Alzheimer Disease Biomarkers as Outcome Measures for Clinical Trials in MCI

Anna Caroli, PhD, Annapaola Prestia, PsyD,† Sara Wade, PhD,‡§ Kewei Chen, PhD,‖ Napatkamon Ayutyanont, PhD,‖ Susan M. Landau, PhD,¶ Cindee M. Madison, MS,¶ Cathleen Haense, MD,# Karl Herholz, MD,** Eric M. Reiman, MD,‖ William J. Jagust, MD,¶ and Giovanni B. Frisoni, MD,†, †† for the Alzheimer's Disease Neuroimaging Initiative*

Background: The aim of this study was to compare the performance and power of the best-established diagnostic biological markers as outcome measures for clinical trials in patients with mild cognitive impairment (MCI).

Methods: Magnetic resonance imaging, F-18 fluorodeoxyglucose positron emission tomography markers, and Alzheimer's Disease Assessment Scale-cognitive subscale were compared in terms of effect size and statistical power over different follow-up periods in 2 MCI groups, selected from Alzheimer's Disease Neuroimaging Initiative data set based on cerebrospinal fluid (abnormal cerebrospinal fluid Aβ1-42 concentration—ABETA +) or magnetic resonance imaging evidence of Alzheimer disease (positivity to hippocampal atrophy—HIPPO +). Biomarkers progression was modeled through mixed effect models. Scaled slope was chosen as measure of effect size. Biomarkers power was estimated using simulation algorithms.

Results: Seventy-four ABETA + and 51 HIPPO + MCI patients were included in the study. Imaging biomarkers of neurodegeneration, especially MR measurements, showed highest performance. For all biomarkers and both MCI groups, power increased with increasing follow-up time, irrespective of biomarker assessment frequency.

Conclusion: These findings provide information about biomarker enrichment and outcome measurements that could be employed to reduce MCI patient samples and treatment duration in future clinical trials.

Key Words: Alzheimer disease, mild cognitive impairment, clinical trials, biomarkers, outcome measures, enrichment biomarkers, biomarkers power

(*Alzheimer Dis Assoc Disord* 2015;29:101–109)

Received for publication May 13, 2014; accepted October 17, 2014.

From the *Medical Imaging Unit, Biomedical Engineering Department, IRCCS Istituto di Ricerche Farmacologiche Mario Negri, Bergamo; †Laboratory of Epidemiology and Neuroimaging, IRCCS Fatebenefratelli, Brescia; §Department of Decision Science, Bocconi University, Milan, Italy; ‡Department of Engineering, University of Cambridge; **Institute of Brain, Behaviour, and Mental Health University of Cambridge, Cambridge; ‖Banner Alzheimer's Institute, Phoenix, AZ; ¶Helen Wills Neuroscience Institute, University of California, Berkeley, CA; #Hannover Medical School, Clinic for Nuclear Medicine, Hannover, Germany; and ††Departments of Internal Medicine and Psychiatry, University Hospitals and University of Geneva, Geneva, Switzerland.

Data collection and sharing of ADNI data for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Alzheimer's Association; Alzheimer's Drug Discovery Foundation; 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 (http://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 NIH Grants P30 AG010129 and K01 AG030514.

Supported in part by the National Institute of Mental Health (R01MH57899), the National Institute on Aging (R01AG031581 and P30AG19610), and the state of Arizona.

The authors declare no conflicts of interest.

Reprints: Giovanni B. Frisoni, MD, Laboratory of Epidemiology and Neuroimaging, IRCCS Fatebenefratelli, IRCCS Centro San Giovanni di Dio-Fatebenefratelli, via Pilastroni 4, Brescia 25125, Italy (e-mail: gfrisoni@fatebenefratelli.it).

Supplemental Digital Content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Website, www.alzheimerjournal.com.

Numerous clinical trials have been conducted to evaluate the efficacy of new drugs for Alzheimer disease (AD). Unfortunately, all of them failed to demonstrate meaningful clinical benefit, causing debate as to methods and therapeutic targets,[1] and emphasizing the need for more carefully designed trials using quantifiable biomarkers of disease progression beyond a core of cognitive symptoms to target patients in the mild or presymptomatic phases of AD.[2]

It is generally estimated that 10% to 20% of participants enrolled in AD trials using standard clinical criteria do not have AD, potentially diluting treatment effects. As drug development and assessment programs move into the presymptomatic population, inclusion criteria become even more important.[3] Evidence of abnormal amyloid and/or neurodegeneration biomarkers increase the likelihood of developing AD from preclinical disease stage[3,4] and are expected to significantly enrich the enrolled population of individuals who will likely progress to AD if left untreated.[5] Indeed, the European Medicines Agency (EMA) has qualified both cerebrospinal fluid (CSF) Aβ1-42 and structural magnetic resonance imaging (MRI) measured hippocampal volume as enrichment biomarkers to enroll mild and moderate as well as predemented AD subjects in

regulatory clinical trials (EMA/CHMP/SAWP/893622/2011 and EMA/CHMP/SAWP/809208/2011 qualification opinions, available at http://www.ema.europa.eu/ema/ by searching the document library).

In addition, diagnostic biological markers may serve as surrogate outcome measures in clinical trials,[6] and could replace previously adopted clinical endpoints, which are limited by substantial measurement variation, low sensitivity to change during early disease, and long follow-up periods.[7] The adoption of biomarkers precisely measuring biological change may increase the statistical power, thus requiring fewer participants studied for shorter durations, and notably reducing the costs of the trials.[8]

Previous studies suggested that F-18 fluorodeoxyglucose positron emission tomography (FDG-PET)[9] and MRI biomarkers[10] could be used as effective outcome measures in clinical trials. Questions regarding which biomarkers are best to use, and how, are currently far from resolved, and the choice must take into consideration the type of therapeutic intervention, the clinical stage of AD, the time dependence of biomarker changes during disease progression, as well as biomarker costs and availability.[11]

The aim of this study was to investigate and compare the performance and power of the best-established diagnostic biological markers as outcome measures for clinical trials in patients with mild cognitive impairment (MCI) and CSF or MRI biomarker evidence of AD, using longitudinal data available in the Alzheimer's Disease Neuroimaging Initiative (ADNI) data set.

