ELSEVIER

Neurobiology of Aging 31 (2010) 1443–1451

NEUROBIOLOGY OF AGING

www.elsevier.com/locate/neuaging

Enrichment through biomarkers in clinical trials of Alzheimer's drugs in patients with mild cognitive impairment

M. Lorenzi[a], M. Donohue[b], D. Paternicò[a], C. Scarpazza[a], S. Ostrowitzki[c], O. Blin[d], E. Irving[e], G.B. Frisoni[a,f,*], The Alzheimer's Disease Neuroimaging Initiative[†]

[a] *LENITEM Laboratory of Epidemiology, Neuroimaging and Telemedicine, IRCCS San Giovanni di Dio-FBF, Brescia, Italy*
[b] *Division of Biostatistics and Bioinformatics, University of California, San Diego, USA*
[c] *Translational Medicine, Neuroscience, F. Hoffmann-La Roche Ltd, Basel, Switzerland*
[d] *Clinical Investigation Centre (CIC-UPCET) and Department of Clinical Pharmacology, UMR-CNRS, 6193 Institute of Cognitive Neurosciences, CHU, Timone, Marseille, France*
[e] *Neurosciences CEDD, GlaxoSmithKline, Harlow, Essex, UK*
[f] *AFaR Associazione Fatebenefratelli per la Ricerca, Rome, Italy*

Received 15 February, 2010; received in revised form 29 April 2010; accepted 29 April 2010

Abstract

Clinical trials of disease modifying drugs for Alzheimer's disease (AD) in patients with mild cognitive impairment (MCI) might benefit from enrichment with true AD cases. Four hundred five MCI patients (143 converters and 262 nonconverters to AD within 2 years) of the Alzheimer's disease Neuroimaging Initative (ADNI) were used. Markers for enrichment were hippocampal atrophy on magnetic resonance (MRI), temporoparietal hypometabolism on FDG PET, cerebrospinal fluid (CSF) biomarkers (Abeta42, tau, and phospho-tau), and cortical amyloid deposition (11C–PIB positron emission tomography (PET)). Two separate enrichment strategies were tested to A) maximize the proportion of MCI converters screened in, and B) minimize the proportion of MCI converters screened out. Based on strategy A, when compared with no enrichment and ADAS-Cog as an outcome measure (sample size of 834), enrichment with 18F-FDG PET and hippocampal volume lowered samples size to 260 and 277 cases per arm, but at the cost of screening out 1,597 and 434 cases per arm. When compared with no enrichment and clinical dementia rating (CDR-SOB) as an outcome measure (sample size of 674), enrichment with hippocampal volume and Abeta42 lowered sample sizes to 191 and 291 cases per arm, with 639 and 157 screened out cases. Strategy B reduced the number of screened out cases (740 for [11C]-PIB PET, 101 hippocampal volume, 82 ADAS-COG and 330 for [18F]-FDG PET) but at the expense of decreased power and a relative increase size (740 for [11C]-PIB PET, 676 for hippocampal volume, 744 for ADAS-Cog, and 517 for [18F]-FDG PET). Enrichment comes at the price of an often relevant proportion of screened out cases, and in clinical trial settings, the balance between enrichment of screened in and loss of screened out patients should be critically discussed.

Keywords: Mild cognitive impairment; Conversion; Marker disease; Clinical trials; Enrichment

Drugs aimed to modify the course of Alzheimer's disease (AD) are under active development. These drugs might be maximally effective when prescribed early in the course of the disease. Amnestic mild cognitive impairment (MCI) is currently the earliest stage when patients with AD can be captured for clinical trial purposes, but the diagnostic category of MCI is contaminated by a sizable proportion (up to 50%) of patients who do not have AD. Indeed, all clinical trials with antidementia drugs that have been carried out in the MCI populations have failed to demonstrate a significant

* Corresponding author at: Tel.: +39 030 3501361; fax: +3902700435727.
E-mail address: gfrisoni@fatebenefratelli.it (G. Frisoni).

† Data used in the preparation of this article were obtained from the Alzheimer's disease Neuroimaging Initiative (ADNI) database (www.loni.ucla.edu/ADNI). 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. The complete listing of ADNI investigators is available at www.loni.ucla.edu/ADNI/Collaboration/ADNI_Manuscript_Citations.pdf.

0197-4580/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.neurobiolaging.2010.04.036

treatment effect (Feldman et al., 2007; Loy and Schneider, 2006; Petersen et al., 2005; Raschetti et al., 2007; Salloway et al., 2004), and one of the proposed reasons is contamination by non Alzheimer's cases (Visser et al., 2005).

