCI patients. Fundamentally, the main conclusion of this study is that psychometric scores provide as much information as neurodegenerative imaging biomarkers in prediction of underlying amyloid pathology, not that either imaging
or cognitive biomarkers should be regarded as having perfect diagnostic accuracy. This conclusion strengthens the argument made in previous studies that cognitive tests are a crucial component in multivariate predictive models for conversion from MCI to AD by demonstrating that cognitive scores predict molecular AD pathology, not just cognition-based diagnoses of AD. Therefore, cognitive tests should be considered just as important a biomarker for AD pathology as other neurodegenerative biomarkers, which have already been recognized by the National Institute on Aging Alzheimer’s Association (NIA-AA) work group for MCI diagnosis. Finally, while the AUC values are relatively modest, the odds ratios suggest that poorer performance on the best cognitive predictors are associated with approximately a three-fold risk of underlying AD pathology, which may influence counseling of patients.

4.4.3. Effect of ApoE

One intriguing result in this study is the marked difference in prediction accuracy in ApoE ε4 positive vs. ε4 negative subjects. This finding is consistent with previous work showing that cognitive function is more closely linked to Aβ status within ε4 negative than within ε4 positive subjects (Vemuri et al., 2010; Kantarci et al., 2012). The mechanism behind this effect is not clear, but may be that the effects of Aβ on cognitive function are modulated by ApoE isoforms. However, an important confounding factor is the highly unbalanced nature of the samples: The ε4 negative group had 79 Aβ+ and 120 Aβ- subjects, whereas the ε4 positive group had 178 Aβ+ and only 29 Aβ- subjects. The relative paucity of ε4 positive but Aβ- subjects may contribute to the lower performance of the predictive model in the ε4 positive group. Thus, it is possible that the strong association of Aβ with ε4 status obscures the association with cognitive measures.
4.4.4. Psychometric Scores as Functional Biomarkers

It is worth pointing out that the current algorithm for determining the likelihood of “MCI due to AD” in the recently proposed criteria treats neurodegenerative and molecular markers as dissociable modalities of evidence. In a sense, psychometric tests can be considered another type of downstream neurodegenerative measure. Thus, it may seem somewhat odd to use one type of biomarker (neurodegenerative) to predict another (molecular) in this context if these measures provide orthogonal information. However, these measures are obviously related and multiple studies have demonstrated the significant predictive value for conversion to clinical AD in patients either with “positive” CSF or PET amyloid studies or neurodegenerative markers (Albert et al., 2011; Rowe et al., 2013; Heister et al., 2011).

Nonetheless, one reason for the modest ability of cognitive measures to predict amyloid status is that MCI Aβ+ likely is associated with variable levels of impairment. This is almost certainly an issue for any neurodegenerative biomarker given the range of disease severity within the MCI category. Indeed, neurodegenerative biomarkers, in addition to providing some currency on the underlying pathology (e.g. cerebral amyloid), also are informative on disease stage and enhance prediction of the timing of transitions to dementia, as has been suggested in the literature (Jack Jr et al., 2010; Buchhave et al., 2012; Dickerson and Wolk, 2013). Thus, relatively poor performance on cognitive measures within the MCI category increases both the likelihood that the underlying process is AD and that progression to dementia is more likely to occur in the near future, which may help provide additional context for clinicians in their assessment of these patients.

The choice of CSF Aβ as the proxy or standard for AD pathology in the present analysis also reflects the notion that it is a more specific measure of AD pathology than neurodegenerative markers given the defining nature of cerebral amyloid in the pathologic criteria for AD. Indeed, more and more therapeutic trials, including in MCI, are using a positive amyloid study as inclusion criteria (Sperling et al., 2011a). Thus, examination of psychometric
measures within the MCI category may contribute to increasing the likelihood that a given patient may qualify for such a study on that basis.

4.5. Conclusion

In an MCI population, psychometric scores predict presence of CSF-based amyloid pathology that overlaps with predictions obtainable from FDG-PET and structural MR images. Thus, psychometric measures may be preferable in the cross-sectional context to provide initial screening on the likelihood of prodromal AD. The ability of cognitive scores to predict the existence of underlying AD pathology indicates that in addition to using cognitive test cutoffs to establish the existence of MCI, the severity of the test scores is as reliable an indicator as imaging biomarkers of neurodegeneration that the cognitive impairment is due to AD pathology. Thus, these measures could be included in the MCI algorithm as a type of neurodegenerative marker that could further help clinicians prognosticate in the clinical setting.
CHAPTER 5

White matter hyperintensities are more highly associated with preclinical Alzheimer’s disease than imaging and cognitive markers of neurodegeneration

5.1. Introduction

Due to the recent failures of several clinical trials in treating symptomatic Alzheimer’s Disease (AD) (Sperling et al. 2011b), focus in therapeutic trials is shifting from reversing the effects of AD to preventing cognitive decline due to AD at the preclinical stage, before any noticeable cognitive change has occurred (Sperling et al. 2014). Preclinical AD is defined based on the presence of cerebral amyloidosis, detected by either amyloid PET or measurement of cerebrospinal A$_{1-42}$ Presence of preclinical AD does not necessarily imply
that clinical AD will result, but does appear to come with a higher risk of developing clinical AD (Vos et al., 2013). Because of the importance of preclinical AD, an accurate and thorough understanding of the cognitive and brain changes at this stage is critical. Further, predictors of preclinical AD are potentially valuable in the context of clinical trials to enrich populations prior to use of more expensive or invasive amyloid measurement.

