ANNALS of Clinical and Translational Neurology
Open Access
ANA American Neurological Association since 1875

RESEARCH ARTICLE

Biomarkers and cognitive endpoints to optimize trials in Alzheimer's disease

Philip S. Insel[1,2], Niklas Mattsson[1,2,3], R. Scott Mackin[1,4], John Kornak[5], Rachel Nosheny[1], Duygu Tosun-Turgut[1,2], Michael C. Donohue[6,7], Paul S. Aisen[7], Michael W. Weiner[1,2] & Alzheimer's Disease Neuroimaging Initiative[a]

[1]Department of Veterans Affairs Medical Center, Center for Imaging of Neurodegenerative Diseases, San Francisco, California
[2]Department of Radiology and Biomedical Imaging, University of California, San Francisco, California
[3]Clinical Neurochemistry Laboratory, Institute of Neuroscience and Physiology, The Sahlgrenska Academy at the University of Gothenburg, Mölndal, Sweden
[4]Department of Psychiatry, University of California, San Francisco, California
[5]Department of Epidemiology and Biostatistics, University of California, San Francisco, California
[6]Division of Biostatistics and Bioinformatics, Department of Family and Preventive Medicine, University of California, San Diego, California
[7]Department of Neurosciences, University of California, San Diego, California

Correspondence
Philip S. Insel, Department of Veterans Affairs Medical Center, Center for Imaging of Neurodegenerative Diseases, San Francisco, CA 94121. Tel: +1 858 652 8480; Fax: +1 415 668 2864; E-mail: philipinsel@gmail.com

Funding Information
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: Alzheimer's Association; Alzheimer's Drug Discovery Foundation; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; GE Healthcare; Innogenetics, N.V.; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Synarc Inc.; and Takeda Pharmaceutical 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

Abstract

Objective: To find the combination of candidate biomarkers and cognitive endpoints to maximize statistical power and minimize cost of clinical trials of healthy elders at risk for cognitive decline due to Alzheimer's disease. **Methods**: Four-hundred and twelve cognitively normal participants were followed over 7 years. Nonlinear methods were used to estimate the longitudinal trajectories of several cognitive outcomes including delayed memory recall, executive function, processing speed, and several cognitive composites by subgroups selected on the basis of biomarkers, including *APOE-ε4* allele carriers, cerebrospinal fluid biomarkers ($A\beta_{42}$, total tau, and phosphorylated tau), and those with small hippocampi. **Results**: Derived cognitive composites combining Alzheimer's Disease Assessment Scale (ADAS)-cog scores with additional delayed memory recall and executive function components captured decline more robustly across biomarker groups than any measure of a single cognitive domain or ADAS-cog alone. Substantial increases in power resulted when including only participants positive for three or more biomarkers in simulations of clinical trials. **Interpretation**: Clinical trial power may be improved by selecting participants on the basis of amyloid and neurodegeneration biomarkers and carefully tailoring primary cognitive endpoints to reflect the expected decline specific to these individuals.

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. This research was also supported by National Institutes of Health grants P30 AG010129 and K01 AG030514.

Received: 25 January 2015; Revised: 8 February 2015; Accepted: 9 February 2015

Annals of Clinical and Translational Neurology **2015; 2(5): 534–547**

doi: 10.1002/acn3.192

[a]Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (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.

Introduction

Several recent clinical trials targeting β-amyloid (Aβ) pathology in Alzheimer's disease (AD) have shown evidence of reducing the accumulation of Aβ plaques, but without proving a slowing of cognitive decline.[1–3] One possible explanation of these failures is that a treatment targeting Aβ deposition, an early-stage process, may have little effect on patients in later stages when only a weak correlation between plaques and cognition remains.[4] In later stages of AD, measures of neurofibrillary tangles and neuronal loss, rather than Aβ pathology, are associated with continued cognitive worsening, suggesting that although plaques may initiate the disease process, it is tau pathology and atrophy that drive cognitive decline once AD has been diagnosed.[5–7] These failures led to a wave of trials that attempt anti-Aβ treatment in early stages, prior to the onset of cognitive symptoms. Identifying a successful treatment requires demonstrating a slowing of cognitive decline, a task that becomes difficult and costly when the target population has yet to demonstrate decline, and may not decline in the near future.[8] Minimizing costs of such trials will require accurately predicting future deterioration in a currently asymptomatic population as well as identifying clinical endpoints that capture the earliest evidence of cognitive decline.

To identify at-risk individuals, The A4 Study, a trial of solanezumab, an anti-Aβ treatment, recruits cognitively normal participants with Aβ-positive positron emission tomography (PET) scans.[9] Another trial, The Alzheimer's Prevention Initiative APOE4 Treatment Trial will recruit cognitively normal subjects homozygous for the *APOE-ε4* allele, a genetic feature increasing the risk of Aβ deposition and cognitive decline.[10] Other inclusion criteria, depending on the target of treatment, could be based on alternative histological features of AD including tau pathology or hippocampal atrophy.[11] Recruiting participants for combinations of pathologies such as amyloid, tau and/or genetic markers may increase the likelihood of near-term cognitive decline, consequently reducing sample sizes and trial costs.

Identifying the endpoint most sensitive to the earliest cognitive changes is another challenge. Recent guidance from the FDA endorses the need to identify optimal cognitive endpoints for trials in these populations.[12] Delayed memory recall, a domain repeatedly shown to decline early, should comprise a substantial portion of any cognitive composite to capture changes of a cognitively normal population.[13–16] Several nonmemory domains also shown to decline early include executive function, attention, and pro-

cessing speed.[17–22] Global composites that assess multiple cognitive domains, such as the Alzheimer's Disease Assessment Scale (ADAS-cog), have also been shown to capture early decline.[23–26] An optimized trial design will include endpoints tailored to the domains expected to decline based on the biomarkers used for study inclusion criteria.

Our primary goal in this analysis was to identify the combination of biomarker-based inclusion criteria and cognitive endpoints that will demonstrate the most longitudinal decline, in order to maximize power of treatment trials in a cognitively normal population. We consider several well-established biomarkers associated with cognitive decline: *APOE-ε4* allele, low cerebrospinal fluid (CSF) Aβ_{42}, high total tau (t-tau) and phosphorylated tau (p-tau), and small hippocampal volume; and cognitive measures associated with early change: delayed memory recall, executive function, processing speed, and global composites.

Methods

Participants

Data were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). ADNI is the result of efforts of many coinvestigators, and subjects have been recruited from over 50 sites across the US and Canada (see www.adni-info.org). The population in this study included ADNI-1 and ADNI-2 participants who were classified as cognitively normal controls at screening, were tested for CSF biomarkers and presence of the *APOE-ε4* allele, had successful baseline FreeSurfer, The General Hospital Corporation, Boston MA, USA processing of MR images, and completed a battery of neuropsychological exams.

CSF biomarker concentrations

Each CSF sample was collected at study baseline by lumbar puncture, and shipped on dry ice to the ADNI Biomarker Core laboratory at the University of Pennsylvania Medical Center for long-term storage at −80°C. CSF biomarkers were measured using the multiplex xMAP Luminex platform (Luminex Corp, Austin, TX) with the Research Use Only INNOBIA AlzBio3 kit (Fujirebio/Innogenetics, Ghent, Belgium).[27,28] Passing-Bablok regression was used to anchor CSF samples from the ADNI2 and GO cohort to samples from the ADNI1 cohort (full details: http://adni.bitbucket.org/upennbiomk5.html).