METHODS

Subjects

Patients enrolled and data used in the preparation of this article were obtained from the ADNI database (adni.loni.usc.edu). Information on ADNI are available as supplementary material, (Supplemental Digital Content, http://links.lww.com/WAD/A106, section 1.1). At baseline, all subjects receive a comprehensive neuropsychological evaluation; they undergo blood drawing (for ApoE genotyping) and structural MR. Subsets of subjects undergo lumbar puncture (for CSF sampling) or PIB-PET, and half of the subjects undergo FDG-PET. All subjects undergo yearly follow-up visits. Moreover, MCI patients are examined every 6 months to assess conversion to dementia.

The syndromic diagnosis of MCI was made according to Petersen et al[12] criteria, and the cognitive profile was consistent with single and multiple domain amnestic MCI.

Patient cohorts enrolled in the current study include MCI patients with both baseline and at least 1 follow-up FDG-PET scan (on May 22, 2011), either with abnormal CSF Aβ1-42 concentration (hereafter named as "ABETA + " MCI, n = 74) or positive to hippocampal atrophy (hereafter named as "HIPPO + " MCI, n = 51). CSF Aβ1-42 positivity was defined based on a previously published cutoff (baseline CSF Aβ1-42 < 192 g/mL[13]). Positivity to hippocampal atrophy was defined as baseline hippocampal volume (the smallest between left and right ones, expressed in *W* scores) below the fifth percentile of its distribution in 143 ADNI cognitively healthy elders.[14] A full list of patients included in the study is available in Table e-1 (Supplemental Digital Content, http://links.lww.com/WAD/A106).

AD Biomarkers

Markers of Cortical Hypometabolism

FDG PET imaging was performed at all of the ADNI PET sites in North America according to previously described acquisition protocols.[15]

For all available FDG-PET scans, AD-related hypometabolism was assessed using 3 FDG-PET data analysis techniques—the PMOD Alzheimer discrimination analysis tool (PALZ)[16,17] (http://www.pmod.com), the hypometabolic convergence index (HCI),[18] and a set of meta-analytically derived regions of interest reflecting AD hypometabolism pattern (metaROI) method.[19] All metrics are based on voxel-by-voxel analysis of FDG-PET images and provide a single measure of AD-related hypometabolism. Details about their different processing procedures are available as supplementary material (Supplemental Digital Content, http://links.lww.com/WAD/A106, section 1.2).

Structural Markers

Brain T1-weighted MRI was performed in all of the ADNI MRI sites in North America, as previously described.[20]

For all available MR scans, hippocampal volumes were automatically segmented using Freesurfer software,[21] and brain atrophy rates were measured using the KN boundary shift integral (KN-BSI) method.[22] Details about structural markers processing procedures are available as supplementary material (Supplemental Digital Content, http://links.lww.com/WAD/A106, section 1.3).

CSF Aβ1-42

CSF was obtained by lumbar puncture performed with a 20- or 24-G spinal needle between L4 and L5 or L3 and L4, and collected in polypropylene tubes. CSF samples were thawed for 1 hour, gentle mixed, aliquoted, and frozen on dry ice at -80°C. CSF Aβ1-42 protein concentration was determined by xMAP Luminex platform (Luminex Corp., Austin, TX) with Innogenetics (INNOBIA AlzBio3, Ghent, Belgium) immunoassay kit–based reagents.[13]

Statistical Analysis

Longitudinal biomarker progression was modeled through a mixed effect model. To make results comparable across biomarkers, biomarker data were preliminarily standardized based on the baseline mean and SD. In this way, all biomarkers start at time 0 with similar behavior. If necessary, they were polarized to obtain positive slopes.

For each MCI patient group, biomarker, and time set, the models were fit separately based on all available data, and Bayesian methods were used for inference. Priors in the Bayesian model were specified using an empirical Bayes approach; in particular, conjugate priors were selected with parameters corresponding to the maximum likelihood estimates.

Scaled slope, defined as average slope divided by SD of the random effects for the slope, was chosen as measure of effect size for each biomarker, as it provides a way to compare biomarkers based on their overall slope and, at the same time, penalize for large variability in the slopes across patients. For each MCI patient group and follow-up time set, scaled slopes were estimated, and biomarkers were ranked accordingly. With the Bayesian method, we are able to quantify uncertainty in the ranking through its estimated probability and to compare scaled slopes across biomarkers

(Supplemental Digital Content, http://links.lww.com/WAD/A106, section 1.4).

To compare biomarker performance across different follow-up times, a Bayesian power analysis was performed to find the optimal sample size needed in a hypothetical clinical trial to detect 20% reduction in the slope for $\alpha = 0.05$.

All statistical analyses were performed using *R* software,[23] version 2.14.1. A detailed description of the model, standardization procedure (particularly for KN-BSI), priors, inference algorithm, estimated quantities, and power analysis can be found in the supplementary material (Supplemental Digital Content, http://links.lww.com/WAD/A106, and methods section 1.4).

RESULTS

Seventy-four MCI patients with abnormally low amyloid concentration in the CSF (ABETA + MCI, age = 75 ± 7 y, 36% females) and 51 patients with abnormally low hippocampal volume on MRI (HIPPO + MCI, age = 75 ± 7, 37% females) were included in the current study. All MCI patients had baseline clinical data and FDG-PET biomarkers; all but 9 MCI ABETA + patients had baseline MRI biomarkers. Most of the patients included in the study underwent 6-month follow-up visits with biomarker assessment up to 24 months after baseline. Clinical features and diagnostic biological markers available at different time points in the 2 MCI patient groups are shown in Table 1. MCI ABETA + and HIPPO + patients were not different in any clinical feature and biomarker except for hippocampal volume, MCI HIPPO + patients showing significantly smaller volumes at any time point ($P < 0.001$).

