ORIGINAL ARTICLE

Neuroimaging Enrichment Strategy for Secondary Prevention Trials in Alzheimer Disease

Linda K. McEvoy, PhD,* Steven D. Edland, PhD,†‡ Dominic Holland, PhD,†
Donald J. Hagler, Jr, PhD,* J. Cooper Roddey, PhD,† Christine Fennema-Notestine, PhD,*§
David P. Salmon, PhD,† Alain K. Koyama, BS,† Paul S. Aisen, MD, PhD,†
James B. Brewer, MD, PhD,*† and Anders M. Dale, PhD*† For the Alzheimer's Disease
Neuroimaging Initiative

Abstract: We examined the improvement in statistical power that could be obtained in therapeutic trials for early (predemential) Alzheimer disease by constraining enrollment to individuals with amnestic mild cognitive impairment (MCI) and an atrophy pattern on a screening magnetic resonance imaging (MRI) scan previously found to be predictive of clinical decline, or to individuals with MCI and the apolipoprotein E ε4 genetic risk factor for Alzheimer disease. Treatable effects were defined as absolute change versus change relative to healthy controls (HCs). Data from 168 HC and 299 MCI participants were analyzed to determine sample sizes required to detect 25% slowing in mean rate of decline using global function, cognitive function, and structural measures as outcome variables. Reductions in estimated sample sizes of 10% to 43% were observed using the genetic enrichment strategy; reductions of 43% to 60% were observed with the neuroimaging enrichment strategy. Sample sizes needed to detect slowing in rate of atrophy in MCI relative to HC were dramatically larger than those needed to detect absolute change in atrophy rates. Constraining enrollment to MCI subjects with predictive atrophy on a screening MRI scan could improve the efficiency of clinical trials. Failure to take into account normal age-related changes risks under-powering trials designed to test disease-modifying properties of potential treatments.

Key Words: mild cognitive impairment, magnetic resonance imaging, atrophy, clinical trials methodology/study design, apolipoprotein E ε4

(*Alzheimer Dis Assoc Disord* 2010;24:269–277)

There is growing concern that therapeutic interventions aimed at slowing or halting the neurobiologic processes underlying Alzheimer disease (AD) may be only minimally effective when administered to individuals meeting clinical criteria for dementia. Clinical trials of such therapies are more likely to succeed if tested in individuals who are in an early, predemential stage of AD but are nonetheless likely to experience clinical deterioration over a relatively short period of time, if left untreated. Individuals with mild cognitive impairment (MCI)[1,2] represent a possible target population for clinical trials as MCI is associated with an increased risk of progression to a diagnosis of probable AD, with rates of 5% to 16% per year relative to 1% to 2% for the general population.[3–5] However, MCI is a heterogeneous condition, not a perfect predictor of AD: some individuals deteriorate rapidly, others remain stable for many years, and some revert to normal cognitive status. The ability to restrict enrollment in clinical trials to the subset of MCI individuals most likely to have preclinical AD and to experience rapid decline could improve the statistical power of clinical trials, enabling the use of smaller sample sizes and shorter trial periods—factors that could substantially reduce the cost or required duration of the trial.

Here we investigate the relative benefit in statistical power that could be achieved by a genetic or a structural

neuroimaging enrichment strategy. The apolipoprotein E (APOE) ε4 allele is a well-known risk factor for AD and confers a higher risk of developing AD in a dose-specific manner.[6] Constraining enrollment to those MCI participants with at least one APOE ε4 allele is thus one potential method for enhancing the power of a clinical trial. Prior research suggests that structural atrophy, evident on magnetic resonance imaging (MRI) data, also confers a higher risk of rapid clinical decline in individuals with MCI.[7–9] Thus another potential enrichment strategy is to constrain enrollment to MCI individuals who show, on a baseline structural MRI, an atrophy pattern earlier shown to be predictive of rapid clinical decline.[8] We assessed the impact of these potential enrichment strategies on statistical power for 2 commonly used outcome variables in AD treatment trials: the Clinical Dementia Rating Scale Sum of Boxes score (CDR-SB),[10] and the Cognitive subscale of the AD Assessment Scale (ADAS-Cog).[11] We also examined 4 structural measures as potential outcome variables, including nonspecific measures that have been shown to be sensitive to progressive degeneration in AD (whole brain and ventricular volumes),[12–14] and mesial temporal structures (hippocampus and entorhinal cortex) that are early targets of AD pathology[15] and sensitive to changes in early and prodromal AD.[12,16–20] Sample sizes were estimated with and without considering the effects of normal aging.

METHODS

Alzheimer's Disease Neuroimaging Initiative

Data used in this study were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI).[21,22] ADNI is an ongoing longitudinal study carried out by a broad consortium of academic institutions and private corporations designed to test whether serial MRI, positron emission tomography, other biologic markers, and clinical and neuropsychologic assessment can be combined to effectively measure the progression of MCI and early AD. Determination of sensitive and specific markers of early AD progression may help researchers and clinicians develop new treatments and monitor their effectiveness, and lessen the time and cost of clinical trials. ADNI was launched in 2003 by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, the Food and Drug Administration, private pharmaceutical companies, and nonprofit organizations, as a $60 million, 5-year public-private partnership.

ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations. ADNI has recruited 229 cognitively normal individuals to be followed for 3 years, 398 people with amnestic MCI to be followed for 3 years, and 192 people with mild AD to be followed for 2 years (see www.adni-info.org). The research protocol was approved by each local institutional review board and written informed consent was obtained from each participant or participant's guardian.

Participants

Data from 168 healthy controls (HCs) and 299 amnestic MCI participants were analyzed. ADNI general eligibility criteria are described at http://www.adni-info.org/index.php?option = com_content&task = view&id = 9&Itemid = 43.

In brief, subjects are 55 to 90 years of age, generally healthy, nondepressed, have a modified Hachinski score of 4 or less, and a study partner able to provide an independent evaluation of functioning. Inclusion criteria for MCI subjects include Mini-Mental State Examination scores between 24 and 30, a subjective memory complaint, objective memory loss measured by education-adjusted scores on the Wechsler Memory Scale Logical Memory Test-II, a CDR of 0.5, preserved activities of daily living, and an absence of dementia.[23] Full details of the ADNI study are publicly available at: http://www.adni-info.org/.