MRI acquisition and processing

At each site, ADNI-1 participants underwent the standardized 1.5 T MRI protocol of ADNI. Image quality and preprocessing were performed at a designated MRI Center, as described in Jack et al.[29] Similarly, ADNI-2 participants underwent the standardized 3T MRI protocol of ADNI-2. Detailed descriptions of both protocols can be found at www.adni-info.org.

Cognitive outcomes

Cognitive measures assessed included the 11-item Alzheimer's Disease Assessment Scale-cognitive subscale (ADAS11),[30] delayed memory recall from the Wechsler Memory Scale (Logical Memory II),[31] delayed Rey Auditory Verbal Learning Test (dAVLT),[32] Digit Symbol Substitution,[31] and the Trail Making Test part B.[33]

Statistical analysis

At baseline, each participant was categorized as positive or negative for each biomarker, including *APOE-ε4* (*APOE+* was defined as the presence of at least one *APOE-ε4* allele), CSF Aβ_{42}, t-tau, p-tau, and hippocampal volume. Participants were considered positive for CSF Aβ_{42} (Aβ+) if CSF Aβ_{42} was less than the a priori threshold 192 ng/L.[27] Participants were considered biomarker-positive for t-tau (t-tau+) or p-tau (p-tau+) if they were in the highest quartile of each respective measure, while the hippocampal volume positivity (hippocampus+) threshold was defined to be the lowest quartile, after left and right hemispheres were averaged and the effect of intracranial volume had been removed via residualization.[34] ADNI-1 and ADNI-2 participants were categorized separately for hippocampal positivity, due to the methodological differences between ADNI-1 and ADNI-2.

In subsequent analyses, these binary categories were used to further classify participants into multiple pathology groups: participants who proved negative for all five biomarkers (group 0), participants who proved positive for only one biomarker (group 1), positive for two biomarkers (group 2), and participants who proved positive for three, four, or five biomarkers (group 3+). In a separate analysis, and because measures of Aβ are invasive (lumbar puncture) or costly (PET imaging, not used in this study due to fewer samples and shorter follow-up time), we considered inclusion criteria based on *APOE-ε4* positivity and hippocampal volume alone: *APOE+* participants with small hippocampi were compared with *APOE−* participants with larger hippocampi. Baseline associations between the biomarker groups and demographic/clinical factors were assessed using Fisher's Exact test for gender and Wilcoxon or Kruskal–Wallis tests for continuous variables (age and education).

The single and multiple pathology classifications were then used to predict longitudinal change in each of the cognitive outcomes and in two derived composites. The

two composites were (1) ADAS11, Trails B and Logical Memory II, and (2) ADAS11, Trails B, and dAVLT, differing only by type of additional delayed memory recall component. Each individual cognitive component was first standardized by mean-centering and scaling to the standard deviation of the scores. The standardized components were then summed and standardized again to form the composites.

Up to 7 years of repeated measures of each outcome were available, with the exception of Digit Symbol Substitution, which was collected for up to 4 years. Cognitive tests were administered annually with an additional test at month 6 for all measures except Logical Memory II. Longitudinal cognitive measures were modeled using linear mixed-effects regression with a random intercept and slope and an unstructured covariance matrix for the random effects. To capture departures from linearity in the trajectory of cognition, continuous time from baseline test was parameterized using restricted cubic splines.[35] For outcomes with 7 years of follow up, spline knots were placed at 1, 3, and 5 years; for Digit Symbol Substitution (4 years of follow up), knots were placed at 0.5, 1.5, and 3 years. With three knots, time is modeled with two parameters. Differences in trajectories among the pathology groups were tested using interactions between the two parameters for time and the pathology group factor, $\beta_{\text{time}_1:\text{group}} + \beta_{\text{time}_2:\text{group}}$. Pathology groups were tested within biomarker for trajectory differences, for example, *APOE*+ versus *APOE*−, and also between pairs of biomarkers, for example, *APOE*+ versus Aβ+. When testing pairs of biomarkers, groups were restricted to being positive on only one biomarker (*APOE*+ and Aβ− vs. Aβ+ and *APOE*−, in the above example). Likelihood ratio tests were used to compare models with and without interactions between biomarker and time. The main effect for group was used to compare biomarker groups at baseline. All models were adjusted for age, gender, and years of education. Significance of tests was reported using two-sided *P*-values.

The association between biomarker groups and missing data was modeled using generalized mixed-effects regression with a binomial indicator for a missing visit. Separate estimates for each biomarker group were tested to evaluate whether biomarker positivity was associated with increased odds of missing data.

We also evaluated the ability of the biomarker groups to improve in each cognitive outcome over the initial year (or 6 months for Digit Symbol Substitution) of testing. Biomarker groups were considered to improve if they demonstrated a statistically significant positive slope from baseline through month 12.

Hypothesis tests for improvement over 1 year and comparisons among different types of biomarker positivity were adjusted for multiplicity using a Hochberg correction.[36] Because of the substantial evidence in the literature demonstrating the deleterious associations between the biomarkers in this study and cognition, we did not correct for multiplicity when testing for differences between biomarker-positive versus negative, for example, *APOE*+ versus *APOE*−. Consistent with this literature, all biomarker-positive groups in this study declined more than their negative counterparts, making results unlikely due to type I error.

Finally, we compared the power to detect a hypothetical drug effect in a clinical trial scenario for each combination of biomarker and cognitive outcome. Using the estimates of change from baseline to 3, 4, 5, and 6 years for the biomarker groups and the estimates of the variance of the residual error, subject-specific intercepts and slopes, and the correlation between the intercepts and slopes, we estimated the power to detect a 25% decrease in the difference between the change in the biomarker-positive group and the biomarker-negative group. Sampling from the above estimates and assuming 750 subjects per arm, we simulated 500 longitudinal clinical trials for each biomarker/outcome combination. Power was estimated as the proportion of significant two-sided *P*-values for the drug effect, using a mixed-model repeated-measures design.[37]

Results

Cohort characteristics

Four-hundred and twelve participants from either ADNI-1 or ADNI-2 who were enrolled into cognitively normal cohorts were included in the analysis. *APOE*-ε4 allele carrier status was available for all 412 participants. Baseline hippocampal volumes were available for 402 participants, and baseline CSF Aβ_{42}, t-tau, and p-tau were available for 221, 218, and 220 participants, respectively. When limited to participants with data available for all five biomarkers ($N = 212$), results show that 74 participants were negative for all five biomarkers, 60 were positive for one biomarker, 40 for two biomarkers, and 38 for three, four, or five biomarkers. Participants in the multiple pathology groups were mostly positive for multiple CSF biomarkers or a combination of CSF biomarkers and *APOE*, whereas the largest contributor to the single pathology group was small hippocampal volume. Small sample sizes precluded subdivision of the 3+ biomarker group for further analysis. See Table S1 for frequencies of pathology combinations.

Biomarker-positive subjects tended to be older, with the exception of *APOE*-ε4 carriers. Participants with smaller hippocampi at baseline were more educated. None of

the biomarker groups was associated with gender. Baseline demographics are summarized in Table 1.

