TRANSLATIONAL BIOMARKER STUDIES OF ALZHEIMER'S DISEASE AND MILD COGNITIVE IMPAIRMENT

By

Igor O. Korolev

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of requirements
for the degree of

Neuroscience – Doctor of Philosophy

2013

ABSTRACT

TRANSLATIONAL BIOMARKER STUDIES OF ALZHEIMER'S DISEASE AND MILD COGNITIVE IMPAIRMENT

By

Igor O. Korolev

INTRODUCTION: Alzheimer's disease is a degenerative brain disorder and the most common cause of dementia, affecting more than 5 million Americans and 30 million people worldwide. In this dissertation, I describe two studies investigating imaging and non-imaging biomarkers for the diagnosis of Alzheimer's disease (AD). Two groups of patients were studied: individuals with a clinical diagnosis of AD and individuals with mild cognitive impairment (MCI), an intermediate state between normal cognitive aging and AD. Patients with MCI have an increased risk of developing dementia (annual incidence of 5-15%), particularly secondary to AD. Although many MCI patients develop dementia, other individuals with MCI stay cognitively stable or even regain normal cognitive status.

METHODS/PRINCIPAL FINDINGS: In the first study, diffusion tensor imaging (DTI) was used to study the effects of AD and MCI on the integrity of limbic white matter pathways. Declines in the integrity of the fornix and the descending portion of the cingulum bundle were detectable in both AD and MCI patients. Decreased integrity of the descending cingulum was associated with decreased glucose metabolism in the posterior cingulate cortex (PCC), the earliest detectable sign of AD on positron emission tomography (PET) scans. These findings suggest that the integrity of limbic white matter pathways, as measured by DTI, may serve as a useful biomarker for the diagnosis of AD. These findings also support the "disconnection hypothesis" as a mechanism for the PCC hypometabolism observed in patients with incipient AD. In the second study, we used statistical pattern classification methods and data from the

Alzheimer's Disease Neuroimaging Initiative to develop a prognostic model of dementia for patients with MCI. More than 750 variables spanning clinical, magnetic resonance imaging (MRI), and plasma proteomic data were considered as potential predictors of progression from MCI to dementia. A model based on a small number of clinical and MRI predictors was found to have good predictive performance. I describe the characteristics of the model, including its advantages and limitations. The prognostic model of dementia developed here provides a non-invasive, cost-effective approach that can be used to (1) improve the selection of MCI patients in clinical trials and (2) identify high-risk MCI patients for early anti-AD treatment.

CONCLUSION: It is now recognized that Alzheimer's disease may begin years prior to the onset of dementia, emphasizing the importance of early diagnosis and timely therapeutic intervention. This dissertation describes promising biomarker approaches for the diagnosis of Alzheimer's disease.

DEDICATION

To my grandmother Khana

ACKNOWLEDGMENTS

Many people have contributed to my training as a scientist during my doctoral studies at Michigan State University. I am particularly grateful to the following people. *Dr. Andrea Bozoki*, my major advisor, for her mentorship, contributing to my development as a clinical researcher, and providing stipend support. *Dr. Laura Symonds*, a member of my dissertation guidance committee, for her mentorship, support and encouragement, and valuable insight into the dissertation process. *Dr. David Zhu*, a dissertation committee member, for his mentorship on neuroimaging and for helping me hone my analytical skills. *Dr. Rong Jin*, a dissertation committee member, for his guidance on the use of machine learning methods. *Dr. Kevin Berger*, a dissertation committee member, for his advice on neuroimaging. *Dr. Justin McCormick*, director of the DO/PhD Program, for his support and encouragement, feedback on the dissertation, and commitment to my training as a physician-scientist. *Bethany Heinlen*, administrator in the DO/PhD Program, for her support and encouragement, feedback on the dissertation, and dedication to my training. *Suzanne Kohler*, for her support and encouragement as well as feedback on the dissertation. I would also like to thank *Dr. David Kaufman* and the Department of Neurology, the DO/PhD Program, and the Graduate School for providing stipend support and their commitment to my training. I am also thankful to the faculty and staff of the Neuroscience Program for their help and support and for contributing to my development as a neuroscientist, especially: *Dr. James Galligan* (director of the program), *Dr. John Johnson*, *Dr. Cynthia Jordan*, *Dr. Marc Breedlove*, *Dr. David Kreulen*, *Jim Stockmeyer*, and *Shari Stockmeyer*. My pursuit to become a scientist may never have started if it was not for the early influences that I had from teachers and mentors in high school and college. I would like to thank *Reen Gibb*, my chemistry teacher at Brookline High School, for showing me that science can be fun and exciting. I am also grateful to *Dr. Timothy Rose*, my

chemistry professor at Brandeis University, for helping me further develop my passion for science and medicine. I would also like to thank *Dr. Michael Kahana*, my research mentor and senior honors thesis advisor at Brandeis University and the University of Pennsylvania, for helping me develop my passion in neuroscience. Last, but not least, I would have been unable to tackle the challenges of my training as a physician-scientist and this dissertation research without the consistent support and encouragement from my family.

TABLE OF CONTENTS

LIST OF TABLES....x

LIST OF FIGURES....xi

Chapter 1. Introduction....1
1.1 Alzheimer's Disease (AD)....2
1.1.1 Alois Alzheimer and Auguste D....2
1.1.2 Dementia....2
1.1.3 Epidemiology of AD....3
1.1.4 Neuropathology of AD....4
1.1.5 Diagnosis and Treatment of AD....7
1.2 Mild Cognitive Impairment (MCI)....8
1.2.1 The MCI Concept....8
1.2.2 Epidemiology of MCI....10
1.3 Neuroimaging of AD and MCI....11
1.4 Motivations for the Present Work....13
REFERENCES....16

Chapter 2. Disruption of Limbic White Matter Pathways in Mild Cognitive Impairment and Alzheimer's Disease: A DTI / FDG-PET Study....21
2.1 Abstract....22
2.2 Introduction....23
2.3 Methods....25
2.3.1 Participants....25
2.3.2 Image Acquisition....27
2.3.3 Data Processing....27
2.3.4 Statistical Analysis....32
2.4 Results....32
2.4.1 Demographic Information....32
2.4.2 DTI Data....32
2.4.3 FDG-PET Data....35
2.4.4 DTI and PET Correlation with MMSE Performance....37
2.5 Discussion....39
REFERENCES....44

Chapter 3. Predicting Progression from Mild Cognitive Impairment to Alzheimer's Dementia using Probabilistic Pattern Classification....49
3.1 Introduction....50
3.2 Materials and Methods....52
3.2.1 Alzheimer's Disease Neuroimaging Initiative (ADNI)....53
3.2.2 Subjects....54
3.2.3 Data Collection and Follow-up....56
3.2.4 Clinical Data....56
3.2.5 Structural MRI Data....58

3.2.6 Plasma Proteomic Data....60
3.2.7 Data Transformation....61
3.2.8 Feature Selection....62
3.2.9 Classification Approach....64
3.2.10 Experimental Design....66
3.2.11 Statistical Analysis and Cross-Validation....68
3.3 Results....72
3.3.1 Subject Characteristics....72
3.3.2 Predictive Performance of Single- and Multi-Source Models....72
3.3.3 Predictors of MCI-to-Dementia Progression....77
3.3.4 Influence of Patient Characteristics on Model Performance....80
3.3.5 Probabilistic Classification and Staging of MCI Patients....81
3.4 Discussion....87
3.4.1 Predictive Utility of Clinical, MRI, and Plasma Proteomic Biomarkers....87
3.4.2 Effect of Multiple Kernel Learning on Model Performance....89
3.4.3 Statistical Considerations: Overfitting and Selection Bias....89
3.4.4 The Importance of Patient Heterogeneity in Pattern Classification....90
3.4.5 Probabilistic Pattern Classification....91
3.4.6 Limitations....92
REFERENCES....94

Chapter 4. Conclusions....102
4.1 Summary of Key Findings and Significance....103
4.2 Predictive Models of Dementia: Challenges and Future Directions....104
4.2.1 Clinical Assessments and Clinical Diagnosis of AD....104
4.2.2 Building the Optimal Prognostic Model of Dementia....106
4.2.3 Patient Heterogeneity....108
REFERENCES....110

LIST OF TABLES

Table 2.1. Patient demographic information and MMSE scores.....26

Table 2.2. Means and standard deviations of tract FA and volume.....34

Table 2.3. ROC analysis statistics.....35

Table 2.4. Means and standard deviations of FDG-PET Z-scores.....36

Table 2.5. Relationship between descending cingulum integrity and regional metabolism.....37

Table 3.1. Subject characteristics at baseline.....55

Table 3.2. Cross-validated performance estimates for models 1-9.....73

LIST OF FIGURES

Figure 1.1. Progressive development of Alzheimer's disease (AD).....9

Figure 1.2. Mild cognitive impairment (MCI) is a heterogeneous syndrome.....11

Figure 2.1. DTI tractography.....29

Figure 2.2. FDG-PET Z-score map of cortical metabolism in representative subjects with MCI and early AD.....31

Figure 2.3. Examples of fornix and descending cingulum fiber tracts.....34

Figure 2.4. Posterior cingulate hypometabolism is associated with descending cingulum integrity.....37

Figure 2.5. MMSE performance is associated with limbic tract integrity and cortical metabolism.....38

Figure 3.1. Data sources (features) analyzed in this study.....60

Figure 3.2. Nested 10-fold cross-validation (CV) design for model development and evaluation.....71

Figure 3.3. Selected measures of model performance.....75

Figure 3.4. Performance curves for the MKL-Gaussian model.....76

Figure 3.5. Top 10 most frequently selected features as baseline predictors of MCI-to-dementia progression.....78

Figure 3.6. Regional MRI predictors of MCI-to-dementia progression.....79

Figure 3.7. Comparison between N-MCI and P-MCI groups on baseline predictor variables.....83

Figure 3.8. Comparison between included (n=259) and excluded (n=131) MCI subjects on baseline predictor variables.....84

Figure 3.9. Effect of patient characteristics on predictive accuracy.....85

Figure 3.10. Model accuracy as a function of predictive confidence.....86

CHAPTER 1

Introduction

The goal of this introductory chapter is to provide a general overview of Alzheimer's disease (AD) and mild cognitive impairment (MCI), with an emphasis on clinical and neurobiological aspects of these conditions. Then, I describe the motivations behind the studies presented in Chapters 2 and 3.

1.1 Alzheimer's Disease (AD)

1.1.1 Alois Alzheimer and Auguste D.

The German psychiatrist and neuropathologist Dr. Alois Alzheimer is credited with describing for the first time a dementing condition which later became known as Alzheimer's disease (AD). In his seminal 1906 conference lecture and a subsequent 1907 article, Alzheimer described the case of Auguste D., a 51-year old woman with a “peculiar disease of the cerebral cortex” (Alzheimer, 1907; Maurer et al., 1997). Remarkably, the clinical observations and pathological findings that Alzheimer published more than a century ago continue to remain central to our understanding of AD today.

1.1.2 Dementia

AD is the most common cause of dementia, with up to 70% of dementia cases estimated to be due to AD in the United States (Plassman et al., 2007). Other types of dementia include vascular dementia (10-20%), dementia secondary to Parkinson's disease, dementia with Lewy bodies, and frontotemporal dementia. Dementia is a clinical syndrome (a group of co-occurring symptoms) that involves progressive deterioration in cognition. Various

cognitive abilities can be impaired with dementia, including memory, language, reasoning, decision making, visuospatial function, attention, and orientation. In individuals with dementia, cognitive impairments are often accompanied by changes in personality, emotional regulation, and social behaviors. Importantly, the cognitive and behavioral changes that occur with dementia interfere with work, social activities, and relationships and impair a person's ability to perform routine daily activities (e.g. driving, shopping, housekeeping, cooking, managing finances, personal care).

1.1.3 Epidemiology of AD

AD is a critical public health issue in the United States and many other countries around the world, with a significant health, social, and financial burden on society. An estimated five million Americans have AD, with a new diagnosis being made every 68 seconds (Thies and Bleiler, 2013). In the U.S., AD is the fifth leading cause of death among older adults, and ~$200 billion are spent annually on direct care of individuals living with dementia. Worldwide, it is estimated that ~35 million people have AD or other types of dementia, and ~65 million people are expected to have dementia by 2030 (115 million by 2050) (Prince et al., 2013).

AD is a multifactorial disease – with no single cause known – and several modifiable and non-modifiable factors are associated with its development and progression. Age is the greatest risk factor for the development of AD; the likelihood of developing AD increases exponentially with age, approximately

doubling every five years after age 65 (Ott et al., 1995; Querfurth and LaFerla, 2010).The vast majority of individuals suffering from AD are age 65 or older; this is known as “late-onset” or “sporadic” AD (>95% of all cases). Family history of AD and rare genetic mutations are associated with the development of AD before age 65; this is known as “early-onset” or “familial” AD (<5% of all cases) (Holtzman et al., 2011). People with familial forms of AD have an autosomal dominant mutation in either one of the presenelin genes located on chromosomes 1 and 14 or in the amyloid precursor protein (APP) gene. The genetics of sporadic AD are more complex and less well understood. It is known that the epsilon 4 allele of the apolipoprotein E (APOE) gene located on chromosome 19 is a risk factor for the development of sporadic AD (Reiman et al., 2005). The prevalence of AD is higher among females than males, reflecting longer life expectancy of women (Hebert et al., 2001). Lower educational attainment has been associated with increased risk of AD dementia (Ott et al., 1995), consistent with the idea that education serves to increase a person's cognitive reserve and resilience to AD pathology (Stern, 2012). A large body of evidence suggests that cerebrovascular risk factors play a significant role in both the development and progression of AD; people with a history of diabetes, hypertension, obesity, and smoking have a substantially elevated risk of AD (Barnes and Yaffe, 2011).

1.1.4 Neuropathology of AD

AD is a progressive neurodegenerative brain disorder that causes a

significant disruption of normal brain structure and function. At the cellular level, AD is characterized by a progressive loss of cortical neurons – especially pyramidal cells – that mediate higher cognitive functions (Mann, 1996; Norfray and Provenzale, 2004). Substantial evidence also suggests that AD causes synaptic dysfunction early in the disease process, disrupting communication within neural circuits important for memory and other cognitive functions (Selkoe, 2002). AD-related degeneration begins in medial temporal lobe structures, specifically in the entorhinal cortex and hippocampus (Jack et al., 1997); this provides a mechanism for the memory and learning deficits that are classically observed with early clinical manifestations of AD. The degeneration then spreads throughout the temporal association cortex and to parietal areas. As the disease progresses, degeneration can be seen in the frontal cortex and eventually throughout most of the remaining neocortex. Of note is the fact that AD causes pronounced damage to multiple components of the limbic system (Bozoki et al., 2012; Holtzman et al., 2011), including the hippocampal formation and the major fiber tracts that connect it to the cerebral cortex (fornix and cingulum), amygdala, cingulate gyrus, and thalamus. This widespread pattern of neurodegeneration, affecting both limbic and neocortical regions, correlates closely with the array of cognitive deficits and behavioral changes that AD patients exhibit (Holtzman et al., 2011). In addition to cognitive impairment across multiple domains (memory, language, reasoning, executive, and visuospatial function), patients with AD show an impaired ability to perform activities of daily living and often experience psychiatric, emotional, and personality disturbances.

It has been theorized that neuronal damage seen in AD is related to the deposition of abnormal proteins both within and outside of neurons. These are the hallmark pathological lesions of AD known as “plaques and tangles”, which were first discovered by Alois Alzheimer (Alzheimer, 1907). The abnormal proteins are deposited in the cerebral cortex along neural pathways that mediate memory and other cognitive functions; this protein deposition follows a stereotypical pattern of spread that parallels neurodegeneration (Norfray and Provenzale, 2004). “Senile plaques” (SP) are extracellular accumulations of amyloid protein, and consist of insoluble amyloid-β protein (Aβ). Normally, cells throughout life release soluble Aβ after cleavage of the amyloid precursor protein (APP) – a cell surface receptor. Abnormal cleavage of APP coupled with activity of free radicals, which are released from dysfunctional mitochondria, cause precipitation of Aβ into dense beta sheets and formation of SP. Microglia and astrocytes mount an inflammatory response to clear the amyloid aggregates; it is believed that in this process, some of the inflammatory products released for digestion of SP likely cause destruction of adjacent neurons and their neurites (axons and dendrites) (Norfray and Provenzale, 2004; Querfurth and LaFerla, 2010). “Neurofibrillary tangles” (NFT) are intracellular aggregates of abnormally hyper-phosphorylated protein tau, which in normal form serves as a microtubule stabilizing protein and plays a role in intracellular (axonal and vesicular) transport. It is possible that NFT interfere with normal axonal transport of components necessary for proper neuronal function and survival (e.g. synaptic vesicles with neurotransmitters, neurotrophic factors, mitochondria), eventually

causing neurons to die (Norfray and Provenzale, 2004; Querfurth and LaFerla, 2010). Substantial evidence supports the idea that amyloid formation and deposition in the cerebral cortex is one of the earliest pathological processes in AD, preceding the clinical onset of the disease by 10-20 years (Holtzman et al., 2011). Despite this, the temporal sequence of events in the deposition of amyloid plaques and formation of NFT during development of AD remains open to debate. In fact, a recent study suggests that the initial formation of NFT may occur in the brainstem rather than the medial temporal lobe and may precede the appearance of the first amyloid plaques in the neocortex (Braak et al., 2011).

1.1.5 Diagnosis and Treatment of AD

The gold standard for the diagnosis of AD is an autopsy-based (post-mortem) pathological evaluation. The presence and distribution of amyloid plaques and neurofibrillary tangles in the brain is used to establish the diagnosis of “definitive” AD and stage the disease (Braak et al., 2011). However, in living individuals, AD has been diagnosed using standardized clinical criteria for the past three decades (McKhann et al., 1984); the clinical diagnosis is referred to as “probable” AD and is less definitive than the pathological diagnosis. The clinical diagnosis of AD is largely based on medical history, physical and neurological examinations, and neuropsychological evaluation. Brain imaging is used only for investigational purposes or as an adjunct to the clinical criteria, particularly for ruling out structural brain lesions or in the differential diagnosis of dementia in more difficult patient cases. Laboratory testing of blood and cerebrospinal fluid is

similarly performed for research purposes or to rule out alternative causes of cognitive impairment but not to establish a diagnosis of AD.

There is no cure for AD, and drug therapy for the disease is still in its infancy. Approved medications for the treatment of probable AD help to control the symptoms of AD but do not slow down the progression or reverse the course of the disease itself (Holtzman et al., 2011). Several clinical trials of disease-modifying AD drugs are currently being conducted, although no such drugs have been approved for use yet. At present, the mainstay of AD therapy are drugs that target neurotransmitter systems in the brain. AD primarily damages glutamate- and acetylcholine-producing neurons and their associated synapses, and this damage correlates well with early cognitive symptoms of AD (Selkoe, 2002). Acetylcholinesterase inhibitors (e.g. donepezil) help improve memory function and attention in AD patients by interfering with the breakdown of acetylcholine, thereby increasing the levels of the neurotransmitter at the synapse. Memantine belongs to a different class of drugs and helps improve cognitive function in AD patients by modulating NMDA-type glutamate receptors (Holtzman et al., 2011).

1.2 Mild Cognitive Impairment (MCI)

1.2.1 The MCI Concept

MCI is a syndrome characterized by memory and/or other cognitive impairments that exceed those expected due to age-related decline in cognition. MCI is often regarded as a precursor to dementia or a transitional state between healthy cognitive aging and dementia (Figure 1.1). In clinical practice and

research studies, MCI has been defined most commonly using the clinical criteria proposed by Petersen and colleagues (Petersen, 2004). These criteria are as follows: (1) a subjective cognitive complaint, preferably corroborated by an informant; (2) objective memory and/or other cognitive impairments that are abnormal for the individual's age and education as documented using neuropsychological testing; (3) cognitive impairments that represent a decline relative to the individual's previous abilities; (4) normal ability to perform activities of daily living; (5) absence of clinical dementia. Researchers have also proposed several subtypes of MCI based on distinct neuropsychological profiles (Petersen et al., 2009), including amnestic MCI (aMCI), which refers to MCI involving memory-only impairments; non-amnestic MCI (naMCI), which refers to MCI involving only impairments in cognitive domains other than memory (e.g. executive function/attention, language, visuospatial); and multi-domain MCI (mdMCI), which refers to MCI characterized by impairments in both memory and non-memory domains.

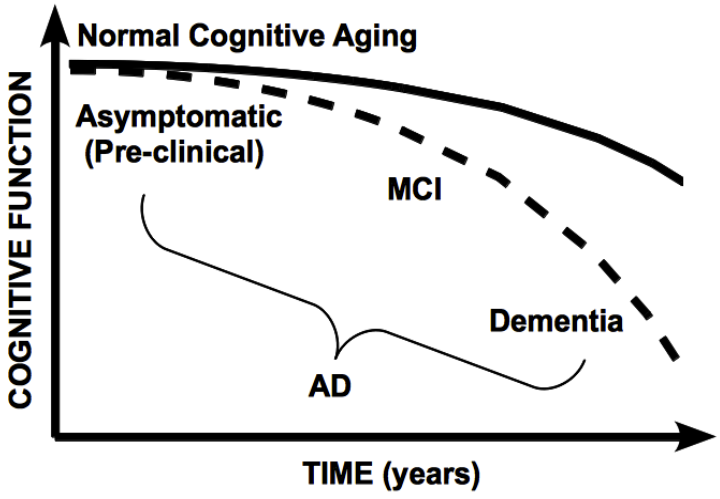


Figure 1.1. Progressive development of Alzheimer's disease (AD). The relationship among pre-clinical, mild cognitive impairment (MCI), and dementia stages of AD (dashed line) is shown relative to normal cognitive aging (solid line).

1.2.2 Epidemiology of MCI

Large population-based epidemiological studies (Busse et al., 2006; Manly et al., 2008; Petersen et al., 2009) in both the U.S. and Europe have estimated that the prevalence of MCI among adults age 65 and older is 3-24%, with higher prevalence in older individuals. Prospective longitudinal studies indicate that patients with MCI exhibit annual rates of progression to dementia of 3-15%, with highest rates for people in specialty clinic-based cohorts as compared to those in community-based cohorts (Farias et al., 2009; Mitchell and Shiri-Feshki, 2008). Overall, rates of progression from MCI to dementia are elevated well above the annual 1-2% incidence rate of dementia in the general older adult population (Petersen et al., 2009). Among MCI patients who convert to dementia, AD is the most prevalent etiology (Busse et al., 2006). However, progression risks vary according to MCI subtype; aMCI and mdMCI subtypes progress more frequently to AD whereas naMCI progresses more frequently to non-AD forms of dementia, including vascular dementia (Manly et al., 2008; Petersen et al., 2009). Furthermore, patients with mdMCI have a greater risk of developing AD than those with aMCI (Bozoki et al., 2001). While many individuals with MCI deteriorate to dementia, a substantial proportion remain cognitively stable or even improve, reverting to normal cognitive status (Manly et al., 2008). Taken as a whole, epidemiological research suggests that MCI is a heterogeneous clinical syndrome (Figure 1.2), due in part to various etiological processes. In addition to neurodegeneration, other factors have been implicated in the etiology of MCI, including: cerebrovascular, psychiatric (depression, drug abuse), demographic

(age, education), lifestyle (smoking, diet), genetic (APOE genotype, family history of dementia), and other medical conditions (brain injury, anti-cholinergic medication side effects), which can all potentially affect cognition in the elderly (Gauthier et al., 2006; He et al., 2009; Petersen et al., 2009).

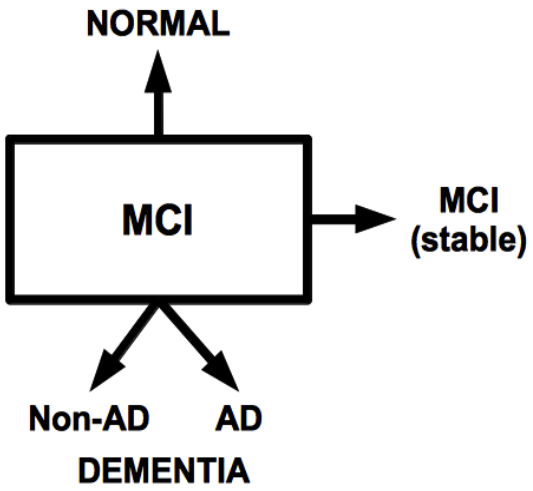


Figure 1.2. Mild cognitive impairment (MCI) is a heterogeneous syndrome. Many patients with MCI eventually develop dementia, either due to Alzheimer's disease (AD) or other causes (e.g. cerebrovascular). However, a substantial proportion of MCI patients stay cognitively stable and some even revert to normal cognitive status.

1.3 Neuroimaging of AD and MCI

In humans, a variety of non-invasive and minimally-invasive neuroimaging techniques have served as indispensable tools for studying different aspects of MCI and AD pathology. Volumetric magnetic resonance imaging (MRI) has been used to study regional patterns of brain atrophy in patients with MCI and AD (Jack et al., 1997; Rabinovici et al., 2007; Whitwell et al., 2007). Positron emission tomography utilizing 18F-fluorodeoxy-glucose (FDG-PET) as a radionuclide tracer is a nuclear imaging technique which measures regional brain metabolism. FDG-PET has proven to be of value in distinguishing different forms

of dementia, especially AD versus frontotemporal dementia, and more recently has also been applied to the study of MCI (Mosconi et al., 2008). Diffusion tensor imaging (DTI), another MRI-based technique, has been applied to the study of brain white matter in MCI and AD patients (Bozoki et al., 2012; Zhu et al., 2013). By measuring the diffusion of water molecules, DTI is able to delineate microscopic organization of white matter and allows both detailed anatomical visualization of fiber tracts and quantitative analysis of their integrity (Mori and Zhang, 2006). Many other neuroimaging approaches are being used to study AD and MCI, including functional MRI, PET-based amyloid imaging, and electroencephalography. Two integrative approaches that have great potential for elucidating the mechanisms of AD and MCI are: (1) the integration of information from multiple structural and functional neuroimaging techniques (termed “multi-modal imaging”); and (2) correlation of *in vivo* neuroimaging with cellular, molecular, and genetic characterizations of the brain.