It is widely believed that MCI patients with abnormal brain structure volume or metabolism, or biochemical marker profile are more likely to develop AD than the parent MCI population. A proposal for new diagnostic criteria has been developed that could allow diagnosis of AD at the MCI stage based on atrophy of medial temporal lobe structures (among which is the hippocampus) on structural magnetic resonance imaging (MRI), hypometabolism in the temporoparietal cortex on 18F-FDG positron emission tomography (PET), low Abeta42 or high tau or phospho-tau in the cerebrospinal fluid (CSF), and positivity on amyloid imaging with tracers such as 11C–PIB (Dubois et al., 2007). A corollary of this is that AD markers might be employed in clinical trials of MCI patients to screen out non-AD MCI cases and select a population of MCI enriched with truly AD cases to be randomized.

Of course, the ideal marker is one with 100% sensitivity and specificity, which would support screening out of all non AD and screening in all AD cases. However, this is hardly a realistic scenario in that markers will in all likelihood merely enrich screened out with true negatives and screened in with true positives. In a clinical trial scenario, a good marker will be one with high sensitivity: the ratio between the proportion of AD cases which are screened positive and included, i.e. true positive rate, and the proportion of AD cases which are screened negative and excluded, i.e. false negative rate. Data that allow estimation of this sensitivity and specifity are available from the Alzheimer's disease Neuroimaging Initative (ADNI) (Mueller et al., 2005). The ADNI has studied 399 MCI patients with structural MRI, 18F-FDG PET, CSF studies, and 11C–PIB and followed them to detect conversion to dementia. The aim of the present study was to assess the benefit of the enrichment of MCI patients with true AD cases by means of hippocampal atrophy on MRI, temporoparietal hypometabolism on 18F-FDG PET, CSF biomarkers (Abeta42, tau, and phospho-tau), and cortical amyloid deposition on 11C–PIB. All markers were measured on continuous scales and the optimal threshold for screening has been defined empirically based on the distribution of the marker in the 229 healthy elders of the ADNI database in whom the same markers have been collected.

1. Methods

1.1. Subjects

The subjects of this study were taken from the ADNI database (www.loni.ucla.edu/ADNI/Data) as of 29 September 2009. The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and nonprofit organizations, as a US$60mn, 5-year public-private partnership. The primary goal of ADNI has been to test whether serial magnetic resonance (MR) imaging, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials. The Principal Investigator of this initiative is Michael W. Weiner MD, VA Medical Center and University of California, San Francisco. ADNI is the result of efforts of many coinvestigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the USA and Canada. The initial goal of ADNI was to recruit 800 adults, ages 55 to 90, to participate in the research – approximately 200 cognitively normal older individuals to be followed for 3 years, 400 people with MCI to be followed for 3 years, and 200 people with early AD to be followed for 2 years. For up-to-date information see www.adni-info.org. Table 1 shows the salient features of the MCI converters and nonconverters and healthy elderly controls

Table 1
Subjects' sociodemographic and clinical features and availability of disease markers

	MCI		Healthy elders
	Converters N = 143	Nonconverters N = 262	
Age	74 ± 7	75 ± 8	76 ± 5
Gender (female)	52 (36%)	90 (34%)	110 (49%)
Education	16 ± 3	16 ± 3	16 ± 3
MMSE [range]	27 ± 2 [23–30]	27 ± 2 [24–30]	29 ± 1 [25–30
CDR-SOB	3.54 ± 2.32	1.64 ± 1.33	0.07 ± 0.78
ADAS-cog	21.22 ± 6.31	16.99 ± 6.2	9.49 ± 4.19
Length of follow-up (months) [range]	11 ± 8 [0–36]	21 ± 11 [0–36]	27 ± 10 [0–36]
no. of semiannual assessments [range]	2 ± 1 [1–5]	4 ± 1 [1–5]	4 ± 1 [1–5]
Hippocampal volume	143 (100%)	256 (98%)	225 (100%)
18F-FDG PET	61 (43%)	146 (56%)	105 (56%)
CSF Abeta42, tau, p-tau	73 (51%)	123 (47%)	107 (48%)
11C–PIB PET	19 (13%)	45 (17%)	18 (8%)

used for the present study. Mean participant age, gender, and education do not differ across groups.

1.2. Markers and neuropsychological scores

The measure of hippocampal atrophy was the mean left and right baseline hippocampal volume reported in the ADNI dataset, collected through manual tracing on high resolution 3D MR scans following the protocol of Jack and colleagues (Jack et al., 1995). The measure of temporoparietal hypometabolism was the t-sum developed by Herholz and colleagues (Herholz et al., 2002) on 18F-FDG PET images. This is an adimensional number ranging between 0 and infinity indicative of hypometabolism in the cortical regions found specific to AD including temporoparietal cortex, posterior cingulate and precuneus, frontal association cortex bilaterally. A value of 11,090 was empirically found to be the optimal threshold to distinguish AD patients from healthy elders (Herholz et al., 2002). The values of baseline CSF biomarkers (Abeta42, tau, and phospho-tau) were those reported in the ADNI database, measured through the Luminex xMAP platform. The measure of cortical amyloid deposition was defined as the mean value of the 11C–PIB PET images in the gray matter. This measure was obtained after a number of image processing steps. The [11C]-PIB PET images were first coregistered to the respective MR images and spatial normalized using the parameters determined from the normalization of MR images through the DARTEL procedure (Ashburner, 2007). 11C–PIB PET normalized images were then scaled to the cerebellum and the mean uptake value was finally computed on the regions defined by the gray matter DARTEL template.