During the 7 years of follow up, 53 or 13% of participants converted to mild cognitive impairment (MCI) one of which later converted to AD, and two participants converted directly to AD. Of the 220 participants with CSF data, 6 (8.1%) from the completely biomarker-negative group converted to MCI, 10 (16.7%) from group 1, 8 (20%) from group 2, and 8 (21.1%) from group 3+.

Associations of biomarkers and cognitive outcomes at baseline

The average time to completion of Trails B at baseline was 11.9 sec longer in the Aβ+ group compared with the Aβ− group (95% CI: [2.3, 21.5], $P = 0.015$). The average score at baseline on the Digit Symbol Substitution test was 3.58 points lower in p-tau+ participants compared with p-tau− participants (95% CI: [0.22, 6.94], $P = 0.037$), and 4.81 points lower in participants positive for three or more biomarkers compared to those who were biomarker-negative (95% CI: [0.58, 9.04], $P = 0.028$). Aβ+ participants scored 0.22 standard deviations worse at baseline on composite #2 (ADAS11, Trails B, and dAVLT) compared with Aβ− participants (95% CI: [0.003, 0.43], $P = 0.047$). There were no other statistically significant associations between biomarker groups and baseline cognition.

Longitudinal associations between biomarkers and cognitive outcomes

There were no statistically significant associations between the biomarker groups and missing data over time, however, APOE+ participants were marginally more likely to miss follow-up visits (odds ratio = 1.49, $P = 0.07$) and participants with small hippocampi were marginally less likely to miss follow-up visits (odds ratio = 0.66, $P = 0.09$). Sample sizes over follow up were similar across groups with ~98% retention at month 6, 93% at month 12, 65% at month 24, 44% at month 36, 30% at month 48, 25% at month 60, 25% at month 72, and 15% at month 84. See Table S2 for exact sample sizes for all groups over time.

There were strong nonlinear trajectories over the 7 years of follow up in all measures of cognition (Fig. 1). *APOE*+ participants declined statistically significantly faster on ADAS11, dAVLT, composite #1 and composite #2, and marginally faster ($P < 0.10$) on Trails B (Table 2). Participants with small hippocampi worsened significantly faster over time on Logical Memory II, dAVLT, Trails B, composite #1 and #2, and marginally faster on ADAS11 and Digit Symbol Substitution. Aβ+ participants declined significantly faster on ADAS11, composite #1 and #2, and marginally faster on Logical Memory II. Participants positive for t-tau declined significantly faster on dAVLT and composite #2, and marginally faster on composite #1.

Table 1. Baseline demographics.

	Age		Education (years)		Gender (F)	
	Mean (SD)	*P*	Mean (SD)	*P*	*N* (%)	*P*
APOE-ε4						
− (*N* = 298)	75.3 (5.6)		16.3 (2.8)		143 (48.0)	
+ (*N* = 114)	74.2 (5.9)	0.16	16.0 (2.7)	0.24	61 (53.5)	0.32
Aβ						
− (*N* = 145)	74.5 (5.4)		16.2 (2.7)		67 (46.2)	
+ (*N* = 76)	76.5 (5.3)	<0.01	16.0 (2.8)	0.72	39 (51.3)	0.48
t-tau						
− (*N* = 163)	74.4 (5.2)		16.0 (2.9)		76 (46.6)	
+ (*N* = 55)	77.7 (5.6)	<0.01	16.3 (2.4)	0.71	29 (52.7)	0.44
p-tau						
− (*N* = 165)	74.7 (5.5)		16.0 (2.8)		82 (49.7)	
+ (*N* = 55)	76.8 (5.2)	0.03	16.5 (2.5)	0.32	23 (41.8)	0.35
Hippocampus						
− (*N* = 309)	73.9 (5.5)		16.1 (2.7)		159 (51.5)	
+ (*N* = 93)	77.9 (5.9)	<0.01	16.9 (2.7)	<0.01	42 (45.2)	0.34
Multiple pathologies						
Group 0 (*N* = 74)	73.4 (5.1)		16.0 (2.6)		34 (45.9)	
Group 1 (*N* = 60)	75.3 (5.5)		16.4 (2.8)		27 (45)	
Group 2 (*N* = 40)	74.8 (5.4)		15.7 (3.1)		23 (57.5)	
Group 3+ (*N* = 38)	78.6 (5.0)	<0.01	16.4 (2.3)	0.65	16 (42.1)	0.53

