

Featured Article

The effect of diagnostic criteria on outcome measures in preclinical and prodromal Alzheimer's disease: Implications for trial design

Daniela Bertens^{a,*}, Betty M. Tijms^a, Lisa Vermunt^a, Niels D. Prins^{a,b}, Philip Scheltens^a, Pieter Jelle Visser^{a,c,1}, for the Alzheimer's Disease Neuroimaging Initiative

^aAlzheimer Centre, Department of Neurology, VU University Medical Centre, Amsterdam, The Netherlands

^bAlzheimer Research Center, Amsterdam The Netherlands

^cAlzheimer Centre, School for Mental Health and Neuroscience (MHeNS), University Medical Centre, Maastricht, The Netherlands

Abstract

Introduction: We investigated the influence of different inclusion criteria for preclinical and prodromal Alzheimer's disease (AD) on changes in biomarkers and cognitive markers and on trial sample size estimates.

Methods: We selected 522 cognitively normal subjects and 872 subjects with mild cognitive impairment from the Alzheimer's Disease Neuroimaging Initiative study. Compared inclusion criteria were (1) preclinical or prodromal AD (amyloid marker abnormal); (2) preclinical or prodromal AD stage-1 (amyloid marker abnormal, injury marker normal); and (3) preclinical or prodromal AD stage-2 (amyloid and injury markers abnormal). Outcome measures were amyloid, neuronal injury, and cognitive markers.

Results: In both subjects with preclinical and prodromal AD stage-2, inclusion criteria resulted in the largest observed decline in brain volumetric measures on magnetic resonance imaging and cognitive markers.

Discussion: Inclusion criteria influence the observed rate of worsening in outcome measures. This has implications for trial design.

© 2017 Published by Elsevier Inc. on behalf of the Alzheimer's Association. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Keywords:

Alzheimer's disease; Longitudinal; Biomarkers; Cognitive markers; Preclinical; Prodromal; Sample size estimates; Clinical trial

D.B., B.M.T., and L.V. report no disclosures.

Dr N.D.P. serves on the advisory board of Boehringer Ingelheim, Forum, and Probiobdrug and has provided consultancy services for Sanofi and Takeda. He has been a speaker at symposia organized by Janssen and Novartis. N.D.P. receives research support from Alzheimer Nederland (project number WE.03-2012-02). N.D.P. is CEO and co-owner of the Alzheimer Research Center, Amsterdam, The Netherlands. Dr P.S. has received grant support (for the institution) from GE Healthcare, Danone Research, Piramal, and MERCK. In the past 2 years, he has received consultancy/speaker fees (paid to the institution) from Lilly, GE Healthcare, Novartis, Forum, Sanofi, Nutricia, Probiobdrug, and EIP Pharma. Dr P.J.V. serves as an advisory board member of Eli Lilly and is consultant for Janssen Pharmaceutical. He receives/received research grants from Bristol-Myers Squibb and GE Healthcare, European Commission 6th and 7th Framework

programme, the Innovative Medicines Initiative (IMI), European Union Joint Programme–Neurodegenerative Disease Research (JPND), and Zon-Mw.

¹Data 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.

*Corresponding author. Tel.: +31204448527; Fax: +31204448529.

E-mail address: d.bertens@vumc.nl

1. Introduction

Alzheimer's disease (AD)-modifying therapy, targeting amyloid, is probably most effective when administered early, that is, before the stage of dementia. A number of research criteria have been proposed to identify nondemented subjects with AD based on the presence of AD biomarkers [1–3]. They can be applied in subjects without cognitive impairment (asymptomatic at risk for AD or preclinical AD) and subjects with mild cognitive impairment (MCI) (MCI due to AD or prodromal AD). However, these criteria allow for different combinations of AD pathology biomarkers, and it is unknown whether this impacts on observed changes in outcome measures. For trial design, it is critical to understand how selection criteria for subjects at such early stages of the disease influence change in outcome measures. Previous studies on outcome measures typically had a short follow-up, did not compare the effect of different inclusion criteria, or restricted their analyses to a limited set of outcome measures [4–12].

The aim of our study was to investigate whether changes in outcome measures are dependent on the inclusion criteria for preclinical and prodromal AD used. We studied three definitions for preclinical and prodromal AD: (1) having abnormal amyloid markers; (2) having abnormal amyloid markers and normal neuronal injury markers; and (3) having both abnormal amyloid and neuronal injury markers. As outcome measures, we used biomarkers for amyloid β ($A\beta$) in cerebrospinal fluid (CSF) or on positron emission tomography (PET), CSF tau, fludeoxyglucose PET (FDG-PET), brain atrophy measured with magnetic resonance imaging (MRI), and measures of cognitive functioning. To study the potential effects of different combinations of inclusion criteria and outcome measures on trial design, we calculated sample sizes for a hypothetical 3-year placebo-controlled trial in subjects at predementia AD stages. To study the additive value of biomarkers to define predementia AD, we also calculated slopes and sample sizes for subjects with normal cognition and MCI, regardless of their biomarker status.

2. Methods

2.1. Alzheimer's Disease Neuroimaging Initiative study

We studied data from subjects that participated in the Alzheimer's Disease Neuroimaging Initiative (ADNI) study (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD.

2.2. Participants

We selected all participants with normal cognition ($N = 522$) or MCI ($N = 872$) from ADNI-1, ADNI-2, and

ADNI-GO who had baseline and follow-up data available for at least one visit within a 3-year period for several biomarkers and cognitive tests (explained in more detail below). The ADNI inclusion criteria for participants with normal cognition were absence of memory complaints, a Mini-Mental State Examination (MMSE) [13] score of 24–30, a Clinical Dementia Rating (CDR) [14] score of 0, and no MCI or dementia diagnosis. The inclusion criteria for subjects with MCI were memory complaints, objective memory loss, an MMSE score between 24 and 30, and a CDR score of 0.5. Exclusion criteria were the absence of an informant, a score of >4 on the modified Hachinski scale [15], and score of >5 on the Geriatric Depression Scale [16], additional diseases expected to interfere with the study, use of investigational agents, multiple trial participation, and findings showing other reasons for cognitive problems. Permitted medication had to be stable for at least 4 weeks before screening. We downloaded ADNI data at 31st March, 2014.

2.3. Subject classification based on AD biomarkers

Subjects were classified as preclinical or prodromal AD with the use of AD biomarkers for amyloidosis and/or neuronal injury (see below), as proposed by International Work Group-2 (IWG-2) or National Institute on Aging and Alzheimer's Association (NIA-AA) research criteria [1–3]. As a marker for amyloidosis we used CSF $A\beta_{1-42}$ or ^{18}F -AV-45-PET, and as marker of neuronal injury we used CSF tau or FDG-PET. If both modalities were present for a given subject, we used their PET measures because they are more commonly used in practice. Subjects with normal cognition were classified as preclinical AD when they had abnormal amyloid, without taking into account neuronal injury markers; as preclinical AD stage-1 if they had abnormal amyloid and a normal injury marker; and as preclinical AD stage-2 if both the amyloid and injury markers were abnormal. MCI subjects were similarly classified as prodromal AD if the amyloid marker was abnormal, without taking into account neuronal injury markers; as prodromal AD stage-1 if the amyloid marker was abnormal but the injury marker normal; and as prodromal AD stage-2 if both the amyloid and injury marker were abnormal. Fig. 1 gives an overview of classification of subjects according to these criteria.

