Neuropsychological Testing Predicts Cerebrospinal Fluid Amyloid-β in Mild Cognitive Impairment

Benjamin M. Kandel,*, Brian B. Avants, James C. Gee, Steven E. Arnold, David A. Wold
and for the Alzheimer’s Disease Neuroimaging Initiative

*Penn Image Computing and Science Laboratory and Department of Bioengineering, University of Pennsylvania, Philadelphia, PA, USA
Penn Image Computing and Science Laboratory and Department of Radiology, Perelman School of Medicine of the University of Pennsylvania, Philadelphia, PA, USA
Department of Psychiatry, Perelman School of Medicine of the University of Pennsylvania, Philadelphia, PA, USA
Department of Neurology, Perelman School of Medicine of the University of Pennsylvania, Philadelphia, PA, USA

Handling Associate Editor: Andrew Saykin

Accepted 26 March 2015

Abstract

Background: Psychometric tests predict conversion of mild cognitive impairment (MCI) to probable Alzheimer’s disease (AD). Because the definition of clinical AD relies on those same psychometric tests, the ability of these tests to identify underlying AD pathology remains unclear.

Objective: To determine the degree to which psychometric testing predicts molecular evidence of AD amyloid pathology, as indicated by cerebrospinal fluid (CSF) amyloid-β (Aβ), in patients with MCI, as compared to neuroimaging biomarkers.

Methods: We identified 408 MCI subjects with CSF Aβ levels, psychometric test data, FDG-PET scans, and acceptable volumetric MR scans from the Alzheimer’s Disease Neuroimaging Initiative (ADNI). We used psychometric tests and imaging biomarkers in univariate and multivariate models to predict Aβ status.

Results: The 30-min delayed recall score of the Rey Auditory Verbal Learning Test was the best predictor of Aβ status among the psychometric tests, achieving an AUC of 0.67 ± 0.02 and odds ratio of 2.5 ± 0.4. FDG-PET was the best imaging-based biomarker (AUC 0.67 ± 0.03, OR 3.2 ± 1.2) followed by hippocampal volume (AUC 0.64 ± 0.02, OR 2.4 ± 0.3). A multivariate analysis based on the psychometric tests improved on the univariate predictors, achieving an AUC of 0.68 ± 0.03 (OR 3.38 ± 1.2).

Conclusion: Psychometric tests performed as well as imaging biomarkers to predict presence of molecular markers of AD pathology in MCI patients and should be considered in the determination of the likelihood that MCI is due to AD.

Keywords: Alzheimer’s disease, magnetic resonance imaging, mild cognitive impairment, positron emission tomography

*Data used in preparation of this article were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.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 report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/ wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

*Correspondence to: Benjamin M. Kandel, B.A., Penn Image Computing and Science Laboratory and Department of Bioengineering, University of Pennsylvania, 3600 Market St, Ste. 370, Philadelphia, PA 19104, USA. Tel.: +1 314 610 7256; E-mail: bkandel@seas.upenn.edu.

ISSN 1387-2877/15/$35.00 © 2015 – IOS Press and the authors. All rights reserved
INTRODUCTION

Recent guidelines for diagnosing mild cognitive impairment (MCI) due to Alzheimer’s disease (AD) have emphasized the importance of psychometric testing for establishing the existence of MCI and subsequently relying on biomarkers based on imaging and biofluids to assess the likelihood that the existing cognitive impairment is “due to AD” relative to a different cause [1]. In particular, cognitive testing is a component of the “core clinical criteria” for MCI, which requires that impairment greater than expected for age must be present in at least one cognitive domain. Once clinical categorization of MCI is established, the guidelines suggest that the likelihood that the cognitive phenotype is “due to AD” should rely on various imaging and molecular biomarkers (each classified as either a biomarker of neurodegeneration or cerebral amyloid), without specifically taking into account the severity of the cognitive deficit within the MCI category.

