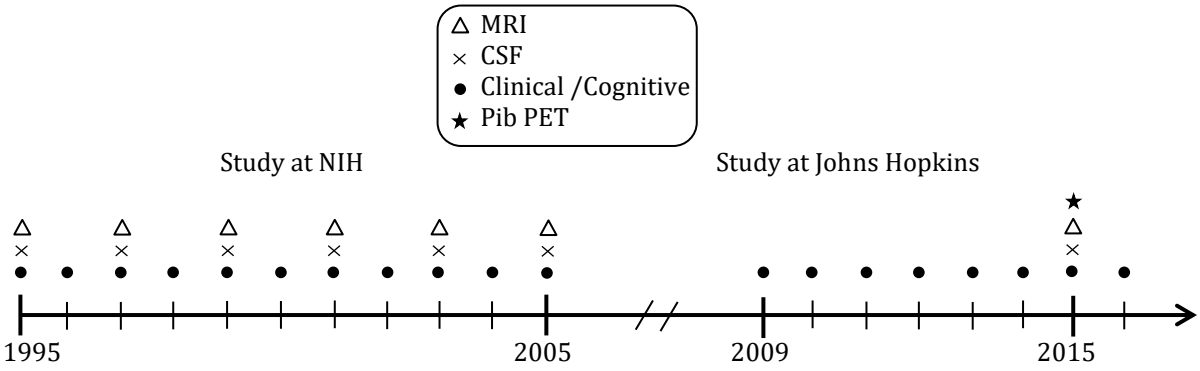
Supplementary Materials

Table of Contents

Figure 1- Schematic Representation of Study Design



Section 1 – Reasons for Exclusion of Subjects from Analyses

Subjects were excluded from analyses for the following reasons: (1) subjects had not yet re-enrolled in the study or had withdrawn (n=28); (2) the estimated age of onset of clinical symptoms was determined to be at or prior to baseline, based on the report of the subject and an informant (n=14); (3) subjects had no CSF data at baseline (n= 42); (4) subjects had no MRI data at baseline (n=23); (5) subjects were missing the baseline cognitive variables included in the analyses (n=18).

Section 2 – Details of Diagnostic Procedures

Each subject included in these analyses received a consensus diagnosis by the staff of the BIOCARD Clinical Core at Johns Hopkins. This research team included: neurologists, neuropsychologists, research nurses and research assistants. During each study visit, each subject had received a comprehensive cognitive assessment and a Clinical Dementia Rating (CDR), as well as a comprehensive medical evaluation (including a medical, neurologic and psychiatric

assessment). For the cases with evidence of clinical or cognitive dysfunction, a clinical summary was prepared that included information about demographics, family history of dementia, work history, past history of medical, psychiatric and neurologic disease, current medication use and results from the neurologic and psychiatric evaluation at the visit. The reports of clinical symptoms from the CDR interview with the subject and collateral source (e.g., spouse, child, friend) were summarized, and the results of the neuropsychological testing were reviewed (see Albert et al., 2014 for the complete battery).

The diagnostic process for each case was handled in a similar manner. Two sources of information were used to determine if the subject met clinical criteria for the syndromes of MCI or dementia: (1) the CDR interview conducted with the subject and the collateral source was used to determine if there was evidence that the subject was demonstrating changes in cognition in daily life, (2) cognitive tests scores (and their comparison to established norms) were used to determine if there was evidence of significant decline in cognitive performance over time. If a subject was deemed to be impaired, the decision about the likely etiology of the syndrome was based on the medical, neurologic, and psychiatric information collected at each visit, as well as medical records obtained from the subject, where necessary. More than one etiology could be endorsed for each subject (e.g., Alzheimer's disease and vascular disease). One of four possible diagnostic categories was selected at each visit for each subject: (1) Normal, (2) Mild Cognitive Impairment, (3) Impaired Not MCI or (4) Dementia. The decision about the estimated age of onset of clinical symptoms was determined separately, and was based on responses from the subject and collateral source during the CDR interview regarding approximately when the relevant clinical symptoms began to develop. These diagnostic procedures are comparable to those implemented by the Alzheimer's Disease Centers program supported by the National Institute on Aging.