2.4. Baseline assessment and longitudinal assessment

Subjects underwent a standardized assessment that included neurological, physical, and neuropsychological examinations, collection of CSF and blood, and performance of MRI and PET scanning. For 32 cognitively normal and 23 MCI subjects, amyloid assessment was performed at follow-up only; and for these subjects, we used the first follow-up assessment with this measure as the baseline visit. The protocols for data collection are described in detail at http://www.loni.ucla.edu/ADNI/Data/ADNI_Data.shtml. Cognitive measures were collected at baseline and at

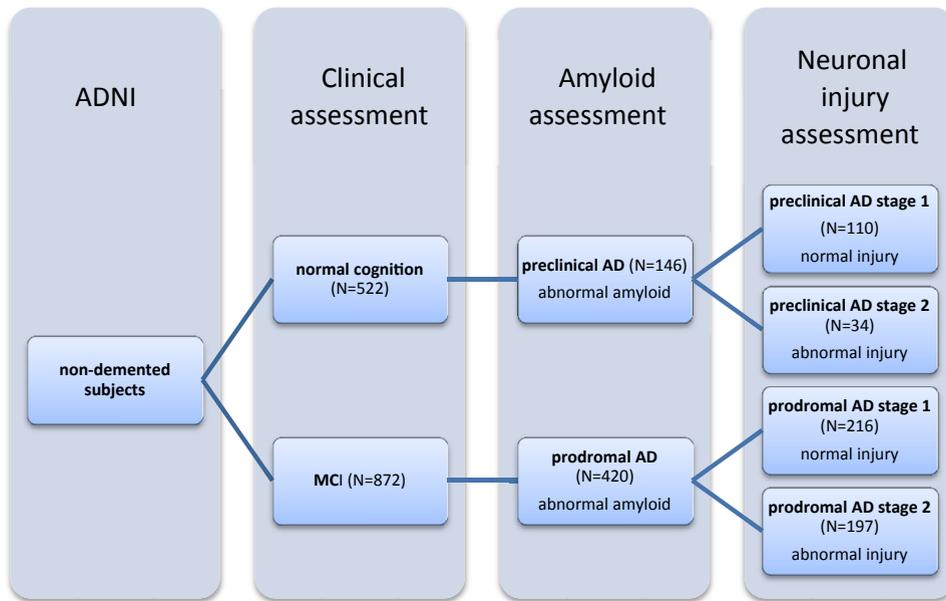


Fig. 1. Schematic overview of the groups according to subclassification, applying the research criteria. Abbreviations: ADNI, Alzheimer's Disease Neuroimaging Initiative; MCI, mild cognitive impairment; AD, Alzheimer's disease. Subject classification based on AD biomarkers: Preclinical AD, $n = 146$; 49 based on CSF measures only, 80 based on PET, and 17 subjects with both modalities present. Preclinical AD stage-1, $n = 110$; 33 based on CSF measures only, 60 based on PET, and 17 with both modalities present. Preclinical AD stage-2, $n = 34$; 16 based on CSF measures only, 17 based on PET, and 1 with both modalities present. For two cognitively normal subjects, we did not have any information of their injury status so they could not be further classified into stage-1 or stage-2. Prodromal AD, $n = 420$; 149 based on CSF measures only, 148 based on PET, and 123 with both modalities present. Prodromal AD stage-1, $n = 216$; 63 based on CSF measures only, 88 based on PET, and 65 with both modalities present. Prodromal AD stage-2, $n = 197$; 85 based on CSF measures, 59 based on PET, 53 with both modalities present. For seven MCI subjects, we did not have any information of their injury status, so they could not be further classified into stage-1 or stage-2.

6 monthly follow-up assessments; biomarkers were collected at baseline and annually.

2.5. CSF analysis

CSF samples were available at baseline for 174 subjects with normal cognition and in 398 subjects with MCI. CSF samples were collected by lumbar puncture and shipped on dry ice to the Penn ADNI Biomarker Core Laboratory at the University of Pennsylvania, Philadelphia for storage until further analysis. CSF was analyzed using a multiplex xMAP Luminex platform (Luminex Corp) with immunoassay kit-based reagents (INNO-BIA AlzBio3; Innogenetics; www.adni-info.org) as described elsewhere [17]. Baseline and follow-up samples were analyzed in the same batch. The cutoff value for abnormal CSF $A\beta_{1-42}$ levels was <192 pg/mL and for tau >92 pg/mL.

2.6. Amyloid-PET analyses

Amyloid ^{18}F -AV-45-PET was available for 286 cognitively normal and 485 MCI subjects. Data were acquired 50 minutes after injection for 20 minutes. In case of motion artifacts, another 20 minutes of scanning was acquired. For each subject, a florbetapir composite standard uptake value ratio (SUVr) was created as a mean binding of four cortical regions (frontal, anterior/posterior cingulate, parietal cortex, and temporal cortex as determined with FreeSurfer v4.5.0

and after coregistration of PET and MRI data with SPM5), divided by the reference region (whole cerebellum). An uptake in the measure above 1.11 was considered to be abnormal [18].

2.7. FDG-PET analyses

FDG-PET was available for 402 cognitively normal and 674 MCI subjects. FDG data were acquired 30 to 60 minutes after injection. After preprocessing, images were spatially normalized in SPM5 to Montreal Neurological Institute (MNI) PET template. Meta-analytically derived regions-of-interest (MetaROIs) were calculated that includes FDG uptake in bilateral angular gyrus, posterior cingular gyrus, and bilateral inferior temporal gyrus. Each MetaROI was normalized to a reference region composed of the pons and vermis. Total FDG uptake was calculated as a mean of the five individual MetaROIs (www.adni-info.org). FDG uptake on PET below 1.21 was considered as abnormal [19].

2.8. MRI analyses

Whole brain structural scans were acquired with 1.5T or 3.0T MRI scanners. We analyzed three MRI-based outcome measures: whole-brain, ventricular, and hippocampal volumes. For measurement of whole-brain and ventricular volumes, the boundary shift integral was used [20,21]. Hippocampal volumes were measured, using FreeSurfer

version 4.3 for ADNI-1 and FreeSurfer 5.1 for ADNI-GO and ADNI-2 [22]. Each scan was segmented according to an atlas defined by FreeSurfer. For ADNI-GO and ADNI-2, two T1 weighted images were acquired, of which we selected the nonaccelerated acquisition scans. Hippocampal volume was measured bilaterally, and the average volume over left and right was used for the present analyses. To correct for interindividual differences in head size, we used the total intracranial volume measure from FreeSurfer. From the ADNI database, baseline hippocampal volume measures were available for 474 subjects with normal cognition and 840 measures for subjects with MCI. Ventricular and whole-brain gray matter volumes were available in 364 subjects with normal cognition and 805 subjects with MCI.

2.9. Cognitive assessment

We used the CDR sum of boxes (CDR-sob), MMSE, and Alzheimer's Disease Assessment Scale-cognitive (ADAS-Cog 11 item) [23] to assess cognition. In addition, we calculated the Alzheimer Disease Cooperative Study Preclinical Alzheimer Cognitive Composite (ADCS-PACC), which is a composite score of the total score of the delayed word recall on the ADAS-Cog subscale, the delayed recall score on the logical memory subscale II from the Wechsler Memory Scale, the digit symbol substitution test score from the Wechsler Adult Intelligence Scale-revised, and the total MMSE score [24]. Because the digit symbol substitution test score was only available for ADNI-1 subjects, we also constructed a PACC-like score, without this test.

2.10. Data analysis

Statistical analyses were performed with SPSS, version 20.0 for the Macintosh.