Medial temporal lobe (MTL) atrophy, involving the hippocampus and entorhinal cortex in particular, is the earliest and most prominent MRI feature evident in AD and predicts progression from MCI to AD (Devanand et al., 2007). On volumetric MRI, AD patients also show marked enlargement of the lateral ventricles, portions of which are adjacent to the MTL (Nestor et al., 2008). In contrast, the earliest sign of AD detectable on an FDG-PET scan is the hypometabolism of the posterior cingulate cortex and precuneus; such hypometabolism is also detectable in the MCI stage of the disease (Mosconi et al., 2008). These structural and functional changes in regions of the limbic

system serve as biological substrates for learning and memory impairment, the cardinal symptom of AD. As the disease progresses, atrophy (as seen on MRI) and hypometabolism (as seen on FDG-PET) develop throughout the temporal, parietal, and frontal association cortices while primary sensorimotor and visual cortices remain relatively unaffected. In addition to cortical atrophy, AD is also characterized by degeneration of white matter within the MTL and subcortical regions and atrophy of the corpus callosum, as detectable using DTI and MRI (Bozoki et al., 2012; Salat et al., 2009).

Neuroimaging techniques not only provide a non-invasive approach to the study of AD and MCI neurobiology but also a platform for the development of potential diagnostic and prognostic biomarkers. Recognizing the importance of neuroimaging in the study of MCI and dementia, the National Institute on Aging (NIA) sponsored the ongoing Alzheimer's Disease Neuroimaging Initiative (ADNI) beginning in 2004 (Weiner et al., 2012). ADNI is the largest project of its kind that seeks to collect longitudinal neuroimaging data along with clinical data, neuropsychological assessments, and biological specimens (e.g. blood and cerebrospinal fluid) from MCI, AD, and healthy older subjects. The key goal of ADNI is to rapidly advance the knowledge on dementia, AD, MCI, and normal cognitive aging and to develop novel AD biomarkers and diagnostic approaches.

1.4 Motivations for the Present Work

The development of novel diagnostic and prognostic approaches that can serve as adjuncts to the clinical evaluation of AD and MCI is of major interest to the biomedical community (Perrin et al., 2009). Early diagnosis of AD and timely

therapeutic intervention is critical given that the disease may begin years or even decades prior to the onset of dementia. The importance of early treatment, prior to the onset of dementia, can be highlighted by clinical trial failures of several disease-modifying drugs over the last few years (Selkoe, 2011). These failed phase 2 and 3 clinical trials have typically been conducted in patients with mild-to-moderate AD dementia, at a stage when the disease process is likely irreversible and brain damage is too great. As such, greater emphasis has recently been placed on conducting clinical trials in non-demented people at risk for developing AD, such as individuals with MCI.

In this dissertation, I describe two studies investigating imaging and non-imaging biomarkers for diagnosis and prognosis of AD. In the first study (Chapter 2), DTI was used to study the effects of AD and MCI on the integrity of limbic white matter pathways. We examined whether DTI could be used to reliably detect alterations in the integrity of the fornix and cingulum (reflecting early AD-related neurodegeneration of the medial temporal lobe) in patients with AD and MCI. We also examined whether integrity of the cingulum is associated with metabolism in the posterior cingulate cortex, the first brain region to exhibit hypometabolism on FDG-PET in patients with AD. In the second study (Chapter 3), we used statistical pattern classification methods along with clinical, structural MRI, and plasma proteomic data from the ADNI database (adni.loni.ucla.edu) to develop a prognostic model of dementia designed to predict future progression from MCI to AD dementia. Classification studies are important because they directly assess the diagnostic/prognostic utility of various potential biomarkers at

the individual patient level and can help advance personalized medicine. The prognostic model developed here could be used to improve the selection of MCI patients in clinical trials and to identify high-risk patients for anti-AD prophylaxis.

REFERENCES

REFERENCES

Alzheimer, A. (1907). Über eine eigenartige Erkrankung der Hirnrinde. Allg Z Psychiatr *64*, 146–148.

Barnes, D.E., and Yaffe, K. (2011). The projected effect of risk factor reduction on Alzheimer's disease prevalence. Lancet Neurol *10*, 819–828.

Bozoki, A., Giordani, B., Heidebrink, J.L., and Foster, N.L. (2001). Mild cognitive impairments predict dementia in nondemented elderly patients with memory loss. Arch. Neurol. *58*, 411–416.

Bozoki, A.C., Korolev, I.O., Davis, N.C., Hoisington, L.A., and Berger, K.L. (2012). Disruption of limbic white matter pathways in mild cognitive impairment and Alzheimer's disease: a DTI/FDG-PET study. Hum Brain Mapp *33*, 1792–1802.

Braak, H., Thal, D.R., Ghebremedhin, E., and Del Tredici, K. (2011). Stages of the pathologic process in Alzheimer disease: age categories from 1 to 100 years. J. Neuropathol. Exp. Neurol. *70*, 960–969.

Busse, A., Hensel, A., Guhne, U., Angermeyer, M.C., and Riedel-Heller, S.G. (2006). Mild cognitive impairment: Long-term course of four clinical subtypes. Neurology *67*, 2176–2185.

Devanand, D.P., Pradhaban, G., Liu, X., Khandji, A., De Santi, S., Segal, S., Rusinek, H., Pelton, G.H., Honig, L.S., Mayeux, R., et al. (2007). Hippocampal and entorhinal atrophy in mild cognitive impairment: prediction of Alzheimer disease. Neurology *68*, 828–836.

Farias, S.T., Mungas, D., Reed, B.R., Harvey, D., and DeCarli, C. (2009). Progression of mild cognitive impairment to dementia in clinic- vs community-based cohorts. Arch. Neurol. *66*, 1151–1157.

Gauthier, S., Reisberg, B., Zaudig, M., Petersen, R.C., Ritchie, K., Broich, K., Belleville, S., Brodaty, H., Bennett, D., Chertkow, H., et al. (2006). Mild cognitive impairment. Lancet *367*, 1262–1270.

He, J., Farias, S., Martinez, O., Reed, B., Mungas, D., and Decarli, C. (2009). Differences in brain volume, hippocampal volume, cerebrovascular risk factors, and apolipoprotein E4 among mild cognitive impairment subtypes. Arch. Neurol. *66*, 1393–1399.

Hebert, L.E., Scherr, P.A., McCann, J.J., Beckett, L.A., and Evans, D.A. (2001). Is the Risk of Developing Alzheimer's Disease Greater for Women than for Men? Am. J. Epidemiol. *153*, 132–136.

Holtzman, D.M., Morris, J.C., and Goate, A.M. (2011). Alzheimer's disease: the challenge of the second century. Sci Transl Med *3*, 77sr1.

Jack, C.R., Petersen, R.C., Xu, Y.C., Waring, S.C., O'Brien, P.C., Tangalos, E.G., Smith, G.E., Ivnik, R.J., and Kokmen, E. (1997). Medial temporal atrophy on MRI in normal aging and very mild Alzheimer's disease. Neurology *49*, 786–794.

Manly, J.J., Tang, M.-X., Schupf, N., Stern, Y., Vonsattel, J.-P.G., and Mayeux, R. (2008). Frequency and course of mild cognitive impairment in a multiethnic community. Ann. Neurol *63*, 494–506.

Mann, D.M. (1996). Pyramidal nerve cell loss in Alzheimer's disease. Neurodegeneration *5*, 423–427.

Maurer, K., Volk, S., and Gerbaldo, H. (1997). Auguste D and Alzheimer's disease. Lancet *349*, 1546–1549.

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., and Stadlan, E.M. (1984). Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology *34*, 939–944.

Mitchell, A.J., and Shiri-Feshki, M. (2008). Temporal trends in the long term risk of progression of mild cognitive impairment: a pooled analysis. J. Neurol. Neurosurg. Psychiatr *79*, 1386–1391.

Mori, S., and Zhang, J. (2006). Principles of diffusion tensor imaging and its applications to basic neuroscience research. Neuron *51*, 527–539.

Mosconi, L., Tsui, W.H., Herholz, K., Pupi, A., Drzezga, A., Lucignani, G., Reiman, E.M., Holthoff, V., Kalbe, E., Sorbi, S., et al. (2008). Multicenter standardized 18F-FDG PET diagnosis of mild cognitive impairment, Alzheimer's disease, and other dementias. J. Nucl. Med *49*, 390–398.

Nestor, S.M., Rupsingh, R., Borrie, M., Smith, M., Accomazzi, V., Wells, J.L., Fogarty, J., and Bartha, R. (2008). Ventricular enlargement as a possible measure of Alzheimer's disease progression validated using the Alzheimer's disease neuroimaging initiative database. Brain *131*, 2443–2454.

Norfray, J.F., and Provenzale, J.M. (2004). Alzheimer's disease: neuropathologic findings and recent advances in imaging. AJR Am J Roentgenol *182*, 3–13.

Ott, A., Breteler, M.M., Van Harskamp, F., Claus, J.J., Van der Cammen, T.J., Grobbee, D.E., and Hofman, A. (1995). Prevalence of Alzheimer's disease and vascular dementia: association with education. The Rotterdam study. BMJ *310*, 970–973.

Perrin, R.J., Fagan, A.M., and Holtzman, D.M. (2009). Multimodal techniques for

diagnosis and prognosis of Alzheimer's disease. Nature *461*, 916–922.

Petersen, R.C. (2004). Mild cognitive impairment as a diagnostic entity. J. Intern. Med *256*, 183–194.

Petersen, R.C., Roberts, R.O., Knopman, D.S., Boeve, B.F., Geda, Y.E., Ivnik, R.J., Smith, G.E., and Jack, C.R. (2009). Mild cognitive impairment: ten years later. Arch. Neurol *66*, 1447–1455.

Plassman, B.L., Langa, K.M., Fisher, G.G., Heeringa, S.G., Weir, D.R., Ofstedal, M.B., Burke, J.R., Hurd, M.D., Potter, G.G., Rodgers, W.L., et al. (2007). Prevalence of Dementia in the United States: The Aging, Demographics, and Memory Study. Neuroepidemiology *29*, 125–132.

Prince, M., Bryce, R., Albanese, E., Wimo, A., Ribeiro, W., and Ferri, C.P. (2013). The global prevalence of dementia: a systematic review and meta-analysis. Alzheimers Dement *9*, 63–75.e2.

Querfurth, H.W., and LaFerla, F.M. (2010). Alzheimer's disease. N. Engl. J. Med *362*, 329–344.

Rabinovici, G.D., Seeley, W.W., Kim, E.J., Gorno-Tempini, M.L., Rascovsky, K., Pagliaro, T.A., Allison, S.C., Halabi, C., Kramer, J.H., Johnson, J.K., et al. (2007). Distinct MRI Atrophy Patterns in Autopsy-Proven Alzheimer's Disease and Frontotemporal Lobar Degeneration. Am J Alzheimers Dis Other Demen *22*, 474–488.

Reiman, E.M., Chen, K., Alexander, G.E., Caselli, R.J., Bandy, D., Osborne, D., Saunders, A.M., and Hardy, J. (2005). Correlations between apolipoprotein E epsilon4 gene dose and brain-imaging measurements of regional hypometabolism. Proc. Natl. Acad. Sci. U.S.A *102*, 8299–8302.

Salat, D.H., Greve, D.N., Pacheco, J.L., Quinn, B.T., Helmer, K.G., Buckner, R.L., and Fischl, B. (2009). Regional white matter volume differences in nondemented aging and Alzheimer's disease. Neuroimage *44*, 1247–1258.

Selkoe, D.J. (2002). Alzheimer's disease is a synaptic failure. Science *298*, 789–791.

Selkoe, D.J. (2011). Resolving controversies on the path to Alzheimer's therapeutics. Nat Med *17*, 1060–1065.

Stern, Y. (2012). Cognitive reserve in ageing and Alzheimer's disease. Lancet Neurol *11*, 1006–1012.

Thies, W., and Bleiler, L. (2013). 2013 Alzheimer's disease facts and figures. Alzheimers Dement *9*, 208–245.

Weiner, M.W., Veitch, D.P., Aisen, P.S., Beckett, L.A., Cairns, N.J., Green, R.C., Harvey, D., Jack, C.R., Jagust, W., Liu, E., et al. (2012). The Alzheimer's Disease Neuroimaging Initiative: a review of papers published since its inception. Alzheimers Dement *8*, S1–68.

Whitwell, J.L., Petersen, R.C., Negash, S., Weigand, S.D., Kantarci, K., Ivnik, R.J., Knopman, D.S., Boeve, B.F., Smith, G.E., and Jack, C.R., Jr (2007). Patterns of atrophy differ among specific subtypes of mild cognitive impairment. Arch. Neurol. *64*, 1130–1138.

Zhu, D.C., Majumdar, S., Korolev, I.O., Berger, K.L., and Bozoki, A.C. (2013). Alzheimer's disease and amnestic mild cognitive impairment weaken connections within the default-mode network: a multi-modal imaging study. J. Alzheimers Dis. *34*, 969–984.

CHAPTER 2

Disruption of Limbic White Matter Pathways in Mild Cognitive Impairment and Alzheimer's Disease: A DTI / FDG-PET Study[#]

[#] This chapter is written in the style of a journal article and has been published in the journal *Human Brain Mapping*. The content of the article is reproduced with permissions from John Wiley and Sons, Inc. (Copyright 2011, Wiley Periodicals, Inc). The citation for the original article is as follows:

Bozoki AC*, Korolev IO*, Davis NC, Hoisington LA, Berger KL. Disruption of limbic white matter pathways in mild cognitive impairment and Alzheimer's disease: a DTI/FDG-PET study. *Hum Brain Mapp*. 2012;33(8):1792–1802. (*Equal contributions)

2.1 Abstract

Background: Alzheimer's disease (AD) and mild cognitive impairment (MCI) affect the limbic system, causing medial temporal lobe (MTL) atrophy and posterior cingulate cortex (PCC) hypometabolism. Additionally, diffusion tensor imaging (DTI) studies have demonstrated that MCI and AD involve alterations in cerebral white matter (WM) integrity. *Objectives*: To test if 1) patients with MCI and AD exhibit decreases in the integrity of limbic WM pathways; 2) disconnection between PCC and MTL, manifested as disruption of the cingulum bundle, contributes to PCC hypometabolism during incipient AD. *Methods*: We measured fractional anisotropy (FA) and volume of the fornix and cingulum using DTI in 23 individuals with MCI, 21 with mild-to-moderate AD, and 16 normal control (NC) subjects. We also measured PCC metabolism using FDG-PET in AD and MCI patients. *Results*: Fornix FA and volume were reduced in MCI and AD to a similar extent. Descending cingulum FA was reduced in AD while volume was reduced in MCI and even more so in AD. Both FA and volume of the fornix and descending cingulum reliably discriminated between NC and AD. Fornix FA and descending cingulum volume also reliably discriminated between NC and MCI. Only descending cingulum volume reliably discriminated between MCI and AD. In the combined MCI-AD cohort, PCC metabolism directly correlated with both FA and volume of the descending cingulum. *Conclusions*: Disruption of limbic WM pathways is evident during both MCI and AD. Disconnection of the PCC from MTL at the cingulum bundle contributes to PCC hypometabolism during incipient AD.

2.2 Introduction

Alzheimer's disease (AD) is the most common cause of dementia among older adults and traditionally conceptualized as a disease of brain grey matter, with particular involvement of the medial temporal lobe (MTL), temporoparietal cortex, and posterior cingulate cortex (PCC). However, studies with diffusion tensor imaging (DTI), an MRI technique that assesses microstructural organization of white matter (WM) based on measurements of water diffusion, have demonstrated that AD also involves alterations in WM integrity. Patients with probable AD have a reduction in the integrity of WM tracts connecting regions of association cortex, such as the splenium of the corpus callosum, superior longitudinal fasciculus, and cingulum bundle, while integrity of the pyramidal tracts (e.g. posterior limb of the internal capsule) is generally preserved (Rose et al., 2000; Bozzali et al. 2002; Fellgiebel et al. 2008). The likely pathogenic mechanism of these WM changes in AD is Wallerian degeneration of fiber tracts due to neuronal loss in cortical associative areas affected by AD pathology (Braak and Braak 1996; Bozzali et al., 2002).

DTI has also been extended to study older adults with Mild Cognitive Impairment (MCI), who have an increased risk of developing dementia relative to their healthy counterparts. Fellgiebel et al. (2005) showed that fractional anisotropy (FA), a measure of WM integrity, differed significantly between controls and both MCI and AD patients in regions affected by AD neurodegeneration. They also showed that performance of the MCI-AD group on the delayed verbal recall test was correlated with posterior cingulum bundle FA. Other studies have shown that FA of cingulum fibers are significantly reduced in MCI, and even more so in AD (Liu et al. 2009; Zhang et al., 2007; Choo et

al., 2008). DTI studies of limbic WM in MCI and AD show some discrepancy, however. Kiuchi et al. (2009), using DTI tractography, found decreased FA in the posterior cingulate fasciculus of MCI patients (relative to normal controls) but no difference between MCI and AD groups. In contrast, Mielke et al. (2009), utilizing a region-of-interest (ROI) technique, found FA differences between MCI and AD subjects in a different region of the cingulum and the fornix but no differences between NC and MCI groups in these areas. In general, a number of studies in which limbic FA of all 3 groups are examined (normal controls, MCI and AD) have not been able to demonstrate a 3-way distinction; either the normal and MCI groups are different from the ADs, but not from each other, or MCI and AD reveal similar changes in FA, different from the controls.

In addition to FA alone, some recent studies have employed multimodal MRI to obtain simultaneous measures of FA and volume. The most common way to do this has been with an ROI-based assessment of FA combined with voxel-based morphometry, as was employed by Walhovd et al (2009). This study showed that higher diagnostic accuracy was achieved when multiple methods and ROIs were combined, though in their hands, morphometry showed the best diagnostic (predictive) sensitivity. In our study, as in some others (Zhang et al., 2007; Thomas et al., 2008; Mielke et al., 2009), a tract-based approach is used to obtain the volumetric information as well as the mean FA of that volume.

WM degeneration has been proposed as a mechanism for the well-documented functional deficit of the PCC observed on FDG-PET scans, a finding often present during early stages of AD in parallel with MTL atrophy (Minoshima et al., 1997). This hippocampal-neocortical “disconnection” hypothesis posits that PCC hypometabolism

occurring during incipient AD may be due to a distant effect of neuronal dysfunction in the MTL via disconnection at the cingulum bundle, an idea first put forth in the early 1990s (Jobst et al., 1992) with more recent imaging support from FDG-PET (Nestor et al., 2003) and combined FDG-PET/MRI studies (Meguro et al., 2001; Chételat et al., 2008; Villain et al., 2008),. Consistent with this hypothesis, Zhou et al (2008) used DTI and fMRI to show that there is decreased structural connectivity from both the PCC and hippocampus to the whole brain as well as a reduction in functional connectivity between these two regions in MCI and AD. Few studies, however, have tested the "disconnection" hypothesis directly.

The present study was designed to (1) address the discrepancies from prior DTI studies of limbic WM in MCI and AD; and (2) directly test the "disconnection" hypothesis of PCC hypometabolism via a combined structural-functional approach. Specifically, we hypothesized that patients with both MCI and AD should exhibit decreases in the integrity of limbic WM pathways, reflective of the neuronal dysfunction and atrophy in the MTL. We used DTI to investigate changes in FA and volume of the fornix and cingulum on a continuum from normal aging to MCI to AD. We also investigated the relationship between integrity of the cingulum bundle and PCC hypometabolism with combined DTI and FDG-PET.

2.3 Methods

2.3.1 Participants

MCI and AD subjects were recruited from the Cognitive Disorders Clinic at Michigan State University between 2005 and 2007. Healthy senior participants were recruited from a database of individuals who volunteered to participate in response to

fliers and paid advertisements in the Lansing, Michigan community. They were selected based on demographics in order to match the MCI and AD subjects for age, gender, and educational attainment. The final group consisted of 23 patients with MCI, 21 patients with mild to moderate probable AD and 16 normal controls (NC). Table 2.1 shows demographic information and Mini Mental State Examination (MMSE) scores (Folstein et al., 1975) for the three subject groups.

	NC	MCI	AD
Age	65.9 (8.5)	70.8 (7.9)	71.6 (10.6)
Gender (M/F)	6/10	12/11	6/15
Education (years)	16.0 (3.3)	15.6 (3.2)	14.3 (3.3)
MMSE score	29.5 (0.7)	26.8 (2.5)	22.0 (4.5)

Table 2.1. Patient demographic information and MMSE scores.

Patients from the Cognitive Disorders Clinic underwent a standardized assessment for dementia. Based on results of this evaluation, all patients were classified according to NINCDS-ADRDA criteria for the clinical diagnosis of probable AD (McKhann et al, 1984; Tierney et al., 1988), and Petersen's modified criteria for the diagnosis of MCI (Petersen, 2004). Exclusion criteria included a history of loss of consciousness for >10 minutes, stroke, psychosis, bipolar disorder, alcoholism, or substance abuse. For healthy seniors, any neurologic disease was considered exclusionary; for MCI and AD subjects, neurologic co-morbidities were exclusionary. For all subjects, a history of depression was acceptable only if the condition was currently in remission, with or without medication. Medications, including cholinesterase inhibitors, were permitted. All subjects or their authorized legal guardians provided written informed

consent. The procedures in the present study were approved by the Michigan State University Committee on Research Involving Human Subjects.

2.3.2 Image Acquisition

All subjects underwent DTI scans. However, FDG-PET scans were administered only to MCI and AD patients, for whom the imaging was justified on clinical grounds, but not normal control subjects, to avoid unnecessary exposure to the radioactive tracer. Imaging was conducted according to the following protocol.

DTI: General Electric Signa Excite 3 Tesla scanner was used for data acquisition. For each scan (scan time = 4 minutes 50 seconds), 40 axial slices were collected using a spin echo EPI pulse sequence with TE = 69.3 ms, TR = 10000 ms, in-plane resolution = 3 mm, slice thickness = 3 mm, inter-slice gap = 0 mm, 240 mm FOV (80 x 80 matrix), and NEX = 4. Diffusion encoding was accomplished in six non-collinear directions with b-value of 1000 s/mm^2. DTI images were interpolated on the scanner to a voxel size of 0.9375 x 0.9375 x 3 mm^3.

FDG-PET: General Electric Discovery STE scanner was used for data acquisition. Images were acquired following 10 mCi injection of ^{18}F- fluorodeoxyglucose (FDG) and 30 minute uptake time using 3-D acquisition over 9 minutes in three 3-minute dynamic frames. CT attenuation correction was used.

2.3.3 Data Processing

DTI: After transferring raw DTI data to a separate computer workstation, they were interpolated to an isotropic voxel size of 1.5 mm^3 using tools from the FMRIB Software Library (Smith et al., 2004); during this step, DTI data were also converted

from the native DICOM format to Analyze format. Data were further processed using MRI Studio (www.mristudio.org) (Jiang et al., 2006) to calculate whole-brain diffusion tensor and FA maps for subsequent fiber tracking. 3-D reconstructions of the fornix, cingulum bundle, its associated parietal projection fibers, and posterior limb of the internal capsule (a control region) were accomplished by using computational tractography algorithms implemented in MRI Studio. For each tract, multiple 2-D region of interest (ROI) “seeds”, selected based on a priori knowledge of tract anatomy, were placed on DTI maps created within MRI Studio (Figure 2.1 a-c); an FA “colormap”, indicating fiber orientation, aided in the placement of ROI seeds. Fibers were then tracked voxel-by-voxel along the primary eigenvector (i.e. vector of the largest eigenvalue), constrained by the ROI seeds. An FA threshold of ≥0.2 and curvature threshold of ≥60º were used to further constrain the course of fiber tracts in the brain. Some post-hoc “pruning” of tracts was necessary in order to remove extraneous fibers if they were determined to be inconsistent with known tract anatomy. Figure 2.1d shows 3-D reconstructions of the fornix, cingulum, and internal capsule from a representative subject. A neurologist (author A. B) checked the tractography results for each subject. Fiber tracts and FA maps were subsequently re-oriented into standard orientation relative to the AC-PC and horizontal transverse axes of an anatomical template via 12-mode affine transformation. These re-oriented fiber tracts and FA maps were then exported in a binary format for further analysis in MATLAB (MathWorks, Natick, MA), which was used to “crop” specific tract ROI’s based on a standardized set of anatomical landmarks (Figure 2.1d).

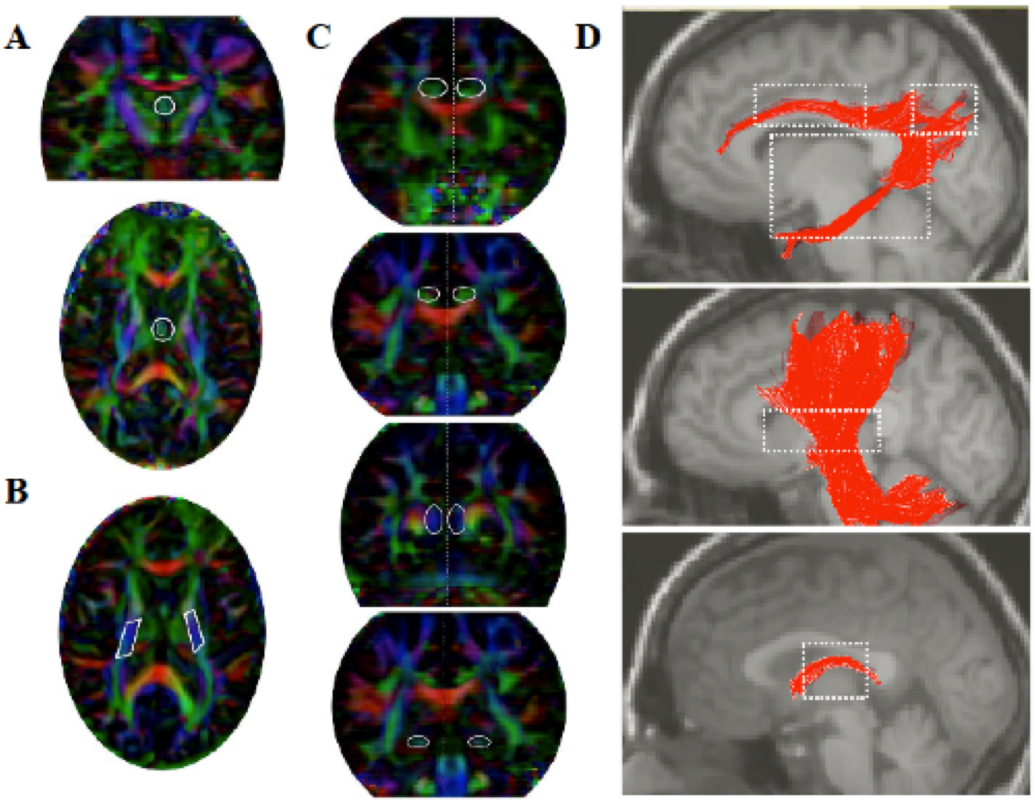


Figure 2.1. DTI tractography. ROI seeds (white solid lines) displayed on FA colormaps as used for tracking the fornix (A), internal capsule (B), and cingulum (C). 3-D reconstructions of the cingulum bundle (D, top), fornix (D, bottom), and internal capsule (D, middle) from a representative normal control subject are shown overlaid on T1-weighted sagittal MRI, with specific tract ROIs that were defined for quantitative analysis marked (dotted white line; see text for details). For interpretation of the references to color in this and all other figures, the reader is referred to the electronic version of this dissertation.