Although imaging-derived biomarkers for diagnosis of AD and prediction of conversion from MCI to AD have been the subject of intensive research [2–4], how these biomarkers can be used most effectively in the presence of alternative sources of clinical information about a subject’s status, such as cognitive testing, is not settled. Several recent studies have examined the relative utility of cognitive testing, imaging, or molecular biomarkers for predicting conversion from MCI to AD [5–9]. These studies have generally found that cognitive testing performs similarly to other biomarkers, but a potential criticism of these study designs is that using psychometric measurements to predict conversion to AD is circular, as the diagnosis of AD is thereby relying on biomarkers based on imaging and molecular pathology; thus, incorporating a molecular neurodegenerative marker like tau may confound the results. Moreover, we wanted to determine the relative and combined predictive value of psychometric testing with neuroimaging biomarkers of neuronal injury or neurodegeneration.

In particular, we examined several cognitive measures, including verbal memory, given their putative sensitivity to prodromal AD. We used diverse imaging-derived biomarkers to accurately represent both standard and developing measurement approaches. Further, we chose structural magnetic resonance imaging (MRI) and FDG-PET measures given their emphasis in the MCI guidelines. For MRI data, we used an automated hippocampal volume measurement, several cortical-thickness measurements including a summary measure of several regions associated with AD-related tissue loss [12, 13], and multivariate analysis of voxelwise measurements of cortical thickness [14, 15]. Hippocampal volume is considered to be one of the most established biomarkers of AD with numerous studies demonstrating its predictive value in MCI. We also used FDG-PET data from a set of regions (meta-region of interest, ROI) previously determined to be sensitive to early AD and prediction of clinical conversion to AD in MCI cohorts [16]. To obtain such a wide variety of clinical data in a sufficiently large population, we utilized the Alzheimer’s Disease Neuroimaging Initiative (ADNI) dataset. If cognitive measures perform similarly to both more standard and developing imaging biomarkers in prediction of AD pathology with MCI patients, they can provide a cost-effective and easily accessible method for assessing the likelihood of prodromal AD in patients with MCI.

METHODS

Clinical data

Subjects

This study was a retrospective analysis of data obtained from the ADNI database (http://adni.loni.usc.edu). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies, and nonprofit organizations, as a $60 million, 5-year public-private partnership. The primary goal of ADNI has...
been to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials.

Data used in this article were downloaded from the ADNI website in January 2014. We included only MCI subjects with complete datasets for the current analysis, including CSF Aβ levels, all neuropsychological tests examined, and FDG-PET. Only those subjects with Freesurfer cortical and hippocampal segmentations of acceptable quality, as determined by the publicly available Freesurfer dataset available through ADNI, were included.

In the ADNI study, MCI is split into two groups, early MCI (EMCI) and late MCI (LMCI). Diagnostic criteria for both EMCI and LMCI subjects were as follows: Mini-Mental State Examination (MMSE) scores between 24–30 (inclusive), a subjective memory concern reported by subject, informant, or clinician, a Clinical Dementia Rating of 0.5, absence of significant levels of impairment in other cognitive domains, essentially preserved activities of daily living, and an absence of dementia. They also were required to have objective memory loss measured by education-adjusted scores on delayed recall of one paragraph from the Wechsler Memory Scale Logical Memory II, which is further determined EMCI (≥ 16 years: 9–11; 8–15 years: 5–9; 0–7 years: 3–6) or LMCI (≥ 16 years: ≥ 8; 8–15 years: ≤ 4; 0–7 years: ≤ 2) status. In this manuscript, MCI refers to both EMCI and LMCI.

The ADNI study includes a variety of cognitive tests around the United States and Canada, and a full list is available at http://adni.loni.usc.edu/about/centers-cores/study-sites/. Recruitment for the ADNI study aimed to achieve a balance of normal controls, MCI, and AD subjects. For ADNI, a random subsample of subjects was selected for FDG imaging; in ADNI 2/GO, all subjects enrolled received FDG imaging. For up-to-date information on specific inclusion and exclusion criteria, please see http://www.adni-info.org.