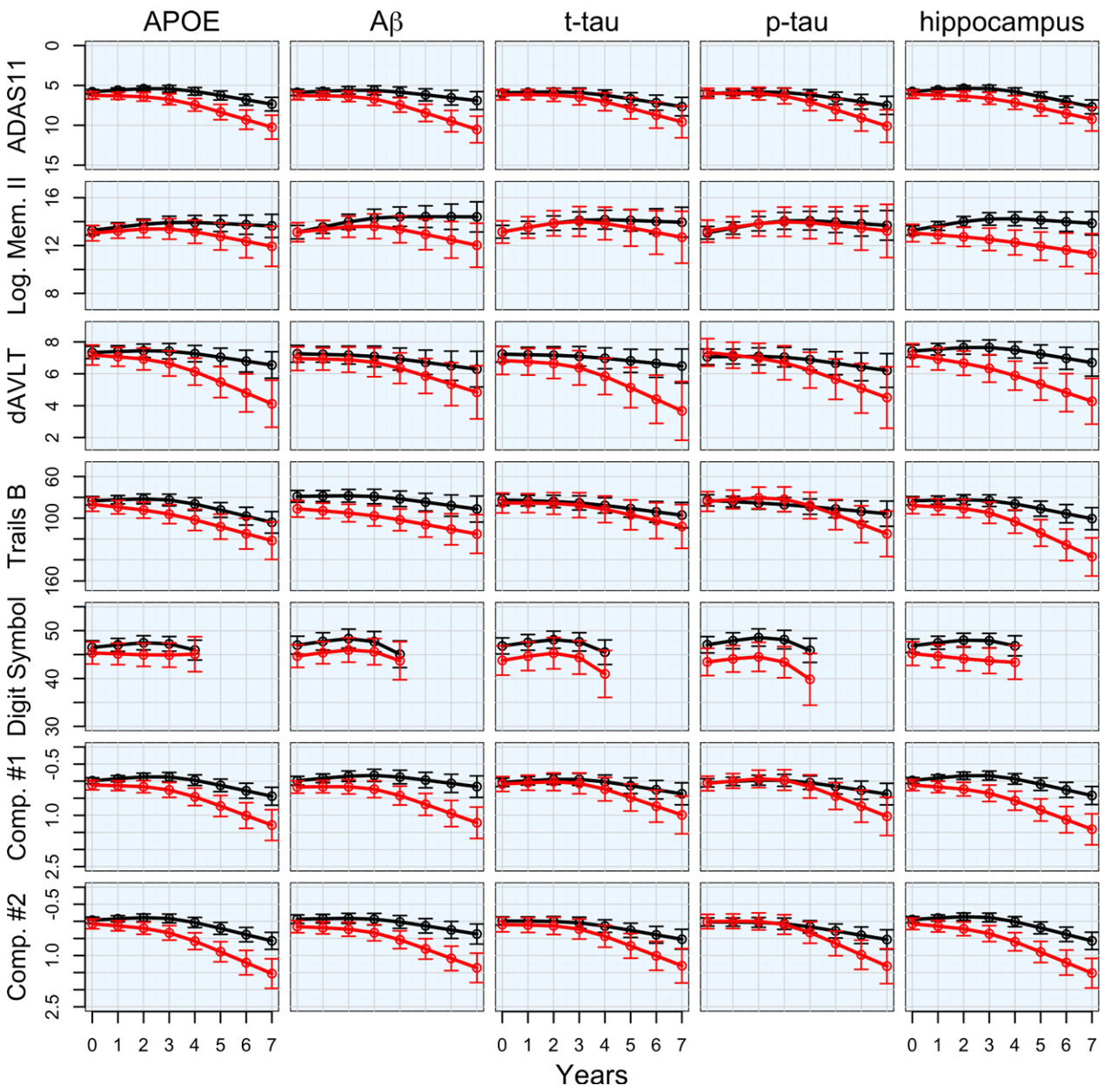

Figure 1. Plots of estimated curves for biomarker-positive (red) versus biomarker-negative (black) for all binary groups and all cognitive outcomes, over time. Estimation of curves included all participants regardless of length of follow-up time. Individual rows show the same cognitive outcome in its original scale, that is, the first row is all ADAS11. Individual columns show the same biomarker group, that is, the first column is all *APOE*+ versus *APOE*−. Composite #1 comprises ADAS11, Trails B, and Logical Memory II. Composite #2 comprises ADAS11, Trails B, and dAVLT. ADAS, Alzheimer's Disease Assessment Scale; dAVLT, delayed Rey Auditory Verbal Learning Test.

Participants positive for p-tau worsened significantly faster over time on composite #1 and #2, and marginally faster on ADAS11 and Trails B. Figure 2 shows each cognitive measure plotted within each biomarker-positive group.

The global test for any difference over time among the multiple pathology groups was significant for dAVLT, composite #1 and #2, and marginally significant for ADAS11 and Trails B, with biomarker-positive groups declining faster. The significance of the global test was driven mostly by differences between the biomarker-positive groups (1, 2, and 3+) compared with the biomarker-negative group (0). Although steeper decline was seen in the 3+ group compared with the other biomarker-positive groups, these differences were not significant after multiple comparison adjustment (Fig. 3).

Table 2. Differences in biomarker trajectories over time for cognitive outcomes.

Outcome	Biomarker (+ vs. −)	*N* participants (*N* obs.)	Trajectory difference χ^2	*P*
ADAS11	APOE	412 (2007)	9.28	0.010
	Aβ	221 (1100)	10.95	0.004
	t-tau	218 (1090)	2.46	0.293
	p-tau	220 (1097)	5.74	0.057
	Hippocampus	400 (1966)	4.68	0.100
Logical memory II	APOE	412 (1617)	2.69	0.261
	Aβ	221 (885)	5.57	0.062
	t-tau	218 (878)	1.45	0.483
	p-tau	220 (883)	0.29	0.863
	Hippocampus	402 (1582)	10.47	0.005
dAVLT	APOE	412 (2006)	8.48	0.014
	Aβ	221 (1101)	1.63	0.442
	t-tau	218 (1091)	6.15	0.046
	p-tau	220 (1098)	4.08	0.130
	Hippocampus	400 (1965)	9.99	0.007
Trails B	APOE	412 (2002)	4.68	0.096
	Aβ	221 (1097)	1.93	0.381
	t-tau	218 (1087)	0.67	0.715
	p-tau	220 (1094)	5.36	0.068
	Hippocampus	400 (1961)	11.27	0.004
Digit symbol	APOE	229 (1096)	2.10	0.350
	Aβ	114 (565)	0.64	0.727
	t-tau	114 (565)	0.23	0.892
	p-tau	114 (565)	0.71	0.702
	Hippocampus	214 (1058)	5.23	0.073
Composite #1	APOE	412 (1599)	9.04	0.011
(ADAS11, Trails B Logical Memory II)	Aβ	221 (877)	12.24	0.002
	t-tau	218 (870)	4.77	0.092
	p-tau	220 (875)	7.56	0.023
	Hippocampus	400 (1562)	14.17	0.001
Composite #2	APOE	412 (1988)	14.50	0.001
(ADAS11, Trails B, dAVLT)	Aβ	221 (1093)	11.55	0.003
	t-tau	218 (1083)	7.93	0.019
	p-tau	220 (1090)	11.60	0.003
	Hippocampus	400 (1947)	14.31	0.001

ADAS, Alzheimer's Disease Assessment Scale; dAVLT, delayed Rey Auditory Verbal Learning Test.

There were few differences when comparing biomarker positivity of different types after adjusting for a large number of comparisons (10 pairwise tests within each cognitive test), and only one remained significant. On Trails B, *APOE*+ participants (who were also p-tau−) declined significantly faster compared to p-tau+ (and *APOE*−) participants ($\chi^2 = 15.14$, $P_{ADJ} = 0.005$).

Practice effects

Slopes over the first year of follow up (between baseline and the first spline knot) were tested to identify which cognitive measures could show initial improvement over time as a result of prior exposure to the test, that is, practice or learning effects. For ADAS11 during the first year, *APOE*-ε4 noncarriers improved at a rate of 0.21 pts/yr (95% CI: [0.05, 0.37], $P_{ADJ} = 0.037$) and participants with larger hippocampi improved at a rate of 0.22 pts/yr (95% CI: [0.06, 0.38], $P_{ADJ} = 0.032$). Improving trajectories can be seen in the biomarker-negative groups in Figure 1. All pathology-negative groups improved on Logical Memory II over the first year: *APOE*− ($\beta = 0.26$, 95% CI: [0.07, 0.44], $P_{ADJ} = 0.007$), Aβ− ($\beta = 0.44$ [0.18, 0.71], $P_{ADJ} = 0.004$), t-tau− ($\beta = 0.37$ [0.12, 0.62], $P_{ADJ} = 0.007$), p-tau− ($\beta = 0.39$ [0.14, 0.64], $P_{ADJ} = 0.006$), hippocampus− ($\beta = 0.37$ [0.19, 0.56], $P_{ADJ} < 0.001$). Similarly, all pathology-negative groups improved on Digit Symbol Substitution over the initial 6 months: *APOE*− ($\beta = 1.06$ [0.28, 1.84], $P_{ADJ} = 0.016$), Aβ− ($\beta = 1.69$ [0.66, 2.72], $P_{ADJ} = 0.007$), t-tau− ($\beta = 1.49$ [0.56, 2.42], $P_{ADJ} = 0.005$),