The landmark boundaries for each ROI, identified on the non-diffusion weighted DTI image for each subject, were as follows: The *fornix body* was defined as that portion of the tract posterior to the fornix columns and anterior to the crura. The cingulum bundle was divided into three portions: *superior cingulum* was defined as that

portion of the tract posterior to the genu and anterior to the splenium of corpus callosum; *descending cingulum* was defined as that portion of the tract below superior aspect of the fornix body and above inferior aspect of the temporal horn; *parietal projection fibers* were defined as that portion of the tract above superior aspect of the fornix body and posterior to the descending cingulum. *Posterior limb of the internal capsule* was defined as that portion of the tract below superior aspect of the fornix body and above superior aspect of the midbrain as well as posterior to the genu and anterior to the splenium of corpus callosum. Then, average FA and volume were computed for each ROI, as measures of tract integrity, for subsequent statistical analysis. Certain steps of the processing stream were performed manually (seed placement, post-hoc pruning of fibers, and definition of ROI boundaries) by an operator (author N. D.), who was blinded to subjects' diagnostic group.

FDG-PET: FDG-PET data were analyzed using an automated voxel-based approach involving the 3-D stereotactic surface projection technique (Minoshima et al., 1995), as implemented within Cortex ID software (GE Healthcare). Subject images were first reconstructed using a filter back-projection method (matrix 128^2, FOV 30 cm^2) to match the resolution of the program's age-matched normal subject database. Then, pons-normalized Z-scores (representing standardized glucose metabolism) were computed for posterior cingulate, medial parietal, parietal association, temporal association and visual cortices for each subject relative to the age-matched, normal subject database. These regions of interest were defined using a PET atlas based on a stereotactic localization technique, as developed by Minoshima et al. (1994). Figure 2.2 shows representative MCI and AD subjects analyzed with the above method.

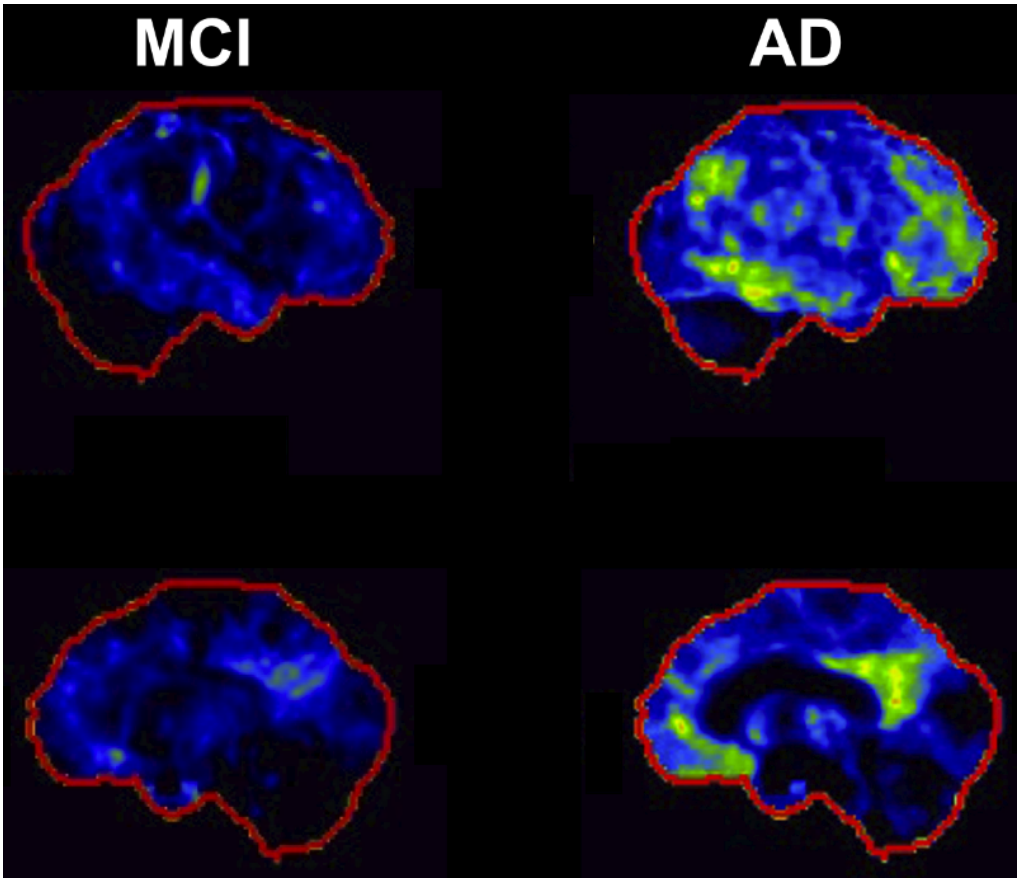


Figure 2.2. FDG-PET Z-score map of cortical metabolism in representative subjects with MCI and early AD. Data were analyzed using an automated voxel-based approach involving the 3-D stereotactic surface projection technique (Minoshima et al., 1994; 1995). Pons-normalized Z-scores (representing standardized glucose metabolism) were computed for each MCI and AD patient relative to the age-matched normal subject database. In the MCI subject's scan, spatial clustering of Z-scores can be seen in posterior cingulate and medial parietal cortices. In the AD subject's scan, a more diffuse hypometabolism pattern can be seen that extends to temporoparietal and frontal cortices. Images show right lateral (top) and medial (bottom) views of the brain. Light blue and green-yellow colors represent pronounced hypometabolism.

2.3.4 Statistical analysis

SPSS Statistics (version 17; IBM, Chicago, IL) was used for statistical analysis. Univariate analysis of variance (ANOVA) and post-hoc t-tests were used to assess the effect of group on tract FA and volume and PET Z-scores. Receiver operating characteristic (ROC) analysis was used to determine the discriminative power of different DTI measurements. Pearson (bivariate) correlations were used to assess the relationship between tract FA (or volume) and PET Z-scores as well as the relationship between DTI /PET measurements and MMSE performance. All p-values were adjusted for multiple comparisons using Bonferroni correction.

2.4 Results

2.4.1 Demographic Information

Age, gender composition, and years of education (Table 2.1) did not significantly differ among the three groups. As expected, MMSE scores (Table 2.1) differed among the three groups [$F(2,54) = 27.9$, $p<0.001$], with NC exhibiting greater MMSE scores than both MCI ($p<0.001$) and AD ($p<0.001$) and MCI exhibiting greater MMSE scores than AD ($p<0.01$).

2.4.2 DTI data

FA and volume: Table 2.2 shows the means and standard deviations of tract FA and volume for NC, MCI, and AD groups as well as ANOVA results. Figure 2.3 shows example fiber tracts (fornix and descending cingulum) from the three groups. Although DTI measurements differed between hemispheres for some regions, this lateralization

was independent of group membership, and we therefore combined data across both hemispheres in our analysis.

Fornix: FA differed across groups, with NC exhibiting greater FA than both MCI (p=0.001) and AD (p<0.001); however, FA did not differ between MCI and AD. Similarly, volume differed across groups, with NC exhibiting greater volume than both MCI (p=0.02) and AD (p<0.001); however, volume did not differ between MCI and AD.

Descending cingulum: FA differed across groups, with NC exhibiting greater FA than AD (p=0.001); however, MCI did not differ in this regard from NC or AD. Volume also differed across groups, with NC exhibiting greater volume than both MCI and AD (both p<0.001) and MCI exhibiting greater volume than AD (p=0.004).

Parietal projection fibers: FA did not differ across groups but volume did, with NC exhibiting greater volume than AD (p=0.001). MCI did not differ in this regard from NC or AD.

Superior cingulum and *internal capsule (control)*: Neither FA nor volume of these structures differed across groups.

Age as a covariate: Due to the five-year mean age difference between NC and MCI/AD groups, we re-analyzed FA and volume data with age as a covariate, and the results above did not change.

Region	NC	MCI	AD	*F* (2,57)
FA				
Fornix	0.31 (0.02)	0.28 (0.01)	0.28 (0.02)	12.2**
Superior cingulum	0.40 (0.02)	0.39 (0.02)	0.38 (0.03)	n.s.
Parietal projection fibers	0.35 (0.01)	0.34 (0.02)	0.33 (0.02)	n.s.
Descending cingulum	0.36 (0.02)	0.34 (0.02)	0.33 (0.02)	7.3*
Internal capsule	0.51 (0.04)	0.50 (0.04)	0.49 (0.03)	n.s.
Volume (ml)				
Fornix	1.8 (0.5)	1.3 (0.5)	1.1 (0.5)	8.5*
Superior cingulum	3.4 (0.7)	3.3 (0.8)	3.2 (0.8)	n.s.
Parietal projection fibers	2.0 (0.8)	1.7 (0.9)	1.1 (0.6)	7.2*
Descending cingulum	4.6 (0.8)	3.5 (0.8)	2.7 (0.8)	25.3**
Internal capsule	7.0 (1.4)	7.3 (2.3)	6.7 (1.8)	n.s.

Table 2.2. Means and standard deviations of tract FA and volume. *F*-statistic for group ANOVA shown (Bonferroni-corrected *p*-values: * $p<0.05$; ** $p<0.01$; n.s., not significant).

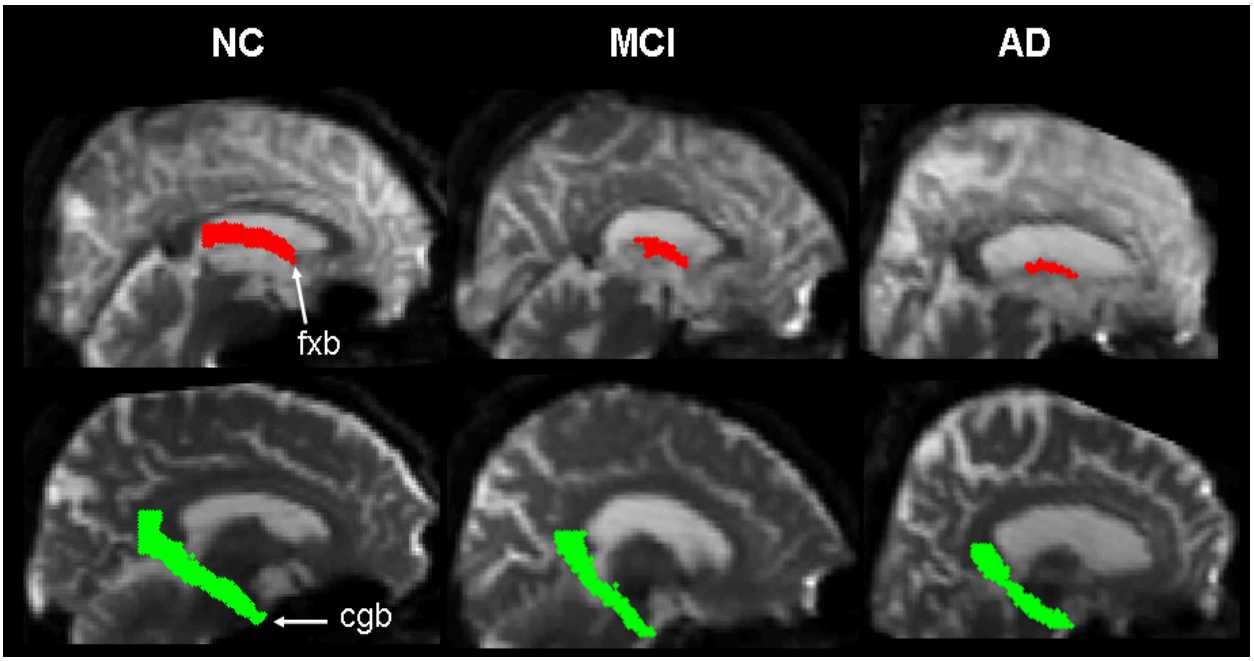


Figure 2.3. Examples of fornix and descending cingulum fiber tracts. Fornix body (fxb) is shown in red and right descending cingulum bundle (cgb) in green for three representative subjects, one from each of the groups (NC, MCI, AD). B0 DTI image is shown in the background (midsagittal slice at the level of fornix body and parasagittal slice at the level of the descending cingulum).

ROC analysis: Table 2.3 shows the area under the curve (AUC), a statistic indicating the discriminative power, for tract FA and volume measurements. Both the FA and volume of fornix, parietal projection fibers and descending cingulum discriminated between NC and AD groups with moderately high power (AUC range: 0.80-0.96, all $p<0.05$). FA of the fornix and volume of the descending cingulum also reliably discriminated between NC and MCI groups (all $p<0.05$). Only descending cingulum volume reliably discriminated between MCI and AD groups ($p<0.05$).

Region	NC vs. AD	NC vs. MCI	MCI vs. AD
FA			
Fornix	0.86*	0.84*	0.56
Superior cingulum	0.67	0.60	0.59
Parietal projection fibers	0.80*	0.73	0.56
Descending cingulum	0.81*	0.62	0.69
Volume			
Fornix	0.86*	0.73	0.66
Superior cingulum	0.58	0.57	0.52
Parietal projection fibers	0.82*	0.66	0.69
Descending cingulum	0.96*	0.79*	0.79*

Table 2.3. ROC analysis statistics. Area under the curve (AUC) as a measure of discriminative power (Bonferroni-corrected *p*-values: * $p<0.05$). Asymptotic *p*-values for the ROC curve were computed based on an assumption that the AUC distribution is nonparametric and the null hypothesis that AUC = 0.5.

2.4.3 FDG-PET data

Table 2.4 shows the means and standard deviations of PET Z-scores for MCI and AD groups. Although Z-scores differed between hemispheres for some regions, this lateralization was independent of group membership, and we combined data across both hemispheres in our analysis. Relative to a database of normal control subjects, the MCI-AD cohort exhibited decreased Z-scores in parietal, temporal, and posterior cingulate cortices (all $p<0.01$) but not visual cortex (control region). The AD group had

lower Z-scores than the MCI group in parietal and temporal association cortices and medial parietal cortex (all p<0.05, uncorrected), but these effects did not survive after correcting for multiple comparisons. The AD group showed lower Z-scores than the MCI group when parietal, temporal, and posterior cingulate ROIs were averaged together (t(41)=2.4, p<0.05).

Cortical Region	MCI	AD
Parietal association	-1.0 (0.8)*	-2.0 (1.6)*
Temporal association	-1.1 (0.6)*	-1.8 (1.3)*
Posterior cingulate	-1.4 (0.5)*	-1.7 (0.8)*
Medial parietal	-1.1 (0.7)*	-1.8 (1.3)*
Visual	0.0 (0.9)	-0.5 (1.3)

Table 2.4. Means and standard deviations of FDG-PET Z-scores. Z-scores represent standardized glucose metabolism relative to the normal subject database (Bonferroni-corrected *p*-values: * *p*<0.01).

Correlation with DTI: We examined the relationship between posterior cingulate hypometabolism (a feature seen early in the course of AD) and integrity of superior and descending cingulum (major input and output pathways of the posterior cingulate). Pearson correlations (Figure 2.4) across the combined MCI-AD cohort revealed that posterior cingulate PET Z-scores (i.e. metabolism) correlated with both FA and volume of the descending cingulum (both r=0.35, *p*<0.05) but not with that of the superior cingulum. To further examine the specificity of this structure-function relationship, we asked whether descending cingulum FA and volume correlated with PET Z-scores of any other closely associated regions involved in AD. As shown in Table 2.5, descending cingulum integrity (FA and volume) selectively correlated with posterior cingulate Z-scores and not that of the medial parietal, parietal association, or temporal association cortices.

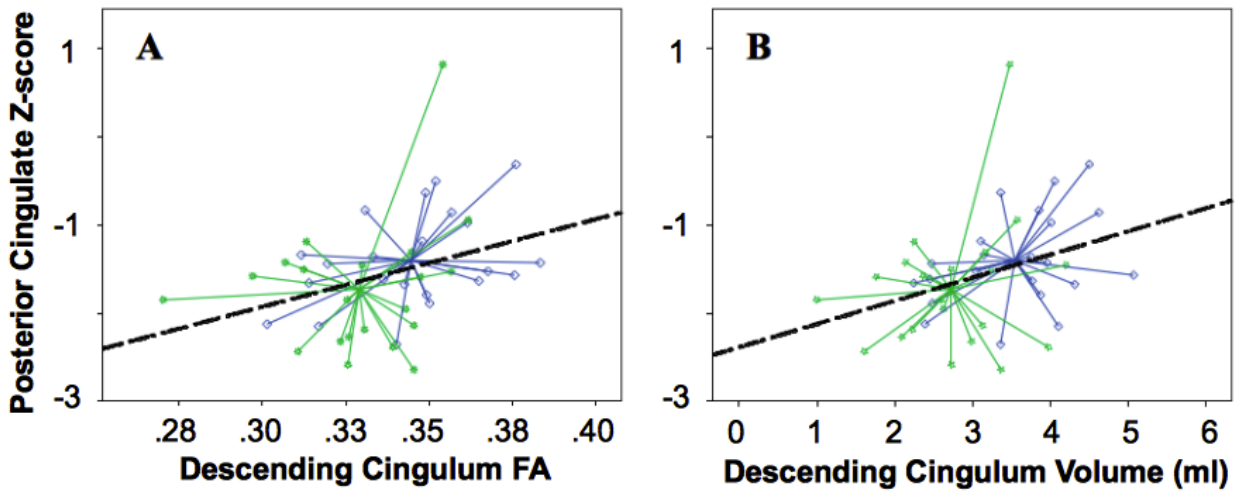


Figure 2.4. Posterior cingulate hypometabolism is associated with descending cingulum integrity. Scatter plots show correlations between posterior cingulate PET Z-scores and descending cingulum FA (A) and volume (B) in the MCI-AD cohort. Both r=0.35, $p<0.05$ (Bonferroni-corrected). Fit lines are shown. Group centroids are shown separately for MCI (blue diamonds) and AD (green stars) subjects.

	Descending Cingulum	
Regional PET Z-scores	*FA*	*Volume*
Parietal association	0.11	0.17
Temporal association	0.20	0.15
Posterior cingulate	0.35*	0.35*
Medial parietal	0.15	0.23

Table 2.5. Relationship between descending cingulum integrity and regional metabolism. Pearson correlation coefficients are shown, indicating strength of the relationship between descending cingulum FA / volume and regional PET Z-scores for the posterior cingulate, parietal, and temporal cortices (* $p<0.05$, Bonferroni-corrected).

2.4.4 DTI and PET correlation with MMSE performance

For regional measures of tract integrity and metabolism that exhibited differences across groups, we tested their association with MMSE performance. Specifically, MMSE

scores correlated with fornix FA (r=0.42, p<0.01) and volume (r=0.45, p<0.01), descending cingulum FA (r=0.45, p<0.01) and volume (r=0.52, p<0.01), parietal projection volume (r=0.48, p<0.01) as well as posterior cingulate (r=0.38, p=0.06), parietal association (r=0.43, p<0.05), temporal association (r=0.50, p<0.01) and medial parietal (r=0.41, p<0.05) PET Z-scores. Figure 2.5 shows scatter plots for some of these relationships. We also asked to what extent PCC metabolism and descending cingulum integrity independently predicted MMSE performance. Partial correlation analysis revealed that the relationship between MMSE scores and descending cingulum FA (r=0.37, p<0.05) and volume (r=0.45, p<0.01) remained significant after controlling for PCC metabolism. In contrast, the relationship between MMSE scores and PCC metabolism was not statistically significant (r=0.23, p=0.2) after controlling for descending cingulum FA and volume, even when using a more liberal α-level of 0.1.

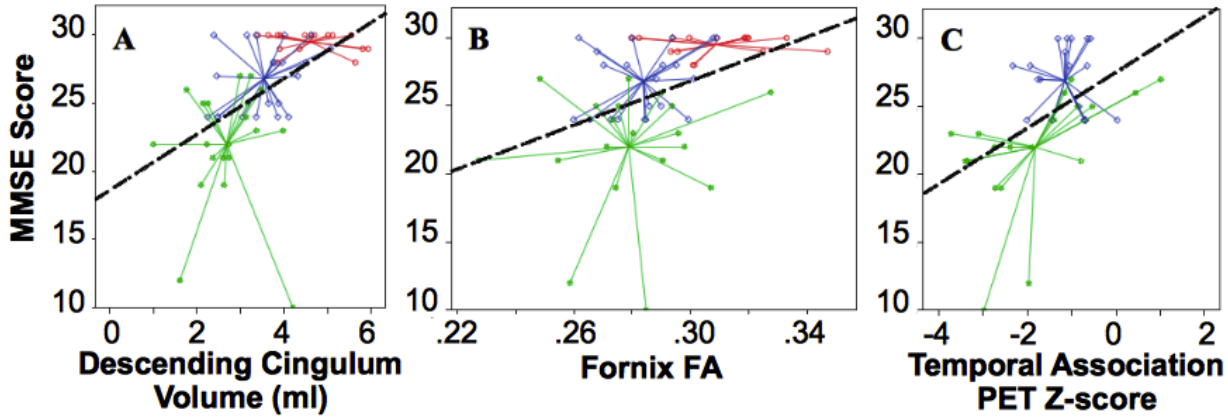


Figure 2.5. MMSE performance is associated with limbic tract integrity and cortical metabolism. Scatter plots show correlations between MMSE scores and descending cingulum volume (A; r=0.52, p<0.01), fornix FA (B; r=0.42, p<0.01) and temporal association cortex PET Z-scores (C; r=0.50, p<0.01) in the MCI-AD cohort. Fit lines and Bonferroni-corrected p-values are shown. Group centroids are shown separately for NC (red circles), MCI (blue diamonds) and AD (green stars) subjects.

2.5 Discussion

The results of this study suggest that fornix integrity declines during the transition from normal aging to MCI. In contrast, descending cingulum integrity declines during both the transition from normal aging to MCI and the transition from MCI to AD. The posterior limb of the internal capsule, which is part of the motor system, did not show significant changes in its integrity across the normal-to-AD continuum. Similarly, the superior cingulum, which is also part of the limbic system but less intimately associated with the MTL memory system, was also unaffected. Thus, our results indicate that white matter degeneration shows selectivity during early-to-moderate stages of AD, affecting pathways which mediate memory function (i.e. fornix and descending cingulum) more prominently than pathways related to non-memory function (e.g. superior cingulum and internal capsule). This is consistent with previous findings that intracortical (e.g. limbic) projecting fibers are affected preferentially early in the course of AD, at a point in time when extracortical fiber tracts are still relatively preserved (Braak and Braak, 1996; Teipel et al.,2007). Furthermore, our ROC analysis indicates that the integrity of fornix and descending cingulum (measured as FA and/or volume) can distinguish among NC, MCI, and AD groups. Our findings are consistent with other recent DTI studies that have reported declines in fornix and cingulum FA in both MCI and AD patients (Xie et al., 2005; Teipel et al., 2007; Zhang et al., 2007; Kiuchi et al., 2009; Liu et al., 2009; Mielke et al., 2009).

Our DTI findings are consistent with the idea that degeneration of the fornix and cingulum bundle, which serve as major input and output pathways of the hippocampal formation (Duvernoy, 1998), likely occurs secondary to progressive hippocampal

damage in MCI and AD. Furthermore, the observed decline in fornix and cingulum integrity on a continuum from normal-to-MCI-to-AD was further corroborated by significant positive relationships between fornix/cingulum integrity and MMSE scores, as a surrogate of disease severity.

Although we found differences in PET metabolism between the combined MCI/AD cohort and the normal subject database, we did not see robust differences between individual patient groups. We believe that this is related to the heterogeneity of our MCI sample, which included amnestic and non-amnestic MCI individuals, as well as those with multi-domain deficits. We deliberately sought to include all MCI subtypes both because it better reflects the actual patient population with this condition, and because it is far from clear in the existing literature whether these subtypes represent different underlying etiologies.

Results from the DTI-PET correlation analysis indicate that PCC hypometabolism is, at least in part, related to descending cingulum integrity in the combined MCI/AD cohort (Figure 2.4). In our study, this structure-function relationship exhibited anatomical specificity (as opposed to being a global effect) given that such a relationship was absent for the superior cingulum and parietotemporal regions (Table 2.5). This finding lends support to the hippocampal-neocortical “disconnection” hypothesis of PCC hypometabolism in early AD, which posits that PCC hypometabolism may be due to a distant effect of neuronal dysfunction in the MTL via disconnection at the cingulum bundle (Jobst et al., 1992; Meguro et al., 2001; Nestor et al., 2003; Chételat et al., 2008).

Nevertheless, the structure-function relationship between descending cingulum and PCC was modest in the present study, accounting for ~12% of the variance (r=0.35). This implies that there may be other concurrent mechanisms contributing to PCC hypometabolism, independent of its connection to the MTL via the cingulum bundle. Local damage (i.e. neuronal dysfunction and/or atrophy) in the PCC serves as one such alternate mechanism for causing PCC hypometabolism (Baron et al., 2001; Chételat et al., 2002, 2009). Damage in the PCC may result, for instance, from local β-amyloid accumulation that reaches critical levels during early stages of AD, as measured by PET imaging (Buckner et al., 2005; Sperling et al., 2009). Interestingly, we found that descending cingulum integrity predicted MMSE performance independently of PCC metabolism but not vice-versa. This suggests that the association between cognitive decline (as measured by MMSE performance) and PCC dysfunction in our study was secondary to disruption of the descending cingulum, consistent with the "disconnection" hypothesis.

To our knowledge, few studies have tested this hypothesis directly by correlating measurements of the cingulum bundle with either structural or functional properties of the PCC**.** Villain and colleagues (2008; 2010) used voxel-based morphometric analysis of volumetric MRI data and FDG-PET to show that hippocampal atrophy was related to cingulum bundle atrophy, which itself was associated with PCC hypometabolism in MCI and early AD patients. Our findings complement those of Villain et al. by providing DTI-based evidence for the disconnection hypothesis. Similar to our study, Choo and colleagues (2008) also investigated the disconnection hypothesis using DTI in MCI and AD patients, but they used volumetric MRI to assess structural damage in the PCC.