Psychometric testing

We aimed to include a battery of psychometric tests that would cover a broad range of cognitive domains, with special focus on memory due to its importance in AD. For memory, we included components of the Rey Auditory Verbal Learning Test (AVLT) [17] given its richness of measures for various aspects of mnemonic processing (e.g., immediate versus delayed recall versus delayed recognition); for assessment of cognitive speed, sequencing, and executive function, the Trail Making Test [18] (Trails A and Trails B) was used; for language/semantics, category fluency [19] (Animals) and the Boston Naming Test [20] were examined; and as a measure of global cognition, the MMSE was utilized [21]. We examined several of the AVLT measures, which depend on differential aspects of episodic and working memory [22]. The AVLT consists of five learning trials in which a list of 15 words is read and the subject is asked to immediately recall as many items as possible. After an interference list of 15 novel words is read and recalled, subjects are then asked to recall words from the initial list (5-min delayed recall). A 30-min delayed recall trial and recognition test follow. For the recognition test, subjects are presented with a list of the 15 studied words and 15 nonstudied foils and are asked to circle all words previously studied. To account for false alarms (FA) to nonstudied items, we calculated a measure of discriminability, d-prime (d′), in a standard fashion based on classic signal detection theory [23]. Because d’ is undefined when either proportion is 0 or 1, we used standard formulas to convert these values: Hits = (no. of hits)/[(no. of hits) + (no. of studied items + 1)] and FA = (no. of FA) + (no. of untried items + 1). For the current study, we analyzed performance on the fifth immediate memory trial (AVLT Trial 5 Recall), 5- and 30-min delayed recall (AVLT 5-min Recall, AVLT 30-min Recall), and recognition memory discrimination (AVLT Recognition Discrimination). In addition, we calculated a retention score, which is the number of items remembered after a 30-min delay (AVLT 30-min Recall) divided by the number of items remembered during the last immediate memory trial (AVLT Trial 5 Recall).

Determination of amyloid and ApoE status

CSF-based molecular biomarkers were processed by the University of Pennsylvania/ADNI Biomarker Core Laboratory as previously described [10, 24]. An Aβ1–42 value of less than or equal to 191 pg/ml was considered to be “positive” for the presence of amyloid pathology based on a prior autopsy-based study performed at the University of Pennsylvania [10]. For analyses involving ApoE status, subjects were dichotomized into ApoE e4 positive and negative groups. ApoE e4 positive status is defined as having at least one ApoE e4 allele.
Processing of neuroimaging data included both analyses made publicly available by ADNI and in-house image processing. The following analyses were based on preprocessed data downloaded from the ADNI website. FDG-PET scans were acquired and analyzed in accordance with a standard protocol [16]. Mean FDG uptake was averaged over 5 ROIs that are sensitive to AD-related changes in metabolism, including right and left angular gyri, right and left inferior temporal regions, and bilateral posterior cingulate. These regions were averaged into a meta ROI and normalized to an ROI focused on the pons and cerebellar vermis to give a summary FDG PET measure. Cortical thickness and hippocampal measurement of the MRI scans were performed according to the standard ADNI FreeSurfer [25] processing pipeline, and downloaded from the ADNI website. Only images that passed ADNI quality control for the temporal, occipital, temporal, and parietal lobe were included. Cortical thickness in the caudal portion of the middle frontal gyrus, medial portion of the orbital frontal cortex, inferior parietal lobule, lateral portion of the occipital cortex, inferior temporal gyrus, entorhinal cortex, temporal pole, and the isthmus of the cingulate cortex were averaged to form a meta ROI thought sensitive to early AD related neurodegeneration, as previously suggested [26].