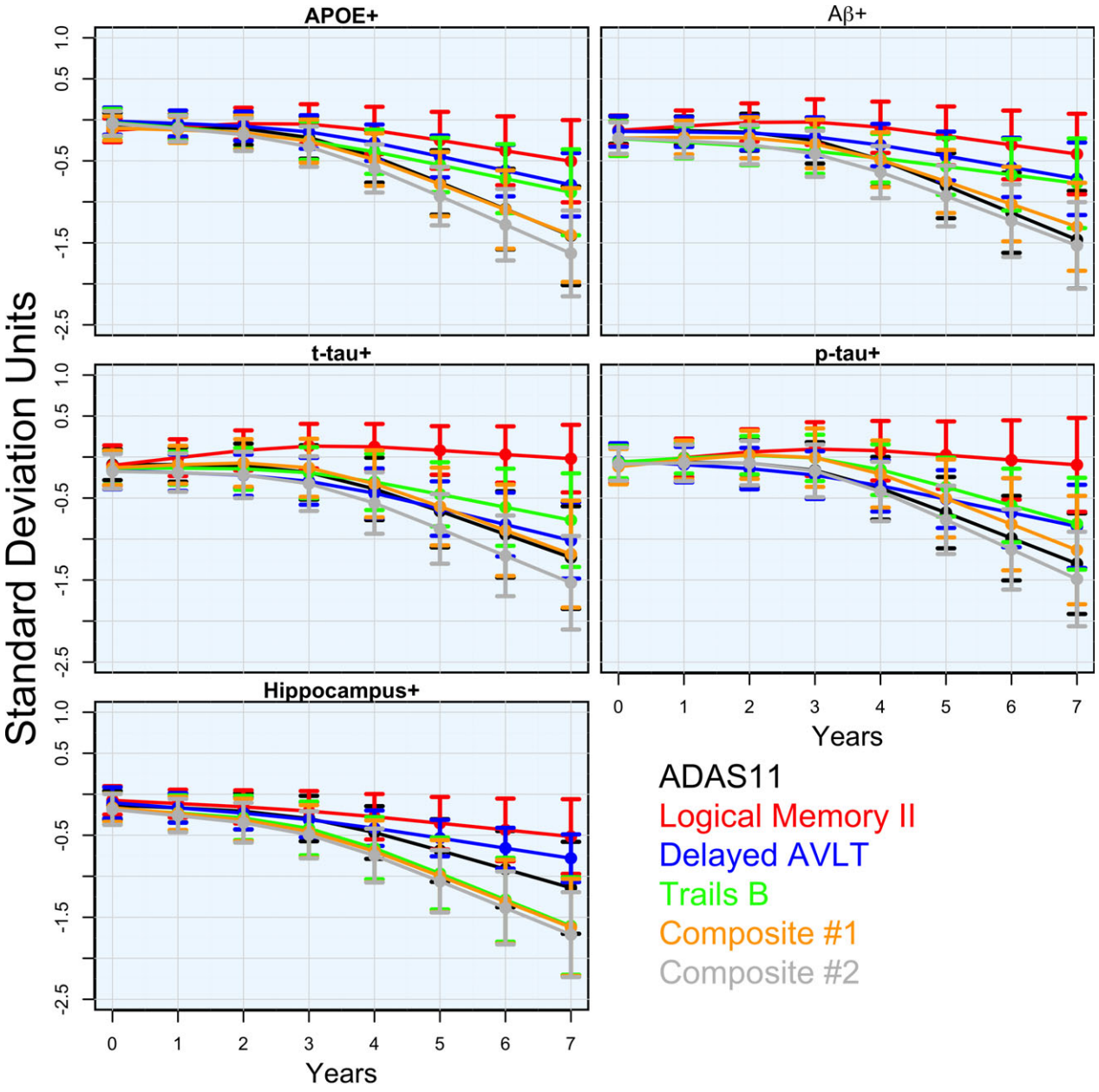

Figure 2. Comparison of cognitive outcomes plotted within each biomarker-positive group. All measures are standardized (centered and scaled) for comparability. Composite #1: ADAS11, Trails B, and Logical Memory II. Composite #2: ADAS11, Trails B, and dAVLT. ADAS, Alzheimer's Disease Assessment Scale; dAVLT, delayed Rey Auditory Verbal Learning Test.

p-tau− (β = 1.60 [0.66, 2.55], P_{ADJ} = 0.004), hippocampus− (β = 0.99 [0.20, 1.78], P_{ADJ} = 0.016). For composite #1, the *APOE*− group improved (β = −0.06 [−0.10, −0.02], P_{ADJ} = 0.011), Aβ− participants improved (β = −0.07 [−0.13, −0.01], P_{ADJ} = 0.048), and hippocampus− participants improved (β = −0.07 [−0.12, −0.03], P_{ADJ} = 0.004). There were no significant practice effects in composite #2, dAVLT, or Trails B, after multiple comparison adjustment.

Power analysis

Using the effect size and variance component estimates from the longitudinal models, we simulated clinical trials to reflect the changes observed over time in the ADNI cohort for each biomarker/outcome combination with 7 years of follow up. Holding the sample size constant at 750 participants/arm and selecting participants for inclusion in the trial based on biomarker positivity, each