Subjects were included in this study if MRI data from the baseline scan and from at least one of the 6, 12, 18, or 24-month follow-up intervals were available and met local quality review procedures. Seven HCs who converted to a diagnosis of MCI or AD during the time course of the study were excluded. Data included here reflect data available as of August 24, 2009 and processed locally as of September 5, 2009. Of the 168 HCs, 93% had acceptable 6-month follow-up MRI data, 86% had acceptable 12-month follow-up data, and 70% had acceptable 24-month follow-up data (HCs were not evaluated at 18 mo in the ADNI). Of the 299 MCI subjects, 89% had acceptable 6-month follow-up data, 79% had acceptable 12-month follow-up data, 71% had acceptable 18-month follow-up data, and 61% had acceptable 24-month follow-up data.

Procedures

Raw DICOM MRI scans (including 2 T_1-weighted volumes per subject per time visit) were downloaded from the public ADNI site (http://www.loni.ucla.edu/ADNI/Data/index.shtml). These data were collected across a variety of scanners with protocols individualized for each scanner,[22] as defined at http://www.loni.ucla.edu/ADNI/Research/Cores/index.shtml. In our laboratory, MRI data were reviewed for quality, automatically corrected for spatial distortion due to gradient nonlinearity[24] and B1 field inhomogeneity,[25] registered, and the 2 volumes per subject were averaged to improve signal-to-noise. Volumetric segmentation[26,27] and cortical surface reconstruction[28–31] and parcellation methods[32,33] based on the FreeSurfer software package were used to quantify regional atrophy. The morphometric procedures used to quantify regional thickness and volumes in baseline scans are described in detail elsewhere.[34] Results of automated labeling of the hippocampus and entorhinal cortex are shown for a representative MCI participant in Figure 1. Approximately 5% of HC and 11% of MCI subjects' baseline MRI data failed local quality review, primarily because of extreme white matter disease or atrophy.

To quantify longitudinal change in brain structural measures, all available MRI data from the 6, 12, 18 and 24-month follow-up visits were analyzed. Dual T_1-weighted follow-up scans for each subject at each follow-up time point were corrected for spatial distortion due to gradient nonlinearity and B_1 field inhomogeneity, rigid-body aligned, averaged, then affine aligned with 12 degrees of freedom to the subject's baseline scan. A deformation field was calculated from nonlinear registration[35] and used to align scans at the sub-voxel level. The reduction of site-specific distortion effects and normalization of inhomogeneities improves the accuracy of the morphometric analysis[19] and minimizes the effects of instrumental drift on atrophy rate estimates. The follow-up aligned image
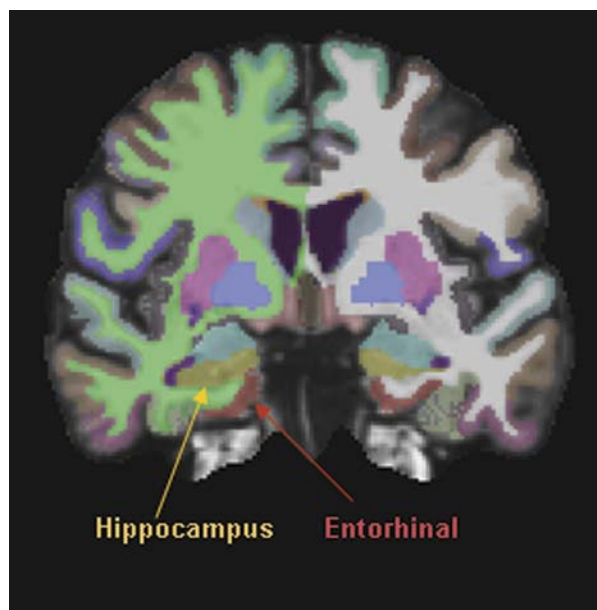
FIGURE 1. section for a participant with mild cognitive impairment showing the results of automated labeling of the hippocampus (shown in gold) and entorhinal cortex (shown in red). These regions are labeled on the left hemisphere.

FIGURE 2. Average thickness difference for AD and MCI subjects relative to HCs on baseline scans. The AD sample, shown here for comparison only, consisted of 161 subjects (81 males, mean age 74.93 ± 7.59 y, mean Mini-Mental State Examination score 23.31 ± 2.0). Top: HC versus AD; middle: HC versus MCI subjects with predictive atrophy; bottom: HC versus MCI subjects without predictive atrophy. Right: lateral; left: mesial views. Yellow and red areas indicate regions of thinning with disease. Scale reflects thickness differences ranging from − 0.3 (yellow) to +0.3 mm (blue). AD indicates Alzheimer disease; HC, healthy control; MCI, mild cognitive impairment.

underwent skull stripping and subcortical segmentation, with labels applied from the baseline scan. For cortical reconstruction, surface coordinates for the white and pial boundaries were derived from the baseline images and mapped onto the follow-up images using the deformation field. Parcellation and labeling from the baseline image was applied to the follow-up image, resulting in a one-to-one correspondence between each vertex in the baseline and follow-up images. Visual quality control, blind to diagnosis, was performed on the volume change field to exclude cases with significant degradation in meaningful registration for at least one region of interest, because of artifacts or major changes in scanner hardware between visits. The most common form of artifact was due to within-scan subject motion. Loss of scans due to motion artifacts may be greatly reduced in future trials by using real-time motion correction procedures.[36,37] Quality control procedures on longitudinal data resulted in rejection of approximately 10% of HC subjects and 12% of MCI subjects who had acceptable baseline MRI data, resulting in the final sample size of 168 HC and 299 MCI.

To identify MCI individuals with the baseline atrophy pattern earlier found to be predictive of decline, we used a linear discriminant model based on that obtained in our prior study.[8] In that study, we trained a linear discriminant classifier on data from 139 HC and 84 AD subjects from the ADNI and found a pattern of baseline regional atrophy that discriminated HC from AD data with a fully cross validated sensitivity of 83% and specificity of 93%.[8] The pattern involved atrophy, relative to HCs, in the hippocampus, entorhinal cortex, middle temporal gyrus, bank of the superior temporal sulcus, isthmus cingulate, superior temporal gyrus, and medial and lateral orbital frontal gyri. Application of this discriminant model to the 299 MCI subjects here (including 151 subjects

from our prior report[8]) resulted in classification of 151 individuals as showing the atrophy pattern predictive of decline at baseline, and 148 who did not (Fig. 2). Characteristics of the HC and the 2 MCI subgroups are shown in Table 1. For the genetic enrichment strategy, 170 MCI participants had at least one APOE ε4 allele and 129 did not. Characteristics of these MCI subgroups are shown in Table 2.