In addition to the image analysis performed by various ADNI investigators, we ran additional analyses of MR images to supplement standard approaches with a state of the art multivariate analysis technique. 1.5T and 3T non-accelerated T1-weighted MPRAGE and SPGR MRI scans of all MCI subjects from ADNI1 and ADNI2/GO were downloaded from http://adni.loni.usc.edu. We computed an alternative measure of cortical thickness using DrReCT [12, 13], and used the AAL labelset [32] to define medial temporal and precuneal ROIs, as these areas are known to atrophy in early AD. We performed a singular value decomposition (SVD) analysis of the whole brain cortical thickness data, as this analysis has proven useful in differentiating AD from frontotemporal dementia and predicting CSF-based biomarkers in this population [28, 29]. The SVD was performed using the princomp function in R, and we retained the top 10 components. A grid search strategy using bootstrapping with 100 repetitions, with half the subjects left out for a validation cohort, was used to determine the optimal number of components to retain.

All statistical analysis was performed using the R programming language, version 3.1.0. For predictive studies, we randomly split the subjects 5 times into training and testing cohorts, retaining half the subjects for training and using the other half for testing in a 5 × 2 cross-validation scheme [30]. All area under the curve (AUC), odds ratios, and positive and negative predictive values are on the testing cohorts. Two-tail t-tests were used to compare AUC values between testing cohorts of different models to calculate a p-value for differences in mean AUC; false discovery rate (FDR) correction was applied to correct for multiple comparisons. For all analyses, patient age, gender, and education were used as additional predictors; for all MCI-based imaging analyses, magnet field strength (1.5 or 3T) was included as a covariate. In addition to univariate predictions of Aβ status from psychometrics and imaging modalities, we performed principal component regression, using three principal components, on all the psychometric scores, as well as the psychometric and imaging values combined. AUC analysis was performed using the ROCR package in R [31].

Subject data was collected between January 2006 and January 2013. A total of 622 MCI subjects with CSF-derived Aβ values were identified, and 407 of those were Aβ positive. Of these, 547 (350 Aβ positive) had FDG scans; 450 (286 Aβ positive) had complete FreeSurfer segmentations without failures; 433 (273 Aβ positive) had intracranial volume available; and 408 subjects (257 Aβ positive) had complete psychometric scores available. There was a mean difference of 15 days between the psychometric tests and imaging studies, with 95% of subjects having the imaging and psychometric tests done within 55 days of each other. The maximum time difference was 138 days. A total of 62 adverse events were reported from the lumbar punctures, most of which were headaches (25 cases) or pain (23 cases), with two subjects reporting nausea and a few reporting a variety of other effects, including bruising, tenderness, and swelling. One adverse event, transient procedural anxiety, occurred during the imaging.

A summary of the demographics of the study population, including the psychometric and imaging information, is given in Table 1. We computed...
a logistic regression relating each psychometric test and modality with Aβ status, while covarying for age, gender, and education (Table 2). The logistic regression results indicated that the psychometric tests and imaging modalities were predictive of Aβ status, even when included in a univariate model.

Predictive models

The associations between the various psychometric scores and Aβ status were strong enough to predict many of the psychometric measures displayed predictive value, varying in range of AUCs from 0.59 to 0.67, immediate and delayed recall measures performed particularly well, reaching an AUC of 0.65 and 0.67 respectively, corresponding to odds ratios of 3.0 and 2.5 (Fig. 1, Table 3). The 30-min delayed recall test was significantly better than both Trails tests, the Boston Naming Test, category fluency, and MMSE.

The standard imaging modalities were similar to each other and the individual psychometric tests in prediction of Aβ status with FDG-PET displaying the highest AUC at 0.67, followed by hippocampal volume at 0.64.
a collection of psychometric scores, and principal components analysis of imaging biomarkers and principal components of PCA, principal components analysis. The resulting components boosted the AUC slightly to 0.59, but the increase was not significant (Table 4). The multivariate analysis of the cognitive tests, however, was statistically significantly better than hippocampal volume, which was not true for any individual cognitive test. Repeating the analysis using only subjects with 3T MR scans did not significantly change the results.

