Journal of Alzheimer's Disease 32 (2012) 373–385
DOI 10.3233/JAD-2012-120832
IOS Press

Enriching Amnestic Mild Cognitive Impairment Populations for Clinical Trials: Optimal Combination of Biomarkers to Predict Conversion to Dementia

Peng Yu[a,*], Robert A. Dean[a], Stephen D. Hall[a], Yuan Qi[b], Gopalan Sethuraman[a], Brian A. Willis[a], Eric R. Siemers[a], Ferenc Martenyi[a], Johannes T. Tauscher[a], Adam J. Schwarz[a]
and for the Alzheimer's Disease Neuroimaging Initiative[1]
[a]*Eli Lilly and Company, Indianapolis, IN, USA*
[b]*Department of Computer Science, Purdue University, West Lafayette, IN, USA*

Accepted 14 June 2012

Abstract. The goal of this study was to identify the optimal combination of magnetic resonance imaging (MRI), [^{18}F]-fluorodeoxyglucose positron emission tomography (FDG-PET), and cerebrospinal fluid (CSF) biomarkers to predict conversion from amnestic mild cognitive impairment (aMCI) to Alzheimer's disease (AD) dementia within two years, for enriching clinical trial populations. Data from 63 subjects in the Alzheimer's Disease Neuroimaging Initiative aMCI cohort who had MRI and FDG-PET imaging along with CSF data at baseline and at least two years clinical follow-up were used. A Bayesian classification method was used to determine which combination of 31 variables (MRI, FDG-PET, CSF measurements, apolipoprotein E (ApoE) genotype, and cognitive scores) provided the most accurate prediction of aMCI to AD conversion. The cost and time trade-offs for the use of these biomarkers as inclusion criteria in clinical trials were evaluated. Using the combination of all biomarkers, ApoE genotype, and cognitive scores, we achieved an accuracy of 81% in predicting aMCI to AD conversion. With only ApoE genotype and cognitive scores, the prediction accuracy decreased to 62%. By comparing individual modalities, we found that MRI measures had the best predictive power (accuracy = 78%), followed by ApoE, FDG-PET, CSF, and the Alzheimer's disease assessment scale-cognitive subscale. The combination of biomarkers from different modalities, measuring complementary aspects of AD pathology, provided the most accurate prediction of aMCI to AD conversion within two years. This was predominantly driven by MRI measures, which emerged as the single most powerful modality. Overall, the combination of MRI, ApoE, and cognitive scores provided the best trade-off between cost and time compared with other biomarker combinations for patient recruitment in clinical trial.

[1]Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://www.loni.ucla.edu/ADNI). 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 list of ADNI investigators is available at http://adni.loni.ucla.edu/research/active-investigators/.

*Correspondence to: Peng Yu, Eli Lilly and Company, Indianapolis, IN 46285, USA. Tel.: +1 317 277 5528; Fax: +1 317 276 6545; E-mail: yu_peng_py@lilly.com.

ISSN 1387-2877/12/$27.50

Keywords: Alzheimer's disease, apolipoprotein E, biomarker, cerebrospinal fluid, conversion, FDG-PET, magnetic resonance imaging, mild cognitive impairment, prodromal, progression

Supplementary data available online: http://dx.doi.org/10.3233/JAD-2012-120832

INTRODUCTION

Patients clinically diagnosed with amnestic mild cognitive impairment (aMCI) have an increased risk of progressing to a clinical diagnosis of Alzheimer's disease (AD) [1] and might thus benefit from disease modifying treatments. Currently, however, the efficiency of clinical treatment trials on pre-demented subjects is limited by the heterogeneity of clinically-defined aMCI cohorts, with only approximately 12% of patients with aMCI converting to AD annually [1]. The remainder improve, remain stable (possibly due to not having underlying AD pathology), or fail to survive long enough to progress to AD [2]. Identification of a subpopulation of aMCI subjects with a higher likelihood of progression to clinical AD within a relatively short time frame (e.g., 1–3 years) would provide clinical study populations enriched for imminent dementia and probably AD pathology and hence with a more consistent disease trajectory than might be selected based on clinical criteria alone. Reliable identification of such study populations is expected to improve the detection of disease modifying treatment effects in clinical trials in the aMCI population.

A number of biochemical and imaging biomarker measurements have been shown to be strongly associated with AD pathology and disease progression. These include: 1) decreased cerebrospinal fluid (CSF) concentrations of amyloid-$\beta_{1\text{-}42}$ ($A\beta_{42}$) isoforms purportedly reflecting central nervous system deposition of amyloid pathology, and increased CSF total tau (tTau) and tau phosphorylated at threonine 181 (p-Tau_{181}) presumably reflecting neuronal injury with microtubule disassembly and development of neurofibrillary tangles; 2) loss of brain volume globally or in specific regions of interest as determined by structural magnetic resonance imaging (MRI) and assumed to reflect parenchymal atrophy; and 3) diminished resting brain glucose metabolic rate measured using [^{18}F]-fluorodeoxyglucose positron emission tomography (FDG-PET). Combined CSF $A\beta_{42}$, tTau, and p-Tau_{181} measures, especially when combined as ratios like the tTau:$A\beta_{42}$ ratio [3, 4] have been shown to be associated with AD pathology [5–7], disease stage [3, 4, 8, 9], subsequent cognitive decline [10], and central amyloid load as determined by amyloid imaging [3]. Anatomically-localized MRI measures have been shown to correlate with disease stage [11], neuronal density [12], postmortem Braak stage [13], and clinical scales [14, 15]. Volumetric loss in selected regions of interest, notably the hippocampus and medial temporal lobes, and volumetric expansion of the ventricles are among the most accurate MRI-based markers of disease stage [16–18]. Finally, FDG-PET measures characterize progressive hypometabolism as a function of disease stage [19, 20].

Moreover, a number of recent studies have indicated that these biomarker measures may individually be associated with, and predictive of, progression from aMCI to clinically diagnosed AD dementia. This opens the possibility of using such biomarkers to enrich or stratify the population of clinical trials targeting a prodromal subpopulation of aMCI subjects at a high risk of progression to AD dementia [21, 22]. Indeed, several groups have shown, in a variety of independent subject cohorts, that measurements derived from structural MRI data [17, 18, 23–27], CSF samples [4, 10, 28–30], and FDG-PET [31, 32] can be predictive of subsequent progression from aMCI to AD. Moreover, apolipoprotein E (ApoE) genotype, in particular the $\varepsilon 4$ allele, has also been shown to be predictive of disease progression [32, 33].

Each of the above biomarkers measures a different aspect of the underlying pathology and may thus be complementary in their ability to predict subsequent clinical decline. All can be included in clinical trials (and are already, although most commonly to date as outcome rather than screening measures). Moreover, use of these methods in combination has also begun to be explored; MRI has been shown to improve the diagnostic prediction of CSF [34, 35] and FDG-PET biomarkers to improve upon the predictive power of ApoE genotype alone [31]. More recently, multimodality models, using a combination of MRI, CSF, FDG-PET, and cognitive functions, have been built to compare AD or aMCI with normal subjects and track disease progression in aMCI [36–40]. Promising results were shown by building classification models in AD and normal and testing the resulting models in aMCI to predict which patients will progress to AD [36, 39, 40]. However, variable selections were seldom conducted to compare the predictive performance of

individual biomarkers and logistic considerations for using these combinations of modalities in clinical trial were not discussed.