Table 2 shows biomarkers effect size (scaled slope) and provides biomarker ranking for each follow-up time set and MCI group along with the estimated probability of the ranking. MR biomarkers showed highest performance for all time sets and both MCI patient groups, followed by the other markers of neurodegeneration and, last, Alzheimer's Disease Assessment Scale-cognitive subscale (ADAS-COG). KN-BSI and hippocampal volume performance were comparable for all time sets and both MCI groups except for 18-month follow-up period (in ABETA + MCI patients only), when KN-BSI significantly outperformed hippocampal volume. ADAS-COG increased performance over longest (24 mo) follow-up periods (in ABETA + MCI patients only, showing significantly better ADAS-COG performance over 24 mo than HIPPO + MCI patients). Among FDG-PET biomarkers, HCI ranked first for all time sets except 6-month observations up to 2 years (HIPPO + MCI patients); metaROI ranked higher than logPALZ for shortest observation periods (12 and 18 mo) and lower than logPALZ for longest follow-up periods.

Estimated probabilities that any biomarker had a larger slope in pairwise comparison with any other biomarker for each follow-up time set and both MCI groups are shown in Table e-2 (Supplemental Digital Content, http://links.lww.com/WAD/A106). Figure 1 visually shows scaled slopes and 95% credible regions of individual biomarkers, estimated for T0-T6-T12 time set given the estimated intercept, in ABETA + (Fig. 1A) and HIPPO + (Fig. 1B) MCI patients.

Figure 2 shows the estimated power of a hypothetical clinical trial designed to detect 20% reduction in biomarker slope as a function of sample size and follow-up time in ABETA + (Fig. 2A) and HIPPO + (Fig. 2B) MCI

TABLE 1. Clinical Features and Diagnostic Biological Markers at Different Time Points in MCI Patients With Abnormal CSF Aβ 1-42 Concentration (MCI ABETA+, n = 74, Age = 75 ± 7 y, 36% Females) or Positive to Hippocampal Atrophy (MCI HIPPO+, n = 51, Age = 75 ± 7, 37% Females)

	T0	n	T6	n	T12	n	T18	n	T24	n
MCI ABETA +										
MMSE	27 ± 2	74	27 ± 2	74	26 ± 3	71	26 ± 2	68	25 ± 3	64
ADAS-COG	12 ± 4	74	13 ± 5	74	13 ± 6	71	14 ± 6	69	15 ± 6	64
logPALZ	1.20 ± 0.67	74	1.30 ± 0.73	74	1.36 ± 0.77	71	1.44 ± 0.79	67	1.52 ± 0.85	60
HCI	1373 ± 886	74	1540 ± 932	74	1645 ± 939	71	1821 ± 1046	67	1962 ± 1169	60
MetaROI	1.17 ± 0.13	74	1.15 ± 0.12	74	1.14 ± 0.13	71	1.14 ± 0.14	67	1.12 ± 0.15	60
Hipp. volume	2963 ± 413	65	2874 ± 462	71	2864 ± 456	65	2839 ± 493	64	2840 ± 498	54
KN-BSI	0 ± 0	74	7.9 ± 7.2	72	14.9 ± 7.5	67	21.1 ± 10.7	64	25.2 ± 12.1	61
MCI HIPPO +										
MMSE	27 ± 2	51	26 ± 3	51	26 ± 3	50	25 ± 3	47	24 ± 4	42
ADAS-COG	13 ± 4	51	14 ± 5	51	14 ± 6	49	16 ± 7	47	16 ± 6	42
logPALZ	1.17 ± 0.69	51	1.23 ± 0.71	51	1.30 ± 0.77	49	1.38 ± 0.83	45	1.46 ± 0.88	41
HCI	1418 ± 893	51	1494 ± 899	51	1602 ± 916	49	1899 ± 1068	45	2022 ± 1111	41
MetaROI	1.16 ± 0.12	51	1.16 ± 0.12	51	1.14 ± 0.11	49	1.12 ± 0.13	45	1.11 ± 0.14	41
Hipp. volume	2544 ± 237	51	2502 ± 235	49	2463 ± 242	50	2436 ± 265	42	2404 ± 276	38
KN-BSI	0 ± 0	51	8.2 ± 7.2	50	12.7 ± 9.9	49	19.6 ± 13.4	44	23.7 ± 14.7	40

Values are mean ± SD.

ADAS-COG indicates Alzheimer's Disease Assessment Scale-cognitive subscale; HCI, hypometabolic convergence index[17]; Hipp. volume, hippocampal volume automatically computed by Freesurfer algorithm; KN-BSI, brain atrophy rate measured by KN boundary shift integral technique[21]; logPALZ, log-transformed PMOD Alzheimer score[16]; MetaROI, FDG-PET summary metric based on meta-analitically derived regions of interest reflecting AD hypometabolism pattern[18]; MMSE, Mini Mental State Examination; T0, baseline; Tn, n-month follow-up.

TABLE 2. Scaled Slope, Defined as Average Slope Divided by SD of the Random Effects for the Slope, With Pertinent 95% Credible Intervals, Estimated for Each Follow-up Time Set in MCI Patients With Abnormal CSF Aβ 1-42 Concentration (MCI ABETA+) or Positive to Hippocampal Atrophy (MCI HIPPO+). MR Biomarkers Showed Highest Effect Size for all Time Sets and Both MCI Patient Groups