Statistical Analyses

Differences between groups on demographic and outcome measures were assessed with χ^2 analyses for categorical variables and analyses of variances for continuous variables. When significant main effects of group were observed in the analysis of variance, post-hoc comparisons were performed to determine differences between the pairs of the groups using Bonferroni corrections for multiple comparisons.

The sample size required to detect 25% slowing in mean rate of decline for a hypothetical disease-modifying treatment versus placebo was estimated for a 24 month, 2-arm, equal allocation trial, with a 6-month assessment interval. Power calculations were performed with the requirement that the trial have 80% power to detect the treatment effect using a 2-sided significance level of 5%. Power calculations are for a linear-mixed effects model

TABLE 1. Characteristics of the HC Group and the 2 MCI Subgroups Based on Presence or Absence of an AD-like Pattern of Regional Atrophy on a Screening MRI

Group	HC (n = 168)	MCI Without Atrophy (n = 148)	MCI With Atrophy (n = 151)	Statistical Comparison
Age (95% CI), y	76.2 ± 5.15 (75.1-77.2)	74.6 ± 7.43 (73.5-75.7)	74.7 ± 7.37 (73.6-75.8)	$F_{(2,464)} = 2.78$; $P = 0.063$
Sex	82 (49%) Female	54 (36%) Female	57 (38%) Female	$\chi^2_{(2)} = 6.11$; $P = 0.047$
Education (95% CI)	16.0 ± 2.8 (15. 6-16.4)	15.6 ± 3.0 (15.1-16.0)	16.0 ± 2.9 (15.5-16.4)	$F_{(2,464)} = 1.03$; $P = 0.36$
APOE ε4+	47 (28%)	76 (51%)	94 (62%)	$\chi^2_{(2)} = 39.6$; $P < 0.001$
MMSE (95% CI)	29.1 ± 1.07 (28.9-29.4)	27.4 ± 1.71 (27.1-27.6)	26.7 ± 1.75 (26.4-26.9)***	$F_{(2,464)} = 112$; $P < 0.001$
ADAS-Cog (95% CI)	6.0 ± 2.84 (5.5-6.6)	10.1 ± 3.94 (9.5-10.7)	13.0 ± 4.21 (12.4-13.6)***	$F_{(2,464)} = 145$; $P < 0.001$
CDR-SB (95% CI)	0.03 ± 0.12 (−0.08-0.14)	1.44 ± 0.77 (1.33-1.56)	1.69 ± 0.97 (1.57-1.80)**	$F_{(2,464)} = 262$; $P < 0.001$

Mean values ± SDs shown for all continuous variables.

Statistical comparison reflects the comparison across the 3 groups. Asterisks in the "MCI With Atrophy" column indicates the significance of the difference between the 2 MCI subgroups. Absence of asterisks in this column indicates nonsignificant differences ($P > 0.05$) between the MCI subgroups.

**$P \leq 0.01$.

***$P \leq 0.001$.

95% CI indicates 95% confidence intervals; ADAS-Cog, score on the cognitive subscale of the Alzheimer's Disease Assessment Scale; CDR-SB, Clinical Dementia Rating Scale, sum of boxes score; HC, healthy control; MCI, mild cognitive impairment; MMSE, Mini-Mental State Examination; MRI, magnetic resonance imaging.

analysis[38] comparing mean rate of decline in the treatment arm versus the untreated arm using the formula:

$$\frac{N}{\text{arm}} = \frac{2\left(\sigma_b^2 + \frac{\sigma_e^2}{\sum_i \left(t_i - \bar{t}\right)^2}\right)\left(z_{1-\alpha/2} + z_{1-\beta}\right)^2}{\Delta^2}$$

where σ_b^2 is the variance of the random slopes within the MCI group and σ_e^2 is the residual error variance of the mixed effects model. The detectable effect size, Δ, is set to 25% of the mean rate of decline observed in the MCI subjects when effects of normal aging are not taken into account ("absolute change"); otherwise the detectable effect size is set to 25% of the mean rate of decline observed in MCI subjects minus that observed in HC subjects ("relative change"). For structural outcome measures, change at each time interval was calculated as percent change from baseline values. Parameters for power calculations were estimated using the "lmer" function in the R statistical package "lme4". The 95% confidence intervals for the

estimated sample sizes were obtained from 1000 bootstrap samples, where bootstrap resampling was with replacement. The number of values per bootstrap sample was equal to the number in the original sample within HCs and MCI cases.

To facilitate comparison with other studies that have estimated sample sizes needed to detect slowing in decline in MCI participants,[39,40] we also calculated sample sizes based on a linear mixed effects model ignoring between-subject variance in rate of change (ie, with random intercepts but taking the group-specific rate of change as a fixed effect). These results are presented in Supplementary Table 1, Supplemental Digital Content 1, http://links.lww.com/WAD/A3.

RESULTS

Sample Size Implications of the Enrichment Strategies

Sample size implications of the genetic and neuro-imaging enrichment strategies for clinical and structural

TABLE 2. Characteristics of the 2 MCI Subgroups Based on Genetic Risk

Group	APOE ε4– (n = 129)	APOE ε4+ (n = 170)
Age (95% CI), y	76.0 ± 8.09 (74.7-77.2)	73.6 ± 6.66 (72.5-74.7)**
Sex	47 (36%) Female	64 (38%) Female
Education (95% CI)	15.8 ± 2.94 (15.2-16.3)	15.8 ± 3.01 (15.3-16.2)
MMSE (95% CI)	27.0 ± 1.78 (26.7-27.4)	27.0 ± 1.75 (26.7-27.3)
ADAS-Cog (95% CI)	10.7 ± 4.29 (10.0-11.5)	12.2 ± 4.24 (11.6-12.8)**
CDR-SB (95% CI)	1.49 ± 0.86 (1.34-1.64)	1.62 ± 0.90 (1.49-1.76)

Absence of asterisks indicates nonsignificant differences ($P > 0.05$) between these MCI subgroups. See footnotes of Table 1 for further information .