### DISCUSSION

**Impact**

The results shown here indicate that a psychometric evaluation can be as useful as FDG-PET or quantitative MR imaging in predicting whether or not a given amnestic MCI patient likely has underlying pathology, we sought to determine, as a secondary analysis, whether the observed effects are modulated by e4 status. We divided the subjects into e4 positive and e4 negative groups and performed the analyses in the same way as before (Table 5). The results were broadly the same in that imaging did not significantly improve diagnostic accuracy over psychometric tests. Nearly all psychometric and neuroimaging biomarkers were more predictive of A\(\beta\) status in e4 negative as compared to e4 positive subjects. This trend was highly statistically significant (p<0.001 using a paired t-test).

**Table 3**

| Area under the curve (AUC), odds ratios, and positive (PPV) and negative predictive values (NPV) predicting A\(\beta\) status from biomarkers |
|---|---|---|---|
| Mini-Mental Status Examination | 0.61 ± 0.03 | 1.94 ± 0.60 | 0.71 ± 0.05 | 0.43 ± 0.05 |
| AVLT Trail 5 Recall | 0.65 ± 0.03 | 3.01 ± 0.36 | 0.75 ± 0.04 | 0.50 ± 0.05 |
| AVLT 30-min Recall | 0.65 ± 0.02 | 2.50 ± 0.44 | 0.73 ± 0.05 | 0.47 ± 0.04 |
| AVLT Recognition Discrimination | 0.64 ± 0.03 | 2.44 ± 0.55 | 0.73 ± 0.02 | 0.48 ± 0.07 |
| Retention | 0.67 ± 0.03 | 2.48 ± 0.48 | 0.73 ± 0.03 | 0.47 ± 0.06 |
| Trail Making Test A | 0.62 ± 0.02 | 2.13 ± 0.46 | 0.73 ± 0.04 | 0.44 ± 0.05 |
| Trail Making Test B | 0.63 ± 0.02 | 2.49 ± 0.48 | 0.75 ± 0.05 | 0.45 ± 0.05 |
| Boston Naming Test | 0.59 ± 0.02 | 1.66 ± 0.17 | 0.70 ± 0.03 | 0.42 ± 0.04 |
| Category Fluency (animals) | 0.60 ± 0.02 | 1.88 ± 0.43 | 0.71 ± 0.05 | 0.42 ± 0.03 |
| Hippocampal volume | 0.64 ± 0.02 | 2.41 ± 0.34 | 0.74 ± 0.04 | 0.46 ± 0.04 |
| Medial Temporal Thickness | 0.59 ± 0.01 | 1.67 ± 0.07 | 0.70 ± 0.04 | 0.42 ± 0.04 |
| Precentral Thickness | 0.59 ± 0.02 | 1.83 ± 0.25 | 0.71 ± 0.03 | 0.43 ± 0.05 |
| Mean Cortical Thickness of AD Meta-ROI | 0.61 ± 0.02 | 1.90 ± 0.31 | 0.71 ± 0.04 | 0.43 ± 0.04 |
| Mean FDG-PET SUVR of AD Meta-ROI | 0.67 ± 0.03 | 3.19 ± 1.22 | 0.76 ± 0.05 | 0.49 ± 0.08 |
| PCA of psychometric scores | 0.68 ± 0.02 | 3.38 ± 1.06 | 0.71 ± 0.03 | 0.56 ± 0.10 |
| PCA of psychometric scores and imaging biomarkers | 0.60 ± 0.02 | 3.18 ± 0.76 | 0.71 ± 0.03 | 0.55 ± 0.08 |
| PCA of cortex-wide cortical thickness | 0.59 ± 0.03 | 1.57 ± 0.21 | 0.67 ± 0.04 | 0.43 ± 0.02 |