Within cognitively normal older adults, two predictors of amyloid status have already been relatively well-established: age and apolipoprotein E (ApoE) status (Jansen WJ et al., 2015; Morris et al., 2010). Beyond these risks, it is possible that other neurodegenerative biomarkers and cognitive changes that presumptively represent the downstream effect of the presence of cerebral Aβ, such as hippocampal atrophy, hypometabolism, and subjective cognitive impairment, may also be sensitive to preclinical AD (Amariglio et al., 2012; Wolk et al., 2013; Papp et al., 2015). While these markers clearly predict conversion from mild cognitive impairment (MCI) to probable AD (Killiany et al., 2000; Davatzikos et al., 2011; Eckerstrm et al., 2013; Ewers et al., 2012; Gomar JJ et al., 2011) and the presence of cerebral amyloid in MCI to varying degrees (Kandel et al., 2015a), their value in preclinical disease is less well established.

One neuroimaging measure that has received less, but growing, attention in relationship to AD is the presence of white matter hyperintensity (WMH) volume. WMH volume has been associated with clinical AD (Yoshita et al., 2006; Provenzano FA et al., 2013), cognitive ability (Son et al., 2012), cortical atrophy (Barnes et al., 2013), and AD pathology in cognitively normal populations (Kester et al., 2014), but no study has examined the association of WMH volume with preclinical AD in the context of more established imaging and cognitive AD biomarkers. Here, we compare the association of a variety of biomarkers, including neurodegenerative, genetic, functional, and cognitive biomarkers, as well as WMH volume, with preclinical AD. This comparison sheds light on the pathogenesis of AD and can inform subsequent studies on longitudinal trajectories of AD biomarkers.
5.2. Methods

5.2.1. Clinical Data

5.2.1.1. Subjects

Data used in the preparation of this article were obtained from two publicly available data repositories: the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu) and the Parkinson’s Progression Markers Initiative (PPMI) database (www.ppmi-info.org/data). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a $60 million, 5-year public-private partnership. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer’s disease (AD). For up-to-date information on the PPMI study, visit www.ppmi-info.org.

Data used in this article were downloaded from the ADNI website in November 2014. We included all cognitively normal subjects from ADNI2 and ADNI-GO who had undergone florbetapir-PET scans to obtain a measure of cerebral amyloidosis, ApoE genotyping, FDG-PET, structural MR imaging, and all cognitive tests examined. Only subjects with Freesurfer cortical and hippocampal segmentations judged acceptable by the structural MR processing core were included. Inclusion criteria for the study and diagnostic criteria for establishing disease state were as previously reported.(Petersen et al. 2010) For up-to-date information on specific inclusion and exclusion criteria, please see www.adni-info.org. Data was also downloaded from the PPMI website, October 2014. Inclusion criteria for this
study data included a baseline diagnosis of cognitively normal, a T1-weighted and Flair MRI, CSF analysis of AD biomarkers, and APOE genotyping. For up-to-date information on the PPMI study, visit www.ppmi-info.org.

5.2.1.2. Psychometric Testing

The following measures were included in the analysis: the Mini-Mental State Examination (Folstein et al., 1975), Rey Auditory Verbal Learning Test (Rey, 1964), immediate and delayed recall of the Logical Memory Test (Wechsler, 1987), the Trail Making Test (Trails A and Trails B) (Reitan, 1958), category fluency (Animals (Butters et al., 1987), and Boston Naming Test (Kaplan et al., 1983). Given the importance of memory in prodromal AD, we examined several of the AVLT measures, which depend on differential aspects of episodic and working memory (Wolk and Dickerson, 2011). For the current study, we analyzed performance on the fifth immediate memory trial (AVLT Trial 5 Recall), 5- and 30-minute delayed recall (AVLT 5-min Recall, AVLT 30-min Recall), and recognition memory discrimination (AVLT Recognition Discrimination). To account for false alarms (FA) to nonstudied items, we calculated a measure of discriminability, d-prime (d'), in a standard fashion (Snodgrass and Corwin, 1988).

In addition to psychometric measures, we also examined a measure of cognitive complaints via the Everyday Cognition (ECog) questionnaire (Farias et al., 2008; Edmonds et al., 2014) using both informant- and self-report data. Informants and participants are separately queried as to the degree to which particular everyday functioning has changed compared to 10 years earlier. Responses for ADNI were obtained on a five-point scale, with increasing values indicating more complaints and 5 indicating “do not know”. The global scores were averaged separately over informant- and self-rated scales, excluding values of 5.
5.2.1.3. Determination of Amyloid Status

Florbetapir-PET was administered in accordance with the ADNI PET protocols available on-line (http://adni.loni.usc.edu/data-samples/pet), and image processing was performed by the ADNI core lab as described previously (Landau et al., 2012). Briefly, a PET scan was acquired 50-70 minutes following injection of florbetapir. Images were smoothed and aligned to an MPRAGE anatomical image to obtain a cortical segmentation. Mean florbetapir uptake in lateral and medial frontal, anterior and posterior cingulate, lateral parietal, and lateral temporal regions was normalized to uptake in the cerebellum to obtain a mean cortical Standardized Uptake Value ratio (SUVr). Cortical florbetapir uptake of greater than or equal to 1.11 was considered “positive” for cortical Aβ (Landau et al., 2012). ApoE genotyping was performed as described on the ADNI website (http://adni.loni.usc.edu/data-samples/genetic-data/).