**$P < 0.01$.

95% CI indicates 95% confidence intervals; ADAS-Cog, score on the cognitive subscale of the Alzheimer's Disease Assessment Scale; APOE ε4, apolipoprotein E ε4; APOE ε4 −, lack of an APOE ε4 allele; APOE ε4 + indicates presence of at least one APOE ε4 allele; CDR-SB, Clinical Dementia Rating Scale, sum of boxes score; MCI, mild cognitive impairment; MMSE, Mini-Mental State Examination.

TABLE 3. Sample Size Implications of the Neuroimaging and Genetic Enrichment Strategies for Clinical Outcome Variables When Effect of Age is (Relative Change) or is not (Absolute Change) Taken Into Account

	Annual Change ± SE (95% CI)	σ_e^2	σ_b^2	N/Arm Absolute Change (95% CI)	N/Arm Relative Change (95% CI)
CDR-SB					
HC	0.04 ± 0.02 (0.00-0.07)				
All MCI subjects	0.67 ± 0.06 (0.56-0.78)	0.65	0.79	437 (330-591)	492 (362-688)
MCI without APOE ε4	0.61 ± 0.08 (0.44-0.77)	0.62	0.76	501 (311-854)	571 (336-1005)
MCI with APOE ε4	0.72 ± 0.08 (0.57-0.87)	0.66	0.81	397 (275-566)	443 (299-659)
MCI without atrophy	0.37 ± 0.07 (0.23-0.52)	0.59	0.73	1203 (689-2344)	1494 (801-3402)
MCI with atrophy	0.97 ± 0.08 (0.82-1.12)	0.70	0.72	191 (133-270)	207 (141-301)
ADAS-Cog					
HC	− 0.34 ± 0.13 (− 0.59-− 0.10)				
All MCI subjects	1.47 ± 0.19 (1.09-1.84)	3.02	2.17	978 (662-1488)	642 (414-1036)
MCI without APOE ε4	1.15 ± 0.28 (0.60-1.79)	3.09	1.96	1442 (677-3760)	857 (413-1714)
MCI with APOE ε4	1.70 ± 0.26 (1.19-2.21)	2.96	2.34	774 (506-1253)	536 (336-882)
MCI without atrophy	0.62 ± 0.23 (0.17-1.07)	2.67	1.76	3873 (1555-26241)	1605 (702-5580)
MCI with atrophy	2.29 ± 0.29 (1.72-2.86)	3.33	2.26	458 (314-691)	346 (228-530)

"Annual Change" shows change in test score over a 1-year period. σ_b^2 = variance of the random slopes within the MCI group; σ_e^2 = the residual error variance of the mixed effects model. Annual change, variance measures, sample size estimates, and CIs, are shown when all MCI subjects are included in the analysis, or when the analysis is restricted to subgroups of MCI subjects based on presence or absence of APOE ε4, or on the presence or absence of structural atrophy on a screening magnetic resonance imaging scan.

95% CI indicates 95% confidence interval; ADAS-Cog, score on the cognitive subscale of the Alzheimer's Disease Assessment Scale; APOE ε4, apolipoprotein E ε4; CDR-SB, Clinical Dementia Rating Scale, Sum of Boxes score; HC, healthy control; MCI, mild cognitive impairment.

outcome measures are summarized in Tables 3 and 4, respectively, and shown in Figure 3. A treatment trial using the CDR-SB as the primary outcome measure and using the

same recruitment criteria as the ADNI would expect a mean increase of 0.67 points per year on the CDR-SB in the placebo arm. If enrollment were constrained to individuals

TABLE 4. Sample Size Implications of the Neuroimaging and Genetic Enrichment Strategies for Structural Outcome Variables, When Effect of Age is (Relative Change) or is Not (Absolute Change) Taken Into Account

	Annual % Change ± SE (95% CI)	σ_e^2	σ_b^2	N/Arm Absolute Change (95% CI)	N/Arm Relative Change (95% CI)
Whole brain					
HC	− 0.37 ± 0.04 (− 0.44-− 0.30)				
All MCI subjects	− 0.77 ± 0.04 (− 0.85-− 0.69)	0.40	0.60	181 (138-227)	679 (422-1097)
MCI without APOE ε4	− 0.64 ± 0.06 (− 0.77-− 0.52)	0.50	0.60	267 (179-394)	1493 (657-5758)
MCI with APOE ε4	− 0.86 ± 0.05 (− 0.96-− 0.76)	0.35	0.59	135 (86-176)	418 (238-620)
MCI without atrophy	− 0.59 ± 0.05 (− 0.68-− 0.50)	0.30	0.50	202 (138-278)	1462 (653-4787)
MCI with atrophy	− 0.94 ± 0.06 (− 1.06-− 0.82)	0.47	0.64	141 (96-192)	386 (232-642)
Inferior lateral ventricle					
HC	4.78 ± 0.42 (3.95-5.61)				
All MCI subjects	10.20 ± 0.50 (9.23-11.18)	3.68	7.82	161 (129-208)	568 (388-920)
MCI without APOE ε4	7.97 ± 0.67 (6.66-9.28)	3.62	6.77	202 (145-268)	1259 (612-3474)
MCI with APOE ε4	11.90 ± 0.69 (10.56-13.25)	3.72	8.21	129 (94-191)	361 (234-592)
MCI without atrophy	7.97 ± 0.62 (6.76-9.19)	3.67	6.73	200 (147-269)	1248 (627-3321)
MCI with atrophy	12.46 ± 0.74 (11.01-13.91)	3.68	8.34	121 (86-185)	319 (208-514)
Hippocampus					
HC	− 0.82 ± 0.08 (− 0.98-− 0.66)				
All MCI subjects	− 1.93 ± 0.10 (− 2.13-− 1.73)	0.85	1.58	186 (149-231)	559 (374-892)
MCI without APOE ε4	− 1.44 ± 0.14 (− 1.72-− 1.16)	0.81	1.43	278 (190-375)	1486 (688-4021)
MCI with APOE ε4	− 2.31 ± 0.14 (− 2.58-− 2.04)	0.88	1.59	133 (99-176)	320 (206-495)
MCI without atrophy	− 1.31 ± 0.12 (− 1.56-− 1.07)	0.81	1.33	294 (207-408)	2077 (900-8104)
MCI with atrophy	− 2.56 ± 0.15 (− 2.84-− 2.27)	0.90	1.57	107 (74-146)	231 (144-354)
Entorhinal					
HC	− 0.75 ± 0.12 (− 0.98-− 0.52)				
All MCI subjects	− 2.50 ± 0.11 (− 2.73-− 2.38)	0.86	1.79	140 (114-173)	285 (202-409)
MCI without APOE ε4	− 1.95 ± 0.16 (− 2.27-− 1.62)	0.88	1.67	206 (148-2784)	543 (316-1013)
MCI with APOEε4	− 2.92 ± 0.15 (− 3.21-− 2.63)	0.85	1.76	100 (75-134)	181 (124-271)
MCI without atrophy	− 1.71 ± 0.15 (− 1.99-− 1.42)	0.90	1.59	245 (174-346)	777 (418-1754)
MCI with atrophy	− 3.29 ± 0.15 (− 3.58-− 3.00)	0.83	1.62	67 (49-91)	113 (77-163)