Within the context of this study, the diagnosis of Impaired Not MCI typically reflected contrasting information from the CDR interview and the cognitive test scores (i.e., the subject or collateral source expressed concerns about cognitive changes in daily life but the cognitive testing did not show changes, or visa versa, the test scores provided evidence for declines in cognition but neither the subject nor the collateral source reported changes in daily life).

Albert M, Soldan A, Gottesman R, et al. Cognitive changes preceding clinical symptom onset of mild cognitive impairment and relationship to ApoE genotype. Current Alzheimer research. 2014; 11:773-784.

Section 3 – CSF Analytic Procedures

The CSF specimens were analyzed using the AlzBio3 kit. The AlzBio3 kit contains monoclonal antibodies specific for Aβ1-42 (4D7A3), t-tau (AT120), and p-tau181p (AT270), each chemically bonded to unique sets of color-coded beads, and analyte-specific detector antibodies (HT7, 3D6). Calibration curves were produced for each biomarker using aqueous buffered solutions that contained the combination of the three biomarkers at concentrations ranging from 54 to 1,799 pg/ml for synthetic Aβ1-42 peptide, 25 to 1,555 pg/ml for recombinant tau, and 15 to 258 pg/ml for a tau synthetic peptide phosphorylated at the threonine 181 position (i.e., the p-

tau181p standard). Each subject had all samples (run in triplicate) analyzed on the same plate. The intra-assay coefficients of variation (CV) for plates used in this study were: 7.7% +/- 5.3 ($A\beta_{1-42}$); 7.1% +/- 4.9 (tau); 6.3% +/- 4.8 (p-tau_{181}). Interassay (plate-to-plate) CVs for a single CSF standard run on all plates used in this study were: 8.9% +/- 6.5 ($A\beta_{1-42}$); 4.7% +/-3.3 (tau), and 4.3% +/- 3.18 (p-tau_{181}). Compared with studies using the same kits and platforms, our absolute results are at the median levels for $A\beta_{1-42}$, tau, and p-tau_{181}. The CVs, plate-to-plate variability, and the dynamic range of our assays are well within published norms (Shaw et al., 2009; Mattson et al., 2009).

Shaw L, Vanderstichele H, Knapik-Czajika M, et al. Alzheimer's Disease Neuroimaging Initiative. Cerebrospinal fluid biomarker signature in Alzheimer's Disease Neuroimaging Initiative subjects. Ann Neuol. 2009; 65: 403-413.

Mattson N, Zetterberg H, Hansson O, et al. CSF biomarkers and incipient Alzheimer disease in patients with mild cognitive impairment. JAMA. 2009; 302: 385-393.

Section 4 – MRI Analytic Procedures

The MRI volumetric regions of interest (ROI) included the entorhinal cortex, hippocampus, and amygdala. ROIs were analyzed using large deformation diffeomorphic metric mapping (LDDMM). For each of the three regions of interest, landmarks were placed manually in each MRI scan to mark the boundaries of the ROI, following previously published protocols (see Csernansky *et al.,* 1998, for the hippocampus, Munn *et al.* , 2007, for the amygdala, and Miller et al., 2013 for the entorhinal cortex). Next, a group template for the entorhinal cortex, hippocampus, and amygdala was created, based on the set of baseline MRI scans. The same set of landmarks was placed into this group template as in the individual subject scans. ROI-LDDMM procedures were then used to map the group template to the individual subject scans, using both landmark matching (Joshi & Miller, 2000) and volume matching (Beg *et al.,* 2005). The resulting segmented binary images for the entorhinal cortex, hippocampus and amygdala were used to calculate the volume of each structure, by hemisphere, by summing the number of voxels within the volume.