	T0-T6-T12-T18-T24		T0-T6-T12-T18		T0-T6-T12		T0-T12-T24	
MCI ABETA +								
ADAS-COG	1.290 (0.869-1.807)	4	0.944 (0.597-1.384)	6	0.720 (0.275-1.374)	6	1.487 (1.177-1.830)	4
logPALZ	1.276 (0.897-1.794)	5	1.193 (0.778-1.802)	5	1.106 (0.681-1.734)	5	1.379 (0.902-2.043)	5
HCI	1.468 (1.106-1.910)	3	1.638 (1.148-2.346)	3	2.839 (1.475-4.949)	3	1.634 (1.148-2.262)	3
MetaROI	1.268 (0.808-1.964)	6	1.337 (0.772-2.292)	4	1.373 (0.733-2.470)	4	1.052 (0.671-1.640)	6
Hipp. volume	2.022 (1.466-2.779)	2	1.610 (1.151-2.255)	2	3.716 (1.908-6.448)	1	2.501 (1.725-3.654)	1
KN-BSI	2.438 (1.975-2.948)	1	2.583 (2.032-3.237)	1	3.661 (2.321-5.935)	2	2.372 (1.901-2.904)	2
Estimated probability of the ranking		0.076		0.096		0.160		0.148
MCI HIPPO +								
ADAS-COG	0.797 (0.548-1.063)	6	0.723 (0.480-0.985)	6	0.384 (0.169-0.612)	6	1.075 (0.780-1.399)	5
logPALZ	1.662 (1.006-2.564)	3	0.739 (0.418-1.156)	5	0.847 (0.496-1.337)	5	1.171 (0.785-1.677)	4
HCI	1.377 (0.977-1.859)	5	1.374 (0.890-2.050)	3	1.378 (0.659-2.508)	3	1.666 (1.087-2.485)	3
MetaROI	1.531 (0.784-2.849)	4	0.861 (0.419-1.648)	4	1.044 (0.435-1.999)	4	1.049 (0.611-1.703)	6
Hipp. volume	2.035 (1.472-2.751)	1	1.614 (1.154-2.202)	2	3.013 (1.598-5.388)	1	2.661 (1.756-4.208)	1
KN-BSI	1.766 (1.362-2.204)	2	1.678 (1.281-2.116)	1	1.560 (1.147-2.043)	2	1.769 (1.346-2.237)	2
Estimated probability of the ranking		0.049		0.065		0.177		0.124

For each time set, biomarkers were ranked in terms of decreasing scaled slope and the estimated probability of the ranking is reported.

ADAS-COG indicates Alzheimer's Disease Assessment Scale-cognitive subscale; HCI, hypometabolic convergence index[17]; Hipp. volume, hippocampal volume automatically computed by Freesurfer algorithm; KN-BSI, brain atrophy rate measured by KN boundary shift integral technique[21]; logPALZ, log-transformed PMOD Alzheimer score[16]; MetaROI, FDG-PET summary metric based on meta-analitically derived regions of interest reflecting AD hypometabolism pattern[18]; T0, baseline; Tn, n-month follow-up.

patients. For all biomarkers and both MCI groups, the power increased with increasing follow-up time, and the main increase was observed from 12 to 18 months of observation; the power showed little increase with 6-monthly biomarker assessment compared with yearly assessment. For each time set and both MCI groups, MR measures showed highest power, with KN-BSI outperforming hippocampal volume, especially in the ABETA + MCI group, followed by HCI, logPALZ, and MetaROI FDG-PET summary metrics. ADAS-COG required higher sample sizes. KN-BSI and hippocampal volume powers reached a plateau around 150 to 200 and 200 to 250 patients per treatment arm, respectively; HCI power increase showed a similar nonlinear trend, reaching a plateau around 300 to 350 patients, whereas for all other biomarkers the power increased in an approximately linear trend, over the sample size range studied. For all time sets and all biomarkers but hippocampal volume, required sample size was higher in MCI HIPPO + than in MCI ABETA + group (Table 3).

DISCUSSION

In the current study we investigated and compared the performance and power of the best-established diagnostic biological markers as outcome measures for clinical trials in MCI patients with CSF or MRI biomarker evidence of AD, over variable follow-up times.

Some preliminary clarifications are needed to fully understand current results and in view of their appropriate use for future clinical trials design. Biomarkers could be used in clinical trials with 2 different objectives: (i) to demonstrate target engagement (which is a necessary but not sufficient condition to drug clinical efficacy and is not addressed in the current study); or (ii) to work as surrogate clinical outcomes, demonstrating disease-modifying effect. We believe the latter to be the best condition in which current findings could be translated, keeping in mind that regulatory agencies have not yet recognized any biomarker as a surrogate clinical outcome measure although biomarkers can presently be used as secondary outcome measures in addition to a measure of clinical efficacy. Moreover, the current study is not aimed at identifying the best marker to be used as a surrogate outcome measure in all future clinical trials, but rather aims at providing information which could drive the choice of outcome measure, which still depends highly on the design of any particular trial.

MRI and FDG-PET imaging outperformed clinical biomarkers, and MRI outperformed FDG-PET measures for all time sets and both MCI patient groups. Among MR biomarkers, KN-BSI, specifically designed as a longitudinal measure to track disease progression, outperformed hippocampal volume, especially among MCI patients screened to be positive for amyloid-β.

There are a number of previous studies focused on determining the effectiveness of different biomarkers as outcomes in MCI clinical trials by calculating sample size estimates based on ADNI data.

Their main limitation (all but[24]) was based on tout-court MCI, rather than on enriched MCI patient groups. As both CSF Aβ1-42 and hippocampal atrophy on MRI have been qualified as enrichment biomarkers to enroll predemented AD subjects in regulatory clinical trials (EMA/CHMP/SAWP/893622/2011 and EMA/CHMP/SAWP/809208/2011 qualification opinions), all future clinical trials will be performed on enriched MCI groups, pointing out the need to have new reliable estimates. Despite use of different selection criteria, the biomarker ranking proposed in this study is in line with previous findings.