Linear mixed models (with covariates for age, sex, and level of education) were used to test the following effects: (1) we tested whether baseline scores differed between preclinical AD stage-1 and stage-2 and between subjects with prodromal AD stage-1 and stage-2 (i.e., differences in intercepts); (2) changes over time were assessed by testing whether slopes differed from 0 and whether they differed between subjects in preclinical AD stage-1 and stage-2 and between subjects in prodromal AD stage-1 and stage-2. For slope analyses, we used an unstructured covariance matrix, assuming a random intercept and fixed slope, and used follow-up time as repeated measure. We assumed a linear change in time (time coded with a quadratic term was not statistically significant). Separate analyses were performed for each criterion used to classify predementia AD. Difference with a P -value $< .05$ were considered to be statistically significant. Sample size was estimated for a hypothetical 3-year randomized-controlled trial with two arms, showing an expected treatment effect of 25% reduction of decline in outcome measures with a power of 80%, a two-sided

alpha of 5%, and a 10% annual dropout rate using the following formula [25]:

$$n/\text{arm} = 2(z_{1-\alpha/2} + z_{1-\beta})^2 \left(\sigma_e^2 / \sum (t_i - t_{\text{mean}}) \right)^2 / \Delta^2$$

with Δ as the difference in mean rate of decline in treatment versus control, σ_e^2 as the residual error variance of the mixed effects model, α as the type I error rate of a two-sided test and $1-\beta$ as the power, t_i as the times i at which measures were made, and t_{mean} as the average follow-up time. The total sample size n required for a trial was then obtained by multiplying this estimation by 2 and adjusted for an annual dropout rate of 10% over the course of 3 years ($n/\text{arm} * 2 * 1.11^3$). Finally, we calculated the numbers needed to be screened, which is the sample size needed in a specific subgroup divided by the prevalence of this group in subjects with the same cognitive status. All analyses were stratified for baseline diagnosis.

3. Results

3.1. Baseline characteristics and longitudinal change in outcome measures in subjects with normal cognition

Table 1 shows the baseline characteristics of the subjects with normal cognition according to different classification criteria. Subjects classified as preclinical AD stage-2 were older ($P = .036$), had smaller hippocampal volumes ($P = .014$), and, by definition, higher CSF tau and lower FDG-PET binding ($P < .001$) in comparison to preclinical stage-1 subjects.

Table 2 shows the change in biomarkers and cognitive markers over time. In subjects with normal cognition, regardless of biomarkers status, all markers showed worsening over time, except for the ADAS-Cog, and both ADCS-PACC composite scores. Subjects with preclinical AD and preclinical AD stage-1 showed increases over time in amyloid PET, CSF tau levels, CDR-sob and ventricular volumes, and decreases in FDG-PET, whole-brain volume, and hippocampal volume. Subjects with preclinical AD stage-2 showed increases over time in ventricular volume and the CDR-sob, and decreases in CSF $A\beta_{1-42}$ levels, whole-brain volume, and hippocampal volume. In subjects with preclinical AD, the rate of hippocampal volume loss and increase in CDR-sob were faster for those classified as belonging to stage-2 than those to stage-1 (Table 2).

3.2. Baseline characteristics and longitudinal change in outcome measures in subjects with MCI

Table 3 shows the baseline characteristics of subjects with MCI according to the different classification criteria. Subjects with prodromal AD stage-2 were more often apolipoprotein $\epsilon 4$ ($APOE \epsilon 4$) positive ($P = .003$), had higher amyloid binding on ^{18}F -AV-45-PET ($P = .0001$), smaller whole-brain and hippocampal volumes ($P = .0001$), worse scores on cognitive

Table 1
Baseline characteristics of subjects with normal cognition according to disease-stage classification at baseline

Baseline characteristics	Total sample cognitively normal (N = 522)	Cognitively normal with normal amyloid and injury markers (N = 221)	Preclinical AD (N = 146)	Preclinical AD stage-1 (N = 110)	Preclinical AD stage-2 (N = 34)	P value differences stage-1 and stage-2
Age (years)	74.24 (5.79)	73.01 (5.76)	74.8 (5.5)	74.2 (5.5)	76.84 (4.94)	.036
Females (%)	51% (0.5)	49% (0.5)	60% (0.5)	66% (0.5)	44% (0.5)	.064
Years of education	16.38 (2.7)	16.50 (2.6)	16.06 (2.7)	15.95 (2.8)	16.47 (2.33)	.62
1/2 APOE ϵ 4 alleles (%) (n = 410)	103/11 (19/2)	29/1 (13/0.5)	46/7 (32/5)	30/6 (28/5)	15/0 (44/0)	.96
CSF A β ₁₋₄₂ (pg/mL) (n = 174)	209.6 (53.5)	243.6 (31.55)	155.27 (31.9)	160.1 (33.8)	144.29 (27.2)	.079
¹⁸ F-AV-45 (SUVr) (n = 286)	1.1 (0.17)	1.01 (0.05)	1.29 (0.17)	1.30 (0.18)	1.29 (0.14)	.65
CSF tau (pg/mL) (n = 172)	68.2 (32.9)	57.4 (19.2)	80 (41)	65.3 (32.9)	126.39 (28.4)	.0001
FDG-PET (SUVr) (n = 402)	1.31 (0.12)	1.34 (0.09)	1.31 (0.12)	1.35 (0.1)	1.17 (0.06)	.0001
Whole-brain volume (cm ³) (n = 364)	1045 (50)	1054 (50)	1043 (49)	1043 (49)	1039 (51)	.83
Hippocampal volume (mm ³) (n = 474)	3709 (385)	3771 (369)	3676 (368)	3710 (384)	3525 (288)	.014
Ventricular volume (cm ³) (n = 363)	46.5 (40.6)	38.3 (29)	49.7 (40.9)	48 (39)	59.5 (47)	.68
CDR sum of boxes (n = 520)	0.05 (0.18)	0.03 (0.12)	0.05 (0.16)	0.059 (0.18)	0.015 (0.09)	.17
MMSE score (n = 521)	29.1 (1.15)	29.1 (1.18)	29.1 (0.95)	29.04 (0.99)	29.12 (1.09)	.64
ADAS-Cog (n = 522)	5.95 (3.03)	5.67 (2.99)	6.08 (2.9)	6.12 (2.73)	6.55 (3.6)	.84
ADCS-PACC (n = 197)	2.63 (1.78)	2.64 (1.74)	2.62 (1.71)	2.63 (1.81)	2.62 (1.56)	.79
ADCS-PACC without digit symbol test (n = 521)	2.15 (1.41)	2.25 (1.40)	2.19 (1.30)	2.15 (1.23)	2.12 (1.74)	.44
Follow-up time (n = 522)	2.14 (1.03)	2.10 (0.96)	2.18 (0.97)	2.12 (1.00)	2.35 (0.81)	.18
Number of visits per subject (n = 522)	5.14 (1.78)	5.30 (1.70)	5.34 (1.76)	5.29 (1.78)	5.44 (1.58)	.64

Abbreviations: A β ₁₋₄₂, amyloid β 1-42; AD, Alzheimer's disease; ADAS-Cog, Alzheimer's Disease Assessment Scale-Cognitive; ADCS-PACC, Alzheimer's Disease Cooperative Study Preclinical Alzheimer Cognitive Composite; CDR, Clinical Dementia Rating scale; CSF, cerebrospinal fluid; FDG, fluoro-deoxyglucose; AV-45, florbetapir; MMSE, Mini-Mental State Examination; PET, positron emission tomography.