They found associations among atrophy of MTL structures, cingulum FA, and PCC atrophy. In contrast to the purely structural imaging approach used by Choo et al., we provide evidence for the disconnection hypothesis using a structural-functional approach correlating DTI with FDG-PET, which is widely used in the clinical setting for differential diagnosis of AD.

Limitations: DTI data were acquired without suppression of the CSF signal, which can contribute to partial volume effects, particularly in WM adjacent to ventricular spaces (Concha et al., 2005). This limitation, coupled with a relatively limited 6 gradient direction and 3 mm2 in-plane resolution acquisition likely contributed to our inability to map the crura of the fornix, thereby restricting our analysis of the fornix to the fornix body. Future studies will proceed with improved technical acquisition parameters as well as CSF signal suppression in order to improve the SNR of limbic DTI data. Another limitation of the present study is its cross-sectional, case-control design, which will limit the generalizability of findings and interpretation regarding causal mechanisms. Follow-up longitudinal studies can help confirm our findings. Finally, we chose to limit this analysis to limbic tracts that are directly related to the hippocampal formation and posterior cingulate gyrus, but an analysis that included a more extensive array of structures from the Papez circuit might have yielded additional results. Although the lack of PET data in normal control subjects could be viewed as a potential limitation, the Cortex ID program uses an age-matched (in 5 year increments) normal subject database to generate Z-scores, and thereby “builds in” a control population, albeit not one with local population demographics.

In conclusion, our results indicate that disruption of limbic white matter pathways occurs during MCI as well as in early-to-moderate AD in a selective manner. This disruption occurs in a graded manner, with more severely affected individuals having a greater degree of fornix and descending cingulum damage. In addition, our results provide evidence in support of the disconnection hypothesis as a mechanism contributing to PCC hypometabolism during incipient AD. Furthermore, this relationship between PCC and descending cingulum is anatomically specific, not involving adjacent cortical regions implicated in AD.

Acknowledgments

This study was supported by an intramural research grant from Michigan State University, a grant from GE Healthcare, and the Department of Radiology at Michigan State University. We also wish to thank Luke Fischer for his significant contribution in analysis of the FDG-PET data, and Dr. Jie Huang, for his technical advice on DTI analysis.

REFERENCES

REFERENCES

Baron JC, Chételat G, Desgranges B, Perchey G, Landeau B, de la Sayette V, et al. In vivo mapping of gray matter loss with voxel-based morphometry in mild Alzheimer's Disease. NeuroImage (2001) 14:298–308.

Bozzali M, Falini A, Franceschi M, Cercignani M, Zuffi M, Scotti G, Comi G, Filippi M. White matter damage in Alzheimer's disease assessed in vivo using diffusion tensor magnetic resonance imaging. J Neurol Neurosurg Psychiatry. 2002 Jun;72(6):742-6.

Braak, H. and E. Braak (1996). "Evolution of the neuropathology of Alzheimer's disease." Acta Neurol Scand Suppl 165: 3-12

Buckner RL, Snyder AZ, Shannon BJ, LaRossa G, Sachs R, Fotenos AF, Sheline YI, Klunk WE, Mathis CA, Morris JC, Mintun MA. Molecular, structural, and functional characterization of Alzheimer's disease: evidence for a relationship between default activity, amyloid, and memory. J Neurosci. 2005 Aug 24;25(34):7709-17.

Chételat G, Desgranges B, de la Sayette V, Viader F, Eustache F, Baron JC. Mapping gray matter loss with voxel-based morphometry in mild cognitive impairment. NeuroReport (2002) 13:1939–43.

Chételat G, Desgranges B, Landeau B, Mézenge F, Poline JB, de la Sayette V, Viader F, Eustache F, Baron JC. (2008). Direct voxel-based comparison between grey matter hypometabolism and atrophy in Alzheimer's disease. Brain. 131:60-71.

Chételat G, Villain N, Desgranges B, Eustache F, Baron JC. Posterior cingulate hypometabolism in early Alzheimer's disease: what is the contribution of local atrophy versus disconnection? Brain. 2009 Dec;132(Pt 12):e133; author reply e134.

Choo IH, Lee DY, Oh JS, Lee JS, Lee DS, Song IC, Youn JC, Kim SG, Kim KW, Jhoo JH, Woo JI. Posterior cingulate cortex atrophy and regional cingulum disruption in mild cognitive impairment and Alzheimer's disease. Neurobiol Aging. 2008 Aug 5.

Concha L, Gross DW, Beaulieu C. Diffusion tensor tractography of the limbic system. Am J Neuroradiol. 2005 Oct;26(9):2267-74.

Duvernoy HM. The human hippocampus*:* functional anatomy, vascularization and serial sections with MRI*.* 2nd ed*.* Berlin: Springer; 1998.

Fellgiebel A, Müller MJ, Wille P, Dellani PR, Scheurich A, Schmidt LG, Stoeter P. Color-coded diffusion-tensor-imaging of posterior cingulate fiber tracts in mild cognitive impairment. Neurobiol Aging. 2005 Aug-Sep;26(8):1193-8. Epub 2005 Jan 12.

Fellgiebel, A., I. Schermuly, et al. (2008). "Functional relevant loss of long association fibre tracts integrity in early Alzheimer's disease." Neuropsychologia 46(6): 1698-706.

Folstein MF, Folstein SE, McHugh PR (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. Journal of psychiatric research 12 (3): 189–98.

Jiang H, van Zijl PC, Kim J, Pearlson GD, Mori S. DtiStudio: resource program for diffusion tensor computation and fiber bundle tracking. Comput Methods Programs Biomed. 2006 Feb;81(2):106-16.

Jobst KA, Smith AD, Barker CS, Wear A, King EM, Smith A, Anslow PA, Molyneux AJ, Shepstone BJ, Soper N, et al. (1992). Association of atrophy of the medial temporal lobe with reduced blood flow in the posterior parietotemporal cortex in patients with a clinical and pathological diagnosis of Alzheimer's disease. J Neurol Neurosurg Psychiatry. 55:190-4.

Kiuchi K, Morikawa M, Taoka T, Nagashima T, Yamauchi T, Makinodan M, Norimoto K, Hashimoto K, Kosaka J, Inoue Y, Inoue M, Kichikawa K, Kishimoto T. Abnormalities of the uncinate fasciculus and posterior cingulate fasciculus in mild cognitive impairment and early Alzheimer's disease: a diffusion tensor tractography study. Brain Res. 2009 Sep 1;1287:184-91.

Liu Y, Spulber G, Lehtimäki KK, Könönen M, Hallikainen I, Gröhn H, Kivipelto M, Hallikainen M, Vanninen R, Soininen H. Diffusion tensor imaging and Tract-Based Spatial Statistics in Alzheimer's disease and mild cognitive impairment. Neurobiol Aging. 2009 Nov 11.

McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology. 1984 Jul;34(7):939-44.

Meguro, K., C. LeMestric, et al. (2001). "Relations between hypometabolism in the posterior association neocortex and hippocampal atrophy in Alzheimer's disease: a PET/MRI correlative study." J Neurol Neurosurg Psychiatry 71(3): 315-21.

Mielke MM, Kozauer NA, Chan KC, George M, Toroney J, Zerrate M, Bandeen-Roche K, Wang MC, Vanzijl P, Pekar JJ, Mori S, Lyketsos CG, Albert M. Regionally-specific diffusion tensor imaging in mild cognitive impairment and Alzheimer's disease. Neuroimage. 2009 May 15;46(1):47-55.

Minoshima S, Koeppe RA, Frey KA, Ishihara M, Kuhl DE. (1994). Stereotactic PET atlas of the human brain: aid for visual interpretation of functional brain images. J Nucl Med. 35(6): 949-54.

Minoshima S, Frey KA, Koeppe RA, Foster NL, Kuhl DE. (1995). A diagnostic approach in Alzheimer's disease using three-dimensional stereotactic surface projections of fluorine-18-FDG PET. J Nucl Med. 36(7): 1238-48.

Minoshima S, Giordani B, Berent S, Frey KA, Foster NL, Kuhl DE. (1997). Metabolic reduction in the posterior cingulate cortex in very early Alzheimer's disease. *Ann Neurol*. 42:85-94.

Nestor PJ, Fryer TD, Ikeda M, Hodges JR. (2003). Retrosplenial cortex (BA 29/30) hypometabolism in mild cognitive impairment (prodromal Alzheimer's disease).Eur J Neurosci. 18:2663-7.

Petersen RC. Mild cognitive impairment as a diagnostic entity. J Intern Med. 2004 Sep;256(3):183-94

Rose SE, Chen F, Chalk JB, Zelaya FO, Strugnell WE, Benson M, Semple J, Doddrell DM. Loss of connectivity in Alzheimer's disease: an evaluation of white matter tract integrity with colour coded MR diffusion tensor imaging. J Neurol Neurosurg Psychiatry. 2000 Oct;69(4):528-30.

Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TE, Johansen-Berg H, Bannister PR, De Luca M, Drobnjak I, Flitney DE, Niazy RK, Saunders J, Vickers J, Zhang Y, De Stefano N, Brady JM, Matthews PM. (2004). Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage. 23 Suppl 1:S208-19.

Sperling RA, Laviolette PS, O'Keefe K, O'Brien J, Rentz DM, Pihlajamaki M, Marshall G, Hyman BT, Selkoe DJ, Hedden T, Buckner RL, Becker JA, Johnson KA. Amyloid deposition is associated with impaired default network function in older persons without dementia. Neuron. 2009 Jul 30;63(2):178-88.

Teipel SJ, Stahl R, Dietrich O, Schoenberg SO, Perneczky R, Bokde AL, Reiser MF, Moller HJ, Hampel H. Multivariate network analysis of fiber tract integrity in Alzheimer's disease. Neuroimage. 2007 Feb 1;34(3):985-95.

Tierney MC, Fisher RH, Lewis AJ, Zorzitto ML, Snow WG, Reid DW, Nieuwstraten P. The NINCDS-ADRDA Work Group criteria for the clinical diagnosis of probable Alzheimer's disease: a clinicopathologic study of 57 cases. Neurology. 1988 Mar;38(3):359-64.

Villain N, Desgranges B, Viader F, de la Sayette V, Mézenge F, Landeau B, Baron JC, Eustache F, Chételat G. (2008). Relationships between hippocampal atrophy, white matter disruption, and gray matter hypometabolism in Alzheimer's disease. J Neurosci. 28:6174-81.

Villain N, Fouquet M, Baron J, Mézenge F, Landeau B, de La Sayette V, Viader F, Eustache F, Desgranges B, Chételat G. Sequential relationships between grey matter and white matter atrophy and brain metabolic abnormalities in early Alzheimer's disease. Brain 2010 (in press).

Walhovd KB, Fjell AM, Amlien I, Grambaite R, Stenset V, Bjørnerud A, Reinvang I, Gjerstad L, Cappelen T, Due-Tønnessen P, Fladby T. Multimodal imaging in mild cognitive impairment: Metabolism, morphometry and diffusion of the temporal-parietal memory network. Neuroimage. 2009 Mar 1;45(1):215-23.

Xie S, Xiao JX, Wang YH, Wu HK, Gong GL, Jiang XX. Evaluation of bilateral cingulum with tractography in patients with Alzheimer's disease. Neuroreport. 2005 Aug 22;16(12):1275-8.

Zhang Y, Schuff N, Jahng GH, Bayne W, Mori S, Schad L, Mueller S, Du AT, Kramer JH, Yaffe K, Chui H, Jagust WJ, Miller BL, Weiner MW. Diffusion tensor imaging of cingulum fibers in mild cognitive impairment and Alzheimer disease. Neurology. 2007 Jan 2;68(1):13-9.

Zhou Y, Dougherty JH Jr, Hubner KF, Bai B, Cannon RL, Hutson RK. Abnormal connectivity in the posterior cingulate and hippocampus in early Alzheimer's disease and mild cognitive impairment. Alzheimers Dement. 2008 Jul;4(4):265-70.

CHAPTER 3

Predicting Progression from Mild Cognitive Impairment to Alzheimer's Dementia using Probabilistic Pattern Classification[#]

Data used in this study were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.ucla.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this chapter. A complete listing of ADNI investigators can be found at: http://adni.loni.ucla.edu/wpcontent/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

3.1 Introduction

Alzheimer's disease (AD) is the leading cause of dementia in the aging population, affecting more than 30 million people worldwide (Barnes and Yaffe, 2011). AD is a degenerative brain disorder that causes a progressive decline in cognitive function, most notably memory loss, and other behavioral changes (Holtzman et al., 2011). Individuals diagnosed with mild cognitive impairment (MCI) have a substantially increased risk of developing clinical AD, and MCI is often considered to be a transitional phase between healthy cognitive aging and dementia (Petersen et al., 2009). Thus, MCI represents a key prognostic and therapeutic target in the management of AD. However, MCI is a heterogeneous syndrome with varying clinical outcomes. Although up to 60% of MCI patients develop dementia within a ten-year period, many people remain cognitively stable or regain normal cognitive function (Manly et al., 2008; Mitchell and Shiri-Feshki, 2009).

Increasing efforts have focused on building predictive models of dementia and the MCI-to-dementia progression using pattern classification methods based on clinical, imaging, genetic, and fluid markers (Chen and Herskovits, 2010; Haller et al., 2011; Klöppel et al., 2012; Perrin et al., 2009). Prognostic classification of MCI at the individual patient level has the potential to improve clinical trial design, identify patients for early treatment, as well as guide clinical and patient decision-making. In this study, we develop a multivariate prognostic model (Steyerberg et al., 2013) for predicting MCI-to-dementia progression using baseline data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) (Weiner et al., 2012). We focus on using widely available, cost-effective, and minimally-invasive data sources, including: (1) clinical data, such as risk factors and cognitive / behavioral assessments; (2) morphometric measures derived

from a structural magnetic resonance imaging (MRI) scan of the brain; and (3) blood plasma-based proteomic data. Much of this data is routinely collected during the clinical workup of dementia and clinical trials.

We used a kernel-based classifier to predict future dementia status of MCI patients by incorporating heterogeneous (clinical, MRI, and proteomic) data. Kernel-based learning algorithms use “kernel functions” to encode the degree of similarity between examples in a dataset based on their features (Ben-Hur et al., 2008; Hofmann et al., 2008), such as individual MCI patients described by their unique biomarker patterns. We applied an extension of this methodology, known as multiple kernel learning (MKL), which allows integration of complementary information derived from different sources or representations of the data using separate kernels (Gönen and Alpaydın, 2011). Recent studies suggest that multiple-kernel classifiers may integrate heterogeneous data more effectively than conventional single-kernel classifiers, improving classification of AD and MCI subjects by as much as 3-11% {(Hinrichs et al., 2011; Zhang and Shen, 2012a, 2012b)}.

The prevailing approach in the literature has been to consider prediction of MCI-to-dementia progression as a non-probabilistic binary classification task, where all patients are unequivocally assigned to either the progressive MCI (P-MCI) or the non-progressive MCI (N-MCI) group (Cui et al., 2011; Davatzikos et al., 2011; Hinrichs et al., 2011; Westman et al., 2012; Zhang et al., 2011). Sir William Osler (1849-1919), a pre-eminent physician of the 20th century, is credited with stating that “medicine is a science of uncertainty and an art of probability” (Westover et al., 2011). In this spirit, we adopt a recently proposed implementation of MKL that generates probabilistic predictions using Bayesian inference (Damoulas and Girolami, 2008). The probability associated with

each individual prediction can be used as a measure of confidence, which in turn can be used to withhold the decision about future dementia status for ambiguous (“low confidence”) MCI cases. This approach is often referred to as classification with a “reject option” {(Herbei and Wegkamp, 2006)}. Alternatively, a probabilistic classifier could be used to develop a predictive model that allows risk stratification at the individual patient level. Thus, we propose that a probabilistic classification approach for predicting MCI-to-dementia progression has greater utility in the context of clinical decision-making than a non-probabilistic approach.

The objectives of this study were to determine: (1) whether clinical, MRI, and plasma protein biomarkers capture complementary information regarding short-term progression from MCI to dementia; (2) if this information is more effectively learned using a multiple-kernel classifier as opposed to a single-kernel classifier; (3) how robust our predictive model is in light of patient heterogeneity; (4) to what extent model performance can be improved when only “high confidence” predictions are allowed; and (5) whether the model's probabilistic predictions reflect any information about the time-to-progression for P-MCI patients.

3.2 Materials and Methods

In this section, we describe the ADNI and the subjects who participated in this study. Then, we describe data collection and processing procedures. Finally, we describe our pattern classification approach, the series of experiments conducted, and statistical analyses. Unless otherwise noted, analyses were done using MATLAB R2010b (The MathWorks, Inc., Natick, MA).

3.2.1 Alzheimer's Disease Neuroimaging Initiative (ADNI)

Data used in this study were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.ucla.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. ADNI is an observational study with both cross-sectional and longitudinal follow-up components. The primary goal of ADNI has been to test whether structural, functional, and molecular neuroimaging, fluid and genetic biomarkers, and clinical and neuropsychological assessments can be combined to measure the progression of MCI and early AD. Sensitive and specific markers of very early AD progression are expected to enhance the development of new treatments and aid in monitoring their efficacy, and to make clinical trials more time- and cost-efficient. The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California – San Francisco. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 subjects. ADNI has been followed by ADNI-GO and ADNI-2. To date, these three phases of ADNI have recruited over 1500 older adults to participate in the research, consisting of cognitively normal (CN) individuals, people with early or late MCI, and people with early AD. Subjects originally recruited for ADNI-1 and ADNI-GO had the option to be followed in ADNI-2. For up-to-date information, see www.adni-info.org.

In this study, we analyzed baseline visit data collected from MCI subjects who were initially recruited during ADNI-1. The various datasets were downloaded on or before the following dates: Clinical data – August 20, 2011; Structural MRI data – August 3, 2011; Plasma proteomic data – June 16, 2012. All subjects and their study partners completed the informed consent process, and the ADNI study protocols were reviewed and approved by the local Institutional Review Board at each ADNI data collection site.

3.2.2 Subjects

The general eligibility, inclusion, and exclusion criteria for ADNI subjects can be found on the ADNI website (www.adni-info.org) and are summarized here. Enrolled subjects were 55 to 90 years old with a minimum 6th-grade level of education; had a study partner able to provide an independent evaluation of functioning; could speak either English or Spanish; had adequate visual and auditory acuity to allow neuropsychological testing; were willing and able to undergo all test procedures, including neuroimaging; and agreed to longitudinal follow-up. All subjects had no significant neurologic disease, major depression, history of schizophrenia or bipolar disorder, recent history of alcohol or substance abuse, and no pacemakers or other objects deemed unsafe for MRI. MCI subjects met the Petersen (Mayo Clinic) diagnostic criteria for amnestic MCI (Petersen, 2004) as follows: (1) a subjective memory complaint; (2) objective memory loss, as measured by age- and education-adjusted scores on Wechsler Memory Scale Logical Memory II, but without significant impairment in other cognitive domains; (3) generally preserved activities of daily living;

and (4) no dementia. MCI subjects also had Mini-Mental State Examination (MMSE) scores of 24-30 and a global score of 0.5 on the Clinical Dementia Rating (CDR) scale.

Characteristic	N-MCI (n=120)	P-MCI (n=139)	*p*-value
Age, years	74.8 ± 7.6	74.8 ± 7.1	>0.5[a]
Education, years	15.7 ± 2.9	15.6 ± 2.9	>0.5[a]
Sex, % female	28.3	38.1	0.097[b]
APOE ε4 carriers, %	41.7	66.2	<0.001[b]
MMSE score	27.6 ± 1.7	26.7 ± 1.7	<0.001[a]
MCI = mild cognitive impairment; N-MCI = non-progressive MCI; P-MCI = progressive MCI; APOE = apolipoprotein E; MMSE = Mini-Mental State Examination			
P-values for differences between N-MCI and P-MCI are based on (a) *t*-test or (b) χ2-test.			

Table 3.1. Subject characteristics at baseline. Values are shown as mean ± standard deviation or percentage.

From a total of 390 individuals with a baseline diagnosis of MCI who were recruited for ADNI-1, 289 subjects met criteria for inclusion as part of either the P-MCI or N-MCI group in this study. Thirty (~10%) of these subjects were further excluded due to partially missing baseline data. Table 3.1 shows the characteristics of the final MCI cohort (n=259). Progressors (P-MCI; n=139) included MCI subjects who progressed to AD-type dementia within 36 months (median: 18 months) of entering the study, as indicated by the National Institute of Neurological and Communicative Disorders and Stroke –Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) criteria for the diagnosis of probable AD (McKhann et al., 1984). Non-progressors (N-MCI; n=120) included MCI subjects who had not progressed to dementia within 36 months of entering the study. This group included subjects who remained cognitively

stable (n=107; did not revert to CN status and did not develop dementia) or those who reverted to CN status and remained dementia-free (n=13).

3.2.3 Data Collection and Follow-up

At study entry (baseline), all subjects underwent a comprehensive clinical evaluation, cognitive/behavioral assessments, and a structural brain MRI scan. Subjects also provided a blood sample for apolipoprotein E (APOE) genotyping and proteomic analysis. Subjects were then followed longitudinally at specific time points (6, 12, 18, 24, 36 months). The clinical status of each MCI subject was re-assessed at each follow-up visit and updated to reflect one of several outcomes (CN, MCI, AD, or other). The N-MCI and P-MCI group designations were based on this follow-up clinical diagnosis and used as the “ground truth” in our classification experiments.

3.2.4 Clinical Data

We considered a total of 186 clinical variables (features) as potential predictors of MCI-to-dementia progression in our classification analyses (Figure 3.1). Clinical features were of two types: risk factors and assessments/markers. Risk factors included demographic, genetic, and medical data, while assessments/markers included total scores and sub-scores on several cognitive, functional, behavioral, and clinical scales. We also included data on the off-label use of AD medications by MCI subjects. Recent studies suggest that cognitive and functional markers may be at least as effective for predicting MCI-to-dementia progression as imaging and fluid biomarkers (Cui et al., 2011; Gomar et al., 2011; Palmqvist et al., 2012).

The following assessments were administered to the subjects in ADNI. The MMSE (Folstein et al., 1975) is widely used as a screening test for dementia and assesses cognitive function in multiple domains, including orientation, language, attention, calculation, constructional praxis, and memory. The CDR (Morris, 1993) is administered as a semi-structured interview with both the patient and an informant in order to assesses the patient's functional and cognitive status in six domains: memory, orientation, judgment and problem-solving, community affairs, home and hobbies, and personal care. The Functional Activities Questionnaire (FAQ) (Pfeffer et al., 1982) assesses the level of independence in performing activities of daily living (e.g. record keeping, managing finances, shopping, meal preparation, remembering dates, transportation). The Geriatric Depression Scale (GDS) (Sheikh and Yesavage, 1986) is a self-report assessment of depressive symptoms and designed to be used as a screening test for depression in older adults. The Neuropsychiatric Inventory Questionnaire (NPI-Q) (Kaufer et al., 2000) is an informant-based assessment of recent psychiatric and behavioral symptoms (e.g. hallucinations, agitation, depression, anxiety, apathy, disinhibition, irritability). The Modified Hachinski Ischemic Scale (HIS) (Rosen et al., 1980) assesses the contribution of cerebrovascular disease to cognitive impairment based on medical history and neurological symptoms and signs. The American National Adult Reading Test (ANART) (Grober et al., 1991) provides a measure of premorbid intelligence by assessing pronunciation of 50 irregular words.

In addition, a battery of neuropsychological tests was administered to further evaluate function in specific cognitive domains. The WMS-III Logical Memory (LM) is a test of episodic memory function that assesses immediate and delayed story recall (Johnson et al., 2003). The Alzheimer's Disease Assessment Scale – Cognitive sub-

scale (ADAS-Cog) (Mohs et al., 1997) assesses multiple aspects of memory and language function as well as attention, orientation, and praxis. The Rey Auditory Verbal Learning Test (RAVLT) (Vakil and Blachstein, 1993) involves learning lists of words and assesses verbal memory. The verbal (category) fluency test (Acevedo et al., 2000) and Boston Naming Test (BNT) (Zec et al., 2007) assess semantic memory and language function. The digit span test (Hester et al., 2004) assesses verbal working memory. The Trail Making Test (TMT) (Tombaugh, 2004) evaluates processing speed (part A) and executive function (part B). The Digit-Symbol Coding Test (DST) (Joy et al., 2004) assesses processing speed, visual working memory, and visual-motor coordination. The Clock-Drawing Test (CDT) (Shulman, 2000) assesses constructional praxis with elements of visuospatial and executive function.

3.2.5 Structural MRI Data

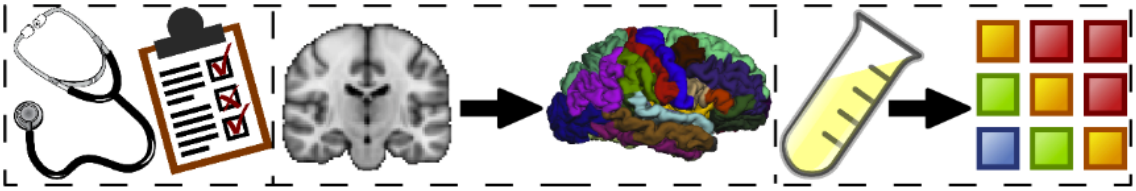
MRI offers a non-invasive, widely available, and more cost-effective alternative for obtaining imaging biomarkers of AD-related neurodegeneration (e.g. atrophy measures) compared to positron emission tomography (PET) (Karow et al., 2010). We considered 452 region-of-interest (ROI)-based morphometric measures computed from individual structural MRIs as potential predictors of MCI-to-dementia progression (Figure 3.1). We generated MRI features for classification using an atlas-based ROI method rather than a voxel-based method in an effort to reduce the dimensionality of the MRI dataset and increase the signal-to-noise ratio of the resulting features.

Subjects received MRI scans at 1.5 Tesla acquired using a variety of scanners (General Electric, Philips, or Siemens) and a standardized protocol (Jack et al., 2008) with the following acquisition parameters: T1-weighted sagittal 3-D MP-RAGE sequence

with 1.25 x 1.25 mm^2 in-plane resolution and 1.2 mm slice thickness, TR = 2400 ms, TI = 1000 ms, TE = 3 ms, flip angle = 8°, 240 x 240 mm^2 FOV, 192 x 192 in-plain matrix size. Raw MRI data underwent quality control and were pre-processed by the ADNI MRI Core to correct for image geometry distortion due to gradient non-linearity and image intensity non-uniformity (see www.adni-info.org for details). In this study, we used these pre-processed, corrected MRI datasets.