The thickness of the entorhinal cortex was also generated. It was modeled by first generating a smooth surface from the segmented gray matter volume. The gray/white surface was then extracted from the closed surface by curvature based dynamic programming delineation of the extreme boundaries so that the surface closest to the white matter was retained (Miller *et al.*, 2013). The laminar thickness was calculated as a single parameter based on the ratio of volume/ surface-area in units of millimeters.

The volumetric measurements of the entorhinal cortex, hippocampus and amygdala were normalized for head size by including total intracranial volume (ICV) as a covariate (Sanfilipo *et al.* , 2004). ICV was calculated using coronal SPGR scans in Freesurfer 5.1.0 (Segonne *et al.* , 2004). Cortical thickness measures were not adjusted for ICV.

A recent study has compared the LDDMM methods used here, with Freesurfer, MALF and FSL and found that LDDMM provides slightly improved segmentation and less variability, using 1.5 T scans, particularly with respect to regions included in these analyses, such as the hippocampus (Tang et al., 2015).

Csernansky JG, Wang L, Swank J, Miller JP, Gado M, McKeel D, et al. Preclinical detection of Alzheimer's disease: hippocampal shape and volume predict dementia onset in the elderly. Neuroimage. 2005; 15;25(3):783-792.

Munn MA, Alexopoulos J, Nishino T, Babb CM, Flake LA, Singer T, et al. Amygdala volume analysis in female twins with major depression. Biol Psychiatry. 2007; 62(5):415-422.

Miller MI, Younes L, Ratnanather JT, Brown T, Trinh H, Postell E, et al. The diffeomorphometry of temporal lobe structures in preclinical Alzheimer's disease. Neuroimage Clin. 2013; 3:352-360.

Beg MF, Miller MI, Trouve A, Younes L. Computing metrics via geodesics on flows of diffeomorphisms. International Journal of Computer Vision. 2005; 61(2):139-157.

Joshi SC, Miller MI. Landmark matching via large deformation diffeomorphisms. IEEE Trans Image Process. 2000; 9(8):1357-1370.

Sanfilipo MP, Benedict RH, Zivadinov R, Bakshi R. Correction for intracranial volume in analysis of whole brain atrophy in multiple sclerosis: the proportion vs. residual method. Neuroimage. 2004; 22 (4):1732-1743.

Segonne F, Dale AM, Busa E, Glessner M, Salat D, Hahn HK, et al. A hybrid approach to the skull stripping problem in MRI. Neuroimage. 2004; 22(3):1060-1075.

Tang X, Crocetti D, Kutten K, Ceritoglu C, Albert M, Mori S, Mostovsky S, Miller MI. Segmentation of brain magnetic resonance images based on multi-atlas likelihood fusion: testing using data with a broad range of anatomical and photometric profiles. Frontiers in Neuroscience, 2015; 9: 61.

Section 5 – Details of Statistical Methods:

The general analytic approach for the ROC analyses can be summarized as follows: Let T be the time from study entry to onset of clinical symptom, t be the pre-fixed failure time of interest, M be the combined biomarker obtained from the proportional hazards model, and c be a cut-off point for classification. The time-dependent ROC curve shows the sensitivity against 1-specificity at various cutoff points c, where sensitivity is defined as $P(M > c \mid T = t)$, and specificity is defined as $P(M > c \mid T > t)$ (Heagerty & Zheng, 2005).

The bootstrap methods used for calculating the confidence intervals involved several steps. For every bootstrapped sample, subjects were sampled with replacement from the dataset, and the number of subjects sampled was equal to the dataset sample size (i.e., N=224). The Cox proportional hazard regression was then fit on the resampled dataset, which gives the linear predictor part of the model as the resampled optimal combined marker. This resampled marker was then used to obtain the time-dependent AUCs. We took 100,000 bootstrapped datasets and used the 2.5th and the 97.5th percentile of 100,000 resampled AUCs to construct the 95% confidence intervals (Efron & Tibshirani, 1994).