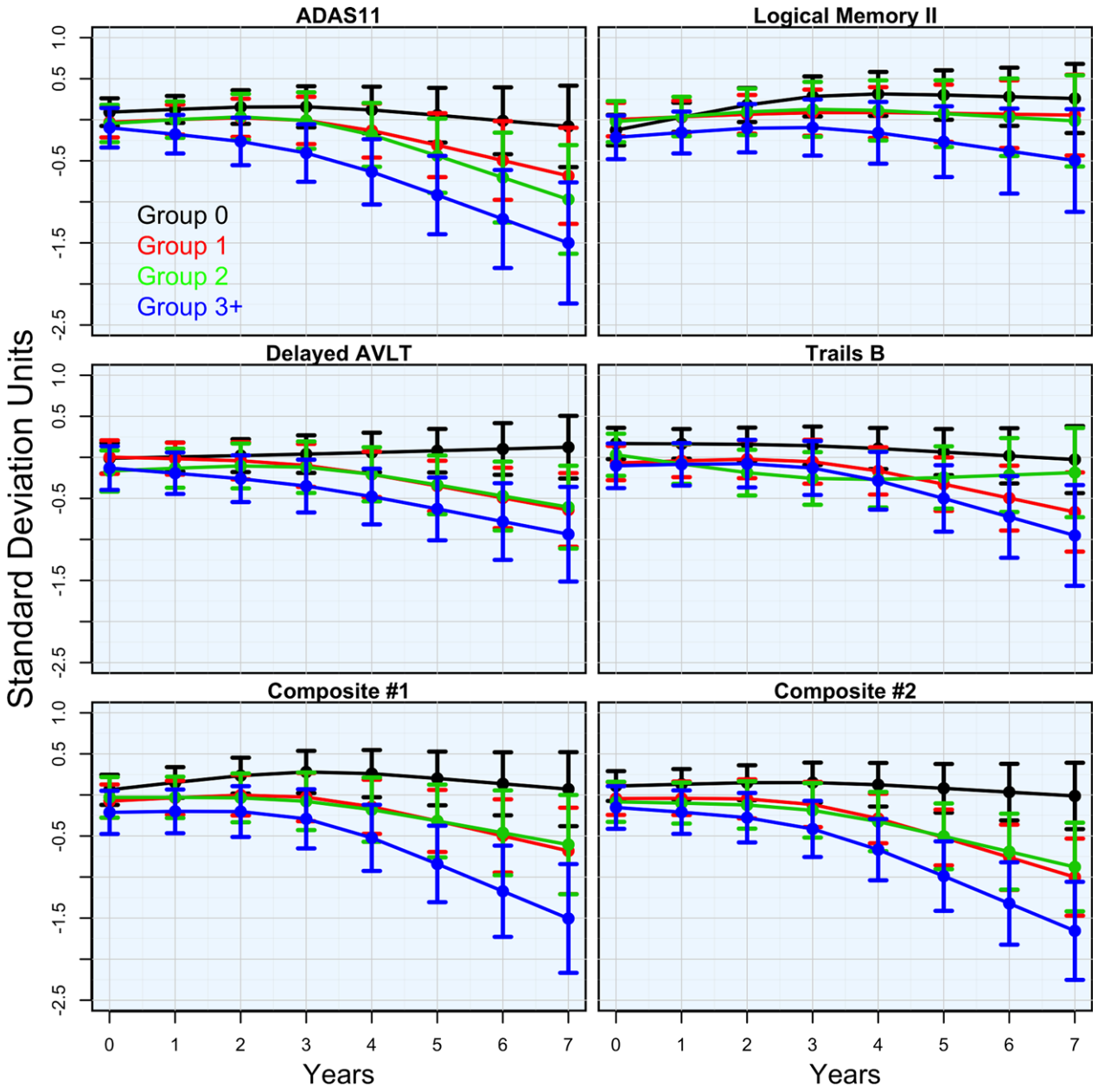

Figure 3. Multiple pathology groups (0, 1, 2, 3+) plotted for each standardized cognitive measures with 7 years of follow up. Composite #1: ADAS11, Trails B, and Logical Memory II. Composite #2: ADAS11, Trails B, and dAVLT. ADAS, Alzheimer's Disease Assessment Scale; dAVLT, delayed Rey Auditory Verbal Learning Test.

outcome was compared in terms of the power to detect a hypothetical drug effect of 25% of the biomarker group difference for trials of varying length. Results are summarized in Table 3. Greater than 80% power was reached when using composite #1 or #2 when including only participants from the 3+ biomarker group, in trials with at least 4 years of follow up. In a post hoc analysis, we estimated the power of a 3-year trial with participants from the 3+ biomarker group, with $N = 1500$ participants/arm to be 73% using composite #1 and 79% using composite #2. In a similar design, but increasing the assumed effect size to a 40% reduction in cognitive decline, a 3-year trial with $N = 650$ participants/arm resulted in 76% power when using composite #1 and 82% power when using composite #2.

In an additional analysis, naïve of CSF biomarker information, we compared *APOE*+ participants with small hippocampi to *APOE*− participants with larger hippocampi.

Table 3. Power (%).

Outcome	Inclusion criteria	Trial length			
		3 years	4 years	5 years	6 years
ADAS11	APOE+	25	40	52	66
	Aβ+	18	39	70	**86**
	t-tau+	10	14	23	35
	p-tau+	11	21	43	63
	Hippocampus+	27	28	33	28
	Group 3+	45	76	**91**	**97**
Logical Memory II	APOE+	6	10	16	21
	Aβ+	12	27	40	58
	t-tau+	5	7	10	18
	p-tau+	6	5	7	9
	Hippocampus+	45	59	55	58
	Group 3+	37	53	66	79
dAVLT	APOE+	16	26	43	60
	Aβ+	6	6	13	16
	t-tau+	6	14	41	66
	p-tau+	16	27	41	54
	Hippocampus+	29	42	54	59
	Group 3+	24	51	76	**90**
Trails B	APOE+	25	26	27	24
	Aβ+	15	17	17	19
	t-tau+	5	6	6	8
	p-tau+	11	5	6	23
	Hippocampus+	12	32	49	70
	Group 3+	5	11	36	71
Composite #1 (ADAS11, Trail B, Logical Memory II)	APOE+	29	42	54	60
	Aβ+	24	50	79	**91**
	t-tau+	6	16	32	51
	p-tau+	5	8	25	53
	Hippocampus+	57	71	78	**80**
	Group 3+	47	**82**	**98**	**100**
Composite #2 (ADAS11, Trail B, dAVLT)	APOE+	42	60	73	**83**
	Aβ+	20	44	71	**86**
	t-tau+	9	19	47	69
	p-tau+	5	17	54	**82**
	Hippocampus+	53	69	77	**82**
	Group 3+	45	**85**	**99**	**100**

ADAS, Alzheimer's Disease Assessment Scale; dAVLT, delayed Rey Auditory Verbal Learning Test.
Numbers greater than or equal to 80 are in bold.

The *APOE*+/hippocampus+ participants performed poorly on the derived composites, resulting in 88% and 89% power in 4-year trials, for composites #1 and #2, respectively. However, with only 24 *APOE*+/hippocampus+ participants, these results should be interpreted with caution.

Discussion

The main findings of this analysis are (1) forming a cognitive composite comprised of measures of delayed memory recall, executive function, and the ADAS11 scale, results in a more sensitive measure of longitudinal decline, thereby providing substantial gains in power when compared with any of the three components alone, (2) increases of 15–20% in power to detect a drug effect can be made by recruiting participants with multiple pathologies at screening, (3) a population with little cognitive impairment can improve on Logical Memory II and Digit Symbol Substitution to the extent where significant positive initial trajectories are observed, and (4) significant decreases in executive function and global cognition at baseline associated with Aβ_{42}, and also a p-tau associated reduction in processing speed, were observable in this population.

Whereas the individual cognitive measures were significantly associated with longitudinal decline in two or at

most three of the five individual pathology groups, composite #2 was strongly associated with all five groups, and composite #1 was significantly associated with four and marginally associated with the fifth. Measuring decline with either derived composite resulted in a consistent increase in power and was more robust to pathology-type, demonstrating that the earliest cognitive changes occur across multiple domains, not only delayed memory recall.

Power suitable for a phase III clinical trial was only reached when restricting inclusion to participants positive for 3+ biomarkers, when using a derived cognitive composite, for at least a 4-year trial. Although the 3+ biomarker group consistently showed the most cognitive decline (Fig. 3), multiple comparison correction and smaller sample sizes in the multiple biomarker groups militated against statistical significance. However, a substantial increase in power resulted when simulating with sample sizes common in phase III trials, using estimates of decline from the multiple biomarker groups. The majority of participants in the 3+ biomarker group were positive for both Aβ and tau biomarkers. This is consistent with the NIA-AA criteria for preclinical AD, which proposes that the combination of biomarker positivity for Aβ and neuronal injury in cognitively healthy people indicates that subjects are closer to cognitive impairment than subjects with isolated Aβ positivity,[38,39] or isolated signs of neuronal injury.[40]

When considering *APOE*4+/hippocampus+ participants, 4-year trial simulations resulted in nearly 90% power. Of the 24 *APOE*+ participants with small hippocampi, CSF data were available for eight. All eight participants were Aβ+ and four were also tau/ptau-positive. It is possible that *APOE*+ participants who also show some evidence of hippocampal atrophy are more likely to harbor Aβ and tau pathology, although this analysis was based on a small sample size.

The two cognitive composites, differing only in type of delayed memory recall component, Logical Memory II versus dAVLT, demonstrated comparable levels of power in participants with multiple biomarker positivity. However, the power seemed to derive from different sources. Logical Memory II was consistently more closely associated with Aβ, whereas dAVLT was more associated with *APOE*-ε4, t-tau, and p-tau, and did not have a significant association with Aβ. The associations between Logical Memory II and the different biomarker groups derived from the biomarker-negative groups improving and the biomarker-positive groups remaining flat or declining slightly. These associations contrast with dAVLT, for which there was little improvement in the biomarker-negative groups and steeper decline in the biomarker-positive groups. Defining power based on a 25% difference between biomarker-positive and negative groups, rather than on a 25% percent slowing of the change in the positive group, had a considerable effect on the power of Logical Memory II given the minimal decline, even in biomarker-positive groups. Although the two delayed memory scales behaved similarly when combined in a composite in participants positive for multiple biomarkers, the differences should be carefully considered in a trial setting, depending on the inclusion criteria and target of the drug. A trial of an anti-Aβ therapy may benefit from using Logical Memory II rather than delayed AVLT, given the difference in the strength of association observed here, although delayed AVLT should be considered when evaluating the association between *APOE* or tau pathology and delayed memory.