Each MRI dataset was post-processed using FreeSurfer v5.0.0 (http://surfer.nmr.mgh.harvard.edu) (Dale et al., 1999; Desikan et al., 2006; Fischl et al., 1999, 2002), an image processing software tool for i) automated model-based reconstruction and segmentation of the brain's cortical surface and subcortical structures and ii) morphometric analysis. The specific MRI processing steps included: removal of non-brain tissue (skull stripping); Talairach transformation of the brain volume into standard anatomical space; intensity normalization; segmentation of cerebral white matter (WM), subcortical gray matter (GM), and ventricles; delineation and 3-D reconstruction of GM/WM and GM/CSF (cerebrospinal fluid) boundaries; and automated topology correction. The cortical surface model was registered to a spherical atlas and used for segmentation of the cerebral cortex into regions based on gyral-sulcal anatomy (i.e. cortical folding pattern). The final segmentation and labeling of brain structures was based on a probabilistic atlas along with intensity and curvature information. Finally, a variety of morphometric measures were computed across 180 anatomically-defined brain regions as MRI features for classification, including cortical and subcortical volumes, mean cortical thickness (and its standard deviation), surface area, and curvature. FreeSurfer-derived morphometric MRI measures have been

validated in studies of normal aging, MCI, and AD (Desikan et al., 2009; Salat et al., 2009; Shen et al., 2010).



CLINICAL STRUCTURAL MRI PLASMA PROTEINS

Data Source (# features)	Description
Clinical Data (186)	
Risk Factors (16)	*Demographic*: age, sex, education *Genetic*: APOE genotype, family hx of dementia *Medical*: CVD factors, BMI, hx of depression, anxiety, alcohol abuse, head trauma, sleep apnea
Assessments/Markers (170)	*Cognitive, functional, and clinical assessments (total scores and sub-scores)*: CDR, FAQ, GDS, NPIQ, MMSE, ADAS-Cog, LM, Clock, RAVLT, Digit Span, VerbFlu, TMT, DST, BNT, ANART, HIS, AD-Meds
MRI Data (452)	*FreeSurfer ROI-based morphometric measures:* volume, cortical thickness (mean and std), surface area, curvature
Proteomic Data (149)	*Levels of plasma proteins reported to be involved in cell-signaling and a variety of disease processes.*

Figure 3.1. Data sources (features) analyzed in this study. hx = history; APOE = apolipoprotein E; BMI = body-mass index; CVD = cerebrovascular disease risk factors (history of diabetes mellitus, coronary artery disease, hypertension, smoking, hyperlipidemia, stroke); CDR = Clinical Dementia Rating; FAQ = Functional Activities Questionnaire; GDS = Geriatric Depression Scale; NPIQ = Neuropsychiatric Inventory Q; MMSE = Mini-Mental State Examination; ADAS-Cog = Alzheimer's Disease Assessment Scale – Cognitive sub-scale; LM = Logical Memory; Clock = clock drawing/copying; RAVLT = Rey Auditory-Verbal Learning Test; VerbFlu = Verbal Fluency; TMT = Trail Making Test; DST = Digit Symbol Coding Test; BNT = Boston Naming Test; ANART = American National Adult Reading Test; HIS = Hachinski Ischemic Scale; AD-Meds = use of AD medications; ROI = region of interest; std = standard deviation.

3.2.6 Plasma Proteomic Data

Plasma-based proteomic biomarkers have been proposed as a less invasive and simpler alternative for the early diagnosis of AD compared to CSF-based biomarkers (Graff-Radford et al., 2007; Ray et al., 2007). However, the utility of plasma biomarkers

in predicting MCI-to-dementia progression remains controversial given the conflicting findings in the literature (Hansson et al., 2010; Johnstone et al., 2012). Thus, in addition to clinical and MRI features, we considered 149 features based on plasma protein levels in this study (Figure 3.1). Plasma samples were analyzed by Rules-Based Medicine (RBM) (Austin, TX) using their Human DiscoveryMAP multiplex immunoassay, which is based on the Luminex xMAP platform (Soares et al., 2012). This immunoassay panel of 190 analytes included proteins previously reported to be involved in cell-signaling and/or associated with a variety of disease processes, including AD, metabolic disorders, inflammation, cancer, and cardiovascular disease. The ADNI team, in collaboration with the Biomarkers Consortium, identified 146 (out of 190) analytes that met quality control standards. We used the cleaned, quality-controlled (QC) dataset containing these 146 analytes, labelled “ADNI Plasma QC Multiplex 11Nov2010”. Further details about the RBM immunoassay and QC procedures can be found in the data primer, “Biomarkers Consortium Project: Use of Targeted Multiplex Proteomic Strategies to Identify Plasma-Based Biomarkers in Alzheimer’s Disease” (available at http://adni.loni.ucla.edu). We also used the plasma levels of amyloid-β proteins (Aβ42, Aβ40, and the Aβ42/Aβ40 ratio), which were assayed by the ADNI Biomarker Core Laboratory at the University of Pennsylvania. Aβ42 and Aβ40 have been identified as the major molecular species contributing to the amyloid (“senile”) plaques, a pathological hallmark of AD (Iwatsubo et al., 1994).

3.2.7 Data Transformation

We applied a series of transformations to the feature data prior to conducting classification analyses. First, volumetric and surface area MRI measures were normalized by the estimated total intracranial volume (Buckner et al., 2004) to correct

for individual differences in head size. Second, each of the 787 features was scaled to have zero mean and unit variance across subjects. Third, each continuous and ordinal feature was discretized into three states (low, intermediate, high) using the mean and standard deviation to define interval boundaries, as described in (Ding and Peng, 2005). Discretized features were used only when conducting information-theoretic feature selection (described below) while non-discretized features were used during model training.

3.2.8 Feature Selection

Feature selection is an important component of the model development process, particularly in the case of high-dimensional pattern classification where the number of features is large and exceeds the number of samples available for classification (787 features and 259 subjects in this study). Many of these features may be irrelevant, redundant, or noisy. Feature selection is a dimensionality reduction strategy that involves identifying a small but informative subset of the original features for classification; it can help avoid model overfitting, improve model performance, and produce models that are easier to interpret and potentially more time- and cost-efficient to develop and use (Saeys et al., 2007). Feature selection techniques include filter- and wrapper-based methods. Filter methods tend to be fast and identify informative features based on inherent statistical properties of the data, independent of any classifier. In contrast, wrapper methods evaluate the merit of various feature subsets based on the performance of a classifier and may select features for classification more effectively, although at a significant cost in terms of speed and greater potential for overfitting.

In this study we adopt a combined filter-wrapper approach to efficiently identify a subset of features that can be used to effectively discriminate between P-MCI and N-MCI. In the first stage of our feature selection procedure, we defined feature subsets of different sizes (ranging from 1 to 50 features) using the Joint Mutual Information (JMI) criterion (Yang and Moody, 1999), as implemented in the FEAST toolbox (http://www.cs.man.ac.uk/~gbrown/fstoolbox) (Brown et al., 2012). In the second stage, we evaluated these feature subsets in terms of cross-validated classification accuracy and determined the optimal feature subset size to be used as a parameter in the final model.

JMI is a multivariate information-theoretic filter method for feature selection and has been shown to perform well in terms of both classification accuracy and stability on a wide range of real-world datasets (Brown et al., 2012). Features are selected based on their JMI score (*J*), defined as:

$$J(X_k) = I(X_k; Y) - \frac{1}{|S|} \sum_{j \in S} \left[I(X_k; X_j) - I(X_k; X_j | Y) \right] \quad (1)$$

where X_k is the feature being considered for selection, X_j is each of the previously selected features in feature subset S, and Y is the outcome/class variable of interest (future dementia status, in this study). The JMI score for a given feature X_k is defined as a linear combination of three mutual information terms I, each of which describes the amount of information shared (or the dependence) between two random variables; these terms correspond to relevance $I(X_k; Y)$,

redundancy $I(X_k; X_j)$, and class-conditional redundancy $I(X_k; X_j|Y)$ (Brown et al., 2012). Thus, JMI-based feature selection favors features that are maximally relevant to the classification task while being minimally redundant and maximally complementary with previously selected features.

3.2.9 Classification Approach

In this study, we use the probabilistic multiple kernel learning (pMKL) classification approach proposed by Damoulas et al. (http://www.dcs.gla.ac.uk/inference/pMKL) (Damoulas and Girolami, 2008, 2009a, 2009b) to build several prognostic models of dementia. pMKL is a kernel-based classifier similar to the widely used support vector machine (SVM) (Ben-Hur et al., 2008; Hofmann et al., 2008). Kernel classifiers rely on the use of kernel functions to map the original feature data into an inner product space that encodes similarity between examples (e.g. patients). The algorithm learns to classify new examples based on this similarity information. Different kernel functions can be used to provide varying definitions of similarity. In this study, we build models with both linear and nonlinear (polynomial and Gaussian) kernels. The similarity between a pair of examples a and b, described by their feature vectors x_a and x_b, can be defined according to each kernel function K as:

$$K(x_a, x_b) = x_a \cdot x_b \quad \text{(linear)} \tag{2}$$

$$K(x_a, x_b) = (x_a \cdot x_b + c)^d \quad \text{(polynomial)} \tag{3}$$

$$K(x_a, x_b) = \exp(-\gamma ||x_a - x_b||^2) \quad \text{(Gaussian)} \qquad (4)$$

where c is a constant, d is the degree of the polynomial, and γ is the kernel width. While an advantage of the linear kernel is that there are no kernel parameters to set, the linear kernel is unable to capture more complex patterns in the data, as can be done by using non-linear kernels.

The pMKL classifier, like an SVM, can be used in either the single-kernel mode or the multiple-kernel mode. In the latter case, referred to as multiple kernel learning (MKL), separate kernels are used to encode information from different sources (representations) of the data (Gönen and Alpaydın, 2011). For illustration, consider a dataset with N examples and S sources of the data, with each example described by the feature vector x_n^s and discrete class (outcome) label $Y_n \in \{1, \ldots, C\}$ where $n = 1, \ldots, N$, $s = 1, \ldots, S$, and C is the number of classes (outcomes). pMKL integrates this information by learning an optimal linear combination of the multiple kernels (Damoulas and Girolami, 2008), such that the $N \times N$ composite kernel is defined as:

$$K^{\beta\Theta}(\boldsymbol{x_a}, \boldsymbol{x_b}) = \sum_{s=1}^{S} \beta_s K^{s\theta_s}(\boldsymbol{x_a^s}, \boldsymbol{x_b^s}) \qquad (5)$$

In Eq. (5), β_s is the kernel weight describing the relative contribution of data source s and θ_s is the kernel parameter that controls the amount of data smoothing (e.g. degree d of the polynomial kernel or width γ of the Gaussian kernel).

The overall pMKL classifier is based on a Generalized Linear Model (GLM) regression framework using the multinomial probit likelihood (Damoulas and Girolami, 2008) given by:

$$P\left(Y_n = i | \boldsymbol{W}, \boldsymbol{k}_n^{\beta\Theta}\right) = E_{p(u)}\left\{\prod_{j \neq i} \Phi\left(u + \left(\boldsymbol{w}_i - \boldsymbol{w}_j\right)\boldsymbol{k}_n^{\beta\Theta}\right)\right\} \quad (6)$$

where E is the expectation with respect to the standard normal distribution $p(u) = N(0,1)$ and Φ is the cumulative distribution function. Eq. (6) computes the probability P that example n belongs to class/outcome i (as opposed to class j) given the feature data (in the form of a kernel matrix $k_n^{\beta\Theta}$) and regression coefficients W. The regression coefficients reflect the weight with which training examples used to construct the model vote for a particular class/outcome. The posterior probability P is determined using Bayesian estimation methods (for details see Damoulas and Girolami, 2009b) and captures the uncertainty or the degree of confidence associated with each prediction. Non-probabilistic classification can be achieved by predicting the class/outcome with the largest posterior probability (>50% for binary classification).

3.2.10 Experimental Design

We built and examined the following series of predictive models, each designed to classify individual patients as belonging to either the N-MCI or the P-MCI group. In experiment 1A, we built a set of models designed to assess the predictive utility of

different data sources, alone and in combination. For this experiment, we built classifiers utilizing the linear kernel. First, single-kernel, single-source classifiers were constructed separately for clinical risk factors (model 1; 'CRF'), clinical assessments / markers (model 2; 'CAM'), MRI markers (model 3; 'MRI'), and plasma proteomic markers (model 4; 'PPM'). Second, a single-kernel, multi-source classifier was constructed where all features across the four data sources (CRF, CAM, MRI, and PPM) were concatenated and considered jointly during feature selection and kernel computation steps (model 5; 'CONCAT'). Third, a multiple-kernel, multi-source classifier was constructed, with a separate linear kernel used to encode the most informative features from each of the four data sources (model 6; 'MKL-Linear'). In experiment 1B, we built a set of multiple-kernel, multi-source classifiers by incorporating different nonlinear kernels in order to capture information regarding more complex interactions among features and to integrate potentially complementary representations of the feature data. These classifiers were constructed by considering all features and data sources and included: i) a three-kernel classifier with a linear, polynomial (d=2 and c=1), and Gaussian ($\gamma=1/D$) kernels, where D is the number of features (model 7; 'MKL-LPG'); ii) a classifier with five polynomial kernels (d=1, 2, 3, 4, 5 and c=1) (model 8; 'MKL-Poly'); and iii) a classifier with five Gaussian kernels ($\gamma=10^{-2}$, 10^{-1}, 10^{0}, 10^{1}, 10^{2}) (model 9; 'MKL-Gaussian').

In experiments 2 and 3, we studied the best performing model identified in experiments 1A and 1B. In experiment 2, we examined the extent to which patient heterogeneity affects model performance. Specifically, we examined the effects of age, sex, educational level, APOE genotype, cerebrovascular risk factors, use of AD medications, and time-to-progression. In experiment 3, we examined the trade-off

between predictive confidence and accuracy, and the proportion of MCI patients for whom predictions can be made at varying degrees of confidence. Predictive confidence was defined as the difference between the predicted probabilities for the two classes/outcomes (N-MCI and P-MCI). We also examined whether there is an association between the predicted probabilities and time-to-progression for P-MCI patients.

3.2.11 Statistical Analysis and Cross-Validation

For each model (1-9), we report cross-validated accuracy, sensitivity (Sn), and specificity (Sp) as measures of predictive performance (Altman and Bland, 1994a), and the area under the curve from receiver operating characteristic analysis (AUC-ROC) as a measure of class discrimination (Altman and Bland, 1994b). We used the balanced accuracy rate (BAR) as the primary measure of model performance, where $BAR = (Sn + Sp)/2$. We also assessed model calibration as a secondary performance measure. Calibration is an important measure of performance for probabilistic classification models and assesses the reliability of the predicted probabilities (Bouwmeester et al., 2012; Kim and Simon, 2011). The probability interval (0-100%) was divided into 10 equal sub-intervals; then, the predicted probability of MCI-to-dementia progression (generated by the pMKL classifier) was compared to the actual probability of progression (fraction of subjects belonging to the P-MCI group) for each of these 10 sub-intervals. The agreement between predicted and actual probabilities was quantified using the concordance correlation coefficient (CCC) (Lin, 1989); the CCC can range from +1 (perfect agreement) to -1 (perfect disagreement), with values of CCC near zero indicating weak or no relationship between predicted and actual probabilities.

By using the 10 probability sub-intervals (minimum recommended in Lin, 1989) we obtained a robust estimate for the CCC while keeping the sub-intervals sufficiently large to maximize the number of subjects within each sub-interval.

Cross-validation (CV) (Kohavi, 1995) refers to various data partitioning techniques commonly used in statistics and machine learning fields when developing predictive models and assessing their performance. The key goal of CV is to obtain an unbiased estimate of a model's predictive performance in circumstances of limited data availability. In essence, CV allows one to estimate how well a model can be expected to make predictions in real-world settings on new data.

We used a nested stratified cross-validation procedure (Figure 3.2) to avoid model overfitting and optimistically-biased estimates of model performance (Cawley and Talbot, 2010; Smialowski et al., 2010; Varma and Simon, 2006). The procedure consisted of two nested CV loops, each implementing 10-fold stratified CV: an outer loop, designed to obtain an unbiased estimate of model performance, and an inner loop, used for model (parameter) selection. In 10-fold stratified CV (Kohavi, 1995), the dataset is randomly partitioned into 10 mutually exclusive parts (folds) of equal size, preserving the proportion of samples in each class as found in the full dataset. Nine out of 10 parts are used to train the classifier, which is then evaluated on the remaining one part. This is repeated until each fold of the dataset has been used once for evaluation, thus resulting in 10 performance estimates per one run of 10-fold CV. During each fold of the outer CV loop, the full dataset (n=259) was split into a 'model development set' (90%) and a 'test set' (10%), which was held out for final model evaluation. Feature selection, model (parameter) selection, and final model construction were repeated independently for each fold of the outer CV loop and based only on the 'model

development set'. The inner CV loop was designed to determine the optimal feature subset size for use in the final model. During each fold of the inner CV loop, data from the 'model development set' were split into a 'training set' (90%) and 'validation set' (10%). Then, JMI-based feature selection was performed based only on the 'training set' to identify subsets with the top *D* most informative features for discriminating between N-MCI and P-MCI, where $D \in \{1, 3, 5, 7, 10, 15, 20, 30, 40, 50\}$. A classifier was then constructed for each of the ten feature subsets using the 'training set' and evaluated on the 'validation set'. The subset size resulting in the highest 10-fold CV accuracy on the 'validation set' was then selected as the optimal feature subset size, $D_{OPTIMAL}$. The final model classifier was constructed using the top $D_{OPTIMAL}$ most informative features (selected via the JMI method) based on the 'model development set' and evaluated on the 'test set'. For better replicability, the above nested 10-fold CV procedure was repeated 10 times, generating 100 performance estimate values for significance testing. We used a modified paired sample *t*-test with 10 degrees of freedom calibrated for 10x10 CV experiments (Bouckaert, 2003) to test for significant differences in performance between model pairs. All statistical tests were considered significant at the $P < 0.05$ level.

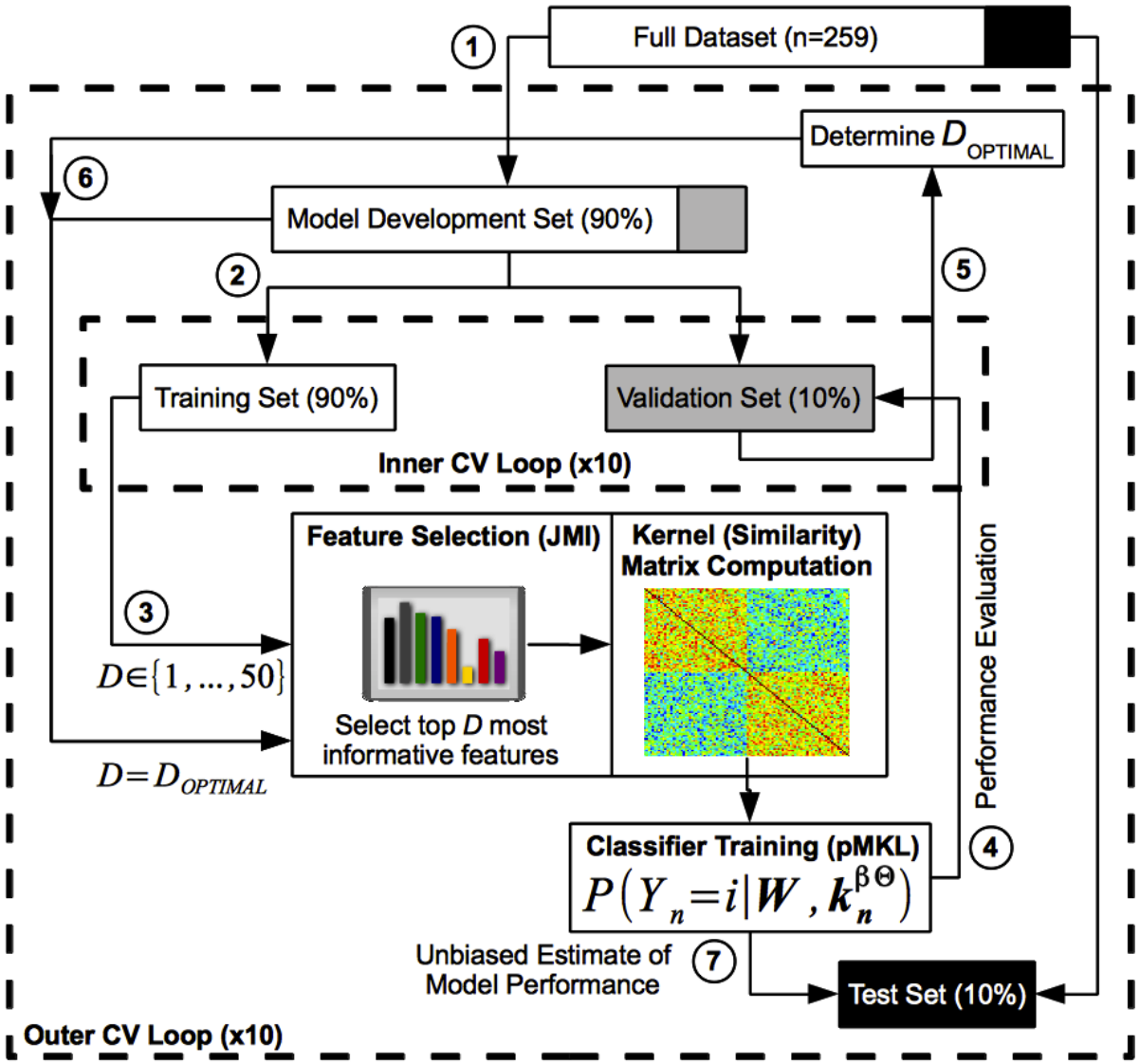


Figure 3.2. Nested 10-fold cross-validation (CV) design for model development and evaluation. See text for details. JMI = Joint Mutual Information; pMKL = probabilistic multiple kernel learning; $D_{OPTIMAL}$ = optimal number of features.

3.3 Results

3.3.1 Subject Characteristics

The N-MCI and P-MCI groups did not defer in terms of age, education, or sex distribution, although there was a trend toward a greater proportion of females among P-MCI subjects (Table 3.1). As expected, there was a greater proportion of APOE $\varepsilon 4$ allele carriers in the P-MCI group compared to the N-MCI group ($P < 0.001$). P-MCI subjects had slightly lower baseline scores on the MMSE than N-MCI subjects ($P < 0.001$); nevertheless, the vast majority of MCI subjects in both groups had baseline MMSE scores considered to be well within the normal range.

3.3.2 Predictive Performance of Single- and Multi-Source Models

Table 3.2 and Figure 3.3 summarize the predictive performance of models 1-9. The predictive accuracies (T-BAR) of all four single-source models (1-4: CRF, CAM, MRI, PPM) exceeded chance-level (all $P < 0.01$, one-sample *t*-test), although they varied from a low of 53.2% for PPM to a high of 76.1% for CAM. The CAM model outperformed the other three single-source models (all $P < 0.001$, paired-sample *t*-test). The CAM and MRI models were well-calibrated, as indicated by high positive concordance correlation coefficients (CCC) (both $P < 0.001$) while the PPM model showed poor calibration (CCC not different from zero, $P > 0.3$). The single-kernel, multi-source model 5 (CONCAT), in which all features across the four data sources were considered jointly, outperformed all single-source models 1-4 (all $P < 0.001$, paired-sample *t*-test), attaining a predictive accuracy of 80.0%. The calibration of the CONCAT model, as measured by the CCC, was statistically similar to that of CAM and MRI models (both $P > 0.3$) and better than that of the PPM model ($P < 0.001$).

Model	V-BAR (%)	T-BAR (%)	Sn (%)	Sp (%)	AUC-ROC	CCC	$D_{OPTIMAL}$ / Total
Single Source							
1. CRF	62.0 ± 1.4	61.8 ± 7.7	65.3 ± 12.7	58.3 ± 11.7	0.61 ± 0.12	#	1 ± 0 / 16
2. CAM	77.9 ± 1.4	76.1 ± 7.2	76.9 ± 9.5	75.3 ± 11.2	0.83 ± 0.07	0.92 ± 0.03	15 ± 10 / 170
3. MRI	71.4 ± 1.6	69.1 ± 8.5	68.5 ± 11.8	69.6 ± 12.4	0.76 ± 0.09	0.91 ± 0.03	10 ± 5 / 452
4. PPM	56.0 ± 2.7	53.2 ± 10.0	51.2 ± 12.9	55.3 ± 14.1	0.54 ± 0.11	0.10 ± 0.31	40 ± 10 / 149
Multi-Source							
5. CONCAT	79.7 ± 1.4	80.0 ± 7.3	80.3 ± 10.6	79.8 ± 10.9	0.86 ± 0.07	0.93 ± 0.02	10 ± 3 / 787
6. MKL-Linear	##	74.9 ± 6.7	74.6 ± 11.7	75.2 ± 11.9	0.84 ± 0.07	0.88 ± 0.04	74 ± 18 / 787
7. MKL-LPG	80.2 ± 1.5	79.7 ± 7.2	81.0 ± 9.8	78.3 ± 12.3	0.87 ± 0.07	0.94 ± 0.02	10 ± 0 / 787
8. MKL-Poly	80.1 ± 1.4	79.5 ± 7.5	82.2 ± 10.2	76.8 ± 12.4	0.87 ± 0.07	0.94 ± 0.02	10 ± 3 / 787
9. MKL-Gaussian	80.3 ± 1.3	79.9 ± 6.8	83.4 ± 9.9	76.4 ± 12.3	0.87 ± 0.07	0.95 ± 0.01	10 ± 3 / 787

CRF = Clinical Risk Factors, CAM = Clinical Assessments/Markers, MRI = Magnetic Resonance Imaging, PPM = Plasma Proteomic Markers, MKL = Multiple Kernel Learning

Robust estimate of CCC could not be obtained for model 1 because only <10 probability sub-intervals could be defined when conducting calibration analysis.

Since no inner CV was necessary other than what was done in models 1-4 to determine the optimal feature subset size, V-BAR was not calculated for this model.