Finally, the Alzheimer's disease assessment scale (ADAS) score was included in the battery of enrichment factors. The score employed in the study is the total score on the modified 13-item ADAS (Petersen et al., 2005), adapted from the Administration and Scoring Manual for the Alzheimer's disease Assessment Scale.

1.3. Data treatment and statistical analysis

Data treatment is summarized in Figure 1. The distribution of the markers was modeled using the fast Fourier transform to convolve the approximation of the empirical distribution with a Gaussian kernel and using linear approximation to evaluate the density at the specific point. The analysis was conducted in the R statistical computing environment (R Development Core Team, 2009) (www.R-project.org). Two different enrichment strategies were tested.

1.3.1. Enrichment strategy A: maximization of the proportion of screened-in mild cognitive impairment converters

Increasingly restrictive thresholds were defined based on the 70th, 85th, 95th, and 99th percentile of the distribution of marker values in healthy elders. The number of MCI converters and nonconverters among screened out and screened in on the biomarker (with 95% confidence interval) (Wilson, 1927) was computed for each threshold. The thresholds associated with the highest proportion of MCI converters among screened in were then chosen (thresholds shown in supplementary table). For temporoparietal hypometabolism, the threshold of 11,090 was also tested following the original results from Herholz et al. (Herholz et al., 2002), and for CSF biomarkers the threshold of 192 pg/mL was tested following Shaw et al. (Shaw et al., 2009).

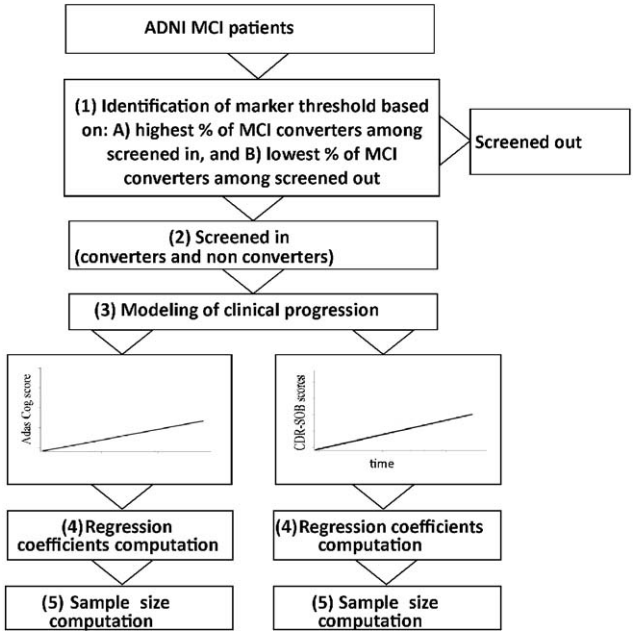


Fig. 1. Flow chart of the statistical procedures and data treatment. (1) The distribution of the markers was modeled in the healthy group and based on different thresholds two enrichment strategies were tested on MCI group: (A) maximization of the proportion of screened in MCI converters and (B) minimization of the proportion of screened out MCI converters; (2) the screened in MCI group (converters and nonconverters) was used to (3) model the progression of the clinical outcomes (ADAS-cog, CDR-SOB) with random effect models. Finally, the resulting coefficients (4) were used to compute sample sizes (5).

1.3.2. Enrichment strategy B: minimization of the proportion of screened-out mild cognitive impairment converters

At each percentile distribution of the control population, the ratio between the number of nonconverters and converters among screened out was computed. The thresholds were chosen according to an efficiency criterium to minimize the proportion of excluded converters among the screened out population. The thresholds optimizing the previous criteria (shown in supplementary table) were then employed to carry out the ensuing power computations.

For each marker, the ratio between the proportion of converters and nonconverters among the screened in was computed using a classical Bayesian approach (Albert, 2009). The distributions of the proportions $p_{converters}$ were inferred using a binomial likelihood and a beta(1,1) as noninformative conju-

gate prior, resulting in a beta distribution for p of parameters $a = n_{converters}+1$ and $b = n_{non\ converters}+1$. The subsequent statistical analyses were then performed on the ratios between $p_{converters}$ and $p_{nonconverters} = 1-p_{converters}$ obtained from drawing 10,000 samples from the posterior distribution.