NOTE. Data are mean (standard deviation) unless specified otherwise.

tests (CDR-sob and ADAS-Cog $P = .0001$, MMSE $P = .001$, ADCS-PACC without digit symbol test $P = .029$), and, by definition, higher CSF tau levels and lower FDG-PET binding ($P = .0001$) than subjects with prodromal AD stage-1.

In the total group of MCI subjects, regardless of their biomarker status, all markers showed worsening over time (Table 4). In all subgroups of MCI patients, most markers became progressively worse over time, except for ¹⁸F-AV-45-PET in the total prodromal AD and prodromal AD stage-2 subjects, and for CSF A β ₁₋₄₂ levels in prodromal AD stage-1 patients. Whole-brain volume loss and hippocampal volume loss and worsening in cognitive test scores over time occurred faster in subjects with prodromal AD stage-2 than in subjects with prodromal AD stage-1 (Table 4).

3.3. Sample size estimations

3.3.1. Sample size estimations for subjects with normal cognition

For subjects with normal cognition, sample size estimates ranged from 81–32,750 (Table 5). The smallest sample sizes were estimated for brain volumetric outcome measures, irrespective of the inclusion criteria used. In preclinical AD stage-1, an intermediately small sample size was estimated with ¹⁸F-AV-45-PET as outcome measure (n = 436). Cognitive outcome measures resulted in the largest sample size estimates, regardless of the inclusion criteria used. Numbers-needed-for-screening were two to three times larger

for subjects with preclinical AD and preclinical AD stage-1 compared with cognitively normal subjects with unspecified biomarker status, when MRI measures, FDG-PET, or CSF tau were used as outcome measures (Supplementary Table 1). Numbers-needed-for-screening were smaller in subjects with preclinical AD and preclinical AD stage-1 when compared to subjects with normal cognition and unspecified biomarker status, when ¹⁸F-AV-45-PET or cognitive measures were used as outcome measure. Numbers-needed-for-screening increased three to nine times in subjects with preclinical AD stage-2, regardless of the outcome measure used.

3.3.2. Sample size estimations for subjects with MCI

In subjects with clinical MCI, sample size estimates ranged from 165 to 9312 with the smallest sample size estimates when whole-brain volume, hippocampal volume, and ventricular volume were used as outcome measures (Table 6). The use of ¹⁸F-AV-45-PET, CSF A β ₁₋₄₂, and CSF tau as an outcome measure resulted in the largest sample size estimations. In prodromal AD, sample size was the smallest for brain volumetric measures, the CDR-sob, and ADCS-PACC. In this group, the largest sample sizes were estimated for the outcome measures CSF A β ₁₋₄₂ and CSF tau. For prodromal AD stage-1 and prodromal AD stage-2, the smallest sample sizes were estimated when using brain volumetric measures as an outcome measure. Numbers-needed-for-screening were the smallest in clinical MCI, followed by subjects with prodromal AD, prodromal AD stage-2, and prodromal AD stage-1 (Supplementary Table 2).

Table 2
Annual change in outcome measures in subjects with normal cognition according to disease-stage classification at baseline

Outcome measures	Cognitively normal (N = 522)	Cognitively normal with normal amyloid and injury markers (N = 221)	Preclinical AD (N = 146)	Preclinical AD stage-1 (N = 110)	Preclinical AD stage-2 (N = 34)	P value slope differences stage-1 and stage-2
CSF A β_{1-42} (pg/mL) (n = 97)	-4.80 (0.97)**	-5.90 (1.41)**	-2.60 (1.35)	-1.01 (1.74)	-5.59 (1.96)*	.12
¹⁸ F-AV-45-PET (SUVR) (n = 135)	0.01 (0.003)**	0.001 (0.002)	0.03 (0.01)**	0.03 (0.01)**	0.01 (0.02)	.096
CSF tau (pg/mL) (n = 97)	2.97 (0.66)**	2.32 (0.71)**	4.12 (1.35)**	4.70 (1.59)**	1.91 (2.51)	.21
FDG-PET (SUVR) (n = 219)	-0.01 (0.001)**	-0.01 (0.003)**	-0.012 (0.004)**	-0.015 (0.005)**	0.002 (0.007)	.05
Whole-brain volume (cm ³) (n = 346)	-8.18 (0.44)**	-7.07 (0.75)**	-8.70 (0.70)**	-7.92 (0.89)**	-10.34 (1.03)**	.073
Hippocampal volume (mm ³) (n = 412)	-52.68 (2.87)**	-45.91 (4.54)**	-58.86 (5.69)**	-43.87 (6.97)**	-90.49 (8.39)**	<.001
Ventricular volume (cm ³) (n = 346)	2.51 (0.13)**	1.87 (0.16)**	3.52 (0.36)**	3.29 (0.45)**	3.99 (0.6)**	.076
CDR sum of boxes (n = 98)	0.09 (0.01)**	0.042 (0.011)**	0.15 (0.02)**	0.09 (0.02)**	0.21 (0.05)**	.022
MMSE score (n = 486)	-0.07 (0.03)*	-0.084 (0.04)	-0.15 (0.06)*	-0.12 (0.06)	-0.09 (0.11)	.89
ADAS-Cog (n = 484)	-0.08 (0.06)	-0.12 (0.09)	-0.05 (0.11)	0.027 (0.12)	0.09 (0.25)	.66
ADCS-PACC (n = 309)	-0.008 (0.041)	0.045 (0.072)	-0.049 (0.081)	-0.049 (0.089)	-0.045 (0.166)	.98
ADCS-PACC without digit symbol test (n = 394)	-0.008 (0.028)	0.030 (0.045)	-0.077 (0.055)	-0.062 (0.059)	-0.025 (0.124)	.99

Abbreviations: A β_{1-42} , amyloid β 1-42; AD, Alzheimer's disease; ADAS-Cog, Alzheimer's Disease Assessment Scale-Cognitive; ADCS-PACC, Alzheimer's Disease Cooperative Study Preclinical Alzheimer Cognitive Composite; CDR, Clinical Dementia Rating scale; CSF, cerebrospinal fluid; FDG, fluoro-deoxyglucose; AV-45, florbetapir; MMSE, Mini-Mental State Examination; PET, positron emission tomography.

NOTE. N after outcome variable indicates the number of subject from the total sample with at least 1 follow-up measure available. Data are mean (standard error). ** $P < .01$, * $P < .05$ indicates a slope different from 0.

4. Discussion

Our main findings are that inclusion criteria, used to identify subjects with predementia AD, influenced the magnitude of change over time in observed outcome measures. Sample size estimates for a hypothetical 3-year clinical trial varied widely according to the combination of inclusion criteria and outcome markers applied. The smallest sample

size needed to show a treatment effect was estimated for subjects with normal cognition or MCI who had both abnormal amyloid and injury markers at baseline using brain volumetric markers as outcome measure.