Heagerty PJ, Zheng Y. Survival model predictive accuracy and ROC curves. Biometrics. 2005; 61: 92-105.

Efron, B, Tibshirani RJ. *An Introduction to the Bootstrap*. CRC Press, 1994.

Section 6: Details of the Efficient Model at 5 Years

In an effort to learn more about the subjects who were misclassified by the Efficient Model at 5 years, we divided all of the participants into four groups, as shown in Supplementary Table 1a (below). Two of the groups consisted of subjects incorrectly classified by the Efficient Model (Groups 1 and 3) and two groups consisted of correctly classified subjects (Groups 2 and 4). Group membership was based on a cutoff value of a 'progression score' derived from the model. The 'progression score' was a linear combination of the weights for each of the variables in the Efficient Model, and the cutoff value of the progression score maximized the sum of the estimated sensitivity and specificity at 5 years. A higher value of the progression score indicated a higher hazard of developing clinical symptoms of MCI at 5 years. The four groups of participants were therefore defined as follows:

Group 1: participants who had progression scores lower than the cutoff value, but who progressed within 5 years (conceptually similar to false negative classifications);
Group 2: participants who had progression scores higher than the cutoff value, and progressed within 5 years (conceptually similar to true positive classifications);
Group 3: participants who had progression scores higher than the cutoff value, but remained normal at 5 years (conceptually similar to false positive classifications);
Group 4: participants who had progression scores lower than the cutoff value, and remained normal at 5 years (conceptually similar to true negative classifications).

The number of subjects in Group 1 was very low (N=3), precluding any strong generalizations, however, these subjects do not appear to differ in baseline age, age at symptom onset, educational attainment, sex, or ApoE-e4 genotype from subjects in Group 2. Group 3 (N=51) included many subjects (35.3%) who became symptomatic after 5 years (mean = 7.9 years to onset of symptoms, range = 5.4 to 13.3 years). In addition, compared to Group 4, participants in Group 3 tended to be slightly older at baseline and were somewhat more likely to be male. These analyses exclude n=6 individuals who were censored due to insufficient follow-up and therefore could not be classified using the ROC methods applied here. A similar pattern of

misclassifications and correct classifications were found for the Full Model (see Supplementary Table 1b).

Supplementary Table 1a

Baseline Characteristics of Participants Misclassified and Correctly Classified by the Efficient Model

	N	Baseline Age (SD)	Age at Symptom Onset (SD)	Years Education (SD)	APOE-e4 carriers	Sex (% Female)
Group 1	3	60.5 (13.5)	65.0 (13.5)	16.3 (3.1)	66.7%	66.7%
Group 2	15	64.4 (9.3)	67.1 (9.3)	16.6 (2.5)	57.1%	60.0%
Group 3	51	62.5 (10.4)	--	16.8 (2.4)	35.3%	51.0%
Progressed Later	18	66.3 (9.4)	74.2 (9.3)	17.1 (2.2)	33.3%	55.6%
Remained Normal	33	60.4 (10.5)	--	16.6 (2.6)	36.4%	48.5%
Group 4	149	55.4 (6.6)	--	17.3 (2.6)	31.5%	65.1%

Supplementary Table 1b

Baseline Characteristics of Participants Misclassified and Correctly Classified by the Full Model

	N	Baseline Age (SD)	Age at Symptom Onset (SD)	Years of Education (SD)	APOE-e4 Carriers	Sex (% Female)
Group 1	3	60.5 (13.5)	65.0 (13.5)	16.3 (3.1)	66.7%	66.7%
Group 2	15	64.4 (9.0)	67.1 (9.3)	16.6 (2.5)	60.0%	60.0%
Group 3	53	61.4 (10.8)	-	16.8 (2.4)	37.7%	50.9%
Progressed Later	18	64.7 (11.7)	72.7 (11.6)	16.9 (2.5)	27.8%	50.0%
Remained Normal	35	59.6 (10.1)	-	16.8 (2.5)	42.9%	51.4%
Group 4	147	55.6 (6.7)	-	17.2 (2.6)	34.0%	65.3%