"Annual Change" shows mean percent volume change from baseline over 1 year. For inferior lateral ventricles, hippocampus and entorhinal cortex, left and right hemisphere volumes were averaged. For other details, see footnotes of Table 3.

95% CI indicates 95% confidence interval; APOE ε4, apolipoprotein E ε4; HC, healthy control; MCI, mild cognitive impairment.
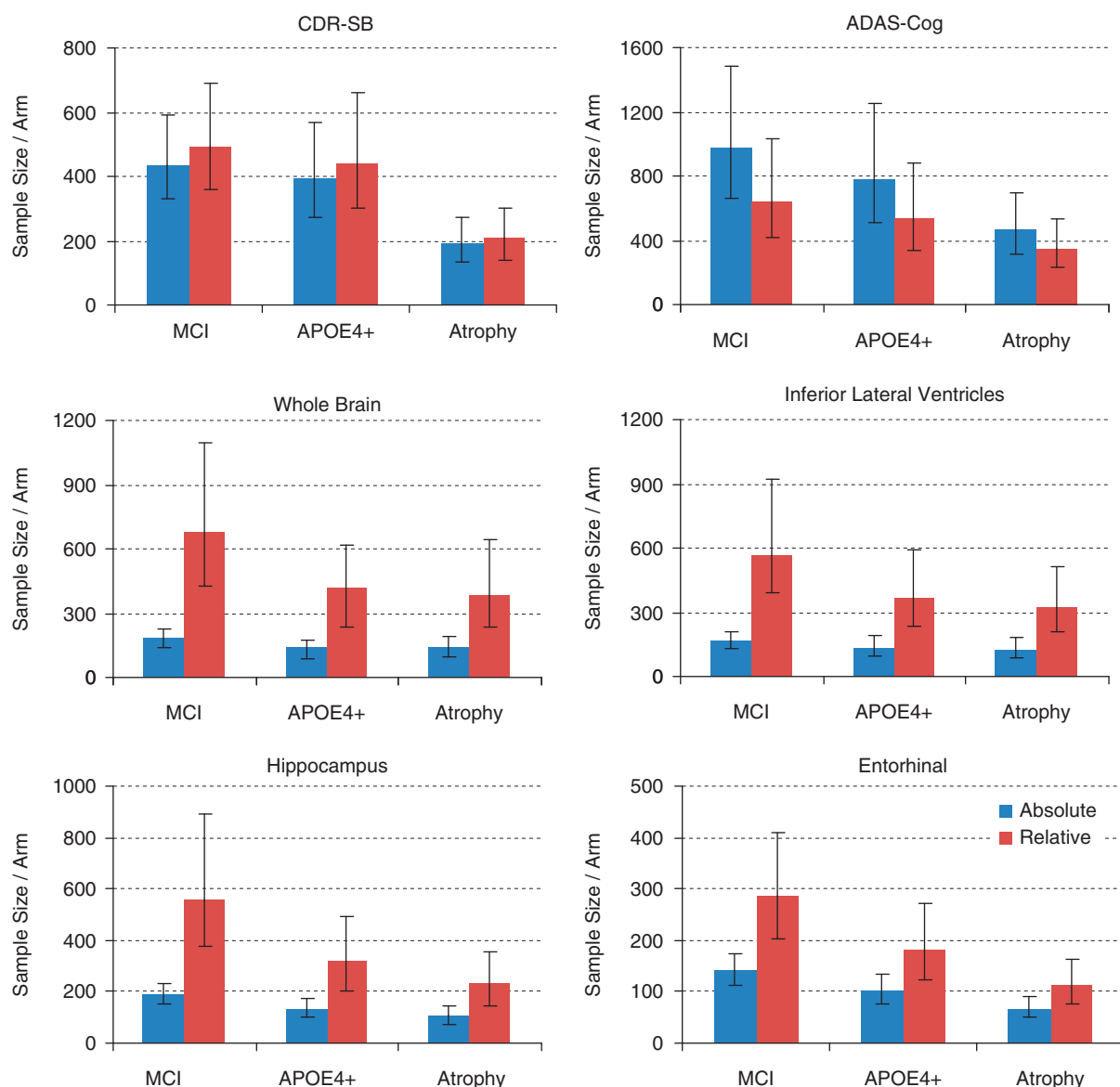
FIGURE 3. Estimated sample size per arm for the 6 outcome variables when age is (red) and is not (blue) taken into account. Results are shown for the genetic enrichment strategy [apolipoprotein E ε4 (APOE4+)] and the neuroimaging enrichment (atrophy) strategy relative to an unenriched sample [mild cognitive impairment (MCI)]. Note that the y-axis differs across outcome variables. Error bars show the 95% confidence intervals. ADAS-Cog indicates cognitive subscale of the Alzheimer's Disease Assessment Scale; CDR-SB, Clinical Dementia Rating Scale sum of boxes score.

with predictive atrophy, an annual increase of 0.97 points would be expected. The resulting increase in power due to the larger potential treatment effect would enable a 57% reduction in sample size (from 492 to 207 subjects/arm), based on detecting a 25% reduction in change relative to HCs. This reduction in sample size is statistically significant, as indicated by the nonoverlapping confidence intervals (Table 3). In contrast, if enrollment were constrained to MCI subjects with the genetic risk factor, the small increase in the potential treatment effect (0.72 points vs. 0.67 points) would permit only a 10% reduction in sample size (from 492 to 443 subjects/arm). Similar results were observed for the ADAS-Cog outcome measure, where a larger potential treatment effect with a neuroimaging

enrichment strategy relative to the full MCI group allowed for a 46% reduction in estimated sample size; the genetic enrichment strategy permitted only a 17% reduction in sample size.