Subjects with normal cognition showed a worsening over time in all markers, except the ADAS-Cog, ADCS-PACC like composite score, and ADCS-PACC without digit

Table 3
Baseline characteristics of subjects with MCI according to disease-stage classification at baseline

Baseline characteristics	Total sample of MCI patients (N = 873)	Prodromal AD (N = 420)	Prodromal AD stage-1 (N = 216)	Prodromal AD stage-2 (N = 197)	P value difference stage-1 and stage-2
Age (years)	72.93 (7.6)	73.38 (7.22)	73.12 (7.25)	73.56 (6.99)	.58
Females (%)	41% (0.5)	42% (0.5)	43% (0.5)	42% (0.5)	.88
Years of education	15.9 (2.8)	15.91 (2.85)	15.90 (2.84)	15.83 (2.89)	.82
1/2 APOE ϵ 4 alleles (%) (n = 856)	338/93 (39/11)	210/63 (50/15)	95/31 (44/15)	114/31 (59/16)	.003
CSF A β_{1-42} (pg/mL) (n = 398)	168.6 (54.2)	137.83 (27.3)	138.9 (31.8)	135.5 (21.9)	.31
¹⁸ F-AV-45(SUVR) (n = 485)	1.2 (0.22)	1.35 (0.18)	1.33 (0.17)	1.40 (0.15)	.0001
CSF tau (pg/mL) (n = 382)	97.6 (57.3)	115.7 (59)	88.8 (39.7)	143.37 (62.1)	.0001
FDG-PET (SUVR) (n = 674)	1.25 (0.13)	1.22 (0.14)	1.29 (0.12)	1.14 (0.1)	.0001
Whole-brain volume (cm ³) (n = 805)	1022 (59)	1023 (58)	1034 (61)	1011 (52)	.0001
Hippocampal volume (mm ³) (n = 840)	3298 (517)	3241 (505)	3349 (532)	3140 (449)	.0001
Ventricular volume (cm ³) (n = 805)	55.3 (50.4)	56.2 (52.5)	55.0 (56.5)	57.3 (48.0)	.77
CDR sum of boxes (n = 873)	1.53 (0.89)	1.62 (0.93)	1.45 (0.91)	1.82 (0.9)	.0001
MMSE score (n = 872)	27.57 (1.81)	27.4 (1.85)	27.6 (1.84)	27.13 (1.81)	.001
ADAS-Cog (n = 869)	10.27 (4.6)	11 (4.7)	10.27 (4.92)	11.96 (4.41)	.0001
ADCS-PACC (n = 372)	-2.56 (2.28)	-2.93 (2.09)	-2.93 (2.23)	-2.91 (2.00)	.81
ADCS-PACC without digit symbol test (n = 869)	-1.20 (2.16)	-1.57 (2.12)	-1.07 (2.07)	-2.16 (2.04)	.029
Follow-up time (n = 873)	2.62 (0.69)	2.64 (0.63)	2.60 (0.64)	2.65 (0.64)	.42
Number of visits per subject (n = 873)	5.94 (1.25)	6.18 (1.14)	6.15 (1.15)	6.14 (1.16)	.92

Abbreviations: A β_{1-42} , amyloid β 1-42; AD, Alzheimer's disease; ADAS-Cog, Alzheimer's Disease Assessment Scale-Cognitive; ADCS-PACC, Alzheimer's Disease Cooperative Study Preclinical Alzheimer Cognitive Composite; CDR, Clinical Dementia Rating scale; CSF, cerebrospinal fluid; FDG, fluoro-deoxyglucose; AV-45, florbetapir; MCI, mild cognitive impairment; MMSE, Mini-Mental State Examination; PET, positron emission tomography.

NOTE. Data are mean (standard deviation) unless specified otherwise.

Table 4

Annual change in outcome measures in subjects with MCI according to disease-stage classification at baseline

Outcome measures	MCI (N = 873)	Prodromal AD (N = 420)	Prodromal AD stage-1 (N = 216)	Prodromal AD stage-2 (N = 197)	P value difference stage-1 and stage-2
CSF A β_{1-42} (pg/mL) (n = 187)	-1.81 (0.66)**	-1.66 (0.64)*	0.13 (1.02)	-3.00 (0.82)**	.13
¹⁸ F-AV-45-PET (SUVR) (n = 234)	0.007 (0.003)*	0.01 (0.005)	0.014 (0.006)*	0.004 (0.01)	.44
CSF tau (pg/mL) (n = 185)	3.24 (0.88)**	4.11 (1.15)**	3.69 (1.27)**	4.26 (1.83)*	.89
FDG-PET (SUVR) (n = 388)	-0.02 (0.002)**	-0.027 (0.003)**	-0.025 (0.003)**	-0.029 (0.003)**	.09
Whole-brain volume (cm ³) (n = 731)	-11.46 (0.35)**	-13.31 (0.49)**	-11.64 (0.66)**	-14.69 (0.72)**	.007
Hippocampal volume (mm ³) (n = 804)	-86.22 (2.28)**	-97.59 (3.78)**	-86.06 (5.58)**	-109.8 (5.10)**	.006
Ventricular volume (cm ³) (n = 731)	4.54 (0.18)**	5.60 (0.30)**	6.04 (0.51)**	5.08 (0.34)**	.08
CDR sum of boxes (n = 811)	0.54 (0.02)**	0.68 (0.03)**	0.52 (0.05)**	0.82 (0.05)**	<.001
MMSE score (n = 823)	-0.73 (0.04)**	-0.91 (0.06)**	-0.63 (0.08)**	-1.16 (0.08)**	<.001
ADAS-Cog (n = 822)	1.14 (0.08)**	1.53 (0.10)**	1.09 (0.14)**	1.99 (0.15)**	<.001
ADCS-PACC (n = 334)	-0.81 (0.05)**	-1.09 (0.08)**	-1.10 (0.14)**	-1.09 (0.10)**	.89
ADCS-PACC without digit symbol test (n = 782)	-0.46 (0.03)**	-0.69 (0.05)**	-0.45 (0.07)**	-0.94 (0.07)**	<.001

Abbreviations: A β_{1-42} , amyloid β 1-42; AD, Alzheimer's disease; ADAS-Cog, Alzheimer's Disease Assessment Scale-Cognitive; ADCS-PACC, Alzheimer's Disease Cooperative Study Preclinical Alzheimer Cognitive Composite; CDR, Clinical Dementia Rating scale; CSF, cerebrospinal fluid; FDG, fluorodeoxyglucose; AV-45, florbetapir; MCI, mild cognitive impairment; MMSE, Mini-Mental State Examination; PET, positron emission tomography.

NOTE. N after outcome variable indicates the number of subject from the total sample with at least 1 follow-up measure available. Data are mean (standard error). ** $P < .01$, * $P < .05$ slope different from 0.

symbol test. When taking into account biomarkers, worsening was typically greater in subjects with preclinical stage-2 than in subjects with preclinical stage-1. In subjects with MCI, all markers became more abnormal over time. Worsening of outcome markers was greater for subjects with prodromal stage-2 than for subjects with prodromal stage-1. Our observation of larger effects in outcome measures for subjects at more advanced disease stages is in line with reports of previous studies [26-30]. Because subjects in stage-2 showed the most worsening over time in all outcome markers, the absolute difference between treated and nontreated groups in our hypothetical trial was the largest as well, and so subsequent sample size estimates were smaller for this group of subjects.

For all definitions of preclinical AD, brain atrophy outcome measures showed the most worsening over time and subsequently resulted in the smallest sample size estimates. This estimate was the smallest for preclinical AD stage-2. Amyloid markers showed some worsening over time, but acceptable sample size estimates based on amyloid were only obtained in preclinical AD stage-2 when CSF A β_{1-42} was used as an outcome measure and in preclinical AD stage-1 when ¹⁸F-AV-45-PET was used. In preclinical AD stage-2, when neuronal injury was defined based on CSF tau, no changes in FDG-PET were observed, which might reflect floor effects in this group of subject and this is in line with previous studies [31,32]. None of the cognitive measures showed decline over time.