5.2.1.4. Neuroimaging Biomarkers

Processing of neuroimaging data was performed by ADNI cores and made publicly available. FDG-PET scans were acquired and analyzed in accordance with a standard protocol (Landau et al., 2011). Mean FDG uptake was averaged over 5 ROI’s that are sensitive to AD-related changes in metabolism, including right and left angular gyri, right and left inferior temporal regions, and bilateral posterior cingulate. Cortical thickness and hippocampal volume measurement based on MRI scans were performed according to the standard ADNI Freesurfer (Dale et al., 1999) processing pipeline, and downloaded from the ADNI website. Only images that passed ADNI quality control for the temporal, occipital, and parietal lobe were included. Cortical thickness in the caudal portion of the middle frontal gyrus, medial portion of the orbital frontal cortex, inferior parietal lobule, lateral portion of the occipital cortex, inferior temporal gyrus, entorhinal cortex, temporal pole, and the isthmus of the cingulate cortex were averaged to form a meta-ROI thought sensitive to early AD related neurodegeneration, as previously suggested (Desikan et al.).
White matter hyperintensity volumes were computed by the ADNI core lab in accordance with previously published protocols (DeCarli et al., 2005). Briefly, FLAIR MR images were corrected for inhomogeneity, and warped to T1 images to provide a segmentation. White matter hyperintensities are seeded at points that are greater than 3.5 standard deviations from the mean signal in white matter, and final segmentation is based on a Bayesian approach, combining spatial priors and tissue class constraints. The WMH segmentation also included segmentations of white matter, gray matter and CSF; the sum of the tissue volumes was used as a surrogate for intracranial volume. For analysis, white matter hyperintensity volumes were normalized to intracranial volume and transformed using the natural logarithm.

5.2.1.5. PPMI Analysis

We used CSF $A\beta_{1-42}$ as a surrogate for cerebral amyloidosis. PPMI has completed two CSF analyses, Project 101 and Project 103, and overall $A\beta_{1-42}$ was significantly elevated in the latter. To maximize sample sizes we used linear regression to transform the elevated CSF $A\beta_{1-42}$ from Project 103 to match values from Project 101, as previously reported (McMillan and Wolk, 2015): Transformed $A\beta_{1-42} = 1.82994 + (A\beta_{1-42} * 0.61562)$. In our transformed CSF series we observed that an $A\beta_{1-42}$ CSF cutoff of 198 pg/mL marked the first point of deviation from the normal distribution of $A\beta_{1-42}$ CSF values, so we selected this cutoff to classify participants as being positive or negative for CSF $A\beta$.

We applied an automated MR image processing pipeline for quantifying white matter hyperintensity volume in the PPMI cohort. The T1-weighted scan of each subject was first preprocessed for intensity inhomogeneity correction (Tustison et al., 2010). A multi-atlas skull stripping algorithm was applied using study-specific atlases for the extraction of the brain tissue (Doshi et al., 2013a). Images with quality issues, such as low T1 resolution, were excluded from the analysis and brain masks with errors were manually corrected. A multi-atlas label fusion method, which uses non-linear registration for transferring atlas
labels to subject space, was applied to form the basis of the white and gray matter segmentations (Doshi et al., 2013b; Ou et al., 2011). Regions of white matter hyperintensity were segmented using a multimodal segmentation method, WMLS, using T1-weighted and FL images (Zacharaki et al., 2008). WMLS is a supervised learning method that trains on lesions manually delineated by an expert radiologist. The lesion segmentation involves data preprocessing via histogram standardization and co-registration, feature extraction, training a voxelwise discriminative model, voxelwise label assignment and false-positive elimination. Quality control was performed on final volumetric data by overlaying each subject’s lesion map on the FLAIR image. None of the lesion masks had errors that would require exclusion. There were minor errors, particularly in the determination of the boundaries of large lesions. However we did not prefer to correct them manually, as the intra- and inter-rater variability associated with manual delineations could potentially bias the results.

5.2.1.6. Statistical Analysis

All statistical analysis was performed using the R programming language, version 3.1.0. Two-tailed two-sample $t$-tests with unequal variances (Welch’s $t$-test) were used to assess differences in demographic characteristics between WMH positive and WMH negative subjects. Logistic regression using a logit link function was used to assess the relationship between white matter hyperintensities and presence of cerebral $\text{A}\beta$. Stepwise forward regression was performed to generate an ideal multivariate linear model, using the Bayesian Information Criterion to regularize the model (Schwarz, 1978). For all analyses, patient age, gender, and education were used as covariates. For hippocampal volume, intracranial vault volume (ICV) was used as an additional covariate.
### 5.3. Results

#### 5.3.0.7. Subject Demographics

<table>
<thead>
<tr>
<th></th>
<th>All Subjects</th>
<th>$A\beta+$</th>
<th>$A\beta-$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of subjects</td>
<td>158</td>
<td>49</td>
<td>109</td>
</tr>
<tr>
<td>Number of males</td>
<td>76</td>
<td>17</td>
<td>59</td>
</tr>
<tr>
<td>Age</td>
<td>73.5±6.1</td>
<td>75.1±5.7</td>
<td>72.7±6.2</td>
</tr>
<tr>
<td>Education</td>
<td>16.4±2.5</td>
<td>15.9±2.4</td>
<td>16.7±2.5</td>
</tr>
<tr>
<td>AVLT Trial 5 Recall</td>
<td>11.4±2.6</td>
<td>11.2±2.8</td>
<td>11.5±2.5</td>
</tr>
<tr>
<td>AVLT 5-min Recall</td>
<td>8.8±3.6</td>
<td>7.9±3.5</td>
<td>9.3±3.6</td>
</tr>
<tr>
<td>AVLT 30-min Recall</td>
<td>7.7±4.0</td>
<td>7.1±3.6</td>
<td>7.9±4.2</td>
</tr>
<tr>
<td>AVLT Recognition Discrimination</td>
<td>3.1±1.0</td>
<td>3.1±1.0</td>
<td>3.2±1.0</td>
</tr>
<tr>
<td>Trail Making Test A</td>
<td>33.3±10.6</td>
<td>35.7±10.4</td>
<td>32.2±10.5</td>
</tr>
<tr>
<td>Trail Making Test B</td>
<td>80.3±38.9</td>
<td>84.7±36.7</td>
<td>78.3±39.9</td>
</tr>
<tr>
<td>Boston Naming Test</td>
<td>28.1±2.3</td>
<td>28.1±2.0</td>
<td>28.2±2.4</td>
</tr>
<tr>
<td>Category Fluency (Animals)</td>
<td>21.4±5.5</td>
<td>21.7±4.8</td>
<td>21.2±5.8</td>
</tr>
<tr>
<td>Mini-Mental Status</td>
<td>29.0±1.3</td>
<td>29.0±1.0</td>
<td>28.9±1.4</td>
</tr>
<tr>
<td>Logical Memory</td>
<td>14.2±3.0</td>
<td>14.0±3.3</td>
<td>14.3±2.8</td>
</tr>
<tr>
<td>Logical Memory, Delayed</td>
<td>13.4±3.1</td>
<td>13.1±2.9</td>
<td>13.5±3.2</td>
</tr>
<tr>
<td>Subject-Reported ECOG Score</td>
<td>1.3±0.3</td>
<td>1.3±0.3</td>
<td>1.3±0.3</td>
</tr>
<tr>
<td>Study Partner-Reported ECOG</td>
<td>1.2±0.3</td>
<td>1.1±0.2</td>
<td>1.2±0.3</td>
</tr>
<tr>
<td>White Matter Hyperintensity Volume (mm$^3$)</td>
<td>3084 (1789-5657)</td>
<td>4836 (3133-7861)</td>
<td>2644 (1402-4864)</td>
</tr>
</tbody>
</table>
Table 10: Summary of demographics, psychometric scores, and imaging data for ADNI subjects. Values are reported as mean ± standard deviation except for white matter hyperintensities, which are reported as median and interquartile ranges.