Table 3.2. Cross-validated performance estimates for models 1-9. For each model, several measures of predictive accuracy are shown, including balanced accuracy rate on the validation set (V-BAR) and the test set (T-BAR), sensitivity (Sn), specificity (Sp), and area under the curve from receiver operating characteristic analysis (AUC-ROC). The concordance correlation coefficient (CCC) is a measure of model calibration and quantifies the agreement between the predicted and actual probabilities of MCI-to-dementia progression. All performance measures, except V-BAR, indicate how a model performed on the test set, which was not used during feature selection, model (parameter) selection, or model training steps. *D*(optimal) is the optimal number of features, a parameter determined via cross-validation (see text); the total number of potential features considered when building each model is shown for reference. Cross-validated performance measures are shown as mean ± standard deviation. *D*(optimal) is shown as median ± median absolute deviation. Model 9 (MKL-Gaussian) was selected as the best performing model and studied further in subsequent analyses.

Next, we compared the predictive performance of multiple-kernel, multi-source models (6-9: MKL-Linear, MKL-LPG, MKL-Poly, MKL-Gaussian) relative to that of the single-kernel, multi-source CONCAT model. In terms of predictive accuracy (T-BAR; Table 3.2 and Figure 3.3A), the CONCAT model performed similarly to MKL-LPG, MKL-Poly, and MKL-Gaussian models (all $P > 0.3$, paired-sample *t*-test) and outperformed the MKL-Linear model ($P < 0.001$). While MKL-LPG and MKL-Poly models were as equally well-calibrated as the CONCAT model (as indicated by the CCC; both $P > 0.2$), the MKL-Linear model was less well calibrated ($P < 0.01$) and the MKL-Gaussian model was better calibrated ($P < 0.05$; Figure 3.3B) than the CONCAT model.

Based on its predictive accuracy and calibration, model 9 (MKL-Gaussian) was selected as the best performing model to be studied in subsequent analyses. Figure 3.4 shows a series of performance curves that further characterize the predictive performance of the MKL-Gaussian model and complement the information presented in Table 3.2 and Figure 3.3. Figure 3.4A shows how model accuracy evolves with increasing number of features. The accuracy increases rapidly and peaks when there are 10 features in the model, with a gradual decline in accuracy observed upon the inclusion of additional features. Figure 3.4B shows the ROC curve and Figure 3.4C shows the calibration curve for the MKL-Gaussian model.

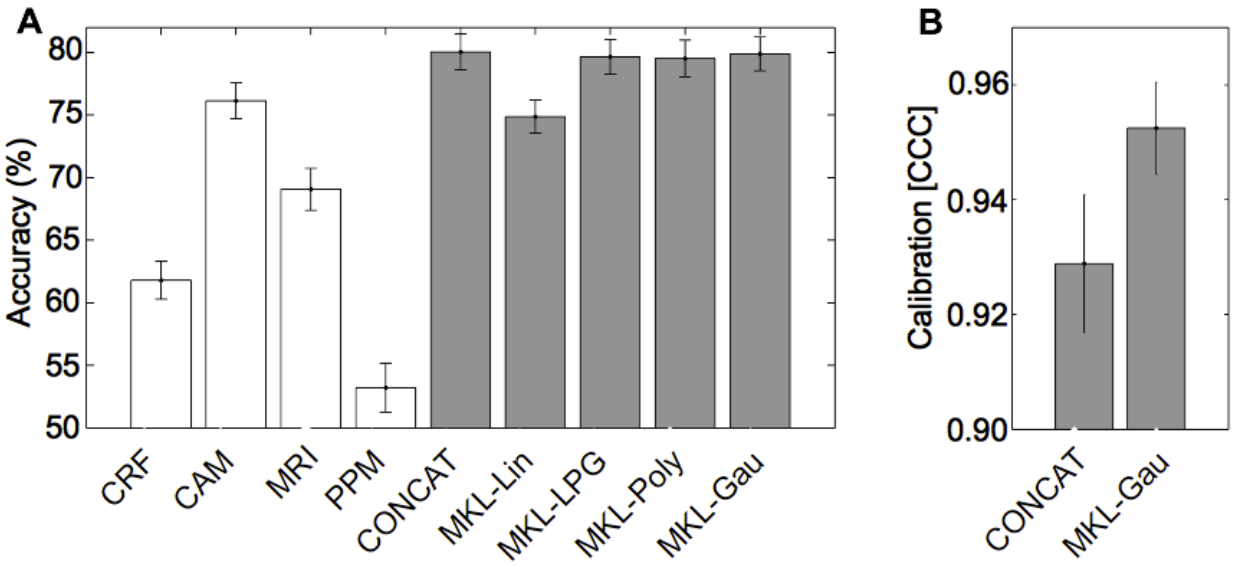


Figure 3.3. Selected measures of model performance. A: Predictive accuracy, defined as the balanced accuracy rate, is shown for single-source models (white bars) and multi-source models (gray bars). B: Calibration, quantified by the concordance correlation coefficient (CCC), is shown for selected models. Values represent means; error bars are 95% confidence intervals.

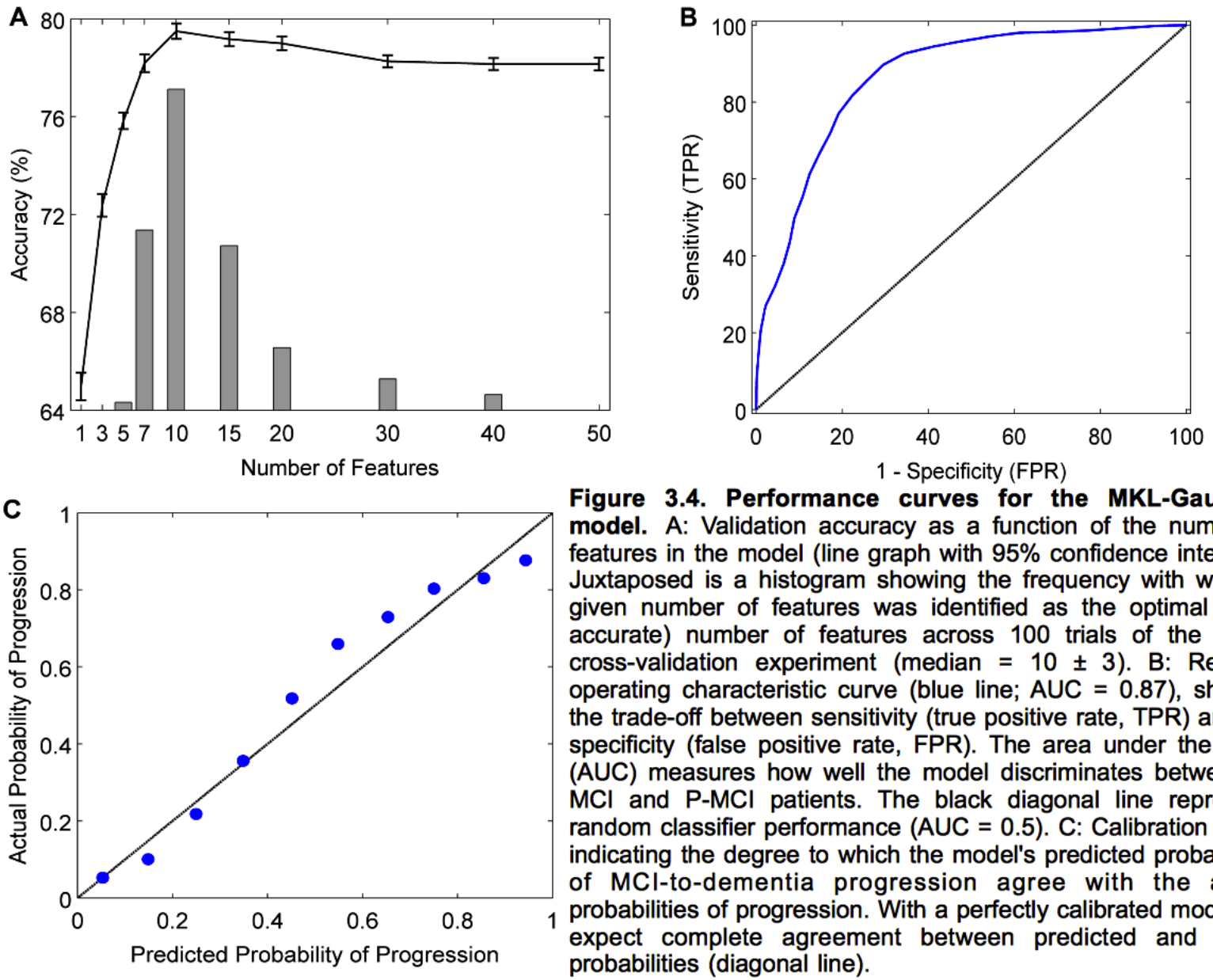


Figure 3.4. Performance curves for the MKL-Gaussian model. A: Validation accuracy as a function of the number of features in the model (line graph with 95% confidence intervals). Juxtaposed is a histogram showing the frequency with which a given number of features was identified as the optimal (most accurate) number of features across 100 trials of the 10x10 cross-validation experiment (median = 10 ± 3). B: Receiver operating characteristic curve (blue line; AUC = 0.87), showing the trade-off between sensitivity (true positive rate, TPR) and 1 – specificity (false positive rate, FPR). The area under the curve (AUC) measures how well the model discriminates between N-MCI and P-MCI patients. The black diagonal line represents random classifier performance (AUC = 0.5). C: Calibration curve, indicating the degree to which the model's predicted probabilities of MCI-to-dementia progression agree with the actual probabilities of progression. With a perfectly calibrated model, we expect complete agreement between predicted and actual probabilities (diagonal line).

3.3.3 Predictors of MCI-to-Dementia Progression

Figure 3.5 shows the top 10 features that were most frequently selected as baseline predictors of MCI-to-dementia progression for each of the single-source models (CRF, CAM, MRI, PPM) and the multi-source MKL-Gaussian model. Figure 3.6 shows the topography of the brain regions selected as predictors in the MRI and MKL-Gaussian models. Among the features considered for selection in the clinical risk factor (CRF) model, only the number of APOE ε4 alleles was selected with a high degree of consistency, with a selection frequency (SF) of 0.87. Other candidate CRF features, including age, were selected infrequently (SF < 0.20). The features most frequently selected in the CAM model included total scores and sub-scores on three assessments: ADAS-Cog, FAQ, and RAVLT (SF range 0.67-1.00). In the MRI model, the most frequently selected features included volume and cortical thickness measures for several brain regions within temporal and parietal lobes (SF range 0.52-1.00). In the plasma proteomic markers (PPM) model, the most frequently selected features included proteins associated with vascular processes, immune function and inflammation, and lipid metabolism (SF range 0.76-0.95).

In the multi-source MKL-Gaussian model (where the optimal number of features selected was 10 ± 3; Figure 3.4A), only CAM and MRI features were selected as predictors (Figure 3.5). CAM features included the 13-item total score and constructional praxis sub-score on the ADAS-Cog, the total score and memory question sub-score on the FAQ, as well as the sum of scores across trials 1-5, trial 5 sub-score, and trial 6 sub-score on the RAVLT (SF range 0.70-1.00). MRI features included left hippocampal volume, left middle temporal cortical thickness, and left inferior parietal cortical thickness (SF range 0.67-0.96; Figure 3.6).

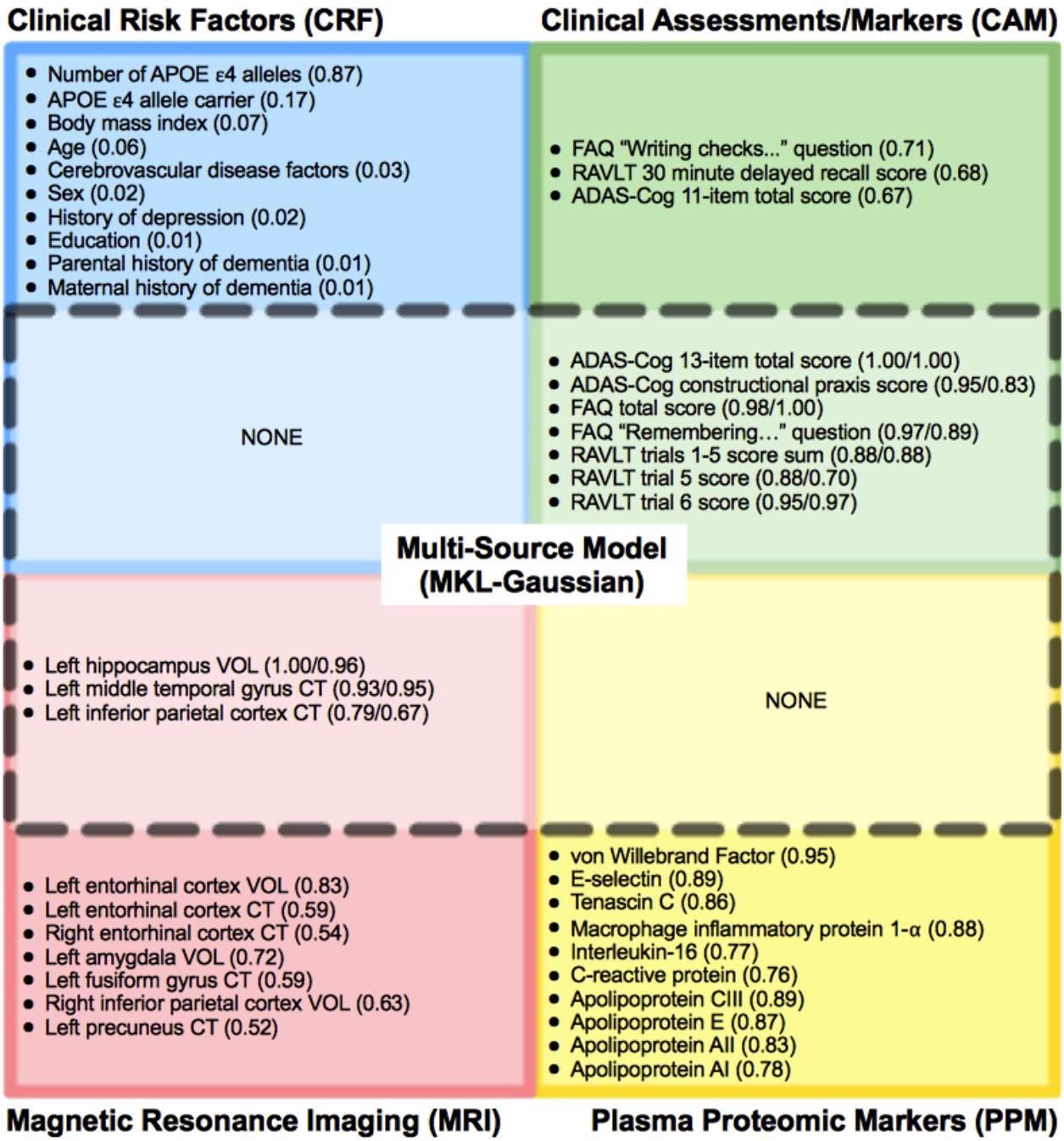


Figure 3.5. Top 10 most frequently selected features as baseline predictors of MCI-to-dementia progression. Features are shown separately for each single-source model: CRF (blue), CAM (green), MRI (red), PPM (yellow). A subset of these features was selected as part of the multi-source MKL-Gaussian model (shown within the dashed area) and included only CAM and MRI features. The selection frequency across 100 trials of the 10x10 cross-validation experiment is shown in parentheses as: (#) for single-source model only or (#/#) for both single/multi-source models. APOE = apolipoprotein E, VOL = volume, CT = cortical thickness, ADAS-Cog = Alzheimer's Disease Assessment Scale – Cognitive sub-scale, FAQ = Functional Activities Questionnaire, RAVLT = Rey Auditory-Verbal Learning Test.

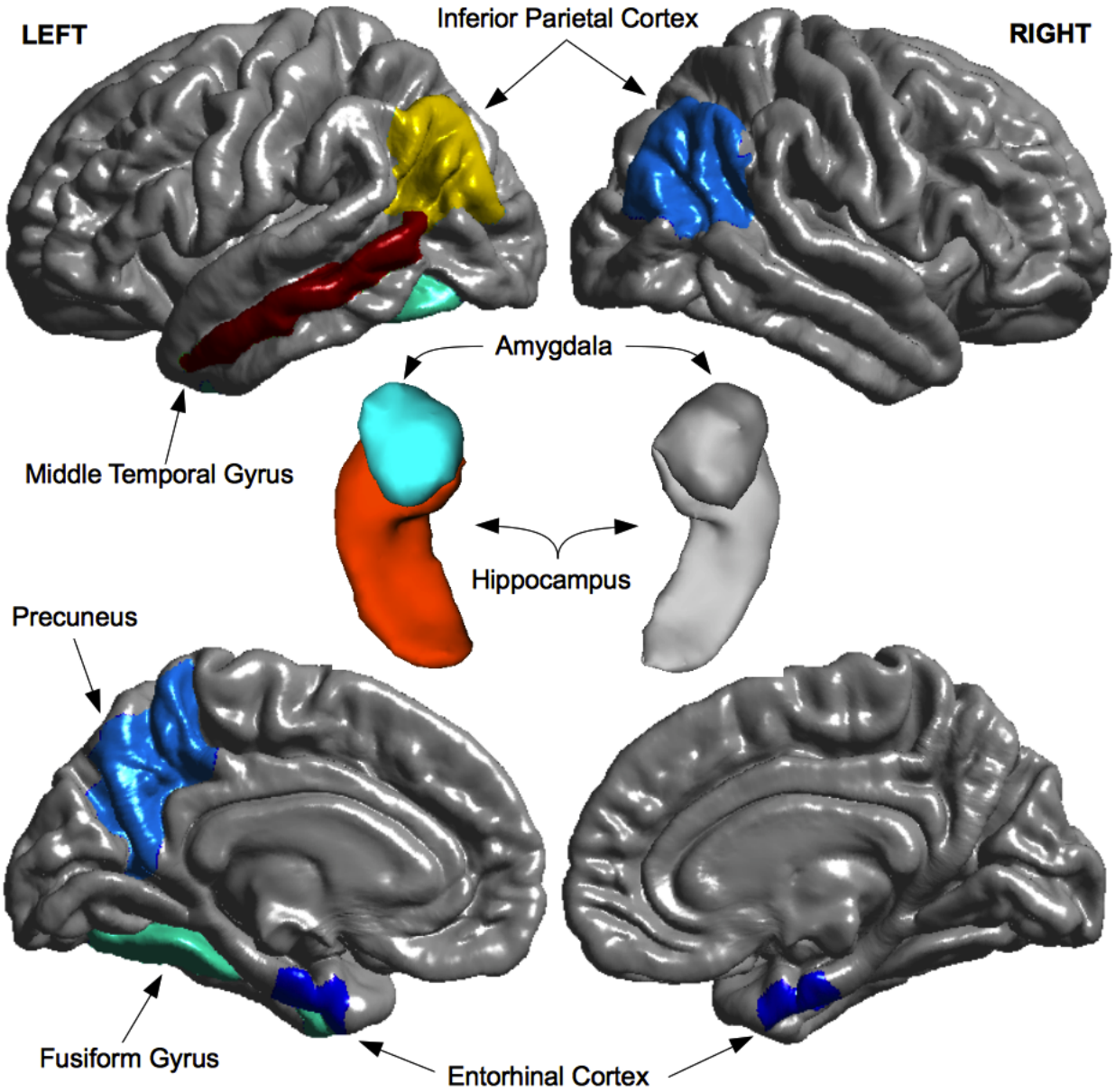


Figure 3.6. Regional MRI predictors of MCI-to-dementia progression. Morphometric measures (volumes and cortical thickness) for brain regions shown in both warm and cool colors were selected as predictors in the single-source MRI model. Morphometric measures for a subset of these regions, shown in warm colors (red, orange, yellow), were also selected as predictors in the multi-source MKL-Gaussian model. Regions of interest are overlaid on top of 3-D model reconstructions of the brain (gray). Top row: lateral view of the cerebral hemispheres. Center: close-up view of the hippocampus-amygdala complex. Bottom row: medial view of the cerebral hemispheres.

As a confirmatory analysis, we compared N-MCI and P-MCI groups on each of the baseline predictors identified in the MKL-Gaussian model (Figure 3.7). As expected, there was a robust statistically significant difference between the two MCI groups for all predictor variables (all $P < 0.001$, independent sample *t*-test). P-MCI subjects were more cognitively and functionally impaired at baseline than N-MCI subjects, as indicated by higher scores on the ADAS-Cog and FAQ. Relative to N-MCI subjects, P-MCI subjects had a more pronounced verbal memory impairment at baseline, as indicated by lower scores on the RAVLT. P-MCI subjects also showed signs of atrophy in temporal and parietal brain regions at baseline, as indicated by reduced hippocampal volume as well as reduced middle temporal and inferior parietal cortical thickness relative to N-MCI subjects.

In this study, we used data from 259 out of 390 MCI subjects in the ADNI-1 database to develop and evaluate predictive models of dementia. The other 131 MCI subjects (~34%) were excluded from the study because they either did not meet our inclusion criteria or due to partially missing data. To examine a potential selection bias that this subject exclusion can introduce into the predictive model, we compared the included and the excluded MCI subjects on each of the baseline predictor variables (Figure 3.8). The included subjects were statistically similar to the excluded subjects on all predictor variables (all $P > 0.4$, independent sample *t*-test).

3.3.4 Influence of Patient Characteristics on Model Performance

To better characterize the best-performing model in this study (MKL-Gaussian), we examined whether the model's predictive accuracy varies as a function of age, sex, education level, APOE genotype, off-label use of AD medications, presence of

cerebrovascular disease risk factors, and history of depression (Figure 3.9A-G). Overall, the model tended to make more accurate predictions about future dementia status for patients with the following characteristics: older age; female sex; higher educational level; APOE ε4 negative genotype; not using AD medications; history of a greater number of conditions considered to be cerebrovascular disease risk factors; or a history of depression. However, the particular effects on the sensitivity and specificity of the model were more complex and depended on the patient characteristic being considered. In the case of P-MCI subjects, predictive accuracy was inversely related to the time-to-progression from MCI to dementia (Figure 3.9H): 0-6 months (93.1%), 6-12 months (89.3%), 12-18 months (87.6%), 18-24 months (74.8%), 24-36 months (71.3%). There was a large difference in terms of accuracy for MCI patients progressing to dementia within 18 months (89.4%) and those progressing after 18 months (73.3%).

3.3.5 Probabilistic Classification and Staging of MCI Patients

We investigated whether probabilistic outputs from the pMKL classifier could be used to improve the predictive accuracy of our prognostic model by permitting only "high confidence" predictions to be made. Figure 3.10 shows how the predictive accuracy varies as we require different levels of confidence to make predictions. As we raise the level of confidence required to make predictions, the accuracy of the model gradually increases. However, this increase in predictive accuracy comes at a cost; with increasing minimum levels of confidence required, the model is able to make such "high confidence" predictions for an increasingly smaller proportion of patients. For example, requiring a minimum predictive confidence level of 0.4 (equivalent to a predicted probability of 0.70), improved model accuracy from 79.9% (83.4% sensitivity, 76.4%

specificity) to 87.4% (91.7% sensitivity, 83.2% specificity). This improved accuracy was achieved by allowing predictions to be made only for the top ~73% most confident patient cases, while designating the predictions for the other ~27% of patient cases as ambiguous or low confidence. We also examined whether probabilistic outputs from the pMKL classifier reflect the time-to-progression information for individual P-MCI subjects, and thus, could potentially be used for staging where along the MCI-AD continuum an MCI subject is. Correlation analysis revealed that there was a small but statistically significant negative association between the predicted probabilities of P-MCI and the time-to-progression (i.e. larger probability of P-MCI was associated with shorter time-to-progression; $r = -0.20$, $P < 0.05$, Spearman correlation).