Table 5

Sample size estimates showing a treatment effect of 25% in a hypothetical 3-year trial in subjects with normal cognition at baseline

Outcome measures	Cognitively normal (N = 522)	Preclinical AD (N = 146)	Preclinical AD stage-1 (N = 110)	Preclinical AD stage-2 (N = 34)
CSF A β_{1-42} (pg/mL)	1205 (617-3311)	-	-	457 (153-6280)
¹⁸ F-AV-45-PET (SUVR)	2756 (1050-19,103)	603 (256-2786)	436 (190-1853)	-
CSF tau (pg/mL)	1440 (698-4532)	1121 (409-9444)	779 (274-7883)	-
FDG-PET (SUVR)	2425 (1327-5775)	2563 (914-24,240)	1622 (623-10,873)	-
Whole-brain volume (cm ³)	304 (249-380)	169 (126-239)	226 (151-375)	86 (60-135)
Hippocampal volume (mm ³)	361 (294-452)	279 (197-426)	514 (297-1095)	81 (58-123)
Ventricular volume (cm ³)	310 (253-388)	271 (188-423)	324 (201-607)	191 (113-389)
CDR sum of boxes	2928 (1874-5206)	1280 (732-2791)	1745 (841-5580)	1046 (431-5342)
MMSE score	32,730 (9837-1057708)	7317 (2373-122846)	-	-
ADAS-Cog	-	-	-	-
ADCS-PACC	-	-	-	-
ADCS-PACC without digit symbol test	-	-	-	-

Abbreviations: A β_{1-42} , amyloid β 1-42; AD, Alzheimer's disease; ADAS-Cog, Alzheimer's Disease Assessment Scale-Cognitive; ADCS-PACC, Alzheimer's Disease Cooperative Study Preclinical Alzheimer Cognitive Composite; CDR, Clinical Dementia Rating scale; CSF, cerebrospinal fluid; FDG, fluorodeoxyglucose; AV-45, florbetapir; MMSE, Mini-Mental State Examination; PET, positron emission tomography.

NOTE. Sample size was estimated for a hypothetical 3-year randomized-controlled trial with two arms, showing an expected treatment effect of 25% reduction of decline in outcome measures with a power of 80%, a two-sided alpha of 5%, and a 10% annual dropout rate. "-" represents not calculated as slope is not significantly different from 0 (see Table 2). Data are mean (95% confidence interval).

Table 6

Sample size estimates showing a treatment effect of 25% in a hypothetical 3-year trial in subjects with MCI according to disease-stage classification at baseline

Outcome measures	MCI (N = 873)	Prodromal AD (N = 420)	Prodromal AD stage-1 (N = 216)	Prodromal AD stage-2 (N = 197)
CSF A β_{1-42} (pg/mL)	6545 (2218–82,165)	5257 (1705–88,214)	-	1469 (619–6964)
¹⁸ F-AV-45-PET (SUVr)	9312 (2850–2,51,663)	-	3925 (1074–5,00,647)	-
CSF tau (pg/mL)	3611 (1539–16,455)	2760 (1148–13,667)	1791 (634–17,633)	3577 (1044–1,62,076)
FDG-PET (SUVr)	888 (662–1254)	447 (330–640)	486 (322–817)	396 (262–665)
Whole-brain volume (cm ³)	165 (147–187)	108 (94–126)	120 (97–152)	102 (85–125)
Hippocampal volume (mm ³)	200 (178–227)	142 (123–167)	197 (155–259)	102 (85–123)
Ventricular volume (cm ³)	274 (236–322)	231 (189–289)	260 (191–373)	185 (144–245)
CDR sum of boxes	636 (536–765)	443 (359–559)	603 (433–897)	355 (273–481)
MMSE score	992 (803–1256)	625 (488–830)	1263 (791–2328)	371 (283–507)
ADAS-Cog	1420 (1105–1893)	703 (548–933)	1346 (854–2431)	418 (318–573)
ADCS-PACC	745 (583–985)	433 (326–604)	488 (308–889)	395 (279–601)
ADCS-PACC without digit symbol test	1509 (1160–2043)	688 (528–932)	1576 (934–3203)	342 (262–465)

Abbreviations: A β_{1-42} , amyloid β 1-42; AD, Alzheimer's disease; ADAS-Cog, Alzheimer's Disease Assessment Scale-Cognitive; ADCS-PACC, Alzheimer's Disease Cooperative Study Preclinical Alzheimer Cognitive Composite; CDR, Clinical Dementia Rating scale; CSF, cerebrospinal fluid; FDG, fluorodeoxyglucose; AV-45, florbetapir; MCI, mild cognitive impairment; MMSE, Mini-Mental State Examination; PET, positron emission tomography.

NOTE. Sample size was estimated for a hypothetical 3-year randomized-controlled trial with two arms, showing an expected treatment effect of 25% reduction of decline in outcome measures with a power of 80%, a two-sided alpha of 5%, and a 10% annual dropout rate. “-” represents not calculated as slope is not significantly different from 0 (see Table 4). Data are mean (95% confidence interval).

Also, for all definitions of prodromal AD in MCI patients, brain atrophy outcome measures showed the most worsening over time and yielded the smallest sample size estimate. The smallest sample size estimate was observed for prodromal AD stage-2. In these subjects, amyloid measures and CSF tau levels showed limited change over time, which resulted in very large sample size estimates. FDG-PET showed some decline and this resulted in reasonable sample size estimates. Of the cognitive measures, the smallest sample size estimate was obtained with the CDR-sob and ADCS-PACC without digit symbol test, which became the most abnormal over a time.

Across all predementia subjects, the inclusion criteria incorporating AD biomarkers considerably reduced sample size estimates. However, this was at the expense of an increase in the number of subjects needed to be screened, since only a subset of subjects with normal cognition or MCI have abnormal AD biomarkers. Still, despite the large numbers-needed-to-be-screened, a previous cost-benefit analysis showed that enriching trials by refining inclusion criteria with the use of a CSF A β_{1-42} /tau in MCI can reduce trial costs with 60%, because fewer subjects are needed to show an effect, and this outweighs the increased costs for screening [33].

The observation that both in subjects with preclinical AD and in subjects with prodromal AD CSF A β_{1-42} decreased in stage-2 but not in stage-1, whereas ¹⁸F-AV-45-PET changed in stage-1 but not in stage-2 suggests that these markers (in part) reflect different disease processes, although it should be noted that the slopes of change did not differ between stage-1 and stage-2 [34].

In our fictive trial, we found reduced sample size estimates when enriching for AD biomarkers, and this is in line with sample size estimates from previous studies that

used a similar fictive trial design approach. In preclinical AD, one study found that with CDR-sob and the MMSE score as outcome measure, the use of an abnormal amyloid marker for inclusion reduced sample size relative to cognitively subjects unselected for biomarkers although sample size estimates in that study were twice as high as in our study [5,11]. Our observed lack of change on the PACC may limit its use as endpoint in preclinical AD trials with a 3-year duration. Longer trial durations are necessary to detect potential treatment effects on cognitive outcomes in preclinical AD (Supplementary Table 20). A recent study using a subset of ADNI subjects with normal cognition showed that when constructing a cognitive composite measure using information from cognitive tests scores of those subjects who show clinical decline, the power to detect changes might be improved [35]. However, in subjects with preclinical AD this optimized composite measure showed only minimal change over time. Two other studies reported similar to our results a reduction of sample size estimates for hippocampal, ventricular and whole-brain volumes, and/or MMSE in prodromal AD compared with unspecified MCI [12,36]. While another study found that sample sizes were smaller for prodromal AD stage-2 than prodromal AD [9]. In addition, APOE ϵ 4 allele carriership as an additional risk marker might lead to even smaller size estimates [24,36,37]. Although these previous studies show that AD biomarkers can decrease sample size estimates, none of these studies have compared the impact on these estimates of amyloid, injury, and clinical outcome markers over different preclinical and prodromal AD stages. Our study further extends these findings, as we covered all predementia stages of AD and demonstrated that also the stages influence sample size estimates.