A total of 184 cognitively normal (CN) subjects with florbetapir-PET were identified from the ADNI database. Of these, 155 subjects had complete psychometric and imaging variables as described in Methods, including acceptable cortical and hippocampal segmentations. A summary of the demographics of the study population, including the psychometric and imaging information, is given in Table 10. Aβ+ subjects were slightly older (M= 75.1, SD=5.7) than Aβ- subjects (M=72.5, SD=6.1), t(100)=2.4, p=0.02, and trended towards having slightly less education (for Aβ+ subjects, M=16 and SD=2.4 years; for Aβ- subjects, M=17 and SD=2.5 years; t(95)=-2, p=0.05).

5.3.0.8. Associative Models

We computed a logistic regression relating each psychometric test and modality with Aβ status, while covarying for age, gender, and education. The logistic regression results indicated that the best univariate predictor of cerebral Aβ was ApoE ε4 status, followed by white matter hyperintensity (WMH) volume. All other imaging and cognitive
measures, including FDG-PET, hippocampal volume, ECog, and any AVLT measure, were not significant predictors of Aβ status. WMH volume was not increased in ApoE ε4-carrying subjects (t(85.8)=-0.018, \( p=0.99 \)). The independence between WMH volume and ApoE genotype implied that both were independent predictors of Aβ status. This was confirmed by running a stepwise forward multivariate regression model, which selected only WMH volume and ApoE status as independent predictors. A boxplot comparing WMH volumes in Aβ+ and Aβ- subjects is shown in Figure 1. There was no significant association between WMH and either \( \tau \) or \( \tau/Aβ \) ratio (\( p=0.56 \) and 0.09, n.s.).

<table>
<thead>
<tr>
<th></th>
<th>Standardized β Estimate</th>
<th>Std. Error</th>
<th>z-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>AVLT Trial 5 Recall</td>
<td>-0.09</td>
<td>0.17</td>
<td>-0.53</td>
<td>0.60</td>
</tr>
<tr>
<td>AVLT 5-min Recall</td>
<td>-0.29</td>
<td>0.18</td>
<td>-1.63</td>
<td>0.10</td>
</tr>
<tr>
<td>AVLT 30-min Recall</td>
<td>-0.17</td>
<td>0.17</td>
<td>-0.95</td>
<td>0.34</td>
</tr>
<tr>
<td>Trail Making Test A</td>
<td>0.19</td>
<td>0.17</td>
<td>1.14</td>
<td>0.26</td>
</tr>
<tr>
<td>Trail Making Test B</td>
<td>0.04</td>
<td>0.17</td>
<td>0.21</td>
<td>0.83</td>
</tr>
<tr>
<td>Boston Naming Test</td>
<td>0.05</td>
<td>0.18</td>
<td>0.29</td>
<td>0.77</td>
</tr>
<tr>
<td>Category Fluency (Animals)</td>
<td>0.24</td>
<td>0.18</td>
<td>1.34</td>
<td>0.18</td>
</tr>
<tr>
<td>Mini-Mental Status</td>
<td>0.08</td>
<td>0.18</td>
<td>0.47</td>
<td>0.64</td>
</tr>
<tr>
<td>Examination</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discrimination</td>
<td>-0.01</td>
<td>0.17</td>
<td>-0.08</td>
<td>0.94</td>
</tr>
<tr>
<td>Logical Memory</td>
<td>-0.02</td>
<td>0.17</td>
<td>-0.11</td>
<td>0.91</td>
</tr>
<tr>
<td>Logical Memory, Delayed</td>
<td>-0.13</td>
<td>0.17</td>
<td>-0.74</td>
<td>0.46</td>
</tr>
<tr>
<td>Subject-Reported ECOG</td>
<td>-0.10</td>
<td>0.18</td>
<td>-0.58</td>
<td>0.56</td>
</tr>
<tr>
<td>ECOG Score Study</td>
<td>-0.17</td>
<td>0.20</td>
<td>-0.85</td>
<td>0.40</td>
</tr>
<tr>
<td>Partner-Reported ECOG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

106
Table 11: Summary of univariate logistic regressions predicting $\text{A}^{\beta}$ status from each psychometric test and imaging biomarker for ADNI subjects. Age, gender, and education level (in years) were included as covariates. All data were scaled before regression to facilitate inspection of regression coefficients. The only variables significant at the $p=0.05$ level were ApoE status and log white matter hyperintensity volume.