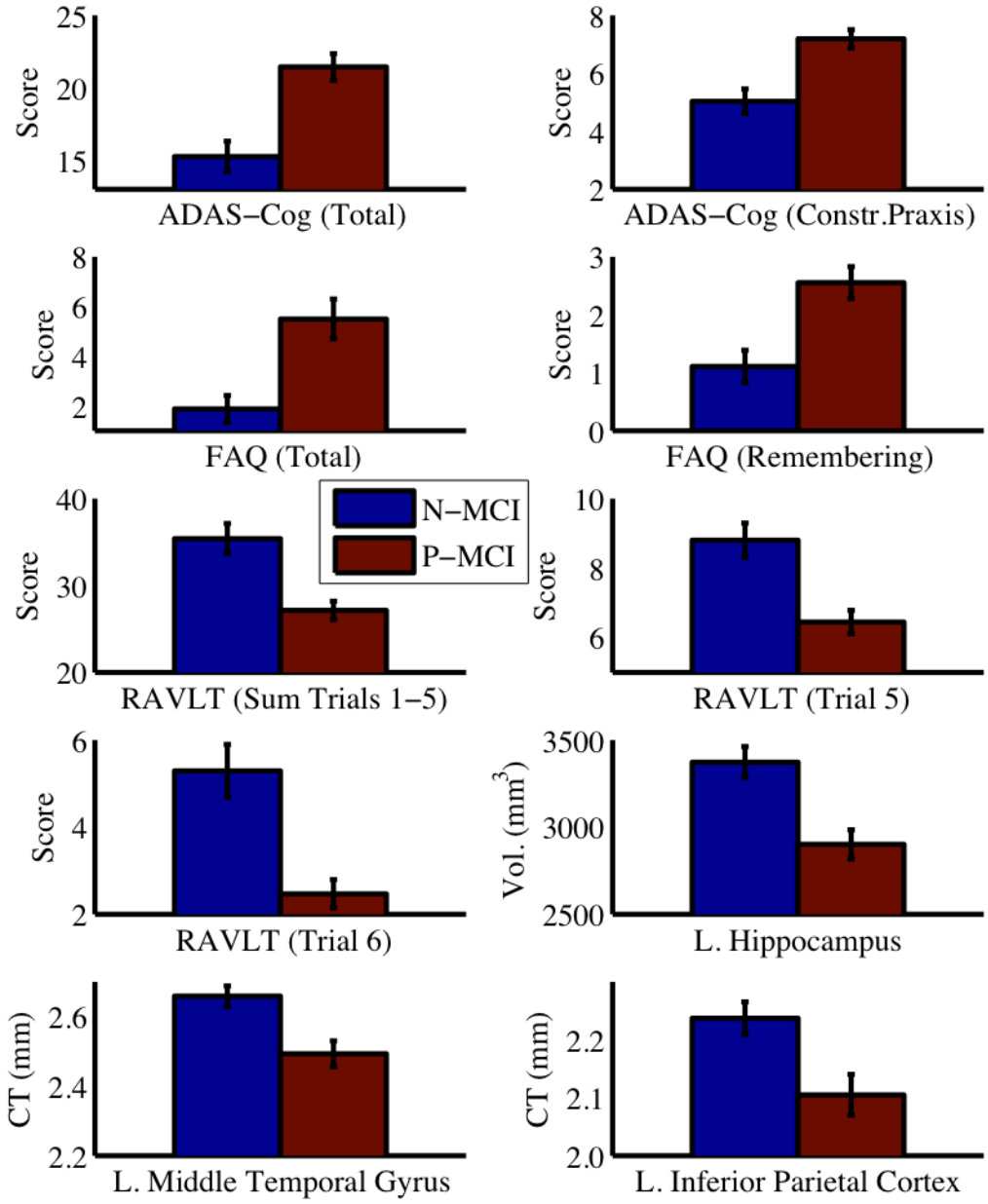


Figure 3.7. Comparison between N-MCI and P-MCI groups on baseline predictor variables. Error bars are 95% confidence intervals. Significant group differences were present for all predictor variables (all $P < 0.001$). Vol. = volume, CT = cortical thickness, ADAS-Cog = Alzheimer's Disease Assessment Scale – Cognitive sub-scale, FAQ = Functional Activities Questionnaire, RAVLT = Rey Auditory-Verbal Learning Test, L. = Left, Constr. = Constructional

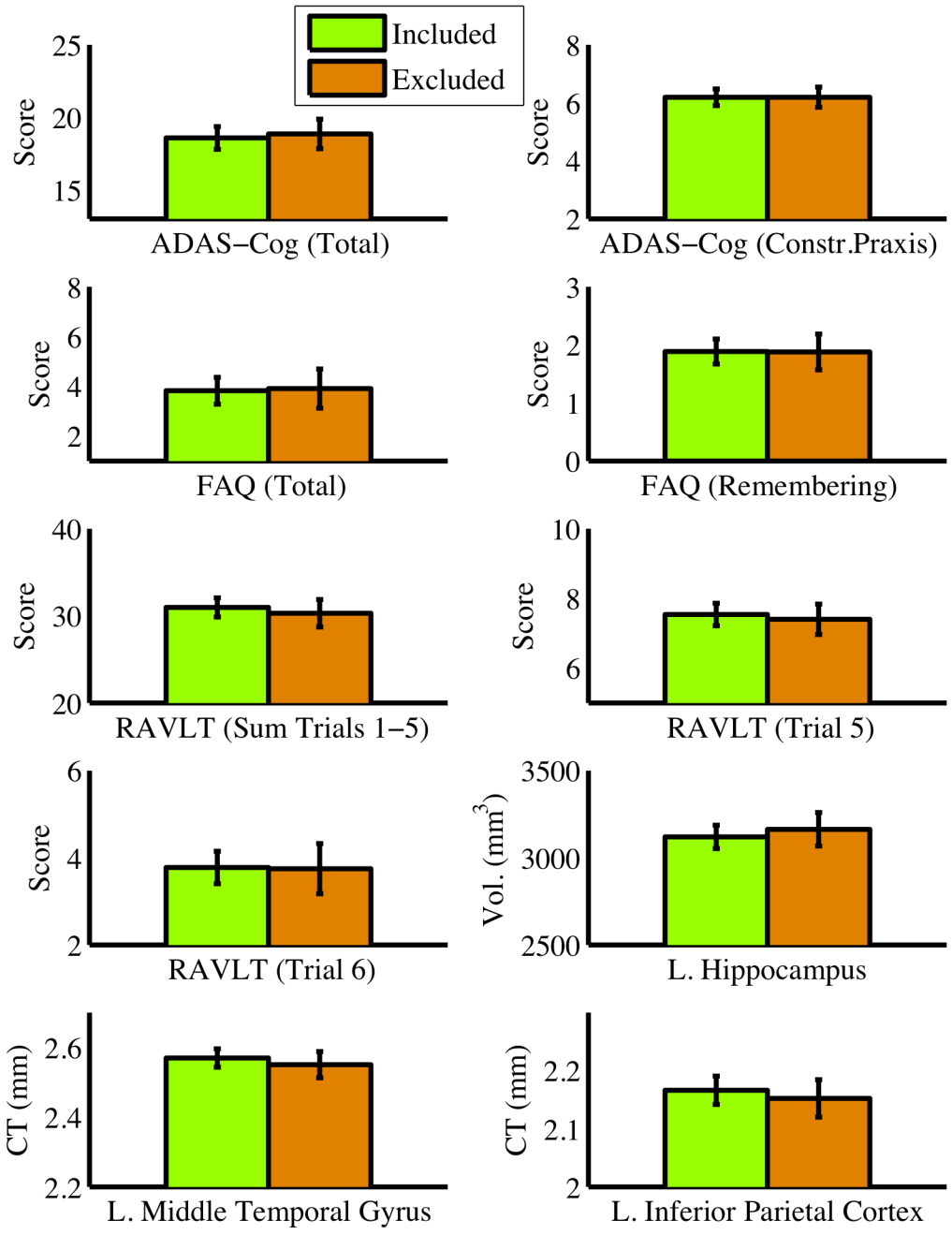


Figure 3.8. Comparison between included (n=259) and excluded (n=131) MCI subjects on baseline predictor variables. Error bars are 95% confidence intervals. No group differences were found for any of the predictor variables (all $P > 0.4$). Vol. = volume, CT = cortical thickness, ADAS-Cog = Alzheimer's Disease Assessment Scale – Cognitive sub-scale, FAQ = Functional Activities Questionnaire, RAVLT = Rey Auditory-Verbal Learning Test, L. = Left, Constr. = Constructional

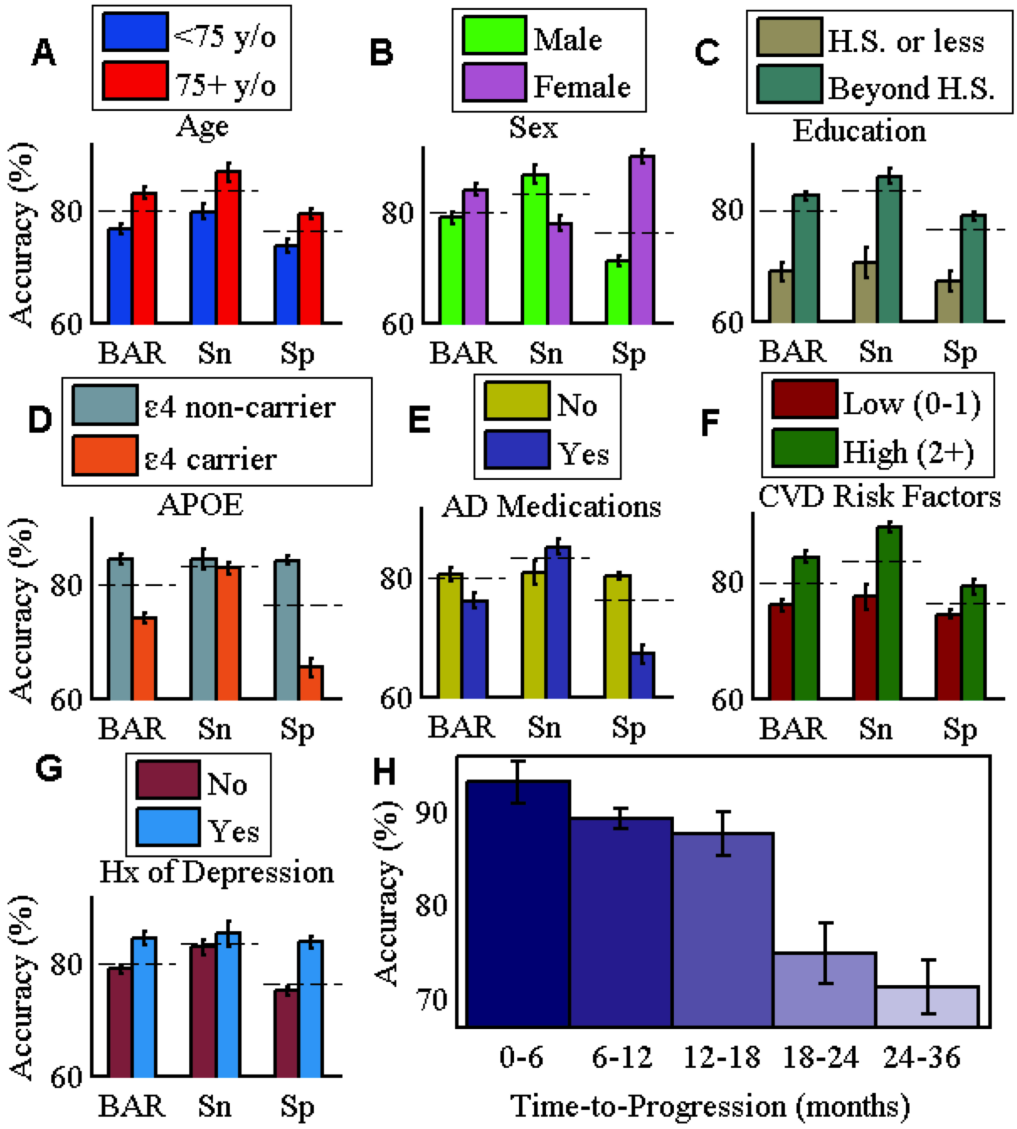


Figure 3.9. Effect of patient characteristics on predictive accuracy. The predictive accuracy of the model (MKL-Gaussian) was not uniform across various patient characteristics. Accuracy varied with baseline demographic (A-C), genetic (D), and clinical (E-G) characteristics. Panel E shows a comparison of MCI patients who were taking medications for Alzheimer's disease (AD) off-label versus those who were not. Panel F compares patients according to the number of pre-existing conditions in their medical history that are considered to be cerebrovascular disease (CVD) risk factors, including diabetes mellitus, coronary artery disease, hypertension, smoking, hyperlipidemia, and stroke. The predictive accuracy of the model varied inversely with time-to-progression for P-MCI patients (H). The overall accuracy of the model (as found in Table 3.2) is shown for reference as a dashed line. Error bars represent 95% confidence intervals across cross-validation trials. BAR = Balanced Accuracy Rate, Sn = Sensitivity, Sp = Specificity, y/o = years old, H.S. = high school, APOE = apolipoprotein E, Hx = history.

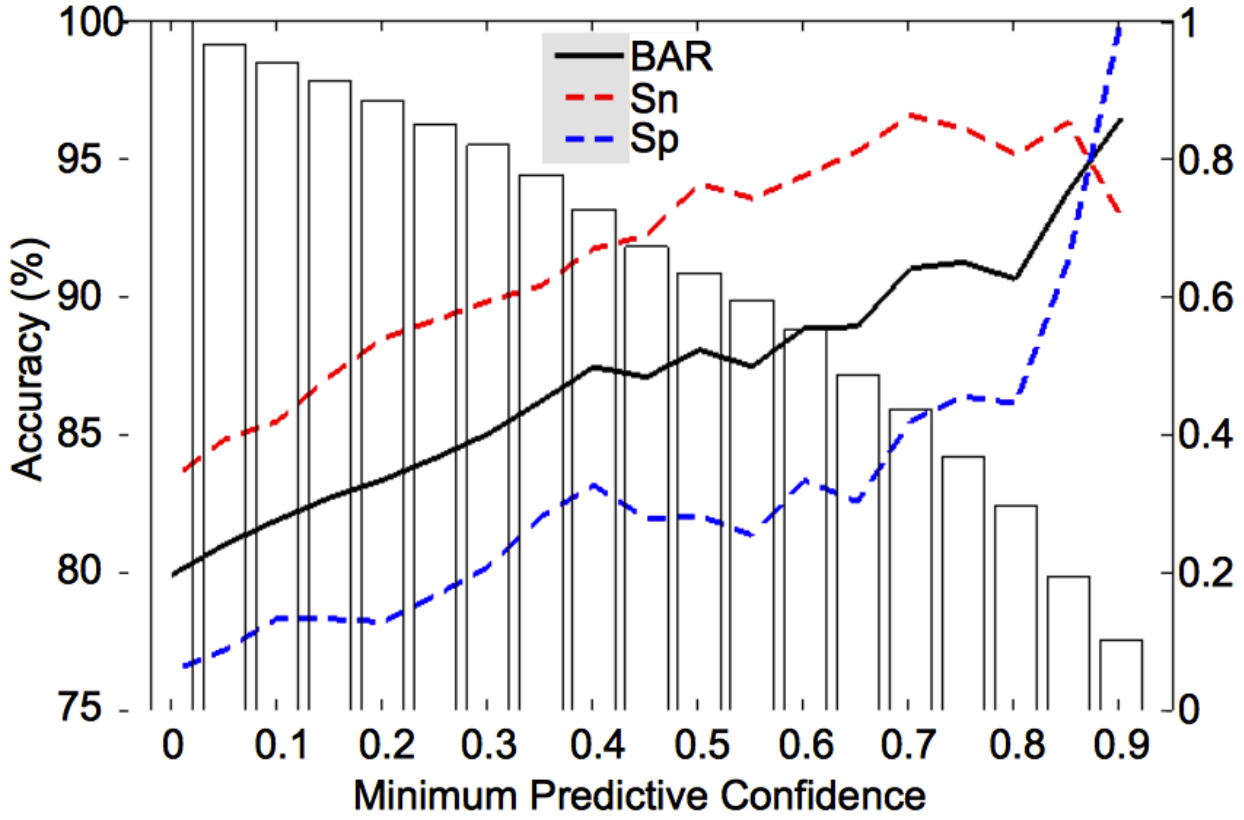


Figure 3.10. Model accuracy as a function of predictive confidence. Increasing the minimum confidence required to make predictions resulted in improved model accuracy (solid and dashed lines; left y-axis), albeit at the cost of a decreasing proportion of MCI patients for whom "high confidence" predictions could be made (white bars; right y-axis). Predictive confidence was defined as the difference between the predicted probabilities for the N-MCI and P-MCI groups. BAR = Balanced Accuracy Rate, Sn = Sensitivity, Sp = Specificity.

3.4 Discussion

In this study, we developed and evaluated several prognostic models for predicting MCI-to-dementia progression during a three-year period. For this task, we analyzed clinical, structural MRI, and plasma proteomic data collected from a large prospectively-followed cohort of 259 MCI patients who were recruited through the collaborative, multi-site Alzheimer's Disease Neuroimaging Initiative (ADNI). We highlight the most important findings from the study, discuss the advantages and limitations of our methodology, and point out some future directions for this work.

3.4.1 Predictive Utility of Clinical, MRI, and Plasma Proteomic Biomarkers

Among the four data sources examined, clinical assessments proved to be the most accurate for predicting MCI-to-dementia progression. Similar to our findings, Gomar et al. (Gomar et al., 2011) showed in a series of logistic regression models that cognitive markers outperform volumetric MRI and CSF biomarkers for predicting progression to dementia over a two-year period. Using SVM classifiers, Cui et al. (Cui et al., 2011) also demonstrated that cognitive and functional measures were more predictive of progression then MRI and CSF measures. Of the four data sources, plasma proteomic biomarkers had the lowest predictive accuracy that was only marginally better than chance level. In addition, the median number of plasma proteomic features selected was substantially larger than the number of features selected for the other data sources (40 versus 15 or less). This suggests that as a potential source of biomarkers, plasma proteomic data has a low signal-to-noise ratio and limited predictive utility for predicting MCI-to-dementia progression. Consistent with our findings but using a different pattern classification strategy, Johnstone and

colleagues (Johnstone et al., 2012) showed that plasma-based proteomic measures could not reliably discriminate between progressive and non-progressive MCI subjects. The predictive utility of structural MRI measures and clinical risk factors was found to be intermediate between that of clinical assessments and plasma proteomic measures.

Among the nine models examined in this study, the model that yielded the best predictive accuracy (MKL-Gaussian) was comprised of clinical assessments and MRI measures, but none of the clinical risk factors or plasma proteomic measures. The combination of clinical assessments and MRI measures outperformed using either source alone for classification of N-MCI and P-MCI subjects, indicating that these sources of data provide complementary information regarding MCI-to-dementia progression. The optimal subset of predictors included baseline total scores and sub-scores on two cognitive tests (ADAS-Cog and RAVLT), a measure of functional status (FAQ), and morphometric measures for three brain regions (left hippocampus, middle temporal gyrus, and inferior parietal cortex). The inclusion of ADAS-Cog scores as predictors in the model, in addition to the RAVLT, suggests that baseline deficits in multiple cognitive domains are predictive of future progression to dementia. This finding is corroborated by previous reports that MCI patients with both memory and non-memory deficits have a greater risk of progression to AD dementia than patients who have memory-only deficits (Bozoki et al., 2001). The inclusion of brain regions in addition to (and outside) of the hippocampus as predictors of MCI-to-dementia progression is consistent with the known pattern of grey matter atrophy (neurodegeneration) associated with AD. The atrophy is known to begin in the medial temporal lobes and, as the disease progresses, spreads to affect the temporal and parietal association cortices. The preferential selection of morphometric MRI features

involving left but not right-sided brain regions suggests that baseline atrophy of left-sided regions is more predictive of MCI-to-dementia progression than atrophy of right-sided regions. Right-sided brain regions appear to provide redundant information about progression. This preferential selection of left-sided temporoparietal brain regions as predictors of progression is consistent with evidence that AD-related atrophy occurs at a faster rate in the left hemisphere (Thompson et al., 2003).

3.4.2 Effect of Multiple Kernel Learning on Model Performance

In this study, the effect of MKL on model performance was modest. Specifically, multi-source, multiple-kernel classifiers did not outperform the multi-source, single-kernel classifier in terms of predictive accuracy. However, we found that the MKL-Gaussian classifier (using five Gaussian kernels) modestly improved the calibration of the resulting model relative to the single-kernel classifier. We used a relatively small number of kernels in our MKL models, and this could account for the limited benefit we observed with MKL. Using a larger number of kernels, as done in some recent studies (e.g. Hinrichs et al., 2011), could yield additional improvements in predictive performance.

3.4.3 Statistical Considerations: Overfitting and Selection Bias

Validation and test accuracies (V-BAR and T-BAR in Table 3.2) were within 3% of each other for all models examined in this study, and in many cases <1% apart, indicating that model overfitting was minimal and that our nested cross-validation procedure was effective. We evaluated feature subsets containing up to 50 features, which corresponds to a sample-to-predictor ratio of ~2:1 (based on the smallest class,

N-MCI, with n=120). From a classical statistical viewpoint, however, at least 10 samples per predictor are often recommended to minimize difficulties with model estimation and overfitting (Bouwmeester et al., 2012). Thus, a model with 15 or less predictors would be preferred given the sample size of our study. The MKL-Gaussian model, designated as the best performing model, met this recommendation with a median of 10 ± 3 features being selected as predictors. We also assessed for potential selection bias that may have occurred secondary to our exclusion of 131 (out of the 390) MCI subjects who did not meet the inclusion criteria or due to missing data. No differences were found between included (n=259) and excluded (n=131) subjects on any of the predictor variables in our model. This null finding provides reassurance that a selection bias due to the exclusion of subjects in this study is unlikely, at least with respect to the predictors of interest.

3.4.4 The Importance of Patient Heterogeneity in Pattern Classification

The best performing model for predicting MCI-to-dementia progression (MKL-Gaussian) did not contain any clinical risk factors, as they were deemed uninformative or redundant in the presence of clinical assessments and structural MRI measures. Nevertheless, we found that many of these patient characteristics have a substantial effect on the predictive accuracy of the model. For some patient characteristics, such as age and educational level, the effect on predictive accuracy was similar for both sensitivity and specificity. For other characteristics, such as APOE genotype, the effect on accuracy was limited to the specificity. Yet for other characteristics, such as sex, there was an opposite effect on the sensitivity and specificity. Overall, these findings suggest that the predictive performance of our model is highly nuanced and depends in

part on patient heterogeneity. In addition, we found that the model could identify P-MCI patients who progressed to AD dementia during the first 18 months after the baseline visit with substantially higher accuracy than patients who progressed after 18 months. Thus, clinical assessments and structural MRI biomarkers can detect AD-related changes most reliably up to 18 months prior to the onset of overt dementia. This finding is consistent with the AD biomarker model proposed by Jack and colleagues (2010), which states that different biomarkers have unique temporal trajectories and may be optimally sensitive to AD-related changes during specific time periods.

3.4.5 Probabilistic Pattern Classification

A unique aspect of this study is our adoption of a probabilistic kernel-based classifier (pMKL) for the prediction of MCI-to-dementia progression. As a kernel-based classifier, pMKL has a few advantages over the SVM (Damoulas and Girolami, 2009a, 2009b). While by design the SVM is a non-probabilistic binary classifier, pMKL directly produces probabilistic predictions without relying on ad-hoc methods to transform classifier outputs into posterior probabilities (e.g. Platt, 1999). Calibration analysis showed that the predicted probabilities of progression, as generated by the pMKL classifier, were in strong agreement with the actual probabilities of progression. This indicates that the probabilistic predictions produced by our model are reliable and could have practical application when predicting future dementia status in patients with MCI.

We investigated whether probabilistic outputs from the pMKL classifier could be used as measures of predictive confidence to improve the accuracy of the model. When using the classifier in the conventional, non-probabilistic mode, where no information about predictive confidence was taken into account, we obtained an overall accuracy of

79.9% (83.4% sensitivity, 76.4% specificity). The estimated risk of progression from amnestic MCI to dementia is 7-15% per year (Mitchell and Shiri-Feshki, 2009; Petersen et al., 2009). By assuming a 30% risk of progression over a three-year period (~10% annually) as the pre-test probability, a prognostic test with this level of sensitivity and specificity would yield a positive post-test probability of 60.2% and a negative post-test probability of 8.5% (via application of the Bayes' rule; Westover et al., 2011). When we used the classifier in the probabilistic mode, where predictions were allowed to be made only for the top ~73% most confident patient cases, the predictive accuracy was 87.4% (91.7% sensitivity, 83.2% specificity). A diagnostic test with this improved level of sensitivity and specificity would yield a positive post-test probability of 70.1% and a negative post-test probability of 4.1%. This means that 60.2% (non-probabilistic version) and 70.1% (probabilistic version) of amnestic MCI patients for whom our prognostic model makes a positive prediction will progress to dementia within a three-year period. In a clinical trial, the use of our prognostic model to enrich the MCI patient sample would result in more than a 50% reduction in the required sample size to detect a drug effect. This would result in a substantially more time-and cost-efficient clinical trial.

3.4.6 Limitations

The present study has a few limitations, some that apply to the ADNI in general and some that are specific to this study. First, our predictive model was designed only for patients with amnestic MCI. The amnestic MCI subtype comprises 2/3 of all MCI cases, and in addition, patients with amnestic MCI have a substantially elevated risk for developing dementia of the Alzheimer's type compared to patients with non-amnestic MCI. From this perspective, it is reasonable that the ADNI – as the first such study of its

kind – focuses on these higher-risk patients. Second, we considered only clinical, structural MRI, and plasma proteomic data in this study, as minimally-invasive, less expensive, and more widely available sources of data. Our pattern classification and analytic strategy could be applied to additional data sources, including PET, DTI, and functional MRI data, CSF and other fluid biomarkers, and genetic expression data. Third, in this study we focus on binary classification, attempting to accurately differentiate between patients who progress to a AD-type dementia and those who do not. For clinical application, a prognostic model of dementia would have a greater utility if it also could predict progression to other types of dementia (e.g. vascular dementia, frontotemporal dementia, dementia with Lewy bodies). In addition, many cases of dementia do not have a single underlying cause but rather are mixed dementias, with contributions from multiple disease processes (e.g. Alzheimer's disease mixed with vascular dementia). In these more complex cases of mixed dementia or where differential diagnosis of dementia is sought, a multi-class classification approach would be more appropriate. The pMKL classifier adopted in this study can naturally be extended to these more challenging prediction tasks. Finally, nested cross-validation procedure allows internal validation of a model (developed and evaluated using the same dataset), as done in this study using the ADNI dataset. Prior to real-world application of the predictive model developed in this study, it must be externally validated on independent datasets (Bouwmeester et al., 2012; Steyerberg et al., 2013).

REFERENCES

REFERENCES

Acevedo, A., Loewenstein, D.A., Barker, W.W., Harwood, D.G., Luis, C., Bravo, M., Hurwitz, D.A., Aguero, H., Greenfield, L., and Duara, R. (2000). Category Fluency Test: Normative data for English- and Spanish-speaking elderly. Journal of the International Neuropsychological Society *6*, 760–769.

Altman, D.G., and Bland, J.M. (1994a). Diagnostic tests. 1: Sensitivity and specificity. BMJ *308*, 1552.

Altman, D.G., and Bland, J.M. (1994b). Diagnostic tests 3: receiver operating characteristic plots. BMJ *309*, 188.

Barnes, D.E., and Yaffe, K. (2011). The projected effect of risk factor reduction on Alzheimer's disease prevalence. Lancet Neurol *10*, 819–828.

Ben-Hur, A., Ong, C.S., Sonnenburg, S., Schölkopf, B., and Rätsch, G. (2008). Support vector machines and kernels for computational biology. PLoS Comput. Biol. *4*, e1000173.

Bouckaert, R.R. (2003). Choosing between Two Learning Algorithms Based on Calibrated Tests. In ICML'03, (Morgan Kaufmann), pp. 51–58.

Bouwmeester, W., Zuithoff, N.P.A., Mallett, S., Geerlings, M.I., Vergouwe, Y., Steyerberg, E.W., Altman, D.G., and Moons, K.G.M. (2012). Reporting and methods in clinical prediction research: a systematic review. PLoS Med. *9*, 1–12.

Bozoki, A., Giordani, B., Heidebrink, J.L., and Foster, N.L. (2001). Mild cognitive impairments predict dementia in nondemented elderly patients with memory loss. Arch. Neurol. *58*, 411–416.

Brown, G., Pocock, A., Zhao, M.-J., and Luján, M. (2012). Conditional Likelihood Maximisation: A Unifying Framework for Information Theoretic Feature Selection. J. Mach. Learn. Res. *13*, 27–66.

Buckner, R.L., Head, D., Parker, J., Fotenos, A.F., Marcus, D., Morris, J.C., and Snyder, A.Z. (2004). A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: reliability and validation against manual measurement of total intracranial volume. Neuroimage *23*, 724–738.

Cawley, G.C., and Talbot, N.L.C. (2010). On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation. Journal of Machine Learning Research *11*, 2079−2107.

Chen, R., and Herskovits, E.H. (2010). Machine-learning techniques for building a diagnostic model for very mild dementia. NeuroImage *52*, 234–244.

Cui, Y., Liu, B., Luo, S., Zhen, X., Fan, M., Liu, T., Zhu, W., Park, M., Jiang, T., and Jin, J.S. (2011). Identification of Conversion from Mild Cognitive Impairment to Alzheimer's Disease Using Multivariate Predictors. PLoS One *6*, e21896.