4.1. Methodological issues

For our power calculation, we made several assumptions that might have influenced our findings. We used linear mixed models with random intercepts and a fixed slope to assess changes over time in outcome measures. The use of a fixed slope might have underestimated sample sizes because it does not take into account variability in slopes between subjects [7,10]. In a post hoc analysis, we found that the use of random slope models indeed resulted in larger sample size estimates, although the model did not converge for several outcome variables (Supplementary Tables 3 and 4). Also, for the present analyses we used effect sizes of 25%, which can be considered to be an upper bound for a relatively small, but clinically relevant effect size. Larger effect sizes provide smaller sample size estimations, and lower numbers needed to screen. For example, with a hypothetical reduction of 35% sample sizes would be 45% smaller (Supplementary Tables 14–17). In addition, we did not correct for age effects in our slopes' analyses, although these are likely to be present in cognitively normal subjects with normal AD markers [7,10]. It can be argued that one should not correct age effects because it cannot be excluded that even in cognitively normal subjects, with normal AD markers at baseline, the change in outcome effects is still reflecting (in part) AD and which might lead to an underestimation of treatment effects. Still, we performed additional post hoc analyses correcting for age, and this resulted in increased sample sizes or, in the case of preclinical AD stage-1, were often not estimable (Supplementary Tables 5 and 6). Of note is that sample size estimates for MCI subjects in stage-1 and stage-2 increased for neuronal injury outcome measures and remained comparable for cognitive outcome measures. A similar effect has been reported by a recent study in another study sample [36]. Possibly this difference reflects that brain structural changes are more intertwined with aging, making it difficult to detangle age from disease processes. Cognitive outcome measures seemed to be less affected by aging, but this might also reflect the much slower pace of decline in preclinical AD for these measures. In our design, we chose to model a hypothetical trial in which a treatment effect would result in a reduced change, to illustrate the effect of different inclusion criteria and outcome measure on sample size estimates. Other sample size estimates would be obtained in designs that assume that treatment might result in biomarker or cognitive improvements. In our analysis, we combined amyloid and injury markers either based on CSF or PET markers, which might have influenced our results because these modalities might reflect different processes. Post hoc analysis in subjects with markers for both modalities showed that the concordance for amyloid status was 85% for subjects with normal cognition and 92% for subjects with MCI (Supplementary Table 7), similar to previous studies [34,38–40]. Concordance for preclinical stage-1 and stage-2 classification in subjects who were amyloid

positive on both CSF and PET was 85% for preclinical AD and 50% for prodromal AD (Supplementary Table 8). Additional exploratory analyses for the subgroup of prodromal AD with classification using only CSF measures provided slope and sample size estimates largely comparable to the total group of stage-1 and stage-2 subjects (Supplementary Table 19). We used FDG-PET as a neuronal injury marker, if amyloid was measured with a PET tracer. However, using two PET, markers may be impractical. We therefore repeated these analyses with hippocampal volume as a neuronal injury marker. We found a concordance of FDG-PET hypometabolism and hippocampal atrophy of 74% for subjects with normal cognition and of 64% for MCI patients (Supplementary Table 9). Slope and sample size estimates were largely similar (Supplementary Tables 10–13), suggesting that MRI-based hippocampal volume can be used as a more practical alternative to FDG-PET for neuronal injury definition.

Our findings show that the definition of predementia AD influences the effects that can be found for outcome measures. This has important implications for trial design. For our hypothetical trial in subjects with preclinical AD, the estimated sample sizes were smallest in subjects at stage-2, using volumetric markers as outcome measurement. In subjects with prodromal AD, the smallest sample size estimates were obtained when including subjects at stage-2, using brain volume and cognitive measures as outcome measurements. Although a change in cognition would be the most clinically relevant outcome, the estimated sample sizes required to demonstrate effects on such markers were about three to four times larger than estimates based on MRI markers. This highlights the need for novel cognitive tests that are more sensitive for cognitive decline in subjects with preclinical and prodromal AD [11,41].

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) 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: AbbVie; Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc; Eisai Inc; Elan Pharmaceuticals, Inc; Eli Lilly and Company; Euroimmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc; Fujirebio; GE Healthcare; IXICO Ltd; Janssen Alzheimer Immunotherapy Research & Development, LLC; Johnson & Johnson Pharmaceutical Research & Development LLC; Lumosity; Lundbeck; Merck & Co., Inc; Meso Scale Diagnostics, LLC; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc; Piramal Imaging;

Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. 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.

The research leading to these results has received support from the Innovative Medicines Initiative Joint Undertaking under grant agreement no 115372, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies' in kind contribution.

Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.trci.2017.08.005>.

RESEARCH IN CONTEXT

1. Systematic review: We searched PubMed for relevant studies using, "preclinical Alzheimer's disease (AD)", "prodromal AD", "sample size", "trials", "normal cognition", "MCI", "MCI due to AD", "biomarkers", and "cognition". Previous studies reported that sample size depends on outcome measures, but it is still unclear how different definitions of predementia AD influence outcome and sample size estimates.
2. Interpretation: Our main finding is that the definition of predementia AD influence rate of decline and sample size estimates for a fictive trial with a treatment that would slow down decline by 25%. The smallest sample size was seen with predementia AD defined by the presence of both abnormal amyloid and injury markers at baseline, as well as brain volumetric markers as the outcome. The sample size was largest for cognitive outcomes.
3. Future directions: The definition of predementia is important in AD-trial design because this influences rate of decline and sample size estimates. Although cognition is often considered to be the most relevant outcome, they are relatively insensitive to capture decline in predementia AD.