**PCA, principal components analysis.**
Table 4

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

<table>
<thead>
<tr>
<th>AVLT Trial 5 recall</th>
<th>AVLT Trial 5 month</th>
<th>AVLT Session recall</th>
<th>Trails A</th>
<th>Trails B</th>
<th>Boston Naming Test</th>
<th>Category Fluency (animals)</th>
<th>MMSE Discrimination</th>
<th>Retention</th>
<th>Medial Temporal Thickness</th>
<th>Precuneus Thickness</th>
<th>Mean FDG</th>
<th>Hippocampal Volume</th>
<th>Thickness of Meta-ROI</th>
<th>PCA of psychometrics</th>
<th>PCA of psychometrics and imaging</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>0.37</td>
<td>0.33</td>
<td>0.02</td>
<td>0.04</td>
<td>0.12</td>
<td>0.25</td>
<td>0.02</td>
<td>0.02</td>
<td>0.29</td>
<td>0.54</td>
<td>0.05</td>
<td>0.12</td>
<td>0.10</td>
<td>0.12</td>
<td>0.05</td>
</tr>
<tr>
<td>0.27</td>
<td>0.24</td>
<td>0.33</td>
<td>0.02</td>
<td>0.04</td>
<td>0.12</td>
<td>0.25</td>
<td>0.02</td>
<td>0.02</td>
<td>0.29</td>
<td>0.54</td>
<td>0.05</td>
<td>0.12</td>
<td>0.10</td>
<td>0.12</td>
<td>0.05</td>
</tr>
<tr>
<td>0.23</td>
<td>0.17</td>
<td>0.03</td>
<td>0.02</td>
<td>0.04</td>
<td>0.12</td>
<td>0.25</td>
<td>0.02</td>
<td>0.02</td>
<td>0.29</td>
<td>0.54</td>
<td>0.05</td>
<td>0.12</td>
<td>0.10</td>
<td>0.12</td>
<td>0.05</td>
</tr>
<tr>
<td>0.39</td>
<td>0.33</td>
<td>0.62</td>
<td>0.03</td>
<td>0.04</td>
<td>0.12</td>
<td>0.25</td>
<td>0.02</td>
<td>0.02</td>
<td>0.29</td>
<td>0.54</td>
<td>0.05</td>
<td>0.12</td>
<td>0.10</td>
<td>0.12</td>
<td>0.05</td>
</tr>
<tr>
<td>0.02</td>
<td>0.01</td>
<td>0.00</td>
<td>0.07</td>
<td>0.08</td>
<td>0.01</td>
<td>0.36</td>
<td>0.01</td>
<td>0.01</td>
<td>0.36</td>
<td>0.16</td>
<td>0.05</td>
<td>0.12</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>0.04</td>
<td>0.03</td>
<td>0.01</td>
<td>0.24</td>
<td>0.50</td>
<td>0.06</td>
<td>0.12</td>
<td>0.01</td>
<td>0.01</td>
<td>0.36</td>
<td>0.16</td>
<td>0.05</td>
<td>0.12</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>0.12</td>
<td>0.10</td>
<td>0.03</td>
<td>0.62</td>
<td>0.31</td>
<td>0.02</td>
<td>0.35</td>
<td>0.12</td>
<td>0.01</td>
<td>0.36</td>
<td>0.16</td>
<td>0.05</td>
<td>0.12</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>0.77</td>
<td>0.77</td>
<td>0.77</td>
<td>0.77</td>
<td>0.77</td>
<td>0.77</td>
<td>0.77</td>
<td>0.77</td>
<td>0.77</td>
<td>0.77</td>
<td>0.77</td>
<td>0.77</td>
<td>0.77</td>
<td>0.77</td>
<td>0.77</td>
<td>0.77</td>
</tr>
<tr>
<td>0.25</td>
<td>0.12</td>
<td>0.02</td>
<td>0.02</td>
<td>0.03</td>
<td>0.02</td>
<td>0.16</td>
<td>0.00</td>
<td>0.00</td>
<td>0.36</td>
<td>0.16</td>
<td>0.05</td>
<td>0.12</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>0.62</td>
<td>0.00</td>
<td>0.05</td>
<td>0.01</td>
<td>0.36</td>
<td>0.02</td>
<td>0.36</td>
<td>0.01</td>
<td>0.01</td>
<td>0.36</td>
<td>0.16</td>
<td>0.05</td>
<td>0.12</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>0.62</td>
<td>0.03</td>
<td>0.01</td>
<td>0.05</td>
<td>0.36</td>
<td>0.02</td>
<td>0.36</td>
<td>0.01</td>
<td>0.01</td>
<td>0.36</td>
<td>0.16</td>
<td>0.05</td>
<td>0.12</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>0.29</td>
<td>0.23</td>
<td>0.76</td>
<td>0.04</td>
<td>0.07</td>
<td>0.01</td>
<td>0.36</td>
<td>0.01</td>
<td>0.01</td>
<td>0.36</td>
<td>0.16</td>
<td>0.05</td>
<td>0.12</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>0.54</td>
<td>0.50</td>
<td>0.10</td>
<td>0.52</td>
<td>0.87</td>
<td>0.05</td>
<td>0.08</td>
<td>0.27</td>
<td>0.06</td>
<td>0.36</td>
<td>0.16</td>
<td>0.05</td>
<td>0.12</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>0.05</td>
<td>0.05</td>
<td>0.01</td>
<td>0.38</td>
<td>0.12</td>
<td>0.25</td>
<td>0.49</td>
<td>0.07</td>
<td>0.11</td>
<td>0.01</td>
<td>0.36</td>
<td>0.12</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>0.12</td>
<td>0.07</td>
<td>0.03</td>
<td>0.02</td>
<td>0.00</td>
<td>0.02</td>
<td>0.01</td>
<td>0.09</td>
<td>0.50</td>
<td>0.00</td>
<td>0.79</td>
<td>0.03</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>0.10</td>
<td>0.05</td>
<td>0.02</td>
<td>0.02</td>
<td>0.00</td>
<td>0.02</td>
<td>0.01</td>
<td>0.07</td>
<td>0.35</td>
<td>0.00</td>
<td>0.65</td>
<td>0.01</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
</tbody>
</table>
AD pathology. The low cost and ready availability of psychometric batteries as compared to imaging studies makes them an attractive and useful alternative to specialized imaging techniques in clinical evaluation. Although the psychometric batteries do not approach the perfect classification between Aβ-positive and Aβ-negative subjects, they can be useful in clinical practice to broadly estimate risk of prodromal AD and, perhaps, guide the process of obtaining additional studies, including molecular biomarkers. For situations in which obtaining an accurate measure of Aβ is paramount, such as evaluating appropriateness of a future anti-amyloid therapy, direct molecular imaging or CSF measurement of Aβ is still necessary, perhaps after initial screening with psychometrics to enrich with amyloid positive patients.