In the present study, we examined the predicative performance of a larger set of biomarkers: MRI, FDG-PET, and CSF measures along with ApoE genotype and baseline cognitive performance. Specifically, the questions we sought to answer in this work were: 1) For the purpose of identifying aMCI subjects who will imminently progress to clinically-diagnosed AD dementia, which combination of the above biomarker, genetic and clinical variables provides the most predictive power?; 2) To what extent do the imaging and CSF biomarkers improve patient enrichment over a selection strategy based on cognitive scores and genotype alone?; 3) Since the acquisition of imaging and CSF data imposes logistical constraints (e.g., site selection) and additional cost, what are the time/cost versus prediction accuracy trade-offs associated with biomarker-driven enrichment.

To address these questions, we employed a Bayesian classification framework to automatically identify most predictive biomarkers across multiple modalities by directly comparing aMCI patients progressed to AD with those remained stable within the follow-up time. We also compared the predictive power and enrichment performance of different modalities (MRI, FDG-PET, CSF, ApoE genotype, cognitive tests) using the classification models built with the Bayesian method. Finally, we built a logistical model to examine the time/cost benefits of using different enrichment strategies in aMCI clinical trials.

MATERIALS AND METHODS

Data

We analyzed Alzheimer's Disease Neuroimaging Initiative (ADNI) data released in June 2010 (http://www.loni.ucla.edu/ADNI). ADNI is a five-year multi-site program funded by a public-private partnership including the National Institute on Aging (NIA), Food and Drug Administration (FDA), pharmaceutical companies, and non-profit organizations to investigate the relationship of neuroimaging, biological, clinical, and neuropsychological assessments to disease progression in AD. 800 subjects were recruited: approximately 200 elderly controls, 400 with aMCI, and 200 with AD. Subjects were followed for 2–3 years and assessed every 6 to 12 months. aMCI subjects had Mini-Mental State Examination (MMSE) scores between 24–30, a memory complaint, objective memory loss measured by education adjusted scores on Wechsler Memory Scale Logical Memory II, a Clinical Dementia Rating (CDR) of 0.5, absence of significant levels of impairment in other cognitive domains, essentially preserved activities of daily living, and an absence of dementia. At each visit, aMCI subjects were assessed whether or not they clinically progressed to AD, remained aMCI, or regressed to normal. Using data measured on these aMCI patients in ADNI, we evaluated the best combination of baseline biomarkers for predicting progression to AD within 2 years.

Biomarker measures

We considered a total of 31 numeric variables, comprising biomarker measures generated by the ADNI-funded laboratories for structural MRI, FDG-PET, and CSF, along with ApoE genotype, ADAS-Cog, and MMSE clinical scales (see Supplementary Table 1; available online: http://www.j-alz.com/issues/32/vol32-2.html#supplementarydata02).

Since multiple research laboratories were funded to analyze the structural MRI data in the ADNI study, and several of these generated closely related measures (e.g., hippocampal volume, cortical parcellation), we used the set of variables from the laboratory that performed well in a comparative power analysis. For each MRI measurement, we calculated the number of subjects needed to detect a 25% reduction of the annual percentage volume change in AD subjects, with 80% power and 5% significance [41]. In a simple setting where we have only one placebo and one treatment arm, the number of subjects per arm, denoted by n, can be calculated by:

$$n = \frac{2\sigma^2(Z_{1-\alpha/2} - Z_\beta)^2}{(d_t - d_p)^2} \quad (1)$$

where d_p and d_t denote the average percentage change in the placebo and treatment groups, σ denotes the standard derivation (assuming equal variance), α denotes the type I error rate (e.g., $\alpha = 0.05$), β denotes the statistical power (e.g., $\beta = 0.8$), and Z_x is the cumulative normal distribution statistic at significance level x. Based on this power analysis, we selected the MRI volumetric measurements generated by the University of California, San Diego (UCSD) for this study [41]. This analysis yielded the 14 volumetric MRI (vMRI) variables listed in Supplementary Table 1.

Different from the MRI images, which were analyzed by multiple labs for a variety of neuroanatomical structures using different methods, the analysis of FDG-PET images has mainly yielded summary

Table 1
Summary of the subset of ADNI subjects used for analysis. Only 2-year change of MMSE is significantly different between aMCI converter and non-converters using 2-sided student's *t*-test

	Converter	Non-converter
# of subjects	25	38
Age	74.00 (5.57)	74.82 (7.16)
Gender (Male)	16	17
ADAS-Cog (11 item)	20.36 (5.40)	18.03 (5.90)
2 year change of ADAS-Cog (11 item)	3.35 (5.00)	1.80 (4.33)
MMSE	26.60 (1.87)	27.29 (1.54)
2 year change of MMSE*	2.95 (3.37)	–0.89 (2.21)
# of ApoE $\varepsilon 4$ carrier	15	17

*p-value < 0.05.

statistics for a few regions-of-interest (ROIs). Therefore, we included 6 variables generated and recommended by all three ADNI funded analysis laboratories [42, 43] as listed in Supplementary Table 1.

Finally, we included all three CSF measurements ($A\beta_{42}$, tTau, and p-Tau_{181}) generated by the ADNI biomarker core, along with ApoE genotype, ADAS-Cog, and MMSE scores (see Supplementary Table 1). Information about the acquisition and measurements of MRI, FDG-PET, CSF, ApoE genotype, ADAS-Cog, and MMSE can be found in ADNI procedure manual (http://www.loni.ucla.edu/ADNI).

MCI subjects

To determine the optimal combination of all the above biomarkers to predict aMCI to AD conversion, we used the aMCI subjects from ADNI who had baseline measurements across all modalities as well as two-year clinical follow-up. Subjects with other, non-AD related underlying pathologies (e.g., frontal lobe dementia, Parkinson's disease) were excluded. Since only a subset of the subjects had CSF and FDG-PET measurements, we had in total 63 aMCI subjects available for the combined biomarker analysis. Of these, 25 subjects converted to AD within 2 years and 38 subjects did not (summary in Table 1, a full list of subject IDs is provided in Supplementary Table 2). Table 1 summarizes the demographic, clinical, and ApoE genotype profiles of the aMCI cohort split into converter and non-converter groups. The 2-year MMSE change was significantly different between the two groups (2-sided Student's *t*-test, $p < 0.05$).

Amyloid imaging was not included, as only about 12% of ADNI subjects had their first amyloid scan taken approximately one year after baseline. The number of aMCI patients with PIB-PET measurements and a two-year follow-up was considered too small for reliable modeling. The conversion rate in this ADNI aMCI sub-population is 43% in 2 years, which is consistent with the overall conversion rate (44%) in the whole ADNI aMCI population. It is noted the ADNI aMCI conversion rate is higher than what has been reported in other studies [1].

Bayesian approach for biomarker selection

Given the 31 biomarker measures across multiple modalities, it is crucial to select the biomarkers that best separate the converters from the non-converters in order to obtain a parsimonious model that avoids over-fitting to noise in the data. In the present study, we adopted a Bayesian approach, predictive Automatic Relevance Determination (pred-ARD) [44], to jointly select measurements that are predictive of the aMCI progression and to train the classifier. Compared to other classification methods (e.g., support vector machines), where variables have to be selected separately from the classification task, this Bayesian method couples the task of variable selection with classification, and jointly selects predictive variables from all the available input measurements. Additional technical details on the Bayesian classification algorithm, pred-ARD, are provided in Supplementary Materials.

Before applying this method, we first reduced the number of variables by eliminating measurements that were highly correlated with others. We grouped the full set of 31 variables into binary clusters using their correlation coefficients. For the purpose of removing highly redundant variables, only one variable in each cluster was retained to form a reduced variable set. Finally, since each variable has different range and unit, we normalized each variable to have a zero mean and variance of 1 across the 63 subjects.