The ADAS11, a composite itself, captured the most decline among the individual cognitive scales with 76% power in a 4-year trial with participants positive for multiple biomarkers. Including a measure of delayed memory and executive function to form composite #2 resulted in a 9% increase in power for a 4-year trial over ADAS11 alone. Trails B, not powerful alone, did result in one of the steepest declines for participants with small hippocampi (Fig. 2), and was also the only individual cognitive test to separate Aβ+ from Aβ− at baseline. It is surprising that Aβ-positivity was significantly associated with executive function at baseline and not with delayed memory, although it remains unclear how early memory impairment relates to executive function with respect to the accumulation of Aβ. Trails B, a measure of executive function, associated with both low CSF Aβ and small baseline hippocampal volume may provide a meaningful increase in power when assessed in conjunction with ADAS11.

Significant improvement, possibly due to prior exposure, was seen in Logical Memory II and Digit Symbol Substitution. Logical Memory II was not administered at month 6, which may have countered further improvement. One limitation of Logical Memory II in ADNI is the administration of the same version of the test at each time point. Using multiple versions of this type of test would be critical if the effect of practice were a concern. However, in a clinically normal population, the successful effect of a drug may be to restore the ability to improve over time, as opposed to slowing decline, on an assessment like Logical Memory II.

There were several significant pathology-related differences in cognition at baseline, none of which was a measure of delayed memory recall alone, as might be expected. Trails B and Digit Symbol substitution, measures of executive function and processing speed, were associated with Aβ and p-tau, respectively, suggesting that the earliest changes caused by AD pathology may not be purely memory related. The reduction in Digit Symbol Substitution scores related to p-tau-positivity provides

evidence that although tau pathology is considered a driver of late-stage cognitive decline, it is also associated with subtle impairment in what is considered a cognitively intact population. This suggests a need to evaluate the effect of an anti-Aβ treatment on tau pathology, even in the earliest stages of disease.

The cognitive composite Alzheimer's Disease Cooperative Study-Preclinical Alzheimer Cognitive Composite (ADCS-PACC),[26] developed for the A4 study incorporates the Mini-mental state examination (MMSE), Digit Symbol Substitution, Free and Cued Selective Reminding Test, and the Delayed Recall score on the Logical Memory IIa subtest. With this composite, 80–90% power was estimated for a 3-year trial of 500 Aβ+ participants/arm. This is a smaller sample size and trial duration compared with our estimates, however, this can in part be explained by the steeper decline of Aβ+ participants (with Aβ measured by PET, compared to CSF in this study) and *APOE*-ε4 carriers observed in two of three of their pilot studies The Australian Imaging, Biomarkers and Lifestyle study of aging (AIBL) and Alzheimer's Disease Cooperative Study-Prevention Instrument study (ADCS-PI), whereas in the third study (ADNI), a drug effect of greater than 40% at 2 years was required to reach 80% power, versus the more conservative 25% treatment effect assumed in this analysis. These differences in power may be explained largely by methodological differences, including recent findings that CSF Aβ-positivity may be associated with earlier-stage changes compared with PET Aβ-postivity.[41] The ADNI cohort is also more educated compared with the other cohorts and may have more cognitive reserve. This could, in part, explain the association between smaller hippocampi and more education at baseline observed in this study. With few studies of cognitive composites, the magnitude of decline expected from a cognitively intact population remains uncertain.

This study has several limitations. Sample sizes available for *APOE* and hippocampal volumes were nearly twice that of CSF biomarkers at baseline, making comparisons of hypothesis tests from separate analyses difficult. Also, while CSF Aβ_{42} has been shown to correlate well with direct measures of Aβ deposition such as PET imaging and autopsy,[42,43] it is possible that CSF t-tau and p-tau do not correlate as closely with tau pathology in the brains of healthy elders.[44] Another limitation is the mixture of MRI methods (1.5 T vs. 3 T) from the two phases of ADNI. The choice of cognitive measures included in the composites was based on a literature review rather than taking a data-driven approach. These choices were an attempt to represent standard tests of the individual domains, although several other scales could have been used from the extensive battery of cognitive tests available in ADNI. We also did not incorporate attrition into our sample size estimates. Our primary goal of identifying the most powerful endpoint/biomarker combinations will not be affected by this omission. However, the required sample size will increase with increasing dropout rates.

In conclusion, recruiting and treating participants with multiple biomarker positivity, especially both Aβ and tau pathologies, may increase the power in trials of a presymptomatic population. This could be especially important when evaluating a treatment with the potential to slow the accumulation of both Aβ and tau, which may be crucial to achieving clinical effects, given the strong correlations between tau pathology and clinical symptoms. Identifying a truly optimal biomarker/endpoint combination will depend on the mechanism of the drug and its capacity to affect the relationship between the target biomarkers and cognition. However, requiring positivity for multiple biomarkers at screening quickly limits the number of eligible participants, highlighting the tradeoff between recruiting from a large pool of lower-risk participants versus a small pool of higher-risk participants. The cost of measuring additional biomarkers at screening is another considerable hurdle and it remains unknown whether these costs will be necessary to identify the cohort required for a successful trial. Also, if the effect of treatment is greater in milder subjects, selecting on the basis of additional biomarkers may actually reduce power. Finally, careful inclusion of both delayed memory recall and nonmemory measures should be considered when selecting a primary cognitive endpoint.

Acknowledgments

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: Alzheimer's Association; Alzheimer's Drug Discovery Foundation; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; GE Healthcare; Innogenetics, N.V.; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Synarc Inc.; and Takeda Pharmaceutical 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. This research was also supported by National Institutes of Health grants P30 AG010129 and K01 AG030514.

Conflict of Interest

None declared.