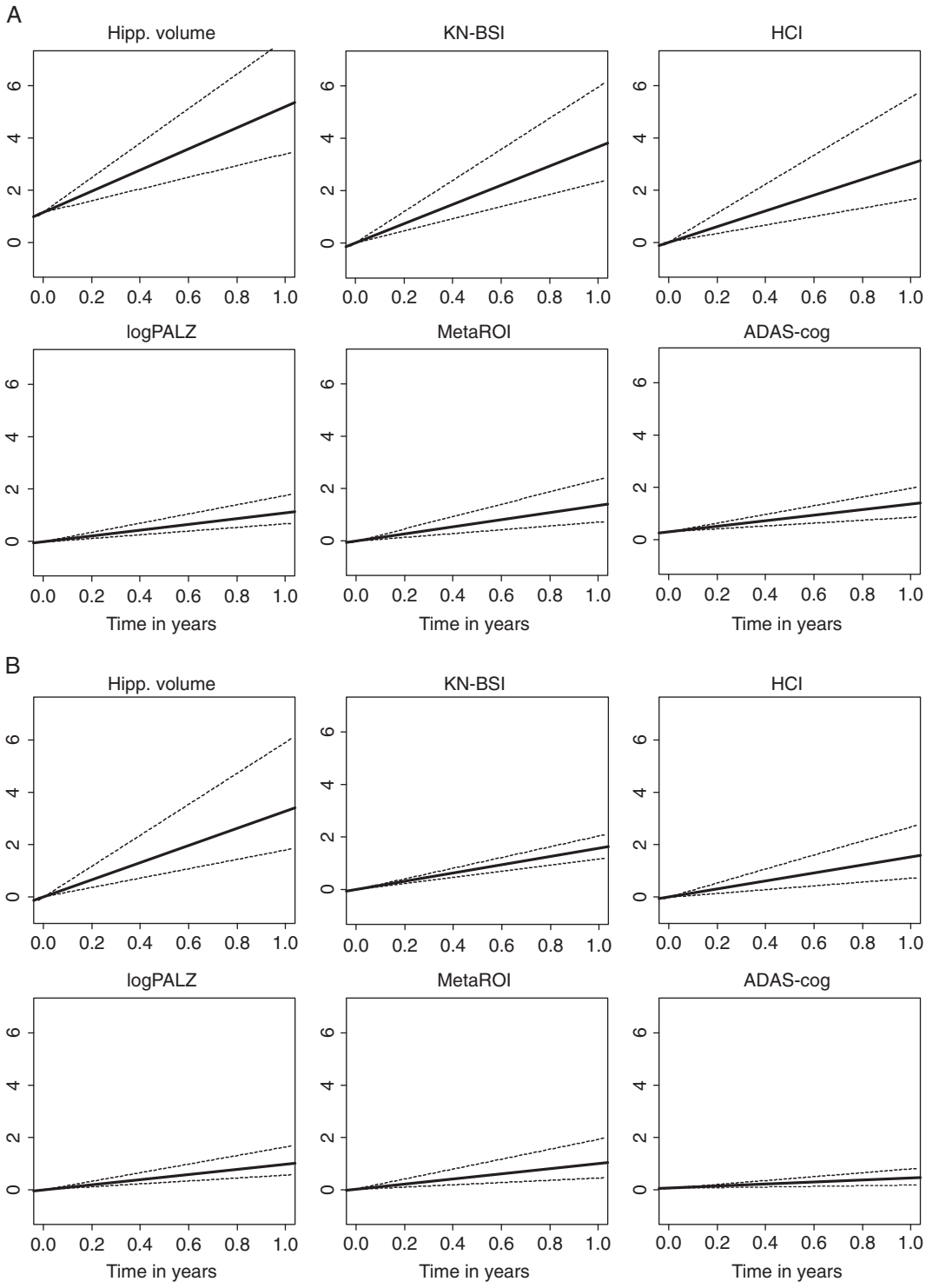

FIGURE 1. Scaled slopes for individual biomarkers with 95% credible regions, estimated in mild cognitive impairment (MCI) patients with abnormal cerebrospinal fluid Aβ 1-42 concentration [ABETA+, (A)] or positive to hippocampal atrophy [HIPPO+, (B)] using all data available in the T0-T6-T12 time set, given the estimated intercept (intercept of 0 for KN-BSI). MR biomarkers showed highest effect size in both MCI groups. ADAS-COG indicates Alzheimer's Disease Assessment Scale-cognitive subscale; HCI, hypometabolic convergence index[17]; Hipp. volume, hippocampal volume automatically computed by Freesurfer algorithm; KN-BSI, brain atrophy rate measured by KN boundary shift integral technique[21]; logPALZ, log-transformed PMOD Alzheimer score[16]; MetaROI, FDG-PET summary metric based on meta-analitically derived regions of interest reflecting AD hypometabolism pattern.[18]