5.3.0.9. Replication in PPMI Data

<table>
<thead>
<tr>
<th></th>
<th>All subjects (mean±standard deviation)</th>
<th>$\text{A}^{\beta}+$</th>
<th>$\text{A}^{\beta}-$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Subjects</td>
<td>58</td>
<td>12</td>
<td>46</td>
</tr>
<tr>
<td>Number of Males</td>
<td>35</td>
<td>7</td>
<td>28</td>
</tr>
<tr>
<td>Age (years)</td>
<td>60±13</td>
<td>62±18</td>
<td>60±11</td>
</tr>
<tr>
<td>Number of ApoE 4 carriers</td>
<td>12</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>White Matter Lesion Volume (mm$^3$)</td>
<td>2654±5757</td>
<td>4757±7509</td>
<td>2105±5168</td>
</tr>
</tbody>
</table>
Due to conflicting results in prior studies about the link between WMH volume and Aβ pathology, we sought to replicate correlation between WMH volume and Aβ pathology with elderly controls from the Parkinson’s Progression Markers Initiative. We identified 240 subjects with FLAIR and T1 images. 207 images passed manual quality control; of these, 58 were control subjects. Log WMH volume normalized to ICV was not associated with age ($\rho=0.12$, $p=0.37$) or gender ($t(38)=-1.6$, $p=0.12$). In contrast to the ADNI cohort, WMH volume was significantly increased in ApoE ε4 carriers ($t(24)=2.3$, $p=0.03$).

After correcting for age and gender, WMH volume was significantly predictive of Aβ status ($\beta$ estimate 1.12 ± 0.57, $z=1.97$, $p=0.048$), even given the limited sample size. A boxplot showing WMH volumes in the PPMI cohort is shown in Figure 23. However, in the PPMI cohort, a stepwise forward regression model included only ApoE ε4 status as a predictor of Aβ, and did not include WMH.

5.3.1. Discussion

This study represents the first comprehensive comparative evaluation of a variety of biomarkers to predict presence of preclinical AD pathology in a cognitively normal population. In particular, the surprising finding that white matter hyperintensities are more highly correlated with AD pathology than any of the standard AD imaging biomarkers or cognitive tests suggests that in the earliest stages of AD, vascular disease, as reflected by WMH, may play a significant role in the development of cerebral amyloidosis or be an early downstream effect of this molecular pathology. The replication of this result in an independent dataset, using different image processing techniques, points to the robustness
Figure 22: Boxplots comparing white matter hyperintensity volumes for the ADNI cohort, normalized to intracranial volume and log-transformed. Aβ+ subjects had significantly higher WMH volumes than Aβ- subjects.

Figure 23: Boxplots comparing log white matter hyperintensity volumes for the PPMI cohort, normalized to intracranial volume and log-transformed. As in the ADNI cohort, Aβ+ subjects had significantly higher WMH volumes than Aβ- subjects.
of the finding. The salient association of WMH with preclinical AD supports earlier studies
(Silbert et al., 2012; Zhuang et al., 2012; Carmichael et al., 2012) that have demonstrated a
link between WMH and cognitive decline at this stage. The strength of this association is in
contrast to conventional thinking about this disease stage which assumes that biomarkers
and, perhaps, subtle cognitive symptoms traditionally used to characterize MCI due to AD
and probable AD are the same ones that would characterize preclinical AD (Jack et al.,
2013; Sperling et al., 2011a). Indeed, AD biomarker cascade models do not generally include
a measure of white matter integrity.

5.3.1.1. White Matter Hyperintensities and Amyloid

The degree of association between WMH volume and amyloid deposition in nondemented
control subjects is controversial. Some studies have not found a link between WMH volume
or other vascular disease markers and amyloid (Hedden et al., 2012; Marchant et al., 2012;
Rutten-Jacobs et al., 2011; Tanskanen et al., 2013; Lo et al., 2012), whereas other studies,
including a large (n=337) Amsterdam study (Kester et al., 2014; Thanprasertsuk et al.,
2014; Stenset et al., 2011), have reported such a correlation. It is possible that differences
in WMH calculation, such as WM histogram normalization, thresholds for defining WMH,
and the incorporation of priors for segmenting WMH’s, may at least partially account
for these discrepancies. Standardization and rigorous comparison of competing methods
for evaluating WMH volume may help to reduce the variation in results. Differential
involvement of periventricular and subcortical WMH may also contribute to differing study
results (de Leeuw et al., 2004), although the relative importance of WMH anatomical
distribution as compared to total WMH volume is still a matter of debate (DeCarli et al.,
2005).

The relationship of cerebrovascular disease to clinical AD and amyloid pathology is likely
complex and an evolving understanding is emerging (Kling et al., 2013; Hughes TM et al.,
2014), but there are clear links between the severity of cerebrovascular disease and the
risk of clinical dementia associated with AD pathology (Schneider et al., 2004). There are a number of overlapping risk factors for cerebrovascular disease and WMH with AD, including hypertension, diabetes, obesity, and tobacco use. For example, poorly controlled hypertension has been found to correlate with amyloid plaque pathology (Hoffman et al., 2009) and circle of Willis atherosclerosis is more highly related to AD pathology than other common proteinopathies (Beach et al., 2007; Yarchoan et al., 2012). Other work has related blood pressure and systemic arterial stiffness, also associated with WMH, to amyloid burden measured by amyloid imaging (Toledo et al., 2012; Hughes et al., 2013). We did not observe statistically significant differences between mean arterial pressure, systolic and diastolic blood pressure, BMI, or history of cardiovascular disease, hypertension, or smoking between $\text{A} \beta^+$ and $\text{A} \beta^-$ subjects in the ADNI cohort. Although we did not have complete medical histories of all subjects in the PPMI cohort, we did not observe differences in blood pressure between $\text{A} \beta^+$ and $\text{A} \beta^-$ subjects in the PPMI cohort either.