Supplementary Table 1c

Model with ApoE-4 status entered at the first step and the CSF p-tau/Abeta ratio entered at the last step

	Variable	**Hazard Ratio of Model (95% CI)**	**Hazard Ratio: p-value**	**AUC of Model (95% CI)**	**Time to Outcome for Model**	**Model Sensitivity**	**Model Specificity**	**Change in AUC vs Prior Step (without CSF ratio) in Model: p-value**
	ApoE-4 + Cognitive + MRI + CSF ptau/abeta ratio			0.823 (0.783, 0.892) 0.825 (0.783, 0.882) 0.813 (0.758, 0.878)	5 years 7 years 10 years	0.759 0.755 0.725	0.735 0.741 0.759	0.150 0.065 0.309
	ApoE-4*	2.042 (1.108, 3.762)	0.022					
	Paired Associates Imm.	0.643 (0.471, 0.877)	0.005					
	Digit Symbol Substitution	0.410 (0.270, 0.623)	< 0.001					
	R. Hippocampus Vol.	0.183 (0.050, 0.665)	0.010					
	R. EC Thickness	0.275 (0.098, 0.773)	0.014					
	CSF P-tau/abeta	1.912 (1.490, 2.454)	< 0.001					

Supplementary Table 1d

Model with ApoE-4 status excluded and the CSF p-tau/Abeta ratio entered at the last step

	Variable	**Hazard Ratio of Model (95% CI)**	**Hazard Ratio: p-value**	**AUC of Model (95% CI)**	**Time to Outcome for Model**	**Model Sensitivity**	**Model Specificity**	**Change in AUC vs Prior Step (without CSF ratio) in Model: p-value**
	Cognitive + MRI + CSF ptau/abeta ratio			0.813 (0.771, 0.887) 0.814 (0.770, 0.871) 0.812 (0.756, 0.874)	5 years 7 years 10 years	0.751 0.763 0.713	0.730 0.718 0.770	0.159 0.072 0.143
	Paired Associates Imm.	0.655 (0.478, 0.899)	0.009					
	Digit Symbol Substitution	0.411 (0.269, 0.627)	< 0.001					
	R. Hippocampus Vol.	0.187 (0.051, 0.680)	0.011					
	R. EC Thickness	0.293 (0.104, 0.824)	0.020					
	CSF P-tau/abeta	1.332 (1.138, 1.558)	< 0.001					

Section 7 – ROC Results for Model with CSF Abeta Entered First

The model in which CSF Abeta was included first, immediately after the demographic variables, was designed to emulate the situation in which amyloid imaging might be used to screen subjects for inclusion in a clinical trial aimed at randomizing only those who were amyloid positive (such as in the A4 Study). The order of the variables in the model was as follows: demographics, CSF Abeta, ApoE status, cognitive test scores, MRI measures, and CSF p-tau. (See Table T1 for the hazard ratios, AUCs, sensitivities and specificities for each set of variables in the model.)

To determine whether adding each domain in the model made a significant difference in predictability, we examined the p-values comparing the AUCs for sets of models, which incrementally added the variables above. The sets of models compared were as follows: (1) Abeta vs [Abeta, ApoE-4], which were not significant (p = 0.195, 0.207, 0.825 at 5, 7 and 10 years post-baseline, respectively) (2) [Abeta, ApoE-4] vs [Abeta, ApoE-4, cognitive], which were all significant (p = 0.006, 0.001, 0.008 at 5, 7 and 10 years post-baseline, respectively); (3) [Abeta, ApoE-4, cognitive] vs [Abeta, ApoE-4, cognitive, MRI] which was significant at 5 years post-baseline (p = 0.054) but not at 7 and 10 years post-baseline (p = 0.081 and 0.170, respectively); (4) [Abeta, ApoE-4, cognitive, MRI] vs [Abeta, ApoE-4, cognitive, MRI, CSF p-tau] which were significant at 5 and 7 years post-baseline (p = 0.035 and 0.044, respectively), but not at 10 years post-baseline (p = 0.070).