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

Further, the fact that even the “standard” cognitive measures examined here displayed some success in determining the likelihood of AD pathology suggests that more research is warranted on designing and evaluating psychometric tests optimized for detection of early AD-related cognitive decline. In particular, measures guided by the cognitive neuroscience literature may be particularly useful in this regard [32]. Finally, the results here indicate that the ability of psychometric scores to identify patients who will progress to AD is not due solely to the fact that those same scores are used to establish presence of probable AD. Instead, it appears that the predictive value of psychometric tests are due, at least in part, to their ability to separate MCI patients into sub-populations with higher and lower prevalence of AD pathology.

Limitations

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

It is also possible that advances in image processing techniques may improve the diagnostic capability of neuroimaging data. Although it is impossible to rule out such advances, the variety of imaging modalities and image processing techniques used here make it less likely that new analytic approaches would improve the predictive power of imaging data enough to supplant psychometric measures as a key method for characterization of MCI patients. Indeed, the current work did use a promising analytic approach involving singular value decomposition across the entire cortical mantle, which had previously demonstrated good predictive value of CSF t-tau/\(A_\beta\) in patients with AD and frontotemporal dementia [28]. Nonetheless, this approach did not display significant advantages over more traditional measures (e.g., hippocampal volume) or psychometric tests. In any case, psychometric tests are more accessible than sophisticated image processing techniques, especially to physicians who do not work in academic medical centers. An obvious limitation of this study is the use of CSF-derived \(A_\beta\) status as a gold standard in the prediction models, as CSF \(A_\beta\) does not perfectly reflect brain AD pathology. While we took this approach to avoid the circularity of longitudinal studies of conversion, a better design would have autopsy-confirmed AD pathology for comparison with the other biomarkers. Nonetheless, CSF \(A_\beta\), along with amyloid PET, are the closest surrogates to histopathologic evaluation presently available and have displayed high sensitivity in autopsy studies [10, 11].