Whether these are common risk factors with dissociable associations or are more directly mechanistically related is unclear. WMH volume may be a marker of cerebral amyloid due to the increased likelihood of deposition of amyloid in vessel walls leading to development of cerebral amyloid angiopathy (Gurol et al., 2013). Alternatively, there are a number of potential mechanisms by which cerebrovascular changes may directly relate to deposition of cerebral amyloid. For example, $\text{A} \beta$ clearance may depend on the perivascular “glymphatic” system which is diminished by reduced arterial pulsatility or stiffness (Iliff et al., 2012). Further, a number of factors associated with cerebrovascular alterations, including hypoperfusion, may accelerate $\text{A} \beta$ production (Iadecola, 2013). An enhanced understanding of these linkages may provide vascular specific therapeutic options at presymptomatic stages.
5.3.1.2. Independent Association of White Matter Hyperintensities and ApoE Status with Cerebral Amyloid

The finding that WMH and ApoE ε4 carrier status are both independently associated with amyloid, and that WMH volume is not associated with ApoE, implies that vascular burden increases the risk for cerebral amyloid over and above this highly significant genetic risk. This finding is in consonance with other studies that have shown that WMH volume is an independent risk factor for incident dementia (Debette et al., 2010), although not all studies have supported this result (Lo et al., 2012). The independence of these factors is particularly interesting given the relationship of ApoE with cardiovascular risk (Gungor et al., 2012; Rodrigue KM et al., 2013). Nonetheless, it is worth noting that we did not observe the same dissociation in the PPMI dataset, although this could be due in part to an issue of power.

5.3.1.3. Lack of Relationship with Traditional AD Neurodegenerative Biomarkers and Cognitive Measures

The current findings did not support the role of neuroimaging biomarkers that have more traditionally been associated with AD in the prodromal and dementia stages of disease, including hippocampal volumes, cortical thickness, and FDG PET. While some prior studies have found associations between cortical thinning or volume changes and amyloid status in cognitively normal individuals (Dickerson et al., 2011; Dickerson and Wolk, 2012; Oh et al., 2011; Mormino et al., 2014; Schott et al., 2010), this has not been a consistent finding (Whitwell et al., 2013). The differences in results may be due to methodological differences, with some studies using rate of atrophy and others cross-sectional measures. The observed inconsistency may also be due to the heterogeneity among the proposed “stages” of preclinical AD, which include a spectrum of neurodegeneration and cognitive change (Sperling et al., 2011a). This also may explain why in some instances cognitive measures have also been reported to be associated with preclinical AD (Sperling et al., 2013).
Rentz et al., 2013; Wolk et al., 2013). It is worth pointing out that while not significant, hippocampal volume was smaller and 5-minute delayed recall on the AVLT was poorer in the group with evidence of cerebral amyloid. Finally, despite work suggesting a link between subjective complaints and amyloid status (Amariglio et al., 2012), we did not find either the patient or informant-based ECog of value for predicting preclinical AD; this finding resonates with results from a recent meta-analysis of amyloid status in cognitively normal adults with and without subjective symptoms (Jansen WJ et al., 2015).

Although small vessel disease is typically associated with executive dysfunction, we did not observe an association between executive function and WMH volume. This lack of association may be because we studied only cognitively normal subjects, so the variance in cognitive ability is relatively limited. It also may be the case that white matter findings associated with cerebral amyloid do not have the same impact on executive function as those unassociated with amyloid and instead accompanied by more pervasive cerebrovascular disease.

5.3.1.4. Limitations

Our study is limited by the cross-sectional nature of the data, and a longitudinal study looking at the relative timing of the development of WMH and cerebral amyloid would be informative of the direction of causality in this relationship. The characteristics of the ADNI cohort may impact the generalizability of the findings. The ADNI cohort is racially and socioeconomically homogeneous and is relatively free of cardiovascular disease and other comorbidities. It is uncertain whether the relationship between WMH and Aβ would be strengthened or weakened in a more heterogeneous cohort with more prevalent cardiovascular pathology. In addition, the lack of standardization both of FLAIR imaging methods, including resolution and WMH quantification techniques makes comparisons of different populations difficult. The clinical applicability of WMH quantification would be greatly enhanced by a standardized measurement method. Additional replication in other
cohorts would bolster confidence in the observed association. Although manual quality control was performed on all hippocampal and WMH segmentations to avoid major failures, it is possible that minor errors affected the results. Finally, although cerebral A has been adopted as the defining feature of preclinical AD, the lack of complete longitudinal data on ADNI 2/GO precludes any definitive conclusion on the impact of WMH on developing clinical AD in the future.

5.4. Conclusions

In our samples of cognitively normal controls, white matter hyperintensities (WMH) are more highly associated with biomarker evidence of cerebral A than any other recognized biomarker or cognitive test. This finding challenges the assumption that biomarkers of neurodegeneration, which are well-established in later stages of disease, are reliably sensitive at the preclinical stage. As the preclinical stage of AD is increasingly recognized as a period of primary importance for preventing and ameliorating incipient neurodegeneration, the importance of accurately understanding correlates of AD pathology at this stage similarly increases. At the same time, the biological mechanism of the correlation between A and WMH is not clear. A more thorough understanding of the nature of the relationship between A and WMH is necessary to establish potential targets for disease-modifying interventions. Nevertheless, it is clear that WMH should be considered as a potential biomarker for preclinical AD in addition to more widely used cognitive tests and imaging biomarkers.
CHAPTER 6

Conclusions and Future Directions

6.1. Summary of Work

In this thesis, I presented the following contributions:

1. A framework for sparse regression that integrates multi-modal imaging into an anatomically and scientifically interpretable model.

2. A novel method that demonstrates that significantly more signal is shared between arterial spin labeling (ASL) imaging and T1-weighted anatomical imaging than previously known, and that both the shared signal and the signal unique to ASL are tightly correlated with age in a pediatric cohort.

3. Evidence that in mild cognitive impairment (MCI), cognitive testing is as accurate as diagnosing AD pathology as multivariate imaging biomarkers are.