For structural outcome measures, both the genetic and neuroimaging enrichment strategies resulted in smaller sample size estimates relative to that obtained using the full MCI group, but greater reductions were observed for the neuroimaging than for the genetic strategy (Table 4 and Fig. 3). Nonoverlapping confidence intervals for estimates obtained with the neuroimaging strategy relative to the full MCI group indicated that the sample size reductions permitted by the neuroimaging enrichment strategy were significant for the hippocampus and entorhinal cortex outcome

measures. Overall, the smallest estimated sample size, when the effects of age were taken into account, was observed for the neuroimaging enrichment strategy with the entorhinal cortex as the outcome variable (113 subjects/arm).

Sample Size Implications of Defining Treatable Effects as Absolute Versus Relative Change

For all structural measures, HCs showed an annual rate of change that ranged from 30% of the amount observed in the full MCI cohort (entorhinal cortex) to 48% (whole brain volume). The reduction in the treatable effect size when change in HCs was subtracted from change observed in MCI participants resulted in substantial increases in sample size estimates (Table 4). For example, the sample size needed to detect slowing of atrophy in whole brain volume in MCI subjects relative to an untreated control group was 37% of the size needed to detect change in excess of that experienced by HCs (181 vs. 679 subjects/arm, respectively; Table 4).

DISCUSSION

Heterogeneity within the MCI population with regard to rate of disease progression is problematic for clinical trials, creating the need for large sample sizes and long follow-up periods to ensure adequate statistical power for assessing disease-modifying treatment effects. Strategies that would enable such trials to enroll a more homogenous sample of MCI individuals who are at higher risk of imminent decline than the general MCI population could enhance the observed treatment effects, thereby increasing the power to observe those effects. Consideration of outcome variables and the definition of treatable effects are also vital to the design of clinical trials.

Enrichment Strategies

We evaluated the reduction in sample size that could be obtained by using 2 potential enrichment strategies relative to a trial using the same enrollment criteria as the ADNI; criteria that have been used in prior therapeutic trials.[4] A genetic enrichment strategy, based on the presence of an APOE ε4 allele, could be easily and inexpensively implemented. However, restricting trial enrollment to those with an APOE ε4 allele could limit generalizability of the results as beneficial therapeutic effects and adverse side-effects may differ as a function of genetic status. Furthermore, our results showed that although sample sizes were generally smaller with this strategy relative to an unenriched trial, the neuroimaging enrichment strategy offered even greater benefit.

Constraining enrollment to MCI individuals with a baseline pattern of atrophy earlier found to be predictive of clinical decline would allow a 58% reduction in sample size using the CDR-SB as the primary outcome variable, or a 60% reduction in sample size, using change relative to controls in the entorhinal cortex as the outcome variable. Such a reduction in sample size would offer substantial savings in a clinical trial, exceeding the cost associated with the neuroimaging analysis required for this strategy. As screening MRIs are routinely used in clinical trials, collection of MRI data poses no additional expense. The semi-automated methods based on the FreeSurfer software package[26–31] used here to quantify regional atrophy in MCI subjects can be efficiently applied to large samples: MRI data processing required approximately 45 minutes per subject for a trained technologist to manually review and edit the cortical surface according to minimal, objective editing rules, with 24 hours' computation time for image construction using a dual quad core Intel(R) Xeon(R) CPU E5420 with a processing speed of 2.50GHz and 16GB ram. Use of several CPU allows processing of multiple subjects' scans to occur in parallel.

Definition of Treatable Effects

We also evaluated the sample size implications of 2 methods of defining treatable effect size: absolute change and change relative to controls. As measures of global function are relatively stable in healthy individuals, use of absolute versus relative change had little impact on sample size estimates using CDR-SB as the outcome variable. For the ADAS-Cog, HCs showed a small improvement in scores over the 1-year period, presumably due to practice effects, whereas MCI subjects deteriorated, resulting in smaller sample size estimates for relative than for absolute change measures. Much larger differences in estimated sample sizes in the opposite direction were observed, however, for structural outcome measures.

Consistent with prior reports, we found that HCs exhibited significant 1-year reduction in whole brain, hippocampal and entorhinal cortex volumes, as well as ventricular expansion.[19,41–45] As a treatment designed to halt or slow the progression of AD is unlikely to affect brain changes associated with normal aging, failure to take into account the magnitude of change experienced by HCs may lead to substantial underestimation in sample sizes needed to detect a beneficial treatment effect on disease-related brain atrophy. Prior studies have reported that failure to control for effects of normal aging would result in sample size estimates approximately 25% to 35% smaller than that required to detect a therapeutic effect on disease-related slowing of whole brain,[14] hippocampal,[46] or entorhinal[46] atrophy in patients with AD. As the proportion of atrophy attributable to general aging is higher in individuals with MCI than in those with AD, we observed even greater differences in sample size estimates. For example, approximately 48% of the change in whole brain volume experienced in the full MCI sample could be attributed to general effects of aging, resulting in a sample size estimate of 73% smaller than that needed to detect a reduction in disease-related whole brain atrophy.