After applying the pre-ARD method to the input variables, we obtain a classification model, where only variables relevant to separating converters from non-converters have nonzero weights. Moreover, we can rank the importance of these selected variables by using their corresponding weights in the classifier. The larger the magnitude of the weight, the more significant the variable is for distinguishing aMCI converters from non-converters. Finally, we can use the resulting classification model to directly predict the probability of conversion for a new patient. The higher the predicted conversion probability, the more likely the patient will convert to aMCI within the given follow-up time. For calculating sensitivity and specificity, we used 50% as cut-off in this study (i.e., subjects with predicted

conversion probability >50% are classified as converters and subjects with conversion probably ≤50% are classified as non-converters). We revisit the question of cut-off selection in the Supplementary Materials.

Because of the limited number of subjects in this study, we used Leave One Out (LOO) cross-validation to calculate the prediction accuracy of the classification method. That is, we built n classification models, using n-1 subjects each time and using the resulting classifier to classify the n-th (left-out) subject as a converter or non-converter. The average LOO prediction accuracy across all $n = 63$ subjects is reported for the cross-validation tests shown in this paper.

Logistical impact of biomarker-based enrichment on clinical trials

The use of biomarker-based screening in clinical trials involves additional cost and time. In this work, we also modeled the logistical impact of using these biomarkers as inclusion criteria in a clinical trial based on their prediction accuracies estimated using this Bayesian method.

For the purpose of comparing different enrichment options, we considered a clinical trial with a duration of γ years and 2 arms of aMCI patients who have an annual aMCI to AD conversion rate of r_c. We wish to determine the number of patients that must be screened and enrolled to achieve the desired statistical power. If Se and Sp are the sensitivity and specificity of the baseline biomarkers for predicting the aMCI to AD conversion, then out of N_s aMCI patients screened, $Se * r_c * \gamma * N_s + (1 - Sp) * (1 - r_c * \gamma) * N_s$ patients will be classified as converters in γ years. However, if we enroll all these predicted converters, only $Se * r_c * \gamma * N_s$ are true positives (will convert to AD), whereas $(1 - Sp) * (1 - r_c * \gamma) * N_s$ are false positives (will not convert). We further assumed these patients are randomized equally into the placebo and treatment arms, and the patients who will convert have a non-zero change of the primary endpoint responding to the treatment, while the patients who will not convert have no treatment response. For simplicity, we also assumed that the change in the primary endpoint has the same variance in the converters and non-converters, and in the treatment and placebo groups. Finally, we denote the yearly drop-out rate as r_d, and we assumed that we only use patients who complete the last visit to calculate efficacy. Using Equation (1), the total number of patients needed to detect a given effect size with a specific type I error rate and statistical power is:

$$N \approx \frac{Se * r_c * \gamma + (1 - Sp) * (1 - r_c * \gamma)}{Se * r_c * \gamma} \quad (2)$$

To recruit these N patients, the number of patients we need to *screen* is

$$N_s = \frac{N}{Se * r_c * \gamma + (1 - Sp) * (1 - r_c * \gamma)} \approx \frac{1}{Se * r_c * \gamma} \quad (3)$$

Equation (3) shows that the number of patients to be screened is inversely proportional to the sensitivity of the screening criteria.

Using Equations (2) and (3), we can compare the number of patients we need to screen and enroll for achieving a certain statistical power using different combinations of biomarkers as inclusion criteria, and evaluate the cost-benefit of different screening strategies. Furthermore, we can estimate the additional time needed for screening N_s patients, if we assume a constant patient recruitment rate. Since N_s is inversely proportional to the product of sensitivity, yearly conversion rate, and the length of the clinical trial (Equation (3)), the length of screening process primarily depends on the sensitivity of different screening strategies, when everything else is held equal.

Using the estimated prediction accuracies, we can calculate the logistical impact of using inclusion criteria in a clinical trial based on different biomarkers as compared with the scenario where no biomarker screening strategies are used. We denoted by N_{no_biomk} the number of subjects needed to obtain a certain statistical power in a clinical trial when no additional biomarker screening were used. The number of subjects N_{biomk} needed to achieve the same power using a biomarker screening strategy can be calculated using Equation (2) as:

$$N_{biomk} = \frac{Se_{no\_biomk} * (Se_{biomk} * r_c * \gamma + (1 - Sp_{biomk}) * (1 - r_c * \gamma))}{Se_{biomk} * (Se_{no\_biomk} * r_c * \gamma + (1 - Sp_{no\_biomk}) * (1 - r_c * \gamma))} N_{no\_biomk} \quad (4)$$

where Se and Sp are the sensitivity and specificity, r_c is the yearly aMCI to AD conversion rate, and γ is the number of years of clinical trial. Using Equation (4), we can compare the cost in recruiting different number of subjects under different biomarker screening strategies. Further, we can compare the differences

Table 2
Variables separating aMCI patients that converted to AD in 2 years and aMCI patients that remained stable, ranked by their predictive power (mean weight) in the classification model

Rank	Variable	Modality
1	RMIDTEMP RINFTEMP	MRI
2	LFUSIFORM	MRI
3	RFUSIFORM	MRI
4	ApoE 44	ApoE
5	VENTRICLES	MRI
6	X2SDSIGPXL X3SDSIGPXL	FDG-PET
7	AVEASSOC AVEFRONT	FDG-PET

in screening time by calculating the number of patients needed to be screened using Equation (3). In this study, our comparison scenario (with no additional biomarker screening) was based on a 2-year clinical trial in aMCI patients with a yearly conversion rate of 15%. In this case, the sensitivity is 100% and specificity is 0% since we accept all aMCI patients in the trial. Based on an ongoing clinical trial in AD patients, the cost of following a patient for $\gamma = 2$ years was estimated to be \$22,000 per person, the cost of acquiring and analyzing MRI, FDG-PET, CSF, and ApoE data was estimated to be \$4000, \$8000, \$700, and \$120 per acquisition, per subject, respectively. These values are approximate, but indicative, and include both acquisition and analysis costs.

RESULTS

Prediction accuracy based on single biomarker variables

We first evaluated individual biomarkers to estimate their predictive power for aMCI to AD conversion. In this step, we built the classification model separately for each of the 31 baseline variables by turning off the variable selection function in the pred-ARD algorithm. We then ranked the individual biomarkers by their LOO classification accuracy (supplementary Figure 1). Several of the top ranked biomarkers were derived from volumetric MRI. The highest accuracy was around 69% by using the right middle temporal lobe volume. Although the aMCI converters had on average small hippocampus at baseline, this measurement alone only provided an accuracy of about 63%. The ApoE genotype was ranked the sixth, with an accuracy of 63%. p-Tau$_{181}$ provided the best prediction accuracy (57%) among the CSF measurements, compared with tTau (55%) and Aβ_{42} (54%).

Prediction accuracy based on combinations of biomarker modalities

We then tested the prediction performance using all biomarkers across all modalities. The dimensionality of input variables was first reduced by removing redundant variables, as described in the Methods. In this step, we grouped the 31 original input variables into 22 groups of one or two members. Variables grouped together were: left and right entorhinal cortex, right middle temporal cortex and inferior temporal cortex, left middle temporal cortex and inferior temporal cortex, left and right inferior lateral ventricle, left and right hippocampus, X2SDSIGPXL and X3SDSIGPXL, AVEASSOC and AVEFRONT, tTau and p-Tau$_{181}$, and TOTAL11 and TOTALMOD (refer to Supplementary Table 1 for variable names). We used the first variable in each of these groups (ranked alphabetically based on variable name) in our analysis.