4. A study suggesting that white matter hyperintensities, a marker of small vessel disease, is a more sensitive marker of AD pathology in the preclinical stage than more traditional AD biomarkers are.
These studies have been published at the Information Processing in Medical Imaging conference (Kandel et al., 2013), as well as *NeuroImage* (Kandel et al., 2015b), the *Journal of Alzheimer’s Disease* (Kandel et al., 2015a), *Methods* (Kandel et al., 2015c), and *Alzheimer’s and Dementia: Diagnosis, Assessment and Disease Monitoring* (Kandel et al., 2016). In addition, work unrelated to the main focus of the dissertation was published at the Medical Image Computing and Computer-Aided Intervention (MICCAI) conference (Kandel et al., 2014).

The two clinical studies suggest that the multifaceted effects of AD on the brain chart a complex course during disease progression, and a single modality does not adequately capture all the aspects of degeneration from disease inception to final conversion to AD. In particular, the finding that WMH, a marker of small vessel disease, is highly correlated with AD pathology in a preclinical population suggests that vascular disease may be an important factor in incipient AD. Cerebral blood flow (CBF), which is related to cardiovascular health, can be measured noninvasively via arterial spin labeling (ASL) MRI. However, the extent to which changes in CBF are reflections of underlying changes in anatomy is not clear.

To explicitly examine this issue, I developed RIPMMARC, which quantifies the amount of information shared between ASL and anatomical imaging. Finally, the sparse regression framework presented here serves as an overarching framework for incorporating multiple imaging modalities into a single interpretable regression analysis.

### 6.2. Sparse Regression: Between Inference and Prediction

The sparse regression framework combines populations of scans from multiple modalities into one linear model while retaining classical statistical interpretability of the resulting
linear model. The scientific insights usually obtained from linear models, including the size and sign of coefficients, as well as the statistical significance of the results, are all present and have their usual interpretation in the output of the method. For example, knowing whether a specific structure tends to increase or decrease in thickness with increasing age is critical for formulating neurobiologically principled interpretations of experimental data. This is a primary advantage of this method over more mainstream machine learning methods, such as support vector machines (SVM’s), random forests, and artificial neural networks (ANN’s), which emphasize predictive accuracy at the cost of model interpretability. At the same time, we do pay a cost: A fundamental premise of statistical hypothesis testing is that the model is specified before the data are seen [Wasserman and Roeder 2009]. As such, any exploratory model fitting, including our proposed sparse regression technique, cannot use the same data for choosing a model as for evaluating its fit. Therefore, we must use a set of training data for generating the model and a testing set for evaluating the data. This training-testing split is standard in machine learning, but runs counter to the spirit of inferential statistics, which seeks to draw conclusions about a broader population from a limited study sample. Because the statistical inference is performed only on the testing dataset, which is a fraction of the data available, the statistical power of studies is decreased. At the same time, the efficiency of the estimator is increased, as the regions of interest are tuned to the sample population and outcome of interest. Nevertheless, increasing the number of subjects available for statistical inference would greatly increase the appeal of the sparse regression method to the broader imaging community. At least two avenues provide a promising path forward for this issue.

One way to maximize the number of subjects used for inferential statistics would be to use sparse regression to obtain regions of interest in a related but distinct dataset from the one under study. Creating standardized data-driven regions of interest would eliminate the need to perform the sparse regression for new studies, so all of the subjects for new studies would be available for statistical analysis without the need for taking a subset of subjects for training. Although this approach is theoretically straightforward, establishing regions
of interest that are robust to heterogeneity in patient population, pulse sequences, scanner model, etc. is not a trivial task and requires careful and rigorous validation across a wide variety of datasets. Despite these challenges, standardized regions of interest have been created for some major diseases. Standardized ROI’s have been investigated for years in AD, for specific areas of interest as PET scanning in AD (Landau et al., 2012) and cortical thinning in early AD (Dickerson et al., 2009). The regions of interest in these examples were created from univariate voxelwise correlations between values obtained from imaging and a clinical outcome. Defining ROI’s based on multivariate approaches such as the sparse regression technique proposed here may help capture a more realistic representation of the brain’s activity and function as an interconnected network, rather than as a collection of points working independently.

A drawback to defining prespecified ROI’s for a given set of outcomes and imaging modalities is that predefining ROI’s is only possible for established target diseases and populations. Because pre-specified ROI’s are by nature generated for a specific study population and disease, they cannot be applied to studies of unrelated pathological processes or vastly different patient populations. Nevertheless, generation of ROI’s derived from multivariate analyses of neuroimaging datasets for at least some common diseases would maintain the advantages of ROI’s specifically developed for multivariate methodologies while still allowing calculation of statistical inference on the entire dataset.

Another avenue that could provide statistics over a set of anatomically interpretable multivariate network of brain regions, without resorting to training and testing splits, is to compare some statistic derived from a model against a null distribution. Classical linear model theory uses the F-statistic to compare the fit of two nested models with different numbers of parameters. In the sparse high-dimensional setting, where there are many possible models to choose from and the model design is not set in advance of the analysis, the F-statistic is far too liberal and would result in including lots of components. Recent advances in significance testing for the Lasso, a popular sparse regression technique,
points to possible avenues for future research. Standard multivariate linear model theory examines the drop in residual sum of squares (RSS) obtained by including additional parameters in a model and compares that to a $\chi^2$ distribution to decide whether or not an additional predictor is significant. When the model design is not fixed, as is the case in high-dimensional settings such as brain mapping, this distribution leads to statistical tests for significance that are far too liberal. Lockhart et al. have proposed a “covariance test statistic” as an alternative to the RSS that follows an exponential distribution (Lockhart et al., 2014). Similar techniques may be used to define an empirical null distribution for brain mapping that can be used to determine the significance of additional brain regions in a sparse multivariate model. Recently, progress has been made towards characterizing the null distribution of weights of support vector machine classifiers in predicting outcomes from brain images (Gaonkar and Davatzikos, 2013), and extending this work to include spatial regularization in the case of linear regression is feasible.