Direct comparison of the current results to prior studies that have examined sample sizes needed to detect slowing in rate of whole brain,[13,14] hippocampal,[17,39] or entorhinal[46] atrophy, or ventricular expansion[13] in AD and MCI patients is hindered by differences in subject populations; number and timing of follow-up MRIs; MRI quantification methods; and statistical analysis methods. Nevertheless, comparison of our results to recent studies that have analyzed data from ADNI's MCI participants reveals that our estimate for detecting slowing in rate of hippocampal atrophy is smaller than that reported by Schuff et al,[39] but consistent with the results of Hua et al,[40] who reported that 88 subjects per arm would be needed to detect 25% slowing in mean rate of temporal lobe atrophy. When we used a statistical model similar to that used by Hua et al (ie, by not fitting random slopes within the MCI group), we found that 73 subjects per arm would be sufficient for detecting 25% slowing in entorhinal atrophy and 95 subjects per arm for detecting slowing of hippocampal atrophy, relative to an untreated MCI group (See Supplementary Table 1, Supplemental Digital Content 1,

http://links.lww.com/WAD/A3). However, failure to take into account between-subject variability in rate of progression may result in substantial sample size underestimation, as can be observed by comparing results in Tables 3 and 4 for the full MCI group with results presented in Supplementary Table 1, Supplemental Digital Content 1, http://links.lww.com/WAD/A3.

Outcome Measures

We investigated the implication of enrichment strategies using symptomatic changes in global and cognitive function, and structural atrophy as outcome measures, rather than attainment of the clinical diagnosis of probable AD (ie, "conversion" to AD) within the period of the trial. As recently reviewed,[47] symptomatic changes in cognition and function may be more effective outcome measures than diagnostic conversion as the definition of conversion is inherently arbitrary in a disease characterized by continuous progression. The lack of diagnostic precision can negatively affect the accuracy and generalizability of trial results.[2,47,48] Furthermore, a continuous primary outcome measure, rather than a dichotomous measure based on conversion to probable AD, may allow for shorter treatment trials.[47]

The ADAS-Cog is routinely used as a primary cognitive outcome measure in AD treatment trials, but has been reported to be suboptimal for use in MCI due to its lack of sensitivity at this mild stage of the disorder.[47] The large sample size estimates needed to detect slowing in annual rate of change obtained here and in prior reports[39,40] are consistent with this. The CDR-SB is another commonly used outcome variable in AD clinical trials, and has been found to be sensitive to longitudinal decline in MCI treatment trials.[4,49] Consistent with this, we found that CDR-SB was more sensitive than the ADAS-Cog, requiring a sample size approximately half that required by the ADAS-Cog to detect a treatment effect.

Structural outcome variables are of interest because of their face validity as markers of disease-modifying properties of a treatment, and to their reduced variability compared with symptomatic measures.[12,17] Of the structural measures assessed here, change in entorhinal cortex was the most sensitive outcome measure. Using an enrichment strategy of constraining enrollment to MCI individuals with atrophy predictive of decline, a sample size of only 113 subjects per arm would be needed to detect a 25% slowing in mean rate of decline relative to that experienced by HCs.

Limitations

This study is limited by the lack of histologic verification of clinical status. It is certain that some HCs have AD pathology.[50,51] To minimize the risk of including HCs with preclinical AD, HC subjects who converted to a diagnosis of MCI or AD during any follow-up visit were excluded. In addition, some individuals with MCI may suffer from pathologies unrelated to AD. Continued follow-up with ADNI's MCI sample may help to address this issue. We noted that the absence of the predictive atrophy pattern at baseline in MCI individuals does not imply that these individuals are not in a preclinical AD state. As a group, the MCI individuals without predictive atrophy experienced progressive clinical and structural decline consistent with AD, but progressed at a slower rate than MCI individuals with predictive atrophy. We also note that we did not consider the effect of study subject dropout in the power calculations reported here. Sample sizes for actual trials should be larger than the numbers reported here to account for loss of power due to dropout.

CONCLUSIONS

Results show that an enrichment strategy for clinical treatment trials of selectively enrolling MCI individuals with evidence of structural atrophy consistent with AD on a screening MRI can improve power to detect a treatment effect on global function and progressive structural deterioration. They also point to the necessity of taking into account the rate of change in healthy elderly when using structural variables as outcome measures to ensure that the trial is adequately powered to detect a disease-modifying effect of treatment.

REFERENCES

1. Petersen RC, Bennett D. Mild cognitive impairment: is it Alzheimer's disease or not? *J Alzheimers Dis.* 2005;7:241–245; discussion 255–262.
2. Morris JC, Storandt M, Miller JP, et al. Mild cognitive impairment represents early-stage Alzheimer disease. *Arch Neurol.* 2001;58:397–405.
3. Petersen RC. Mild cognitive impairment as a diagnostic entity. *J Intern Med.* 2004;256:183–194.
4. Petersen RC, Thomas RG, Grundman M, et al. Vitamin E and donepezil for the treatment of mild cognitive impairment. *N Engl J Med.* 2005;352:2379–2388.
5. Jelic V, Kivipelto M, Winblad B. Clinical trials in mild cognitive impairment: lessons for the future. *J Neurol Neurosurg Psychiatry.* 2006;77:429–438.
6. Richard F, Amouyel P. Genetic susceptibility factors for Alzheimer's disease. *Eur J Pharmacol.* 2001;412:1–12.
7. Fan Y, Batmanghelich N, Clark CM, et al. Spatial patterns of brain atrophy in MCI patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline. *Neuroimage.* 2008;39:1731–1743.
8. McEvoy LK, Fennema-Notestine C, Roddey JC, et al. Alzheimer disease: quantitative structural neuroimaging for detection and prediction of clinical and structural changes in mild cognitive impairment. *Radiology.* 2009;251:195–205.
9. Whitwell JL, Shiung MM, Przybelski SA, et al. MRI patterns of atrophy associated with progression to AD in amnestic mild cognitive impairment. *Neurology.* 2008;70:512–520.
10. Hughes CP, Berg L, Danziger WL, et al. A new clinical scale for the staging of dementia. *Br J Psychiatry.* 1982;140:566–572.
11. Rosen WG, Mohs RC, Davis KL. A new rating scale for Alzheimer's disease. *Am J Psychiatry.* 1984;141:1356–1364.
12. Jack CR Jr, Shiung MM, Gunter JL, et al. Comparison of different MRI brain atrophy rate measures with clinical disease progression in AD. *Neurology.* 2004;62:591–600.
13. Schott JM, Price SL, Frost C, et al. Measuring atrophy in Alzheimer disease: a serial MRI study over 6 and 12 months. *Neurology.* 2005;65:119–124.
14. Fox NC, Cousens S, Scahill R, et al. Using serial registered brain magnetic resonance imaging to measure disease progression in Alzheimer disease: power calculations and estimates of sample size to detect treatment effects. *Arch Neurol.* 2000;57:339–344.
15. Braak H, Braak E. Neuropathological staging of Alzheimer-related changes. *Acta Neuropathol (Berl).* 1991;82:239–259.
16. Jack CR Jr, Petersen RC, Xu Y, et al. Rates of hippocampal atrophy correlate with change in clinical status in aging and AD. *Neurology.* 2000;55:484–489.