References

1. Doody RS, Thomas RG, Farlow M, et al. Phase 3 trials of solanezumab for mild-to-moderate Alzheimer's disease. N Engl J Med 2014;370:311–321.
2. Salloway S, Sperling R, Gilman S, et al. A phase 2 multiple ascending dose trial of bapineuzumab in mild to moderate Alzheimer disease. Neurology 2009;73:2061–2070.
3. Holmes C, Boche D, Wilkinson D, et al. Long-term effects of Aβ42 immunisation in Alzheimer's disease: follow-up of a randomised, placebo-controlled phase I trial. Lancet 2008;372:216–223.
4. Hyman BT, Marzloff K, Arriagada PV. The lack of accumulation of senile plaques or amyloid burden in Alzheimer's disease suggests a dynamic balance between amyloid deposition and resolution. J Neuropathol Exp Neurol 1993;52:594–600.
5. Arriagada PV, Growdon JH, Hedley-Whyte ET, Hyman BT. Neurofibrillary tangles but not senile plaques parallel duration and severity of Alzheimer's disease. Neurology 1992;42:631–639.
6. Ingelsson M, Fukumoto H, Newell KL, et al. Early Aβ accumulation and progressive synaptic loss, gliosis, and tangle formation in AD brain. Neurology 2004;62:925–931.
7. Buchhave P, Minthon L, Zetterberg H, et al. Cerebrospinal fluid levels of β-amyloid 1-42, but not of tau, are fully changed already 5 to 10 years before the onset of Alzheimer dementia. Arch Gen Psychiatry 2012;69:98–106.
8. Rizk-Jackson A, Insel P, Petersen R, et al. Early indications of future cognitive decline: stable versus declining controls. PLoS One 2013;8:e74062.
9. Sperling RA, Rentz DM, Johnson KA, et al. The A4 study: stopping AD before symptoms begin? Sci Transl Med 2014;6:228 fs13.
10. Reiman EM, Langbaum JB, Fleisher AS, et al. Alzheimer's Prevention Initiative: a plan to accelerate the evaluation of presymptomatic treatments. J Alzheimers Dis 2011;26:321–329.
11. Aisen PS, Andrieu S, Sampaio C, et al. Report of the task force on designing clinical trials in early (predementia) AD. Neurology 2011;76:280–286.
12. US Department of Health and Human Services; US Food and Drug Administration; Center for Drug Evaluation and Research. Guidance for industry: Alzheimer's disease: developing drugs for the treatment of early stage disease (draft guidance). Available at: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM338287.pdf. Published February 2013 (accessed 16 September 2014).
13. Grober E, Hall CB, Lipton RB, et al. Memory impairment, executive dysfunction, and intellectual decline in preclinical Alzheimer's disease. J Int Neuropsychol Soc 2008;14:266–278.
14. Mormino EC, Betensky RA, Hedden T, et al. Amyloid and APOE ε4 interact to influence short-term decline in preclinical Alzheimer disease. Neurology 2014;82:1760–1767.
15. Lim YY, Maruff P, Pietrzak RH, et al. Effect of amyloid on memory and non-memory decline from preclinical to clinical Alzheimer's disease. Brain 2014;137:121–131.
16. Baxter LC, Caselli RJ, Johnson SC, et al. Apolipoprotein E ε4 affects new learning in cognitively normal individuals at risk for Alzheimer's disease. Neurobiol Aging 2003;24:947–952.
17. Villemagne VL, Burnham S, Bourgeat P, et al. Amyloid β deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer's disease: a prospective cohort study. Lancet Neurol 2013;12:357–367.
18. Snitz BE, Weissfeld LA, Lopez OL, et al. Cognitive trajectories associated with β-amyloid deposition in the oldest-old without dementia. Neurology 2013;80:1378–1384.
19. Caselli RJ, Dueck AC, Osborne D, et al. Longitudinal modeling of age-related memory decline and the APOE ε4 effect. N Engl J Med 2009;361:255–263.
20. Resnick SM, Sojkova J, Zhou Y, et al. Longitudinal cognitive decline is associated with fibrillar amyloid-beta measured by [11C] PiB. Neurology 2010;74:807–815.
21. Kantarci K, Lowe V, Przybelski SA, et al. APOE modifies the association between Aβ load and cognition in cognitively normal older adults. Neurology 2012;78:232–240.
22. Caselli RJ, Reiman EM, Locke DE, et al. Cognitive domain decline in healthy apolipoprotein E ε4 homozygotes before the diagnosis of mild cognitive impairment. Arch Neurol 2007;64:1306–1311.
23. Landau SM, Mintun MA, Joshi AD, et al. Amyloid deposition, hypometabolism, and longitudinal cognitive decline. Ann Neurol 2012;72:578–586.
24. Roe CM, Fagan AM, Grant EA, et al. Amyloid imaging and CSF biomarkers in predicting cognitive impairment up to 7.5 years later. Neurology 2013;80:1784–1791.

25. Doraiswamy PM, Sperling RA, Coleman RE, et al. Amyloid-β assessed by florbetapir F 18 PET and 18-month cognitive decline: a multicenter study. Neurology 2012;79:1636–1644.
26. Donohue MC, Sperling RA, Salmon DP, et al. The preclinical Alzheimer cognitive composite: measuring amyloid-related decline. JAMA Neurol 2014;71:961–970.
27. Shaw LM, Vanderstichele H, Knapik-Czajka M, et al. Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. Ann Neurol 2009;65:403–413.
28. Olsson A, Vanderstichele H, Andreasen N, et al. Simultaneous measurement of beta-amyloid(1-42), total tau, and phosphorylated tau (Thr181) in cerebrospinal fluid by the xMAP technology. Clin Chem 2005;51: 336–345.
29. Jack CR Jr, Bernstein MA, Fox NC, et al. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. J Magn Reson Imaging 2008;27:685–691.
30. Rosen WG, Mohs RC, Davis KL. A new rating scale for Alzheimer's disease. Am J Psychiatry 1984;141: 1356–1364.
31. Wechsler D. Wechsler adult intelligence scale-revised. New York: Psychological Corporation, 1981.
32. Rey A. L'examen clinique en psychologie. Paris: Presses Universitaires de France, 1964.
33. Reitan R. Validity of the Trail-Making Test as an indication of organic brain damage. Percept Mot Skills 1958;8:271–276.
34. Jack CR Jr, Twomey CK, Zinsmeister AR, Sharbrough FW. Anterior temporal lobes and hippocampal formations: normative volumetric measurements from MR images in young adults. Radiology 1989;172:549–554.
35. Devlin TF, Weeks BJ. Spline functions for logistic regression modeling. Proc 11th Annual SAS Users Group Intnl Conf. Cary NC: SAS Institute, Inc., 1986. pp. 646–651.
36. Hochberg Y. A sharper Bonferroni procedure for multiple tests of significance. Biometrika 1988;75:800–803.
37. Siddiqui O, Hung HJ, O'Neill R. MMRM vs. LOCF: a comprehensive comparison based on simulation study and 25 NDA datasets. J Biopharm Stat 2009;19:227–246.
38. Sperling RA, Aisen PS, Beckett LA, et al. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011;7:280–292.
39. Vos SJ, Xiong C, Visser PJ, et al. Preclinical Alzheimer's disease and its outcome: a longitudinal cohort study. Lancet Neurol 2013;12:957–965.
40. Jack CR, Knopman DS, Weigand SD, et al. An operational approach to National Institute on Aging–Alzheimer's Association criteria for preclinical Alzheimer disease. Ann Neurol 2012;71:765–775.
41. Mattsson N, Insel PS, Donohue MC, et al. Independent information from cerebrospinal fluid amyloid-ß and florbetapir imaging in Alzheimer's disease. Brain 2015;138:772–783.
42. Strozyk D, Blennow K, White LR, Launer LJ. CSF Aβ 42 levels correlate with amyloid-neuropathology in a population-based autopsy study. Neurology 2003;60:652–656.
43. Mattsson N, Insel PS, Landau S, et al. Diagnostic accuracy of CSF Ab42 and florbetapir PET for Alzheimer's disease. Ann Clin Transl Neurol 2014;1:534–543.
44. Seppälä TT, Nerg O, Koivisto AM, et al. CSF biomarkers for Alzheimer disease correlate with cortical brain biopsy findings. Neurology 2012;78:1568–1575.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Frequency of pathology combinations.
Table S2. Longitudinal sample sizes.