In this combination study, we first applied the models to all 22 variables from all modalities. As a result, a combination of 7 was selected by the pred-ARD model to best predict aMCI to AD conversion within 2 years (Table 2). The temporal lobe volumetric measurements were ranked the highest among all the input variables, indicating that these were most predictive of the aMCI to AD conversion. Atrophy in the fusiform gyrus and enlargement in the lateral ventricle were also associated with higher risk of aMCI to AD conversion. ApoE $\varepsilon 4$ homozygotes also had a higher risk of converting to AD. Ranked by modality, MRI measurements had most predictive power, followed by ApoE genotype and FDG-PET measurements. This result was consistent with the ranking of individual measurements shown above.

With the combination of all biomarker modalities, we obtained an overall prediction accuracy of 81% (sensitivity = 80%, specificity = 81%) in cross validation. This performance was better than any single biomarker or variable alone (supplementary Figure 1). We note, however, that the top 7 features selected in this combined analysis were not the same as the top 7 ranked features in the single biomarker analysis (see above). In fact, the prediction accuracy was only 76% when we used the combination of the top 7 features shown in supplementary Figure 1. These results demonstrated that the classification algorithm can automatically select the best combination of biomarkers that predict the conversion from complementary aspects.

Next, we applied the classification model to subsets of variables corresponding to the MRI, FDG-PET,

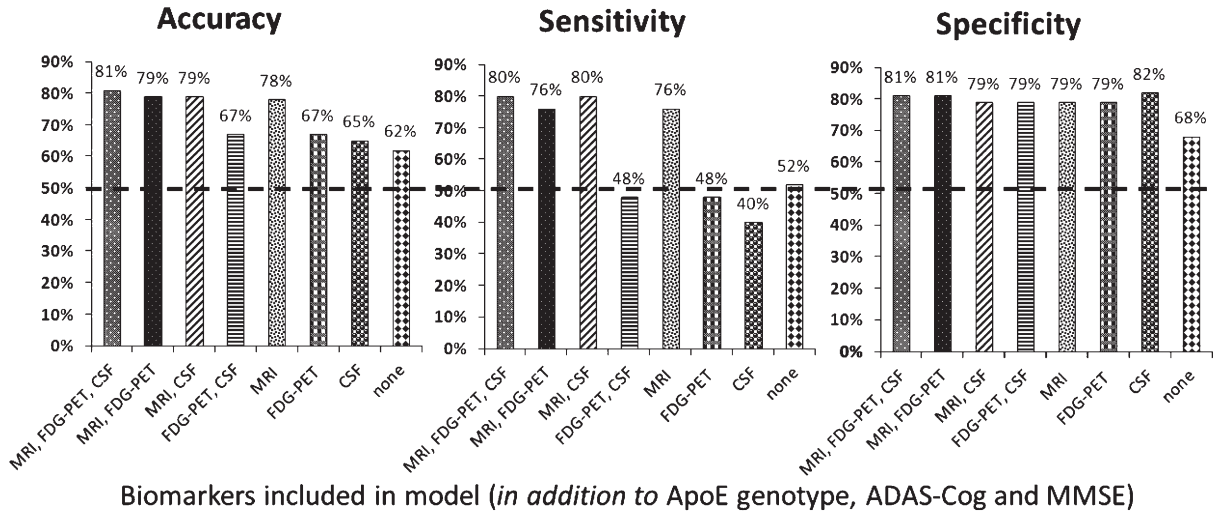

Fig. 1. Classification Accuracy by the combination of biomarker modalities (bars arranged by the order in the legend).

Table 3
Summary of ADAS-Cog and MMSE changes in subjects with estimated conversion probability >50% in the cross validation using different screening strategies. All 63 subjects are included when there is no additional enrichment (last row)

	Modalities	N	2 year ADAS-Cog change, mean (SD)	2 year MMSE change, mean (SD)
Biomarker combination (in addition to ApoE, ADAS-Cog, MMSE)	MRI, FDG-PET, CSF	29	5.38 (6.29)	–2.89 (3.24)
	MRI, CSF	29	4.76 (5.55)	–2.15 (3.37)
	MRI	27	4.94 (6.36)	–3.04 (3.14)
	CSF	17	4.26 (4.54)	–1.18 (2.43)
	ApoE, ADAS-Cog, MMSE	19	4.31 (4.97)	–1.53 (2.62)
	No Enrichment	63	3.36 (5.86)	–1.67 (2.87)

and CSF modalities to understand the added value of each modality (gain in classification accuracy). In this test, the ApoE genotype and cognitive test scores were always included since they are typically available for clinical trials. As shown in Fig. 1, we found that most of the increased predictive power obtained with biomarker data was due to the MRI measures, with FDG-PET and CSF measures providing incremental improvement in classification accuracy. Moreover, most of the increased accuracy was driven by increased sensitivity to clinical progression, with smaller improvements in specificity.

Distribution of prediction probability and cognitive score trajectories

Obtaining faster progression and more homogeneous clinical trajectory is one of the main reasons for considering the enrichment strategy discussed in this paper. Given the noise in the clinical instruments as well as the biomarkers, it must be shown that an enriched population indeed demonstrates the desired qualities. To validate the use of these proposed biomarkers, we calculated the 2-year change of cognitive scores for patients that were predicted to be converters using the different combination of biomarkers. The cognitive scores we considered in this test include the ADAS-Cog (13 item: 0–85) and MMSE. In Table 3, we list the mean 2-year change of these scores in patients who were predicted to have >50% probability to convert to AD in 2 years. Compared with the no enrichment scenario, where all 63 subjects were considered to be converters, patients selected with additional screening have more aggressive progression (more change of ADAS-Cog and MMSE scores in 2 years). Furthermore, among different enrichment strategies, the addition of biomarkers can help to select sub-population with faster progression compared with only ApoE genotype, ADAS-Cog, and MMSE. However, as shown in Table 3, adding CSF alone did not result in a subgroup of faster progressors in terms of the ADAS-Cog and MMSE changes, compared

Table 4
Cost and time estimates for enrichment using different combinations of biomarker-based screening in a two-year clinical trial in aMCI patients (with a probability cut-off of 50%). Cost and time saving estimates are calculated relative to the number of subjects needed to achieve a certain statistical power when no enrichment strategy is used

	Modalities	*Se* (%)	*Sp* (%)	Number of subjects screened	Number of subjects enrolled	Screening cost ($10,000)	Trial cost ($10,000)	Total cost ($10,000)	Additional cost (%)	Screening duration	Additional screening duration (when $T = 2$ years)
Biomarker combination (in addition to ApoE, ADAS-Cog, MMSE)	MRI, FDG-PET, CSF	81	80	1.23N	0.47N	1.58N	1.04N	2.62N	19	1.23T	0.47
	MRI, CSF	80	79	1.25N	0.48N	0.60N	1.06N	1.67N	−24	1.25T	0.50
	MRI	76	79	1.32N	0.49N	0.54N	1.09N	1.63N	−26	1.32T	0.63
	CSF	40	82	2.50N	0.62N	0.21N	1.35N	1.56N	−29	2.50T	3.00
	ApoE, ADAS-Cog, MMSE	52	68	1.92N	0.73N	0.02N	1.61N	1.63N	−26	1.92T	1.85
	No Enrichment	100	0	N	N	0	2.2N	2.2N	0	T	0

with other enrichment strategies. However, as shown in previous results, CSF did increase the overall aMCI to AD prediction accuracy due to the improvement in specificity (i.e., identifying the non-converters correctly).