17. Jack CR Jr, Slomkowski M, Gracon S, et al. MRI as a biomarker of disease progression in a therapeutic trial of milameline for AD. *Neurology*. 2003;60:253–260.
18. Devanand DP, Pradhaban G, Liu X, et al. Hippocampal and entorhinal atrophy in mild cognitive impairment: prediction of Alzheimer disease. *Neurology*. 2007;68:828–836.
19. Stoub TR, Rogalski EJ, Leurgans S, et al. Rate of entorhinal and hippocampal atrophy in incipient and mild AD: relation to memory function. *Neurobiol Aging*. 2008. doi:10.1016/j.neurobiolaging.2008.08.003.
20. McDonald CR, McEvoy LK, Gharapetian L, et al. Regional rates of neocortical atrophy from normal aging to early Alzheimer disease. *Neurology*. 2009;73:457–465.
21. Mueller SG, Weiner MW, Thal LJ, et al. The Alzheimer's disease neuroimaging initiative. *Neuroimaging Clin N Am*. 2005;15:869–877, xi-xii.
22. 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.
23. McKhann G, Drachman D, Folstein M, et al. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*. 1984;34:939–944.
24. Jovicich J, Czanner S, Greve D, et al. Reliability in multi-site structural MRI studies: effects of gradient non-linearity correction on phantom and human data. *Neuroimage*. 2006;30:436–443.
25. Sled JG, Zijdenbos AP, Evans AC. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans Med Imaging*. 1998;17:87–97.
26. 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.
27. Fischl B, Salat DH, Van der Kouwe AJ, et al. Sequence-independent segmentation of magnetic resonance images. *Neuroimage*. 2004;23(suppl 1):S69–S84.
28. Dale AM, Sereno MI. Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: a linear approach. *J Cogn Neurosci*. 1993;5:162–176.
29. Dale AM, Fischl B, Sereno MI. Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage*. 1999;9:179–194.
30. Fischl B, Sereno MI, Dale AM. Cortical surface-based analysis. II: inflation, flattening, and a surface-based coordinate system. *Neuroimage*. 1999;9:195–207.
31. Fischl B, Dale AM. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proc Natl Acad Sci U S A*. 2000;97:11050–11055.
32. Fischl B, Van Der Kouwe A, Destrieux C, et al. Automatically parcellating the human cerebral cortex. *Cereb Cortex*. 2004;14:11–22.
33. Desikan RS, Segonne F, Fischl B, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage*. 2006;31:968–980.
34. Fennema-Notestine C, Hagler DJ Jr, McEvoy LK, et al. Structural MRI biomarkers for preclinical and mild Alzheimer's disease. *Hum Brain Mapp*. 2009;30:3238–3253.
35. Holland D, Hagler DJ Jr, Fennema-Notestine C, et al. Longitudinal nonlinear registration and quantitative analysis of change in whole brain and regions of interest. *Alzheimers Demen: J Alzheimers Assoc*. 2008;4:T288.
36. Van der Kouwe AJ, Benner T, Dale AM. Real-time rigid body motion correction and shimming using cloverleaf navigators. *Magn Reson Med*. 2006;56:1019–1032.
37. White NW, Roddey C, Shankaranaraynan A, et al. PROMO: real-time prospective motion correction in MRI using image-based tracking. *Magn Reson Med*. 2010;63:91–105.
38. Fitzmaurice GM, Laird NM, Ware JH. *Applied Longitudinal Analysis*. Hoboken, NJ: John Wiley & Sons; 2004:8187–234.
39. 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.
40. Hua X, Lee S, Yanovsky I, et al. Optimizing power to track brain degeneration in Alzheimer's disease and mild cognitive impairment with tensor-based morphometry: an ADNI study of 515 subjects. *Neuroimage*. 2009;48:668–681.
41. Du AT, Schuff N, Zhu XP, et al. Atrophy rates of entorhinal cortex in AD and normal aging. *Neurology*. 2003;60:481–486.
42. Resnick SM, Pham DL, Kraut MA, et al. Longitudinal magnetic resonance imaging studies of older adults: a shrinking brain. *J Neurosci*. 2003;23:3295–3301.
43. Scahill RI, Frost C, Jenkins R, et al. A longitudinal study of brain volume changes in normal aging using serial registered magnetic resonance imaging. *Arch Neurol*. 2003;60:989–994.
44. Jack CR Jr, Shiung MM, Weigand SD, et al. Brain atrophy rates predict subsequent clinical conversion in normal elderly and amnestic MCI. *Neurology*. 2005;65:1227–1231.
45. Raz N, Lindenberger U, Rodrigue KM, et al. Regional brain changes in aging healthy adults: general trends, individual differences and modifiers. *Cereb Cortex*. 2005;15:1676–1689.
46. Du AT, Schuff N, Kramer JH, et al. Higher atrophy rate of entorhinal cortex than hippocampus in AD. *Neurology*. 2004;62:422–427.
47. Vellas B, Andrieu S, Sampaio C, et al. Disease-modifying trials in Alzheimer's disease: a European task force consensus. *Lancet Neurol*. 2007;6:56–62.
48. Dubois B, Feldman HH, Jacova C, et al. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. *Lancet Neurol*. 2007;6:734–746.
49. Winblad B, Gauthier S, Scinto L, et al. Safety and efficacy of galantamine in subjects with mild cognitive impairment. *Neurology*. 2008;70:2024–2035.
50. Knopman DS, Parisi JE, Salviati A, et al. Neuropathology of cognitively normal elderly. *J Neuropathol Exp Neurol*. 2003;62:1087–1095.
51. Markesbery WR, Schmitt FA, Kryscio RJ, et al. Neuropathologic substrate of mild cognitive impairment. *Arch Neurol*. 2006;63:38–46.