Dale, A.M., Fischl, B., and Sereno, M.I. (1999). Cortical surface-based analysis. I. Segmentation and surface reconstruction. Neuroimage *9*, 179–194.

Damoulas, T., and Girolami, M.A. (2008). Probabilistic multi-class multi-kernel learning: on protein fold recognition and remote homology detection. Bioinformatics *24*, 1264–1270.

Damoulas, T., and Girolami, M.A. (2009a). Combining feature spaces for classification. Pattern Recognition *42*, 2671–2683.

Damoulas, T., and Girolami, M.A. (2009b). Pattern recognition with a Bayesian kernel combination machine. Pattern Recognition Letters *30*, 46–54.

Davatzikos, C., Bhatt, P., Shaw, L.M., Batmanghelich, K.N., and Trojanowski, J.Q. (2011). Prediction of MCI to AD conversion, via MRI, CSF biomarkers, pattern classification. Neurobiol Aging *32*, 2322.e19–2322.e27.

Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage *31*, 968–980.

Desikan, R.S., Cabral, H.J., Hess, C.P., Dillon, W.P., Glastonbury, C.M., Weiner, M.W., Schmansky, N.J., Greve, D.N., Salat, D.H., Buckner, R.L., et al. (2009). Automated MRI measures identify individuals with mild cognitive impairment and Alzheimer's disease. Brain *132*, 2048–2057.

Ding, C., and Peng, H. (2005). Minimum redundancy feature selection from microarray gene expression data. J Bioinform Comput Biol *3*, 185–205.

Fischl, B., Sereno, M.I., and Dale, A.M. (1999). Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. Neuroimage *9*, 195–207.

Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., Van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., et al. (2002). Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron *33*, 341–355.

Folstein, M.F., Folstein, S.E., and McHugh, P.R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res *12*, 189–198.

Gomar, J.J., Bobes-Bascaran, M.T., Conejero-Goldberg, C., Davies, P., and Goldberg, T.E. (2011). Utility of Combinations of Biomarkers, Cognitive Markers, and Risk Factors to Predict Conversion From Mild Cognitive Impairment to Alzheimer Disease in Patients in the Alzheimer's Disease Neuroimaging Initiative. Arch Gen Psychiatry *68*, 961–969.

Gönen, M., and Alpaydın, E. (2011). Multiple Kernel Learning Algorithms. J. Mach. Learn. Res. *12*, 2211–2268.

Graff-Radford, N.R., Crook, J.E., Lucas, J., Boeve, B.F., Knopman, D.S., Ivnik, R.J., Smith, G.E., Younkin, L.H., Petersen, R.C., and Younkin, S.G. (2007). Association of low plasma Abeta42/Abeta40 ratios with increased imminent risk for mild cognitive impairment and Alzheimer disease. Arch. Neurol. *64*, 354–362.

Grober, E., Sliwinsk, M., and Korey, S.R. (1991). Development and Validation of a Model for Estimating Premorbid Verbal Intelligence in the Elderly. Journal of Clinical and Experimental Neuropsychology *13*, 933–949.

Haller, S., Lovblad, K.O., and Giannakopoulos, P. (2011). Principles of classification analyses in mild cognitive impairment (MCI) and Alzheimer disease. J. Alzheimers Dis. *26 Suppl 3*, 389–394.

Hansson, O., Zetterberg, H., Vanmechelen, E., Vanderstichele, H., Andreasson, U., Londos, E., Wallin, A., Minthon, L., and Blennow, K. (2010). Evaluation of plasma Abeta(40) and Abeta(42) as predictors of conversion to Alzheimer's disease in patients with mild cognitive impairment. Neurobiol. Aging *31*, 357–367.

Herbei, R., and Wegkamp, M.H. (2006). Classification with reject option. Canadian Journal of Statistics *34*, 709–721.

Hester, R.L., Kinsella, G.J., and Ong, B. (2004). Effect of age on forward and backward span tasks. J Int Neuropsychol Soc *10*, 475–481.

Hinrichs, C., Singh, V., Xu, G., and Johnson, S.C. (2011). Predictive Markers for AD in a Multi-Modality Framework: An Analysis of MCI Progression in the ADNI Population. Neuroimage *55*, 574–589.

Hofmann, T., Schölkopf, B., and Smola, A.J. (2008). Kernel methods in machine learning. Ann. Statist. *36*, 1171–1220.

Holtzman, D.M., Morris, J.C., and Goate, A.M. (2011). Alzheimer's disease: the challenge of the second century. Sci Transl Med *3*, 77sr1.

Iwatsubo, T., Odaka, A., Suzuki, N., Mizusawa, H., Nukina, N., and Ihara, Y. (1994). Visualization of A beta 42(43) and A beta 40 in senile plaques with end-specific A beta monoclonals: evidence that an initially deposited species is A beta 42(43). Neuron *13*, 45–53.

Jack, C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P.J., L Whitwell, J., Ward, C., et al. (2008). The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. J Magn Reson Imaging *27*, 685–691.

Johnson, D.K., Storandt, M., and Balota, D.A. (2003). Discourse analysis of logical memory recall in normal aging and in dementia of the Alzheimer type. Neuropsychology *17*, 82–92.

Johnstone, D., Milward, E.A., Berretta, R., and Moscato, P. (2012). Multivariate protein signatures of pre-clinical Alzheimer's disease in the Alzheimer's disease neuroimaging initiative (ADNI) plasma proteome dataset. PLoS ONE *7*, e34341.

Joy, S., Kaplan, E., and Fein, D. (2004). Speed and memory in the WAIS-III Digit Symbol--Coding subtest across the adult lifespan. Arch Clin Neuropsychol *19*, 759–767.

Karow, D.S., McEvoy, L.K., Fennema-Notestine, C., Hagler, D.J., Jennings, R.G., Brewer, J.B., Hoh, C.K., and Dale, A.M. (2010). Relative capability of MR imaging and FDG PET to depict changes associated with prodromal and early Alzheimer disease. Radiology *256*, 932–942.

Kaufer, D.I., Cummings, J.L., Ketchel, P., Smith, V., MacMillan, A., Shelley, T., Lopez, O.L., and DeKosky, S.T. (2000). Validation of the NPI-Q, a brief clinical form of the Neuropsychiatric Inventory. J Neuropsychiatry Clin Neurosci *12*, 233–239.

Kim, K.I., and Simon, R. (2011). Probabilistic classifiers with high-dimensional data. Biostatistics *12*, 399–412.

Klöppel, S., Abdulkadir, A., Jack Jr., C.R., Koutsouleris, N., Mourão-Miranda, J., and Vemuri, P. (2012). Diagnostic neuroimaging across diseases. NeuroImage *61*, 457–463.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2, (San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.), pp. 1137–1143.

Lin, L.I. (1989). A concordance correlation coefficient to evaluate reproducibility. Biometrics *45*, 255–268.

Manly, J.J., Tang, M.-X., Schupf, N., Stern, Y., Vonsattel, J.-P.G., and Mayeux, R. (2008). Frequency and course of mild cognitive impairment in a multiethnic community. Ann. Neurol *63*, 494–506.

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., and Stadlan, E.M. (1984). Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology *34*, 939–944.

Mitchell, A.J., and Shiri-Feshki, M. (2009). Rate of progression of mild cognitive impairment to dementia--meta-analysis of 41 robust inception cohort studies. Acta Psychiatr Scand *119*, 252–265.

Mohs, R.C., Knopman, D., Petersen, R.C., Ferris, S.H., Ernesto, C., Grundman, M., Sano, M., Bieliauskas, L., Geldmacher, D., Clark, C., et al. (1997). Development of cognitive instruments for use in clinical trials of antidementia drugs: additions to the Alzheimer's Disease Assessment Scale that broaden its scope. The Alzheimer's Disease Cooperative Study. Alzheimer Dis Assoc Disord *11 Suppl 2*, S13–21.

Morris, J.C. (1993). The Clinical Dementia Rating (CDR): current version and scoring rules. Neurology *43*, 2412–2414.

Palmqvist, S., Hertze, J., Minthon, L., Wattmo, C., Zetterberg, H., Blennow, K., Londos, E., and Hansson, O. (2012). Comparison of Brief Cognitive Tests and CSF Biomarkers in Predicting Alzheimer's Disease in Mild Cognitive Impairment: Six-Year Follow-Up Study. PLoS ONE *7*, e38639.

Perrin, R.J., Fagan, A.M., and Holtzman, D.M. (2009). Multimodal techniques for diagnosis and prognosis of Alzheimer's disease. Nature *461*, 916–922.

Petersen, R.C. (2004). Mild cognitive impairment as a diagnostic entity. J. Intern. Med *256*, 183–194.

Petersen, R.C., Roberts, R.O., Knopman, D.S., Boeve, B.F., Geda, Y.E., Ivnik, R.J., Smith, G.E., and Jack, C.R. (2009). Mild cognitive impairment: ten years later. Arch. Neurol *66*, 1447–1455.

Pfeffer, R.I., Kurosaki, T.T., Harrah, C.H., Chance, J.M., and Filos, S. (1982). Measurement of Functional Activities in Older Adults in the Community. J Gerontol *37*, 323–329.

Platt, J.C. (1999). Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. In ADVANCES IN LARGE MARGIN CLASSIFIERS, (MIT Press), pp. 61–74.

Ray, S., Britschgi, M., Herbert, C., Takeda-Uchimura, Y., Boxer, A., Blennow, K., Friedman, L.F., Galasko, D.R., Jutel, M., Karydas, A., et al. (2007). Classification and prediction of clinical Alzheimer's diagnosis based on plasma signaling proteins. Nat. Med. *13*, 1359–1362.

Rosen, W.G., Terry, R.D., Fuld, P.A., Katzman, R., and Peck, A. (1980). Pathological verification of ischemic score in differentiation of dementias. Ann. Neurol. *7*, 486–488.

Saeys, Y., Inza, I., and Larrañaga, P. (2007). A review of feature selection techniques in bioinformatics. Bioinformatics *23*, 2507 –2517.

Salat, D.H., Greve, D.N., Pacheco, J.L., Quinn, B.T., Helmer, K.G., Buckner, R.L., and Fischl, B. (2009). Regional white matter volume differences in nondemented aging and Alzheimer's disease. Neuroimage *44*, 1247–1258.

Sheikh, J.I., and Yesavage, J.A. (1986). Geriatric Depression Scale (GDS): recent evidence and development of a shorter version. In Clinical Gerontology : A Guide to Assessment and Intervention, (NY: The Haworth Press), pp. 165–173.

Shen, L., Saykin, A.J., Kim, S., Firpi, H.A., West, J.D., Risacher, S.L., McDonald, B.C., McHugh, T.L., Wishart, H.A., and Flashman, L.A. (2010). Comparison of manual and automated determination of hippocampal volumes in MCI and early AD. Brain Imaging Behav *4*, 86–95.

Shulman, K.I. (2000). Clock-drawing: is it the ideal cognitive screening test? Int J Geriatr Psychiatry *15*, 548–561.

Smialowski, P., Frishman, D., and Kramer, S. (2010). Pitfalls of supervised feature selection. Bioinformatics *26*, 440–443.

Soares, H.D., Potter, W.Z., Pickering, E., Kuhn, M., Immermann, F.W., Shera, D.M., Ferm, M., Dean, R.A., Simon, A.J., Swenson, F., et al. (2012). Plasma Biomarkers Associated With the Apolipoprotein E Genotype and Alzheimer Disease. Archives of Neurology 1–8.

Steyerberg, E.W., Moons, K.G.M., Van der Windt, D.A., Hayden, J.A., Perel, P., Schroter, S., Riley, R.D., Hemingway, H., and Altman, D.G. (2013). Prognosis Research Strategy (PROGRESS) 3: Prognostic Model Research. PLoS Med. *10*, e1001381.

Thompson, P.M., Hayashi, K.M., De Zubicaray, G., Janke, A.L., Rose, S.E., Semple, J., Herman, D., Hong, M.S., Dittmer, S.S., Doddrell, D.M., et al. (2003). Dynamics of gray matter loss in Alzheimer's disease. J. Neurosci. *23*, 994–1005.

Tombaugh, T.N. (2004). Trail Making Test A and B: normative data stratified by age and education. Arch Clin Neuropsychol *19*, 203–214.

Vakil, E., and Blachstein, H. (1993). Rey auditory-verbal learning test: Structure analysis. Journal of Clinical Psychology *49*, 883–890.

Varma, S., and Simon, R. (2006). Bias in error estimation when using cross-validation for model selection. BMC Bioinformatics *7*, 91.

Weiner, M.W., Veitch, D.P., Aisen, P.S., Beckett, L.A., Cairns, N.J., Green, R.C., Harvey, D., Jack, C.R., Jagust, W., Liu, E., et al. (2012). The Alzheimer's Disease Neuroimaging Initiative: a review of papers published since its inception. Alzheimers Dement *8*, S1–68.

Westman, E., Muehlboeck, J.-S., and Simmons, A. (2012). Combining MRI and CSF measures for classification of Alzheimer's disease and prediction of mild cognitive impairment conversion. Neuroimage *62*, 229–238.

Westover, M.B., Westover, K.D., and Bianchi, M.T. (2011). Significance testing as perverse probabilistic reasoning. BMC Med *9*, 20.

Yang, H.H., and Moody, J. (1999). Feature Selection Based on Joint Mutual Information. In Proceedings of International ICSC Symposium on Advances in Intelligent Data Analysis, pp. 22–25.

Zec, R.F., Burkett, N.R., Markwell, S.J., and Larsen, D.L. (2007). Normative data stratified for age, education, and gender on the Boston Naming Test. Clin Neuropsychol *21*, 617–637.

Zhang, D., and Shen, D. (2012a). Multi-Modal Multi-Task Learning for Joint Prediction of Multiple Regression and Classification Variables in Alzheimer's Disease. Neuroimage *59*, 895–907.

Zhang, D., and Shen, D. (2012b). Predicting future clinical changes of MCI patients using longitudinal and multimodal biomarkers. PLoS ONE *7*, e33182.

Zhang, D., Wang, Y., Zhou, L., Yuan, H., and Shen, D. (2011). Multimodal classification of Alzheimer's disease and mild cognitive impairment. Neuroimage *55*, 856–867.

CHAPTER 4

Conclusions

4.1 Summary of Key Findings and Significance

Alzheimer's disease (AD) is the most common cause of dementia among older adults and has been classically viewed as a neurodegenerative disorder that affects the brain's gray matter, with earliest atrophy seen in medial temporal lobe (MTL) structures. In the first study (Chapter 2), we used diffusion tensor imaging (DTI), an MRI-based technique that measures water diffusion within tissues *in vivo*, to show that patients with clinically diagnosed AD exhibit decreased integrity within two major limbic white matter pathways (fornix and cingulum). We also showed that these structural connectivity changes within the limbic system can be detected as early as during mild cognitive impairment (MCI), considered to be a transitional phase between healthy cognitive aging and dementia. Hypometabolism within the posterior cingulate cortex (PCC) is the earliest sign of incipient AD detectable using positron emission tomography (PET). We showed that this metabolic change is directly related to the structural integrity of the cingulum bundle, which connects PCC and regions of the MTL. Thus, our data support the “disconnection hypothesis” of PCC hypometabolism in AD. Finally, this work suggests that limbic white matter integrity, as measured by DTI, could serve as a potential biomarker for the early diagnosis of AD. Future prospective, longitudinal studies in MCI and cognitively normal subjects will need to be conducted to investigate this possibility further.

In the second study (Chapter 3), we used statistical pattern classification methods and data from the Alzheimer's Disease Neuroimaging Initiative to develop a prognostic model of dementia for patients with MCI. More than 750 variables spanning clinical, magnetic resonance imaging (MRI), and plasma proteomic data were considered as potential predictors of progression from MCI to dementia. A model incorporating the performance on three clinical assessments as well as morphometric MRI measures for three temporoparietal brain regions was able to predict progression to dementia in individual MCI patients with ~80% accuracy.

The accuracy of the model could be improved further by taking into account the confidence of the predictions. The prognostic model of dementia developed as part of this dissertation provides a non-invasive, cost-effective approach that can be used to (1) improve the selection of MCI patients in clinical trials and (2) identify high-risk MCI patients for early anti-AD treatment. In the case of clinical trials, use of the prognostic model can help enrich the MCI subject cohort to include only high-risk individuals, resulting in significant cost and time savings for the trial. The model can also be used as a first-line tool to establish a prognosis for patients with amnestic MCI. Based on the model prediction, the patient can be informed whether they belong to the high-risk or low-risk MCI subgroup and can use this information with their healthcare provider to choose a course of action (e.g. treatment or watchful waiting). In cases where the model cannot make a confident prediction, the clinician can then order additional biomarker studies. The model can also serve as a reference standard against which researchers can evaluate the prognostic utility of novel, more invasive, and/or more expensive biomarkers.

4.2 Predictive Models of Dementia: Challenges and Future Directions

4.2.1 Clinical Assessments and Clinical Diagnosis of AD

Consistent with prior research, clinical assessments were more predictive of MCI-to-dementia progression than other types of markers. A possible explanation for this finding is that the baseline clinical assessments are to some degree correlated with the clinical outcomes in patients with amnestic MCI. In ADNI and the majority of other clinical studies of AD and MCI, the criteria used for diagnosis of AD and determination of MCI-to-AD conversion largely rely on assessing the cognitive and functional status of the patient. In some sense, then, it may be logical to expect that patients who are more cognitively and functionally

impaired at baseline are also the ones that are more likely to progress from MCI to dementia. On the one hand this correlation between baseline clinical assessments and clinical outcomes can be viewed as a confound. On the other hand, it encourages a discussion about the seemingly "blurry" line between amnestic MCI and probable AD. The findings in this dissertation suggest that subtle but reliable baseline differences in cognitive and functional status are detectable in individual patients with MCI who are destined to progress to AD dementia. Some researchers have proposed that MCI represents the early stage of AD (Morris et al., 2001). Given the large body of evidence on the heterogeneity of MCI as a clinical condition, this position may be extreme. As a compromise, the prognostic model of dementia based on clinical assessments and MRI markers developed as part of this dissertation could be used to refine the widely used 1984 clinical criteria for the diagnosis of probable AD (McKhann et al., 1984). Specifically, patients diagnosed with MCI (according to the Petersen criteria) and for whom the model assigns the progressive label could be reclassified as having mild probable AD. MCI patients for whom the model assigns the non-progressive label could then retain the diagnosis of MCI. A recently proposed revision to the diagnostic guidelines for AD by the National Institute on Aging-Alzheimer's Association workgroup also recognizes the limitations of the 1984 clinical criteria and emphasizes the incorporation of biomarkers such as MRI measures of atrophy (McKhann et al., 2011).

Another challenge in building predictive models of dementia is the reliance on the clinical diagnosis of AD as the "ground truth" (gold standard). The clinical diagnosis of probable AD has an accuracy of 70-90% relative to the pathological diagnosis (Beach et al., 2012), with greater accuracies being achieved in specialty settings (e.g. memory disorders clinic). This has two implications for the development of pattern classification models. First, models that are developed with data based on the clinical diagnosis of AD cannot be more

accurate than the clinical diagnosis itself. The predictive accuracy of 80-87% for predicting MCI-to-AD dementia progression obtained in the second study is approaching this theoretical limit. Second, the relative inaccuracy of the clinical diagnosis likely introduces additional noise into the model development process. Therefore, it is critical that future research on predictive models of dementia incorporates not only clinically diagnosed AD cases but also pathologically-verified cases in an effort to raise the upper bound on the predictive accuracy that can be achieved using pattern participation methods.

4.2.2 Building the Optimal Prognostic Model of Dementia

Discrimination between progressive and non-progressive MCI patients at the individual level is a significantly more challenging prediction task than discrimination between normal control and AD subjects. It is becoming increasingly clear that no single marker or source of data is adequate to achieve early diagnosis of AD, and that different biomarkers and/or sources of data capture different aspects of AD pathology during different time periods (see (Jack et al., 2010). As a demonstration of this, the prognostic model based on the combination of clinical assessments and structural MRI markers (from the second study) outperformed models based on either source of data alone. Further progress in developing predictive models of dementia/AD can likely be achieved by using both pathologically-verified data sets in combination with multivariate/multimodal modeling approaches.

Of the current biomarker approaches being investigated, the addition of a biomarker data source that can serve as a window into the functional integrity of the brain would be the most likely to further improve the predictive ability of the model developed in this dissertation. A promising functional neuroimaging approach that should be investigated as a potential biomarker for early AD diagnosis is resting-state functional MRI (rs-fMRI). While both FDG-

PET and rs-fMRI measure brain function/metabolism, rs-fMRI is unique in that it can be used to assess the functional connectivity or coherence among multiple brain regions (Zhu et al., 2013). This technique may be particularly appropriate as a potential biomarker because AD causes a disruption of various brain networks important for memory and higher cognitive function. Biomarkers that capture the molecular aspects of AD pathology are likely to also be complementary to clinical assessments and structural/functional neuroimaging. PET-based imaging of amyloid deposition in the brain may provide such complementary information, and studies of its diagnostic utility are currently ongoing. However, imaging of *in vivo* tau pathology will likely offer more specificity for diagnosis of AD when such techniques become available in the future. Functional neuroimaging and molecular amyloid/tau biomarkers may be particularly important for improving the model's ability to identify MCI progressors who develop AD more than 18 months after evaluation, when clinical assessments and morphometric MRI measures appear to have relatively limited predictive value. This is consistent with the idea that subtle functional alterations in the brain likely precede overt AD-related neurodegeneration and that abnormal protein deposition is one of the earliest pathological events in AD.

To enhance the practical utility of the prognostic model, future work should also incorporate both amnestic and non-amnestic MCI cases and develop a model that can predict progression to both AD- and non-AD-type dementia. Moreover, the prognostic model of dementia should be extended in the future to asymptomatic individuals. Another useful extension of the prognostic model would allow the model to stage MCI patients in terms of where along the MCI-to-AD dementia continuum the patient may be. Although the probabilistic predictions generated by the model in this dissertation were associated with time-to-progression, this was a rather small effect. For staging purposes, a model would need to

be explicitly trained to accurately predict time-to-progression, as opposed to training it to classify progressive versus non-progressive MCI subjects (as done in the present work).

4.2.3 Patient Heterogeneity

Finally, the findings in this dissertation suggest that patient heterogeneity is an important consideration when developing predictive models of dementia. It was found that the predictive performance of the model (based on cognitive/functional assessments and structural MRI measures of atrophy) varied with several patient characteristics, including age and APOE genotype. This interaction between predictive accuracy and patient characteristics occurred regardless of whether a given characteristic was selected as a predictor in the model. This dissociation can be explained if the patient characteristics are viewed as moderator variables.

Age, the strongest risk factor for the development of AD, was not informative of MCI-to-dementia progression at the individual patient level nor was there a difference in age between progressive and non-progressive MCI groups. A possible explanation for this is that age may be a strong risk factor for the development of amnestic MCI rather than the subsequent deterioration from MCI to AD dementia. Despite the fact that age was not selected as a predictor of MCI-to-dementia progression, model accuracy tended to be higher for older subjects. There is, however, evidence to suggest that aging itself causes brain atrophy and a decline in cognitive abilities (Raji et al., 2009; Stricker et al., 2011). Thus, aging and incipient AD processes likely interact and jointly may exert a complex influence on cognitive performance and brain structure. In the case of APOE (a well established risk gene for AD), the model made less accurate predictions for MCI subjects who were carriers of the epsilon-4 allele, and this effect was present only for specificity (percentage of non-progressive MCI

cases correctly identified) but not sensitivity of the model. APOE epsilon-4 positivity is associated with regional temporal lobe atrophy in normal control subjects (Wishart et al., 2006), an effect that occurs independent of whether or not the individual develops AD dementia. Thus, the model likely tends to become confused when asked to classify a non-progressive MCI patient who is an APOE epsilon-4 carrier (and thus has greater baseline brain atrophy than the typical non-progressor might).

The effect of patient heterogeneity on model performance must be characterized in future studies that attempt to develop predictive models of dementia. This will help researchers and clinicians better understand the strengths and limitations of the published models. The issue of how to effectively correct for the patient heterogeneity remains an open question. The use of pathologically-verified patient cases for model development and incorporation of multiple, diverse biomarkers may make the models more robust in the face of patient heterogeneity.

REFERENCES

REFERENCES

Beach, T.G., Monsell, S.E., Phillips, L.E., and Kukull, W. (2012). Accuracy of the clinical diagnosis of Alzheimer disease at National Institute on Aging Alzheimer Disease Centers, 2005-2010. J. Neuropathol. Exp. Neurol. *71*, 266–273.

Jack, C.R., Knopman, D.S., Jagust, W.J., Shaw, L.M., Aisen, P.S., Weiner, M.W., Petersen, R.C., and Trojanowski, J.Q. (2010). Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. Lancet Neurol *9*, 119.

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., and Stadlan, E.M. (1984). Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology *34*, 939–944.

McKhann, G.M., Knopman, D.S., Chertkow, H., Hyman, B.T., Jack, C.R., Jr, Kawas, C.H., Klunk, W.E., Koroshetz, W.J., Manly, J.J., Mayeux, R., et al. (2011). The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement *7*, 263–269.

Morris, J.C., Storandt, M., Miller, J.P., McKeel, D.W., Price, J.L., Rubin, E.H., and Berg, L. (2001). Mild cognitive impairment represents early-stage Alzheimer disease. Arch. Neurol. *58*, 397–405.

Raji, C.A., Lopez, O.L., Kuller, L.H., Carmichael, O.T., and Becker, J.T. (2009). Age, Alzheimer disease, and brain structure. Neurology *73*, 1899–1905.

Stricker, N.H., Chang, Y.-L., Fennema-Notestine, C., Delano-Wood, L., Salmon, D.P., Bondi, M.W., and Dale, A.M. (2011). Distinct profiles of brain and cognitive changes in the very old with Alzheimer disease. Neurology *77*, 713–721.

Wishart, H.A., Saykin, A.J., McAllister, T.W., Rabin, L.A., McDonald, B.C., Flashman, L.A., Roth, R.M., Mamourian, A.C., Tsongalis, G.J., and Rhodes, C.H. (2006). Regional brain atrophy in cognitively intact adults with a single APOE ε4 allele. Neurology *67*, 1221–1224.

Zhu, D.C., Majumdar, S., Korolev, I.O., Berger, K.L., and Bozoki, A.C. (2013). Alzheimer's disease and amnestic mild cognitive impairment weaken connections within the default-mode network: a multi-modal imaging study. J. Alzheimers Dis. *34*, 969–984.