In Supplementary Figure 2, we plot the histogram of the predicted conversion probabilities for 63 subjects using different enrichment strategies. With only ApoE genotype, ADAS-Cog, and MMSE, the predicted conversion probability is centered around 50%, which indicates weak predictions. By contrast, with the addition of biomarkers, the conversion probability is distributed around either the top (75–100%) or the bottom (0–25%) quartiles, and is statistically significantly different from the borderline probability (50%) using student's *t*-test ($p < 10^{-20}$). These results demonstrate that, compared with only ApoE and cognitive scores, the biomarker data enabled stronger predictions about the disease progression.

Logistical implications of biomarker-driven inclusion criteria for clinical trials

For clinical trials, baseline cognitive tests and genotype information would typically be available as routine. However, the acquisition of CSF and imaging biomarker data involves additional cost and logistical constraints. For the purposes of enriching a clinical trial in a prodromal AD population for likelihood of imminent progression, it is therefore critical to understand the added value of each modality in a practical sense.

As shown in Table 4, the overall cost in each scenario was a trade-off between the additional cost of screening more patients and the cost saved by recruiting and maintaining fewer patients in the trial. Using vMRI, in addition to ApoE genotype, ADAS-Cog, and MMSE, we needed to screen about 32% more patients compared with the no-screening scenario, but enroll only about 49% of the patients into the 2-year trial. As a result, the overall cost is reduced by 26%. Taking out MRI from this combination did not change the cost saving because the increase in trial cost (due to the reduction of prediction accuracy) is balanced out by the saving in screening, due to the low cost of ApoE genotyping ($120/patient). In contrast, although the combination of all the modalities gave the best prediction accuracy (81%), the high cost of FDG-PET made this combination the most expensive strategy. The least expensive strategy was the combination of CSF with ApoE genotype, ADAS-Cog, and MMSE (29% cost reduction) because of the lower cost of the CSF biomarkers compared with MRI and FDG-PET imaging. When compared with screening based on genotype and cognitive testing alone, the inclusion of MRI, CSF, or both in the screening stage did not markedly alter the overall trial cost.

However, by increasing the number of patients screened, all the biomarker-based enrichment strategies would lengthen the trial due to an increased time required for screening. The increased screening time was determined mainly by the prediction sensitivity. In Table 4, we listed the additional screening time assuming we need 2 years to recruit N subjects, which made the total time to be 4 years (with 2 years trial period). Since MRI measurements mainly improved the diagnostic sensitivity, strategies involving MRI were less time-consuming. Indeed, all strategies involving MRI reduced the screening duration in comparison to screening without any biomarkers. For example, using

MRI, ApoE, ADAS-Cog, and MMSE required only 0.63 year of additional screening time, in contrast to an additional 1.85 years if only ApoE, ADAS-Cog, and MMSE were used. Although we have similar cost savings using both strategies (26%), the addition of MRI resulted in an overall trial duration that was shorter by 1.22 years. In contrast, the least expensive strategy (CSF, ApoE, ADAS-Cog, and MMSE) required the longest screening time (3 additional years) because of the low sensitivity (40%) associated with this combination.

DISCUSSION

We identified a set of biomarker measures for the purpose of identifying aMCI subjects who will progress to clinically-diagnosed AD dementia within 2 years, an approach that has the potential to enable more efficient clinical treatment trials. We found the most accurate (81%) and sensitive (80%) set of measures to be a combination of MRI, FDG-PET summary measures, and CSF ($A\beta_{42}$, p-Tau_{181}, tTau) biomarkers along with ApoE genotype and cognitive scores.

Although the performance of this combination was better than any single modality alone, it was driven predominantly by MRI volumetric measurements (accuracy 78%, sensitivity 76%). Of the individual measurements within each modality, vMRI measurements in the temporal lobe were the strongest individual predictors, followed by ApoE genotype and the measurement of hypometabolic activity from FDG-PET (Table 3). When practicalities such as time and cost were factored in, a more parsimonious combination of MRI, ApoE genotype, and cognitive tests provided a sensitivity of 76%, specificity of 79% and overall cost savings of 26% with only a 30% increase in screening duration. We found that the differences in prediction performance were driven predominantly by differences in sensitivity, with all measures yielding similar specificity of ∼80%.

In contrast, cognitive scores alone yielded an accuracy of 62% and a sensitivity of only 52% in predicting imminent dementia. This may in part reflect the fact that both the MMSE and ADAS-cog scales are optimized for AD populations, with aMCI subjects near the ceiling and floor of each scale, respectively.

The finding of increased prediction accuracy using a combination of biomarker measures reflects their association with different aspects of the underlying pathology and the different temporal relationships between the biomarker changes and disease trajectory [45]. Our results are consistent with a number of other published studies examining combinations of biomarker modalities for the prediction of short term conversion from aMCI to AD. In an independent analysis based on a different subset of ADNI data, MRI was shown to provide superior predictive power to FDG-PET in the same subjects [46]. In studies in which CSF and MRI biomarkers were examined together, medial temporal lobe atrophy was found to improve the progression prediction accuracy obtained by CSF biomarkers alone (74%) to 84% in a sample of 24 MCI subjects [34]. In a study of 192 aMCI and 98 AD subjects from ADNI, a structural abnormality index ("STAND") atrophy pattern score—dominated by changes in the temporal lobes [23]—was found to be a stronger predictor of short-term future cognitive change (∼2 years) than CSF measures with a hazard ratio for time to conversion of 2.6 [15].

MRI measures

The accuracy of MRI measures for predicting aMCI to AD conversion in a short period of time (2 years) in the present study is consistent with other published findings. Hippocampal measures were found to be related to increased risk of progression to AD in a community sample of 80 subjects [18] and in a referral sample of 190 subjects from the ADCS Vitamin E Donepezil trial [27]. In a 1.5 year study, the accuracy in distinguishing between aMCI subjects who developed dementia and subjects who remained stable was 70–80% based on MRI analyzed using deformation-based morphometry [47]. Cortical thickness measures in $N = 49$ referral subjects with a CDR-SB score of 0.5 provided an accuracy of 74% (sensitivity 83%, specificity 65%) in predicting progression in an average follow-up time of 2.5 years [26]. Finally, McEvoy and colleagues recently demonstrated that using MRI atrophy profiles to distinguish "AD-like" from "normal-like" subjects in an enrichment strategy based on baseline atrophy rates can substantially increase the statistical power to detect a treatment effect [25].

The volumetric MRI measurements used in this paper were analyzed by UCSD using a well-validated software package along with careful human curation [41]. Different image processing methods and procedures will result in slightly different numeric values and may thus affect the resulting classification accuracy in conversion prediction. This and the different population samples may explain in large part the differences in sensitivity, specificity and accuracy reported

between the above studies. Nevertheless, the reported accuracy values of 70–80% are consistent with the present report.

FDG-PET measures

In the present study, both FDG-PET measurements and ApoE genotype were found to be strong predictors of aMCI to AD progression. In a previous study of 30 MCI patients, FDG-PET was shown to have a better prediction performance (sensitivity 92%, specificity 89%) compared with ApoE $\varepsilon 4$ (sensitivity 75%, specificity 56%) [31]. The superior performance of FDG-PET measurements (compared with ApoE) in the other dataset may be due to differences in both the subject sample and image analysis methods.