6.3. Perspectives on Different Sources of Imaging Information

We have presented a method for quantifying the amount of information shared between T1-weighted and arterial spin labeling (ASL) MR images. The method has promise for evaluating the extent of information shared between two related imaging modalities, but also raises important methodological issues.

First, the method as implemented imputes the missing perfusion image from a T1 image on a per-subject basis. The reason for imputing on a per-subject basis in the current study was to find an upper bound on the amount of information that is unique between T1 and ASL imaging. Depending on the scientific question at hand, though, obtaining the most
accurate predictions of perfusion possible may not be the primary goal. We consider a perfusion-based biomarker of Alzheimer’s Disease as a pertinent clinical example (Wolk and Detre, 2012; Zhang, 2016). Most subjects who would receive an ASL scan also undergo an anatomically detailed T1 scan, which makes an anatomically informed interpretation of the ASL perfusion image feasible. When examining a known disease state, the amount of perfusion that is not predicted by a model trained on age-matched healthy controls may be more clinically meaningful than the amount of perfusion not predicted by a model trained on an intra-subject basis. As an analogy, we will bring a hypothetical chart relating height to expected weight. It is impossible to interpret the clinical impact of weight on a person’s health without knowing her height, as it is the difference between the actual weight and the weight that would be expected to be associated with her height that is meaningful. If the goal of a chart relating height and weight is to identify subjects with unusually high weights given their heights, one would not want to build such a chart from a sample of obese people, as these measurements would bias the chart in a way that would make overweight subjects not appear unusually heavy. In the same way, it makes sense to draw on normal controls to build a model that aims to highlight lower-than-normal perfusion. Building models based on normal controls that can be used to identify subjects with abnormal values of perfusion given brain structure would greatly enhance the clinical utility of RIPMMARC.

A second avenue for improving the clinical interpretation of RIPMMARC would be to build lobe- or region-specific models for predicting perfusion. Although the whole-brain model presented here demonstrates the feasibility of predicting ASL-derived perfusion from T1 imaging, the significant variability in brain structure between different brain regions suggests that lobe- or region-specific models may be more accurate and biologically sound. Deciding how to divide up the brain for region-specific models is not a trivial task, though, and the RIPMMARC architecture may be sufficient to naturally learn appropriate weights for different brain regions by recognizing the projection on the eigenpatches.
6.4. Future Directions in Predictions of Amyloid

We have presented two studies that focus on the prediction of $A\beta$ from multimodal cognitive and imaging clinical data sources. Both studies demonstrated that techniques that are not conventionally recognized as AD biomarkers have predictive accuracy that is as at least as good as traditional biomarkers, but the effectiveness of each technique varies depending on disease status. In preclinical AD, defined as presence of cerebral amyloid without accompanying cognitive changes, small vessel disease as measured by white matter hyperintensities was the most predictive biomarker. On the other hand, in mild cognitive impairment (MCI), a variety of biomarkers were predictive of cerebral amyloid, but none of them were more predictive of amyloid than cognitive tests. These results raise intriguing questions about the nature of the relationship between imaging and cognitive biomarkers and amyloid.

First, one major limitation of both studies is that they were primarily based on the ADNI cohort. Although groundbreaking in the size of the cohort recruited and the quantity of data collected on its participants, the demographics of the cohort are not representative of the overall elderly demented population. ADNI subjects are in general highly educated, with a median of 16 years of education. ADNI subjects are in general of high socioeconomic status. The top ten most common occupations for ADNI participants include engineer, lawyer, accountant, and physician. In addition, the ADNI participants are overwhelmingly white. Finally, ADNI subjects were selected to be relatively free of vascular disease. All these factors make the ADNI population not fully representative of the population affected by AD. Replicating these studies in populations that are more socioeconomic and ethnically diverse would confirm their applicability to community health clinics. Replication of the findings in a broader population is particularly important for understanding the validity of cognitive tests that assume a strong command of the English language, as this assumption
may not be valid in communities with many immigrants and subjects with limited formal education.

The finding that WMH are more highly associated with cerebral amyloid than other established AD imaging biomarkers is somewhat surprising and is deserving of further study. The study design presented here is cross-sectional, focused on identifying subjects who match the definition of preclinical AD. It still remains to be seen, though, whether the amyloid identified in subjects with WMH leads to dementia, either Alzheimer’s or otherwise. It may be that the WMH is more associated with amyloid in arteries, not the amyloid plaques associated with AD. If that is the case, subjects with WMH, although formally qualifying for the category of preclinical AD, may not actually be much more likely to develop AD than other subjects. To identify whether or not the amyloid associated with WMH leads to AD, longitudinal studies must be conducted to show a clear progression of disease from clinically normal, but with WMH, to having MCI or AD. In addition, it is possible that future advances in either WMH quantification or other imaging modalities may change the relative importance of the modalities for predicting the presence of Alzheimer’s pathology.


130


137


C. McMillan and D. Wolk. Presence of Cerebral Amyloid Modulates Phenotype and Pattern of Neurodegeneration in early Parkinsons Disease (P5.023). *Neurology*, 84 (14 Supplement):P5.023, Apr. 2015. ISSN 0028-3878, 1526-632X. URL http://www.neurology.org/content/84/14_Supplement/P5.023.


N. Saikrishnan, G. Kumar, F. J. Sawaya, S. Lerakis, and A. P. Yoganathan. Accurate Assessment of Aortic Stenosis A Review of Diagnostic Modalities and Hemodynamics. *Circulation*, 129(2):244–253, Jan. 2014. ISSN 0009-7322, 1524-4539. doi: 10.1161/CIRCULATIONAHA.113.002310. URL [http://circ.ahajournals.org/content/129/2/244](http://circ.ahajournals.org/content/129/2/244).