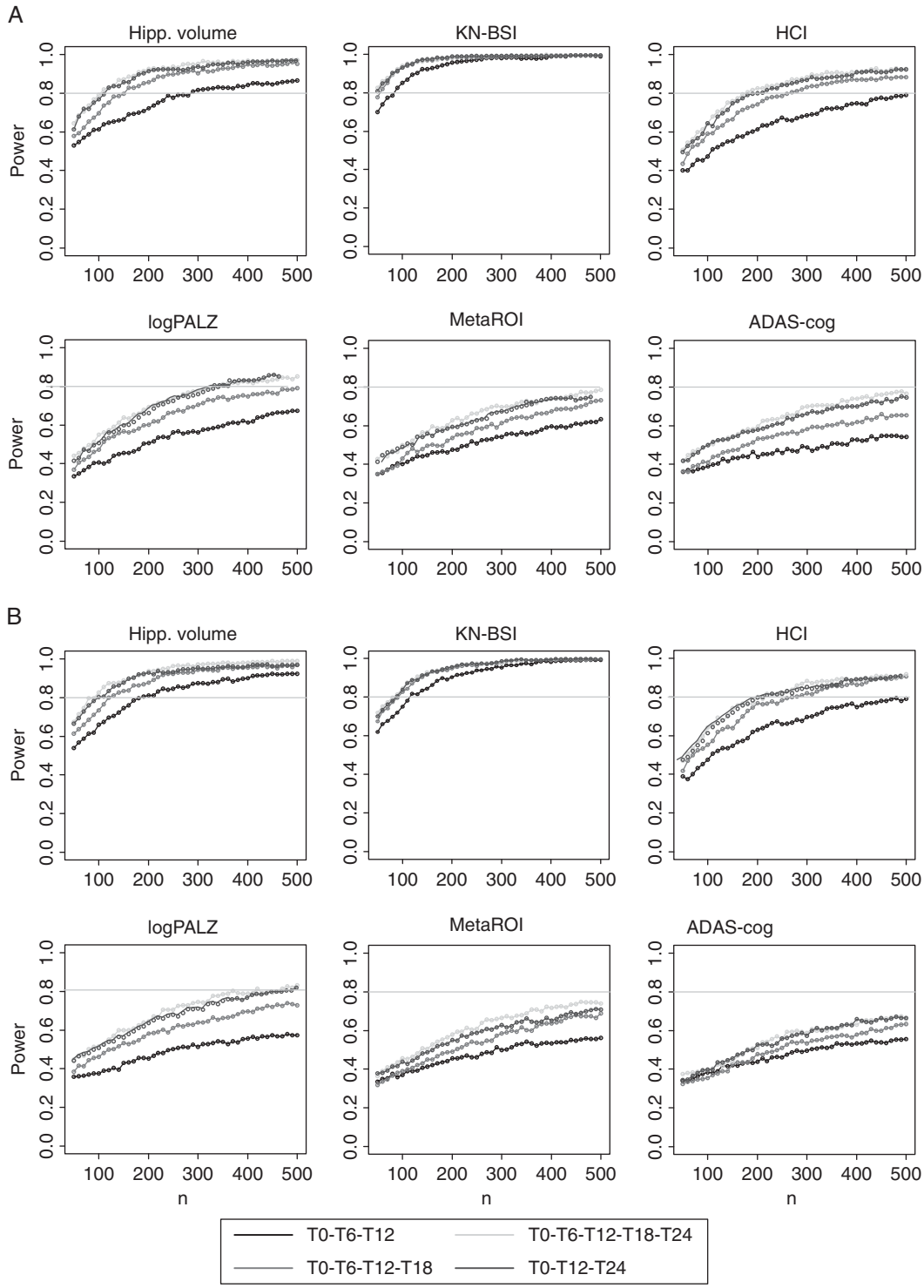

FIGURE 2. Estimated power of a hypothetical clinical trial designed to detect 20% reduction in biomarker slope in mild cognitive impairment (MCI) patients with abnormal cerebrospinal fluid Aβ 1-42 concentration [ABETA+, (A)] or positive to hippocampal atrophy [HIPPO+, (B)], as a function of sample size (per treatment arm) and follow-up time sets. Significance level was set to $\alpha = 0.05$. For all biomarkers and both MCI groups, the power increased with increasing follow-up time, irrespective of biomarker assessment frequency. MR measures showed highest power (with KN-BSI outperforming hippocampal volume), and a nonlinear trend. ADAS-COG indicates Alzheimer's Disease Assessment Scale-cognitive subscale; HCI, hypometabolic convergence index[17]; Hipp. volume, hippocampal volume automatically computed by Freesurfer algorithm; KN-BSI, brain atrophy rate measured by KN boundary shift integral technique[21]; logPALZ, log-transformed PMOD Alzheimer score[16]; MetaROI, FDG-PET summary metric based on meta-analitically derived regions of interest reflecting AD hypometabolism pattern[18]; T0, baseline, Tn, n-month follow-up.

TABLE 3. Sample Size Per Treatment Arm Needed to Obtain 20% Reduction in the Slope, for $\alpha = 0.05$ and $\beta = 0.2$, Estimated Via a Simulation Algorithm in MCI Patients With Abnormal CSF Aβ 1-42 Concentration (MCI ABETA+) or Positive to Hippocampal Atrophy (MCI HIPPO+). In Both MCI Groups, Sample Size Decreased With Increasing Follow-up Time for all Biomarkers, Irrespective of Biomarker Assessment Frequency, and MR Measures Required Lowest Sample Size. For all Time Sets, Sample Size was Higher in MCI HIPPO+ Than in MCI ABETA+ Group (all Biomarkers but Hippocampal Volume)

	T0-T6-T12-T18-T24	T0-T6-T12-T18	T0-T6-T12	T0-T12-T24
MCI ABETA +				
ADAS-COG	568	> 1000	> 1000	991
logPALZ	326	507	> 1000	343
HCI	175	263	512	185
MetaROI	562	865	> 1000	680
Hipp. volume	102	144	282	112
KN-BSI	46	54	78	48
MCI HIPPO +				
ADAS-COG	> 1000	> 1000	> 1000	> 1000
logPALZ	367	792	> 1000	468
HCI	198	263	532	204
MetaROI	649	> 1000	> 1000	969
Hipp. volume	84	120	188	99
KN-BSI	77	87	117	85

ADAS-COG indicates Alzheimer's Disease Assessment Scale-cognitive subscale; HCI, hypometabolic convergence index[17]; Hipp. volume, hippocampal volume automatically computed by Freesurfer algorithm; KN-BSI, brain atrophy rate measured by KN boundary shift integral technique[21]; logPALZ, log-transformed PMOD Alzheimer score[16]; MetaROI, FDG-PET summary metric based on meta-analitically derived regions of interest reflecting AD hypometabolism pattern[18]; T0, baseline; Tn, n-month follow-up.

Indeed, previous studies have found that MRI[25–28] and FDG-PET biomarkers[17–19] clearly outperform cognitive tests as outcome measures of rates of change both in AD and MCI patients, regardless of statistical methods and model assumptions used.[29] Baseline MRI measures, particularly hippocampal volume, outperformed measures of glucose hypometabolism in terms of effect size in preclinical and early AD.[30] Estimated sample sizes were lowest for MRI measures of hippocampal volume,[25,26] and enthorinal cortex[31] followed by those for prespecified FDG ROIs and cognitive scores.

In a recent review, Weiner et al[32] showed that using MRI, FDG-PET, or cognitive biomarkers as outcome measures in MCI clinical trials would require tens to few hundreds, hundreds to few thousands, and thousands of patients respectively to detect a 25% reduction with 80% power and 5% significance. Restricting enrollment to MCI groups enriched based on CSF biomarkers or structural atrophy was shown to reduce sample size by one half.[24]

Despite their overall consistent findings, previous studies found quite different sample size and power estimates due to different methodologies adopted. Moreover, they limited investigation to just a few biomarkers or a single duration of observation. The current study intended to move a step forward by comparing head-to-head the performance and power of the best-established diagnostic biological markers at a time, in 2 enriched MCI cohorts, over different time sets, using a simulation technique, thus providing pieces of information which could be useful to optimize future clinical trial design.