CSF biomarkers

Over 100 separate publications have reported the association of abnormally low $A\beta_{42}$, high tTau and p-Tau_{181}, or changes in various ratios derived from these three parameters with both neuropathologically confirmed and clinically-defined probable AD dementia. Abnormal, AD-like, CSF neurochemical profiles tend to occur early among individuals presenting with MCI who subsequently manifest AD dementia. Numerous previous studies supported the use of CSF biomarkers for AD diagnosis [4, 8, 28–30, 48, 49] and reported MCI to AD conversion prediction accuracies of 85–95%, although with better performance for substantially longer follow-up times.

Longitudinal analyses of AD patients suggest that decreases in CSF $A\beta_{42}$ occur early and more abruptly than observed increases in CSF tTau and p-Tau_{181}. The AD-associated decrease in CSF $A\beta_{42}$ also occurs considerably earlier than disease-associated abnormalities in FDG-PET, structural MRI, and cognitive and behavioral changes. This early manifestation of CSF neurochemistry changes in the AD disease process is a probable factor in our finding that CSF measures did not provide as sensitive a prediction for aMCI to AD dementia progression within a 2 year time window when compared to structural MRI and FDG-PET. The latter modalities may reflect later stage disease processes with more rapid changes close to the aMCI to AD stages [45]. As a result, CSF biomarkers were noted to be poorer predictors for imminent progression from aMCI to AD dementia relative to structural MRI and FDG-PET.

Logistics and trade-offs for clinical trials

The main aim of this study was to examine the relative utility of different biomarker modalities (and combinations thereof) to predict short-term conversion from aMCI to clinical AD dementia for the purpose of enriching clinical treatment trials. These biomarkers can thus be used at screening to enroll only aMCI patients who are predicted to convert to AD within the specified time frame. The rationale for such an enrichment strategy has several aspects: 1) imminent converters are more likely to have the underlying AD pathology at which a disease-modification treatment is targeted; 2) the study population will represent a more homogeneous sample of a particular stage in the disease trajectory, hence increasing the power of clinical outcome measures; and 3) the possibility of using conversion events as an endpoint. However, the acquisition of biomarker data has associated costs in both time and monetary terms. The overall benefit of using biomarker-based enrichment is a trade-off between saving in trial cost, additional screening time and potentially reduced trial duration due to fewer subjects needing to be maintained in the trial. Using the proposed logistical model, we have shown that the additional screening time is determined only by the predication sensitivity of these strategies. Using costs and parameters taken from an ongoing Phase III clinical trial, we thus also compared this trade-off and consider the logistical impact of different biomarker strategies to screen patients in a putative clinical trial.

In the present context, a biomarker strategy is most efficient when the acquisition cost is low and the prediction accuracy is high. Although MRI was the strongest aMCI to AD conversion predictor, it is only beneficial when the prediction accuracy is high enough to compensate for the increased cost ($4000/patient). Based on our classification results and trial simulations, using enrichment based on structural MRI along with ApoE genotype and cognitive tests would require the enrollment of approximately half the number of subjects that would be needed if no additional screening beyond the ADNI aMCI entry criteria were used. An additional 32% of subjects would need to be screened, increasing the enrollment period and lengthening the trial overall, but the overall trial costs would be reduced by 26%. Using CSF measures instead of structural MRI resulted in even greater cost savings (29%), despite the lower prediction accuracy, due to the lower unit cost for the CSF measures ($700/patient). However, the lower prediction sensitivity resulted in

a much longer expected enrollment and overall trial duration. In contrast, the use of a combination of MRI, FDG-PET, ApoE genetotype, and cognitive tests, yielding the highest prediction accuracy (81%) and an optimally enriched trial population, would increase the overall trial cost (∼19%) because of the high costs associated with both MRI and FDG-PET imaging, despite the smallest increase in trial duration compared with all other biomarker combinations. Overall, based on the prediction estimates obtained in the present study, we found that the combination of MRI, ApoE genotype, and cognitive tests performed well in terms of cost saving (26% compared with no enrichment) and relatively modest increase in screening time. Indeed, if our model were applied to genotype and cognitive test screening data alone, the trial duration would actually be increased with no compensatory overall cost savings, compared with the inclusion of MRI.

Importantly, we note that although these results were derived based on certain assumed values related to the trial setup, cost structure, and annual aMCI to AD conversion rate, the mathematical framework we introduced for this analysis is quite general. It can be applied with any defined parameter values to assess the impact of biomarker-based enrichment strategies on both cost and time. The intensive use of biomarkers in clinical trials also brings other logistical issues. For example, many patients, site investigators, and IRBs in some countries do not accept the lumbar puncture procedure required to obtain CSF. If CSF sample collection is a required part of a clinical study, the inability or unwillingness to conduct lumbar puncture may thus constrain site selection. Similarly, the requirement for imaging data requires the involvement and coordination of suitable imaging sites within traveling distance of the clinical centers.

Limitations

In the ADNI study, while almost all subjects had 1.5T MRI at multiple time points, only about 50% of subjects had CSF biomarker data and 50% had FDG-PET scans. Accordingly, the biomarker data available at the time of the present analysis comprised only 63 aMCI subjects with all three of these modalities available and that had been followed for 2 years subsequent to their first (baseline) measurements. The observations made in this study, such as the increase of prediction power by using multiple modalities and the ranking of different modalities in predicting aMCI to AD conversion, need to be further validated in larger datasets. We further note that while this performance may prove useful for enriching populations for clinical trials, it is not yet robust enough for general clinical use.

CONCLUSIONS

Using a Bayesian classification method, we found that a combination of structural MRI and FDG-PET imaging, CSF measurements, and ApoE genotype provided better prediction accuracy (81%; sensitivity 80%) of aMCI to AD conversion than any single modality alone. This performance was primarily driven by vMRI measurements (accuracy 78%, sensitivity 76%). However, the overall utility of using these biomarkers as an enrichment strategy for clinical trials also involves consideration of the impact on both cost and trial duration. Based the ADNI dataset, we found that a more parsimonious combination of vMRI measures, ApoE genotype, and cognitive tests provided both considerable cost saving and a short screening time compared with other screening strategies.

ACKNOWLEDGMENTS

The authors would like to thank Karen Sundell at Eli Lilly and Company for her contribution in clinical trial and biomarker acquisition cost assessment. Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Abbott, AstraZeneca AB, Bayer Schering Pharma AG, Bristol-Myers Squibb, Eisai Global Clinical Development, Elan Corporation, Genentech, GE Healthcare, GlaxoSmithKline, Innogenetics, Johnson and Johnson, Eli Lilly and Co., Medpace, Inc., Merck and Co., Inc., Novartis AG, Pfizer Inc, F. Hoffman-La Roche, Schering-Plough, Synarc, Inc., as well as non-profit partners the Alzheimer's Association and Alzheimer's Drug Discovery Foundation, with participation from the U.S. Food and Drug Administration. Private sector contributions to ADNI 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 California,

Los Angeles. This research was also supported by NIH grants P30 AG010129, K01 AG030514, and the Dana Foundation.

Authors' disclosures available online (http://www.j-alz.com/disclosures/view.php?id=1403).

REFERENCES

[1] Petersen RC (2004) Mild cognitive impairment as a diagnostic entity. *J Int Med* **256**, 183-194.

[2] Mitchell AJ, Shiri-Feshki M (2009) Rate of progression of mild cognitive impairment to dementia - meta-analysis of 41 robust inception cohort studies. *Acta Psychiatric Scand* **119**, 252-265.