Supplementary Table 2

Increment in Prediction of Progression from Normal Cognition to MCI for Models in which CSF Abeta was Entered Immediately After the Demographic Variables

	Variable	Hazard Ratio of Model (95% CI)	Hazard Ratio: p-value	AUC of Model (95% CI)	Time to Outcome for Model	Model Sensitivity	Model Specificity
1	**CSF Abeta**			0.707 (0.652, 0.791) 0.698 (0.649, 0.764) 0.724 (0.665, 0.788)	5 years 7 years 10 years	0.636 0.630 0.622	0.665 0.659 0.713
	Abeta	0.664 (0.509, 0.865)	0.002				
2	**CSF Abeta + ApoE-4**			0.718 (0.662, 0.803) 0.709 (0.658, 0.784) 0.721 (0.665, 0.786)	5 years 7 years 10 years	0.618 0.627 0.573	0.700 0.676 0.747
	Abeta	0.686 (0.522, 0.902)	0.007				
	ApoE-4*	1.592 (0.860, 2.945)	0.139				
3	**CSF Abeta + ApoeE-4 + Cognitive**			0.796 (0.756, 0.865) 0.802 (0.759, 0.862) 0.796 (0.746, 0.859)	5 years 7 years 10 years	0.727 0.731 0.703	0.725 0.729 0.747
	Abeta	0.616 (0.453, 0.838)	0.002				
	ApoE-4*	1.605 (0.856, 3.007)	0.140				
	Paired Associates Imm.	0.608 (0.467, 0.790)	< 0.001				
	Digit Symbol Substitution	0.544 (0.380, 0.778)	0.001				
4	**CSF Abeta + ApoE-4 + Cognitive + MRI**			0.825 (0.789, 0.890) 0.822 (0.786, 0.879) 0.812 (0.763, 0.877)	5 years 7 years 10 years	0.755 0.751 0.738	0.740 0.741 0.747

	Abeta	0.600 (0.433, 0.833)	0.002				
	ApoE-4*	1.703 (0.919, 3.159)	0.091				
	Paired Associates Imm.	0.661(0.506, 0.863)	0.002				
	Digit Symbol Substitution	0.506 (0.354, 0.721)	< 0.001				
	R. Hippocampus Vol.	0.791 (0.600, 1.044)	0.098				
	R EC Thickness	0.655 (0.482, 0.890)	0.007				
5	**CSF Abeta + ApoE4 + Cognitive + MRI + CSF p-tau**			0.850 (0.807, 0.913) 0.843 (0.803, 0.897) 0.831 (0.781, 0.890)	5 years 7 years 10 years	0.804 0.815 0.764	0.740 0.724 0.759
	Abeta	0.785 (0.572, 1.077)	0.133				
	ApoE-4*	1.904 (1.024, 3.541)	0.042				
	Paired Associates Imm.	0.617 (0.469, 0.812)	0.001				
	Digit Symbol Substitution	0.454 (0.315, 0.655)	< 0.001				
	R. Hippocampus Vol.	0.699 (0.526, 0.930)	0.014				
	R. EC Thickness	0.594 (0.429, 0.821)	0.002				
	P-tau	1.779 (1.355, 2.336)	< 0.001				

* ApoE-4 is a binary variable, and thus not standardized as other continuous variables. Therefore its hazard ratio is not comparable to those of continuous variables.

Abbreviations: CI, confidence interval; AUC, Area Under the Curve; ApoE-4, apolipoprotein E-4; Imm., immediate; R., right; Vol., volume; EC, entorhinal cortex; CSF, cerebrospinal fluid; p-tau, phosphorylated tau

Note: Age and Education were entered first in each model.