We found that for all biomarkers and both MCI groups the power increased with increasing duration of follow-up with the main differences observed for study durations of 12 to 18 months. These results are in line with a previous MR study showing that hippocampal atrophy power increases with time of observation.[25] We showed that the power did not significantly increase with 6-monthly biomarker assessment compared with yearly assessment; to our knowledge, no previous study investigated this aspect of study design.

MR measures, beyond having highest estimated power for all follow-up time sets, showed a nonlinear power trend, reaching an early plateau. Conversely, most of the other biomarkers increased in power linearly with increasing sample size. This key finding suggests that, in case MR biomarkers are used as outcome measures, it is useless to increase sample size beyond a given threshold size, as it would result in a negligible increase in power.

From a methodological point of view, there are a number of issues that deserve discussion. All of them were addressed in the pertinent section of the supplementary material (Supplemental Digital Content, http://links.lww.com/WAD/A106, section 2).

The current study has a number of limitations. First, biomarkers included in the study do not represent all potential outcome measures. They were chosen among the best-established biomarkers of AD progression, based on data availability in the ADNI longitudinal data set. Among FDG-PET biomarkers, 3 summary metrics of AD-like hypometabolism (logPALZ, HCI, and metaROI), previously shown to be sensitive measures of change in cognition in AD and MCI patients,[17–19] were included in the study. Unlike most other markers included in this analysis, logPALZ had originally been developed and validated in a completely independent sample.[16] Conclusions related to use of FDG-PET as a biomarker, therefore, are limited to the specific methods used for quantification. It would have been interesting to include in the study FDG-PET biomarkers specifically designed to track progression in MCI (such as empirically pre-defined statistical region of interest [sROI] measure[33]), but this was not possible as only few data were available, especially at follow-up. Among structural MRI biomarkers, we chose to include hippocampal volume, previously shown to parallel and precede cognitive decline,[29,34] and the KN-BSI measure of brain atrophy rate,

a longitudinal measure (unlike hippocampal volume and all of the FDG-PET measures, which are cross-sectional in nature) specifically designed to optimally track disease progression.[22] Only automatically computed hippocampal volumes were considered as no manual tracing was available in the ADNI data set and semiautomated volumes were available only for a few time points. Among clinical markers we chose ADAS-COG, rather than other clinical scores known to be even more sensitive for the early stages of AD (eg, Clinical Dementia Rating scale – Sum of Boxes, CDR-SB), as the former is widely adopted as outcome measure in clinical trials.

It would have been interesting to consider a third MCI group, enriched based on amyloid imaging (which was recently qualified as enrichment biomarker to enroll predemented AD subjects in clinical trials, beyond CSF Aβ 1-42 and hippocampal atrophy), but this was not possible due to paucity of data available (just around 20 ADNI subjects have baseline PIB data, and such a limited number would have resulted in unreliable findings; about 50 MCI patients have 12-mo PIB data, 34 of whom are PIB + , but considering 12-mo follow-up visit as baseline would have entailed to have limited follow-up information and would have biased the comparisons). Moreover, we could investigate an additional cohort represented by patients who have both CSF and MRI evidence of AD; despite a previous study showing significant advances to combining evidence for both biomarkers,[24] we chose not to include it in the study, as we are not aware of any clinical trial using > 1 enrichment biomarker at a time.

KN-BSI and hippocampal volume quality control information, albeit available, were not used, and data were included even in case of failure of the algorithm-specific quality control. Data derived from a failed run of the algorithm may be less reliable but, in contrast, if we used only data passing the algorithm-specific quality control, sample size estimates would need to be inflated to account for the percentage of scans that might fail. Retrospective checks revealed that all data used in the study passed KN-BSI quality control (data were rated at least as borderline acceptable); all baseline data passed Freesu rfer-based hippocampal volume quality control, whereas a limited number of follow-up data (about 12% HIPPO + and 22% ABETA + data) did not pass it.

The probabilities of the biomarker rankings proposed in this study seem low, but are still fairly high compared with the probabilities of chance (if all rankings were equally likely, the probability of a given ranking of the 6 biomarkers would be 0.001389).

Finally, it has been recently pointed out that change in patients should be considered relative to change in healthy controls, rather than in absolute terms, to have reliable sample size estimates for treatments targeting amyloid-related pathology.[31] Even though we basically agree with this observation, as the annualized brain change in healthy controls were shown to be much lower than in MCI and AD patients,[35] we believe that such correction would negligibly modify sample size estimates over a limited time (2 y) period.

In conclusion, in the current study we provided evidence that imaging biomarkers indeed outperform clinical markers of AD progression, widely used in the past as outcome measures for clinical trials, and that MRI measures (especially KN-BSI measure of brain atrophy) are the most effective outcome measures in both ABETA + and HIPPO + MCI groups, even for short (12 mo) duration clinical trials and yearly observations. These findings provide information about the biomarker enrichment and outcome measures that could be employed to reduce MCI patient samples and treatment duration in future clinical trials, and could drive the choice of surrogate outcome measure with respect to the mechanism of action of the drug under trial. In contrast, the assessment of imaging biomarkers in clinical trials has a number of drawbacks which need to be taken into account, in terms of cost, availability, and required experience, and bring a number of unresolved issues regarding reliability and standardization over different centers. Future studies aimed at comparing feasibility and cost-effectiveness of the use of biomarkers in clinical trials and clinical practice are needed.