[3] Fagan AM, Roe CM, Xiong C, Mintun MA, Morris JC, Holtzman DM (2007) Cerebrospinal fluid tau/beta-amyloid(42) ratio as a prediction of cognitive decline in nondemented older adults. *Arch Neurol* **64**, 343-349.

[4] Mattsson N, Zetterberg H, Hansson O, Andreasen N, Parnetti L, Jonsson M, Herukka SK vdFW, Blankenstein MA, Ewers M, Rich K, Kaiser E, Verbeek M, Tsolaki M, Mulugeta E, Rosén E, Aarsland D, Visser PJ, Schröder J, Marcusson J, de Leon M, Hampel H, Scheltens P, Pirttilä T, Wallin A, Jönhagen ME, Minthon L, Winblad B, Blennow K (2009) CSF biomarkers and incipient Alzheimer disease in patients with mild cognitive impairment. *JAMA* **302**, 385-393.

[5] Shaw LM, Vanderstichele H, Knapik-Czajka M, Clark CM, Aisen PS, Petersen RC, Blennow K, Soares H, Simon A, Lewczuk P, Dean R, Siemers E, Potter W, Lee VM, Trojanowski JQ (2009) Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. *Ann Neurol* **65**, 403-413.

[6] Engelborghs S, De Vreese K, Van de Casteele T, Vanderstichele H, Van Everbroeck B, Cras P, Martin JJ, Vanmechelen E, De Deyn PP (2008) Diagnostic performance of a CSF-biomarker panel in autopsy-confirmed dementia. *Neurobiol Aging* **29**, 1143-1159.

[7] Sunderland T, Linker G, Mirza N, Putnam KT, Friedman DL, Kimmel LH, Bergeson J, Manetti GJ, Zimmermann M, Tang B, Bartko JJ, Cohen RM (2003) Decreased beta-amyloid1-42 and increased tau levels in cerebrospinal fluid of patients with Alzheimer disease. *JAMA* **289**, 2094-2103.

[8] Visser PJ, Verhey F, Knol DL, Scheltens P, Wahlund LO, Freund-Levi Y, Tsolaki M, Minthon L, Wallin AK, Hampel H, Burger K, Pirttila T, Soininen H, Rikkert MO, Verbeek MM, Spiru L, Blennow K (2009) Prevalence and prognostic value of CSF markers of Alzheimer's disease pathology in patients with subjective cognitive impairment or mild cognitive impairment in the DESCRIPA study: A prospective cohort study. *Lancet Neurol* **8**, 619-627.

[9] Blennow K, Hampel H, Weiner M, Zetterberg H (2010) Cerebrospinal fluid and plasma biomarkers in Alzheimer disease. *Nat Rev Neurol* **6**, 131-144.

[10] Snider BJ, Fagan AM, Roe C, Shah AR, Grant EA, Xiong C, Morris JC, Holtzman DM (2009) Cerebrospinal fluid biomarkers and rate of cognitive decline in very mild dementia of the Alzheimer type. *Arch Neurol* **66**, 638-645.

[11] Jack CR Jr, Petersen RC, Xu YC, Waring SC, O'Brien PC, Tangalos EG, Smith GE, Ivnik RJ, Kokmen E (1997) Medial temporal atrophy on MRI in normal aging and very mild Alzheimer's disease. *Neurology* **49**, 786-794.

[12] Bobinski M, de Leon MJ, Wegiel J, Desanti S, Convit A, Saint Louis LA, Rusinek H, Wisniewski HM (2000) The histological validation of post mortem magnetic resonance imaging-determined hippocampal volume in Alzheimer's disease. *Neuroscience* **95**, 721-725.

[13] Vemuri P, Whitwell JL, Kantarci K, Josephs KA, Parisi JE, Shiung MS, Knopman DS, Boeve BF, Petersen RC, Dickson DW, Jack CR Jr (2008) Antemortem MRI based Structural Abnormality Index (STAND)-scores correlate with postmortem Braak neurofibrillary tangle stage. *Neuroimage* **42**, 559-567.

[14] McDonald CR, Gharapetian L, McEvoy LK, Fennema-Notestine C, Hagler DJ Jr, Holland D, Dale AM, ADNI (2012) Relationship between regional atrophy rates and cognitive decline in mild cognitive impairment. *Neurobiol Aging* **33**, 242-253.

[15] Vemuri P, Wiste HJ, Weigand SD, Shaw LM, Trojanowski JQ, Weiner MW, Knopman DS, Petersen RC, Jack CR Jr (2009) MRI and CSF biomarkers in normal, MCI, and AD subjects: Predicting future clinical change. *Neurology* **73**, 294-301.

[16] Risacher SL, Shen L, West JD, Kim S, McDonald BC, Beckett LA, Harvey DJ, Jack CR Jr, Weiner MW, Saykin AJ, ADNI (2010) Longitudinal MRI atrophy biomarkers: Relationship to conversion in the ADNI cohort. *Neurobiol Aging* **31**, 1401-1418.

[17] Jack CR Jr, Shiung MM, Weigand SD, O'Brien PC, Gunter JL, Boeve BF, Knopman DS, Smith GE, Ivnik RJ, Tangalos EG, Petersen RC (2005) Brain atrophy rates predict subsequent clinical conversion in normal elderly and amnestic MCI. *Neurology* **65**, 1227-1231.

[18] Jack CR Jr, Petersen RC, Xu YC, O'Brien PC, Smith GE, Ivnik RJ, Boeve BF, Waring SC, Tangalos EG, Kokmen E (1999) Prediction of AD with MRI-based hippocampal volume in mild cognitive impairment. *Neurology* **52**, 1397-1403.

[19] Mosconi L, Berti V, Glodzik L, Pupi A, De Santi S, de Leon MJ (2010) Pre-clinical detection of Alzheimer's disease using FDG-PET, with or without amyloid imaging. *J Alzheimers Dis* **20**, 843-854.

[20] Haense C, Herholz K, Jagust WJ, Heiss WD (2009) Performance of FDG PET for detection of Alzheimer's disease in two independent multicentre samples (NEST-DD and ADNI). *Dement Geriatr Cogn Disord* **28**, 259-266.

[21] Beckett LA, Harvey DJ, Gamst A, Donohue M, Kornak J, Zhang H, Kuo JH (2010) The Alzheimer's Disease Neuroimaging Initiative: Annual change in biomarkers and clinical outcomes. *Alzheimers Dement* **6**, 257-264.

[22] Hampel H, Frank R, Broich K, Teipel SJ, Katz RG, Hardy J, Herholz K, Bokde AL, Jessen F, Hoessler YC, Sanhai WR, Zetterberg H, Woodcock J, Blennow K (2010) Biomarkers for Alzheimer's disease: Academic, industry and regulatory perspectives. *Nat Rev Drug Discov* **9**, 560-574.

[23] Vemuri P, Gunter JL, Senjem ML, Whitwell JL, Kantarci K, Knopman DS, Boeve BF, Petersen RC, Jack CR Jr (2008) Alzheimer's disease diagnosis in individual subjects using structural MR images: Validation studies. *Neuroimage* **39**, 1186-1197.

[24] Desikan RS, Cabral HJ, Settecase F, Hess CP, Dillon WP, Glastonbury CM, Weiner MW, Schmansky NJ, Salat DH, Fischl B (2010) Automated MRI measures predict progression to Alzheimer's disease. *Neurobiol Aging* **31**, 1364-1374.

[25] McEvoy LK, Edland SD, Holland D, Hagler DJ Jr, Roddey JC, Fennema-Notestine C, Salmon DP, Koyama AK, Brewer JB, PS A, Dale AM (2010) Neuroimaging enrichment strategy for secondary prevention trials in Alzheimer disease. *Alzheimer Dis Assoc Disord* **24**, 269-277.