References

- [1] Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, 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. *Alzheimer's Dement* 2011;7:280-92.
- [2] Dubois B, Feldman HH, Jacova C, Hampel H, Molinuevo L, Blennow K, et al. Position Paper Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria. *Lancet Neurol* 2014;13:614-29.
- [3] Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's Dement* 2011;7:270-9.
- [4] Vemuri P, Wiste HJ, Weigand SD, Knopman DS, Trojanowski JQ, Shaw LM, et al. Serial MRI and CSF biomarkers in normal aging, MCI, and AD. *Neurology* 2010;75:143-51.
- [5] Schott JM, Bartlett JW, Fox NC, Barnes J. Increased brain atrophy rates in cognitively normal older adults with low cerebrospinal fluid A β 1-42. *Ann Neurol* 2010;68:825-34.
- [6] Schott JM, Bartlett JW, Barnes J, Leung KK, Ourselin S, Fox NC. Reduced sample sizes for atrophy outcomes in Alzheimer's disease trials: Baseline adjustment. *Neurobiol Aging* 2010;31:1452-62. 1462.e1-2.
- [7] McEvoy LK, Edland SD, Holland D, Hagler DJ, Roddey JC, Fennema-Notestine C, et al. Neuroimaging enrichment strategy for secondary prevention trials in Alzheimer disease. *Alzheimer Dis Assoc Disord* 2010;24:269-77.
- [8] Yu P, Dean RA, Hall SD, Qi Y, Sethuraman G, Willis BA, et al. Enriching amnesic mild cognitive impairment populations for clinical trials: optimal combination of biomarkers to predict conversion to dementia. *J Alzheimers Dis* 2012;32:373-85.
- [9] Holland D, McEvoy LK, Desikan RS, Dale AM. Enrichment and stratification for predementia Alzheimer disease clinical trials. *PLoS One* 2012;7:e47739.
- [10] Holland D, McEvoy LK, Dale AM. Unbiased comparison of sample size estimates from longitudinal structural measures in ADNI. *Hum Brain Mapp* 2012;33:2586-602.
- [11] Grill JD, Di L, Lu PH, Lee C, Ringman J, Apostolova LG, et al. Estimating sample sizes for predementia Alzheimer's trials based on the Alzheimer's Disease Neuroimaging Initiative. *Neurobiol Aging* 2013;34:62-72.
- [12] Barnes J, Bartlett JW, Fox NC, Schott JM. Targeted recruitment using cerebrospinal fluid biomarkers: implications for Alzheimer's disease therapeutic trials. *J Alzheimers Dis* 2013;34:431-7.
- [13] Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 1975;12:189-98.
- [14] Berg L. Clinical Dementia Rating (CDR). *Psychopharmacol Bull* 1988;24:637-9.
- [15] Hachinski VC, Iliff LD, Zilhka E, Du Boulay GH, McAllister VL, Marshall J, et al. Cerebral blood flow in dementia. *Arch Neurol* 1975;32:632-7.
- [16] Sheikh JYJ. Geriatric Depression Scale (GDS): recent evidence and development of a shorter version. In: *Clinical Gerontology: A Guide to Assessment and Intervention*. New York: The Haworth Press; 1986.
- [17] Shaw LM, Vanderstichele H, Knopik-Czajka M, Clark CM, Aisen PS, Petersen RC, et al. Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. *Ann Neurol* 2009; 65:403-13.
- [18] Clark CM, Schneider JA, Bedell BJ, Beach TG, Bilker WB, Mintun MA, et al. Use of florbetapir-PET for imaging beta-amyloid pathology. *JAMA* 2011;305:275-83.

- [19] Landau SM, Harvey D, Madison CM, Koeppe RA, Reiman EM, Foster NL, et al. Associations between cognitive, functional, and FDG-PET measures of decline in AD and MCI. *Neurobiol Aging* 2011;32:1207–18.
- [20] Freeborough PA, Fox NC. The boundary shift integral: an accurate and robust measure of cerebral volume changes from registered repeat MRI. *IEEE Trans Med Imaging* 1997;16:623–9.
- [21] Leung KK, Clarkson MJ, Bartlett JW, Clegg S, Jack CR Jr, Weiner MW, et al. Robust atrophy rate measurement in Alzheimer's disease using multi-site serial MRI: tissue-specific intensity normalization and parameter selection. *Neuroimage* 2010;50:516–23.
- [22] Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* 2002;33:341–55.
- [23] Rosen WG, Mohs RC, Davis KL. A new rating scale for Alzheimer's disease. *Am J Psychiatry* 1984;141:1356–64.
- [24] Hua X, Ching CRK, Mezher A, Gutman BA, Hibar DP, Bhatt P, et al. MRI-based brain atrophy rates in ADNI phase 2: acceleration and enrichment considerations for clinical trials. *Neurobiol Aging* 2016; 37:26–37.
- [25] Ard MC, Edland SD. Power Calculations for Clinical Trials in Alzheimer's Disease. *J Alzheimers Dis* 2011;26:369–77.
- [26] Van Rossum IA, Visser PJ, Knol DL, Van Der Flier WM, Teunissen CE, Barkhof F, et al. Injury markers but not amyloid markers are associated with rapid progression from mild cognitive impairment to dementia in Alzheimer's disease. *J Alzheimers Dis* 2012;29:319–27.
- [27] Desikan RS, McEvoy LK, Thompson WK, Holland D, Brewer JB, Aisen PS, et al. Amyloid- β -Associated Clinical Decline Occurs Only in the Presence of Elevated P-tau. *Arch Neurol* 2012;69:709–13.
- [28] Mormino EC, Betensky RA, Hedden T, Schultz AP, Amariglio RE, Rentz DM, et al. Synergistic effect of β -amyloid and neurodegeneration on cognitive decline in clinically normal individuals. *JAMA Neurol* 2014;71:1379–85.
- [29] Toledo JB, Weiner MW, Wolk DA, Da X, Chen K, Arnold SE, et al. Neuronal injury biomarkers and prognosis in ADNI subjects with normal cognition. *Acta Neuropathol Commun* 2014;2:26.
- [30] Vos SJB, Verhey F, Frölich L, Kornhuber J, Wiltfang J, Maier W, et al. Prevalence and prognosis of Alzheimer's disease at the mild cognitive impairment stage. *Brain* 2015;138:1327–38.
- [31] Fagan AM, Xiong C, Jasielec MS, Bateman RJ, Goate AM, Benzinger TL, et al. Longitudinal change in CSF biomarkers in autosomal-dominant Alzheimer's disease. *Sci Transl Med* 2014; 6:226ra30.
- [32] Bertens D, Knol DL, Scheltens P, Visser PJ. Temporal evolution of biomarkers and cognitive markers in the asymptomatic, MCI, and dementia stage of Alzheimer's disease. *Alzheimers Dement* 2015;11:511–22.
- [33] Van Rossum IA, Vos S, Handels R, Visser PJ. Biomarkers as predictors for conversion from mild cognitive impairment to Alzheimer-type dementia: Implications for trial design. *J Alzheimers Dis* 2010; 20:881–91.
- [34] Mattsson N, Insel PS, Donohue M, Landau S, Jagust WJ, Shaw LM, et al. Independent information from cerebrospinal fluid amyloid- β and florbetapir imaging in Alzheimer's disease. *Brain* 2014; 138:772–83.
- [35] Insel PS, Donohue MC, Mackin RS, Aisen PS, Hansson O, Weiner MW, et al. Cognitive and functional changes associated with Abeta pathology and the progression to mild cognitive impairment. *Neurobiol Aging* 2016;48:172–81.
- [36] Fujishima M, Kawaguchi A, Maikusa N, Kuwano R, Iwatsubo T, Matsuda H. Sample Size Estimation for Alzheimer's Disease Trials from Japanese ADNI Serial Magnetic Resonance Imaging. *J Alzheimers Dis* 2016;56:75–88.
- [37] Mattsson N, Insel PS, Donohue M, Jagust W, Sperling R, Aisen P, et al. Predicting Reduction of Cerebrospinal Fluid β -Amyloid 42 in Cognitively Healthy Controls. *JAMA Neurol* 2015;72:554–60.
- [38] Jagust WJ, Landau SM, Shaw LM, Trojanowski JQ, Koeppe RA, Reiman EM, et al. Relationships between biomarkers in aging and dementia. *Neurology* 2009;73:1193–9.
- [39] Landau SM, Lu M, Joshi AD, Pontecorvo M, Mintun MA, Trojanowski JQ, et al. Comparing positron emission tomography imaging and cerebrospinal fluid measurements of β -amyloid. *Ann Neurol* 2013;74:826–36.
- [40] Palmqvist S, Zetterberg H, Blennow K, Vestberg S, Andreasson U, Brooks DJ, et al. Accuracy of Brain Amyloid Detection in Clinical Practice Using Cerebrospinal Fluid β -Amyloid 42. *JAMA Neurol* 2014;71:1282.
- [41] Donohue MC, Sperling RA, Salmon DP, Rentz DM, Raman R, Thomas RG, et al. The preclinical Alzheimer cognitive composite: measuring amyloid-related decline. *JAMA Neurol* 2014;71:961–70.