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

### Effect of ApoE

One intriguing result in this study is the marked difference in prediction accuracy in ApoE \(e4\) positive versus \(e4\) negative subjects. This finding is consistent with previous work showing that cognitive function is more closely linked to \(A_\beta\) status within \(e4\) negative than within \(e4\) positive subjects [33, 34]. The mechanism behind this effect is not clear, but may be that the effects of \(A_\beta\) on cognitive function are modulated by ApoE isoforms. However, an important confounding factor is the highly unbalanced nature of the samples: The \(e4\) negative group had 79 \(A_\beta^+\) and 120 \(A_\beta^-\) subjects, whereas the \(e4\) positive group had 178 \(A_\beta^+\) and only 29 \(A_\beta^-\) subjects. The relative paucity of \(e4\) positive but \(A_\beta^-\) subjects may contribute to the lower performance of the predictive model in the \(e4\) positive group. Thus, it is possible that the strong association of \(A_\beta\) with \(e4\) status obscures the association with cognitive measures.

### Psychometric scores as functional biomarkers

It is worth pointing out that the current algorithm for determining the likelihood of “MCI due to AD” in the recently proposed criteria treats neurodegenerative and molecular markers as dissociable modalities of evidence. In a sense, psychometric tests can be considered...
another type of downstream neurodegenerative measure. Thus, it may seem somewhat odd to use one type of biomarker (neurodegenerative) to predict another (molecular) in this context if these measures provide orthogonal information. However, these measures are obviously related and multiple studies have demonstrated the significant predictive value for conversion to clinical AD in patients either with “positive” CSF or PET amyloid studies or neurodegenerative markers [1, 35, 36].

Nonetheless, one reason for the modest ability of cognitive measures to predict amyloid status is that MCI Aβ+likely is associated with variable levels of impairment. This is almost certainly an issue for any neurodegenerative biomarker given the range of disease severity within the MCI category. Indeed, neurodegenerative biomarkers, in addition to providing some currency on the underlying pathology (e.g., cerebral amyloid), also are informative on disease stage and enhance prediction of the timing of transitions to dementia, as has been suggested in the literature [37–39]. Thus, relatively poor performance on cognitive measures within the MCI category increases both the likelihood that the underlying process is AD and that progression to dementia is more likely to occur in the near future, which may help provide additional context for clinicians in their assessment of these patients.

The choice of CSF Aβ as the proxy or standard for AD pathology in the present analysis also reflects the notion that it is a more specific measure of AD pathology than neurodegenerative markers, given the defining nature of cerebral amyloid in the pathologic criteria for AD. Indeed, more and more therapeutic trials, including in MCI, are using a qualitative amyloid study as inclusion criteria [40]. Thus, examination of psychometric measures within the MCI category may contribute to increasing the likelihood that a given patient may qualify for such a study on that basis.

**CONCLUSION**

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

**ACKNOWLEDGMENTS**

This research was supported by NIH grants T32-EB009384, AG010124, and AG040271. The funding agencies had no say in the design and conduct of the study; collection, management, analysis, and interpretation of the data; or preparation, review, and approval of the manuscript.

Authors’ disclosures available online (http://j-alz.org/manuscript-disclosures/14-2943r2).

Data collection and sharing for this project was funded by the Alzheimer’s Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Alzheimer’s Association; Alzheimer’s Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; B braingen Idee Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Mesro Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Synarc Inc.; and Takeda Pharmaceuticals Company. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer’s Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.
REFERENCES