REFERENCES

1. Selkoe DJ. Resolving controversies on the path to Alzheimer's therapeutics. *Nat Med*. 2011;17:1060–1065.
2. Selkoe DJ. Preventing Alzheimer's disease. *Science*. 2012;337: 1488–1492.
3. Aisen PS. Clinical trial methodologies for disease-modifying therapeutic approaches. *Neurobiol Aging*. 2011;32(suppl 1): S64–S66.
4. 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.
5. Lorenzi M, Donohue M, Paternicò D, et al. Alzheimer's Disease Neuroimaging Initiative. Enrichment through biomarkers in clinical trials of Alzheimer's drugs in patients with mild cognitive impairment. *Neurobiol Aging*. 2010;31:1443–1451.
6. Cummings JL. Biomarkers in Alzheimer's disease drug development. *Alzheimers Dement*. 2011;7:e13–e44.
7. Vellas B, Andrieu S, Sampaio C, et al. European Task Force Group. Endpoints for trials in Alzheimer's disease: a European task force consensus. *Lancet Neurol*. 2008;7:436–450.
8. Morris JC, Selkoe DJ. Recommendations for the incorporation of biomarkers into Alzheimer clinical trials: an overview. *Neurobiol Aging*. 2011;32(suppl 1):S1–S3.
9. Herholz K. Use of FDG PET as an imaging biomarker in clinical trials of Alzheimer's disease. *Biomark Med*. 2012;6: 431–439.
10. Ciumas C, Montavont A, Ryvlin P. Magnetic resonance imaging in clinical trials. *Curr Opin Neurol*. 2008;21:431–436.
11. Vellas B, Hampel H, Rougé-Bugat ME, et al. Task Force Participants. Alzheimer's disease therapeutic trials: EU/US Task Force report on recruitment, retention, and methodology. *J Nutr Health Aging*. 2012;16:339–345.
12. Petersen RC, Smith GE, Waring SC, et al. Mild cognitive impairment: clinical characterization and outcome. *Arch Neurol*. 1999;56:303–308.
13. Shaw LM, Vanderstichele H, Knapik-Czajka M, et al. Alzheimer's Disease Neuroimaging Initiative. Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. *Ann Neurol*. 2009;65:403–413.
14. Prestia A, Caroli A, van der Flier WM, et al. Prediction of dementia in MCI patients based on core diagnostic markers for Alzheimer's disease. *Neurology*. 2013;80:1048–1056.
15. Jagust WJ, Bandy D, Chen K, et al. The Alzheimer's Disease Neuroimaging Initiative positron emission tomography core. *Alzheimers Dement*. 2010;6:221–229.
16. Herholz K, Salmon E, Perani D, et al. Discrimination between Alzheimer dementia and controls by automated analysis of multicenter FDG PET. *Neuroimage*. 2002;17:302–316.
17. Herholz K, Westwood S, Haense C, et al. Evaluation of a calibrated (18)F-FDG PET score as a biomarker for progression in Alzheimer disease and mild cognitive impairment. *J Nucl Med*. 2011;52:1218–1226.

18. Chen K, Ayutyanont N, Langbaum JB, et al. Alzheimer's Disease Neuroimaging Initiative. Characterizing Alzheimer's disease using a hypometabolic convergence index. *Neuroimage*. 2011;56:52–60.
19. Landau SM, Harvey D, Madison CM, et al. Alzheimer's Disease Neuroimaging Initiative. Associations between cognitive, functional, and FDG-PET measures of decline in AD and MCI. *Neurobiol Aging*. 2011;32:1207–1218.
20. 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.
21. Fischl B, Salat DH, Busa E, et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*. 2002;33:341–355.
22. Leung KK, Clarkson MJ, Bartlett JW, et al. Alzheimer's Disease Neuroimaging Initiative. Robust atrophy rate measurement in Alzheimer's disease using multi-site serial MRI: tissue-specific intensity normalization and parameter selection. *Neuroimage*. 2010;50:516–523.
23. R Development Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2011.
24. Holland D, McEvoy LK, Desikan RS, et al. Alzheimer's Disease Neuroimaging Initiative. Enrichment and stratification for predementia Alzheimer disease clinical trials. *PLoS One*. 2012;7:e47739.
25. Schuff N, Woerner N, Boreta L, et al. MRI of hippocampal volume loss in early Alzheimer's disease in relation to ApoE genotype and biomarkers. *Brain*. 2009;132:1067–1077.
26. Holland D, Brewer JB, Hagler DJ, et al. Subregional neuroanatomical change as a biomarker for Alzheimer's disease. *Proc Natl Acad Sci USA*. 2009;106:20954–20959.
27. Hua X, Hibar DP, Ching CR, et al. Alzheimer's Disease Neuroimaging Initiative. Unbiased tensor-based morphometry: improved robustness and sample size estimates for Alzheimer's disease clinical trials. *Neuroimage*. 2013;66:648–661.
28. Gutman BA, Hua X, Rajagopalan P, et al. Alzheimer's Disease Neuroimaging Initiative. Maximizing power to track Alzheimer's disease and MCI progression by LDA-based weighting of longitudinal ventricular surface features. *Neuroimage*. 2013;70:386–401.
29. Ard MC, Edland SD. Power calculations for clinical trials in Alzheimer's disease. *J Alzheimers Dis*. 2011;26(suppl 3): 369–377.
30. Karow DS, McEvoy LK, Fennema-Notestine C, et al. Alzheimer's Disease Neuroimaging Initiative. Relative capability of MR imaging and FDG PET to depict changes associated with prodromal and early Alzheimer disease. *Radiology*. 2010;256:932–942.
31. Holland D, McEvoy LK, Dale AM, Alzheimer's Disease Neuroimaging Initiative. Unbiased comparison of sample size estimates from longitudinal structural measures in ADNI. *Hum Brain Mapp*. 2012;33:2586–2602.
32. Weiner MW, Veitch DP, Aisen PS, et al. Alzheimer's Disease Neuroimaging Initiative. The Alzheimer's Disease Neuroimaging Initiative: a review of papers published since its inception. *Alzheimers Dement*. 2012;8(suppl):S1–S68.
33. Chen K, Langbaum JB, Fleisher AS, et al. Alzheimer's Disease Neuroimaging Initiative. 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*. 2010;51:654–664.
34. den Heijer T, van der Lijn F, Koudstaal PJ, et al. A 10-year follow-up of hippocampal volume on magnetic resonance imaging in early dementia and cognitive decline. *Brain*. 2010;133:1163–1172.
35. Henneman WJ, Sluimer JD, Barnes J, et al. Hippocampal atrophy rates in Alzheimer disease: added value over whole brain volume measures. *Neurology*. 2009;72:999–1007.