[26] Bakkour A, Morris JC, Dickerson BC (2009) The cortical signature of prodromal AD: Regional thinning predicts mild AD dementia. *Neurology* **72**, 1048-1055.

[27] DeCarli C, Frisoni GB, Clark CM, Harvey D, Grundman M, Petersen RC, Thal LJ, Jin S, Jack CR Jr, Scheltens P (2007) Qualitative estimates of medial temporal atrophy as a predictor of progression from mild cognitive impairment to dementia. *Arch Neurol* **64**, 108-115.

[28] Hansson O, Zetterberg H, Buchhave P, Londos E, Blennow K, Minthon L (2006) Association between CSF biomarkers and incipient Alzheimer's disease in patients with mild cognitive impairment: A follow-up study. *Lancet Neurol* **5**, 228-234.

[29] Zetterberg H, Wahlund LO, Blennow K (2003) Cerebrospinal fluid markers for prediction of Alzheimer's disease. *Neurosci Lett* **352**, 67-69.

[30] DeMeyer G, Shapiro F, Vanderstichele H, Vanmechelen E, Engelborghs S, DeDeyn PO, Coart E, Hansson O, Minthon L, Zetterberg H, Blennow K, Shaw L, Trojanowski JQ, ADNI (2010) Diagnosis-independent Alzheimer disease biomarker signature in cognitively normal elderly people. *Arch Neurol* **67**, 949-956.

[31] Drzezga A, Grimmer T, Riemenschneider M, Lautenschlager N, Siebner H, Alexopoulus P, Minoshima S, Schwaiger M, Kurz A (2005) Prediction of individual clinical outcome in MCI by means of genetic assessment and (18)F-FDG PET. *J Nucl Med* **46**, 1625-1632.

[32] Mosconi L, Perani D, Sorbi S, Herholz K, Nacmias B, Holthoff V, Salmon E, Baron JC, De Cristofaro MT, Padovani A, Borroni B, Franceschi M, Bracco L, Pupi A (2004) MCI conversion to dementia and the APOE genotype: A prediction study with FDG-PET. *Neurology* **63**, 2332-2340.

[33] Stone DJ, Molony C, Suver C, Schadt EE, Potter WZ (2010) ApoE genotyping as a progression-rate biomarker in phase II disease-modification trials for Alzheimer's disease. *Pharmacogenomics J* **10**, 161-164.

[34] Brys M, Glodzik L, Mosconi L, Switalski R, De Santi S, Pirraglia E, Rich K, Kim BC, Mehta P, Zinkowski R, Pratico D, Wallin A, Zetterberg H, Tsui WH, Rusinek H, Blennow K, de Leo MJ (2009) Magnetic resonance imaging improves cerebrospinal fluid biomarkers in the early detection of Alzheimer's disease. *J Alzheimers Dis* **16**, 351-362.

[35] Galluzzi S, Geroldi C, Ghidoni R, Paghera B, Amicucci G, Bonetti M, Zanetti O, Cotelli M, Gennarelli M, Frisoni GB, TOMCWG (2010) The new Alzheimer's criteria in a naturalistic series of patients with mild cognitive impairment. *J Neurol* **257**, 2004-2014.

[36] Hinrichs C, Singh V, Xu G, Johnson SC, ADNI (2011) Predictive markers for AD in a multi-modality framework: An analysis of MCI progression in the ADNI population. *NeuroImage* **55**, 574-589.

[37] Walhovd KB, Fjell AM, Brewer J, McEvoy LK, Fennema-Notestine C, Hagler DJ, Jennings RG Jr, Karow D, Dale AM (2010) *ADNI*, Combining MR imaging, positron-emission tomography, and CSF biomarkers in the diagnosis and prognosis of Alzheimer disease. *Am J Neuroradiol* **31**, 347-354.

[38] Landau SM, Harvey D, Madison CM, Reiman EM, Foster NL, Aisen PS, Petersen RC, Shaw LM, Trojanowski JQ, Jack CR Jr, Weiner MW, Jagust WJ, ADNI (2010) Comparing predictors of conversion and decline in mild cognitive impairment. *Neurology* **75**, 230-238.

[39] Zhang D, Wang Y, Zhou L, Yuan H, Shen D, ADNI (2011) Multimodal classification of Alzheimer's disease and mild cognitive impairment. *NeuroImage* **55**, 856-867.

[40] Cui Y, Liu B, Luo S, Zhen X, Fan M, Liu T, Zhu W, Park M, Jiang T, Jin J, ADNI (2011) Identification of conversion from mild cognitive impairment to Alzheimer's disease using multivariate predictors. *PLoS ONE* **6**, e21896.

[41] Holland D, Brewer JB, Hagler DJ, Fennema-Notestine C, Dale AM, ADNI (2009) Subregional neuroanatomical change as a biomarker for Alzheimer's disease. *Proc Natl Acad Sci U S A* **106**, 20954-20959.

[42] Landau SM, Harvey D, Madison CM, Koeppe RA, Reiman EM, Foster NL, Weiner MW, Jagust WJ, ADNI (2011) Associations between cognitive, functional, and FDG-PET measures of decline in AD and MCI. *Neurobiol Aging* **32**, 1207-1218.

[43] Chen K, Langbaum JB, Fleisher AS, Ayutyanont N, Reschke C, Lee W, Liu X, Bandy D AG, Thompson PM, Foster NL, Harvey DJ, de Leon MJ, Koeppe RA, Jagust WJ, Weiner MW, Reiman EM, Alzheimer's Disease Neuroimaging EM, Initiative (2010) Twelve-month metabolic declines in probable Alzheimer's disease and amnestic mild cognitive impairment assessed using an empirically pre-defined statistical region-of-interest: Findings from the Alzheimer's Disease Neuroimaging Initiative. *Neuroimage* **51**, 654-664.

[44] Qi Y, Minka TP, Picard RW, Ghahramani Z (2004) Predictive automatic relevance determination by expectation propagation. *Proceedings of Twenty-first International Conference on Machine Learning*, p. 85. doi: 10.1145/1015330.1015418.

[45] Jack CR Jr, Knopman DS, Jagust WJ, Shaw LM, Aisen PS, Weiner MW, Petersen RC, Trojanowski JQ (2010) Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *Lancet Neurol* **9**, 119-128.

[46] Karow DS, McEvoy LK, Fennema-Notestine C, Hagler DJ Jr, Jennings RG, Brewer JB, Hoh CK, Dale AM (2010) Relative capability of MR imaging and FDG PET to depict changes associated with prodromal and early Alzheimer disease. *Radiology* **256**, 932-942.

[47] Teipel SJ, Born C, Ewers M, Bokde AL, Reiser MF, Möller HJ, Hampel H (2007) Multivariate deformation-based analysis of brain atrophy to predict Alzheimer's disease in mild cognitive impairment. *Neuroimage* **38**, 13-24.

[48] Andreasen N, Blennow K (2005) CSF biomarkers for mild cognitive impairment and early Alzheimer's disease. *Clin Neurol Neurosurg* **107**, 165-173.

[49] Buchhave P, Minthon L, Zetterberg H, Wallin AK, Blennow K, Hansson O (2012) Cerebrospinal fluid levels of beta-amyloid 1-42, but not of tau, are fully changed already 5 to 10 years before the onset of Alzheimer dementia. *Arch Gen Psychiatry* **69**, 98-106.