Finally, the screened groups enriched with the different markers were used to compute the sample size required to detect a hypothetical 25% difference in the rate of decline in a 2-year placebo controlled randomized clinical trial with 6-month visit intervals. The group resulting from [11C]-PIB enrichment was dropped from further consideration due to inadequate size. Longitudinal ADAS-COG and CDR-SOB scores available from the ADNI dataset were used to fit random effects models to estimate the annual rate of change for each enrichment scenario. The models included random intercept and slope and, based on the parameter estimates with 95% confidence intervals, the sample size required with associated confidence intervals was then computed using the formula of Liu and Liang (1997). We note here that the resulting estimates depend on the composition of the different groups from which the model is fitted. Sample size estimates are inflated to account for a 30% dropout rate over 2 years.

2. Results

Figure 2 shows that the distribution of markers was roughly bell-shaped for all markers in the three groups of healthy elderly controls, MCI converters, and MCI nonconverters. A hint of a bimodal distribution could be appreciated in healthy elders for CSF Abeta42, consistent with the notion that some healthy elders might host presymptomatic forms of the disease. A clear bimodal distribution was present in MCI nonconverters for PIB, consistently with the notion that some of the MCI nonconverters might convert soon, as also supported by the observation of a large share of CSF Abeta42 values in MCI nonconverters lying in the conversion area. Interestingly, a small bell-shaped tail can be appreciated for [11C]-PIB PET in MCI converters lying in the healthy elders area, suggesting that some MCI converters to Alzheimer's disease might on the contrary have other forms of dementia.

The figure also shows that with enrichment strategy A, increasingly restrictive thresholds (from none to the 99th percentile of the distribution of healthy elderly controls) generally lead to select a monotonously increasingly enriched proportion of future converters among those screened-in except in the case of CSF markers, where the correspondence of the distribution curves of MCI converters and nonconverters led to a monotonous increase. The highest proportion of future converters was achieved by hippocampal volume thresholded at the first percentile of the healthy elders distribution, and [11C]-PIB PET thresholded at the 95th percentile, increasing from 38% with no threshold to 59% and 60%, respectively. However, this enrichment was obtained at the expense of a marked increase of screened out rate, up to 77% and 84% of those MCI enrolled.

Lastly, the figure shows the thresholds found with enrichment strategy B. The lowest proportion of screened out converters was achieved by ADAS-cog (7.5%) and [11C]-PIB PET (9%) at the 58th and 85th percentile respectively. For CSF biomarkers, the proportion of screened out MCI converters was monotonously decreasing with decreasing marker values, thus preventing us from identifying an optimal threshold.

Figure 3 shows that with enrichment strategy A the most favorable ratio between MCI converters and nonconverters is achieved by [11C]-PIB PET (ratio of 1.5), but due to the small group size the confidence interval is very large (3.69–0.62) and the point estimate is poorly reliable. A slightly less favorable ratio (1.46) is achieved by hippocampal volume, but with a much more accurate point estimate. A lower favorable ratio (1.14) is achieved by FDG PET. However, in all of these cases the proportion of screened out is remarkably high (84%, 77%, and 86%). All other markers yield ratios below one, ranging between 0.98 (ADAS-Cog) and 0.87 (CSF Abeta42 and p-tau). It should be noted that for five markers ([11C]-PIB PET, hippocampal volume, ADAS-Cog, CSF tau, and CSF Abeta42) the decreasingly favorable ratio of MCI converters to nonconverters was associated with an expected decreased proportion of screened out (from 84% down to 77%, 56%, 38%, and 35%), and for two markers (CSF tau/Abeta42, CSF p-tau) the proportion of screened out was relatively high (46% and 55%) despite an unfavorable ratio of MCI converters to nonconverters. It should be noted that, although all markers led to a significant enrichment (ratios always significantly greater than the reference condition) the ratio of hippocampal volume was significantly greater than all other ratios except PIB PET due to its low group size and wide confidence interval, and [18F]-FDG PET.

With enrichment strategy B, [11C]-PIB PET at the 85th percentile leads again to the most favorable ratio (ratio of 1.00) between MCI screened in converters and nonconverters. This ratio is lower than that obtained with strategy A (ratio of 1.50) as well as for the other markers, associated with ratios ranging between 0.56 and 0.64.

Table 2 shows that enrichment strategy A leads to identify 18F-FDG PET as the marker associated with the lowest sample size per arm for a hypothetical 24-month trial in MCI patients of a disease modifying drug with 25% efficacy and 90% power and ADAS-COG as an outcome measure (260 cases per arm, estimated from the screened group of 28 patients), but at the cost of screening out 1,597 cases per arm. When CDR sum of boxes was used as an outcome measure, the lowest sample size was achieved by hippocampal volume with 191 cases per arm and 639 screened out cases. CSF Abeta42 is associated with the lowest screened

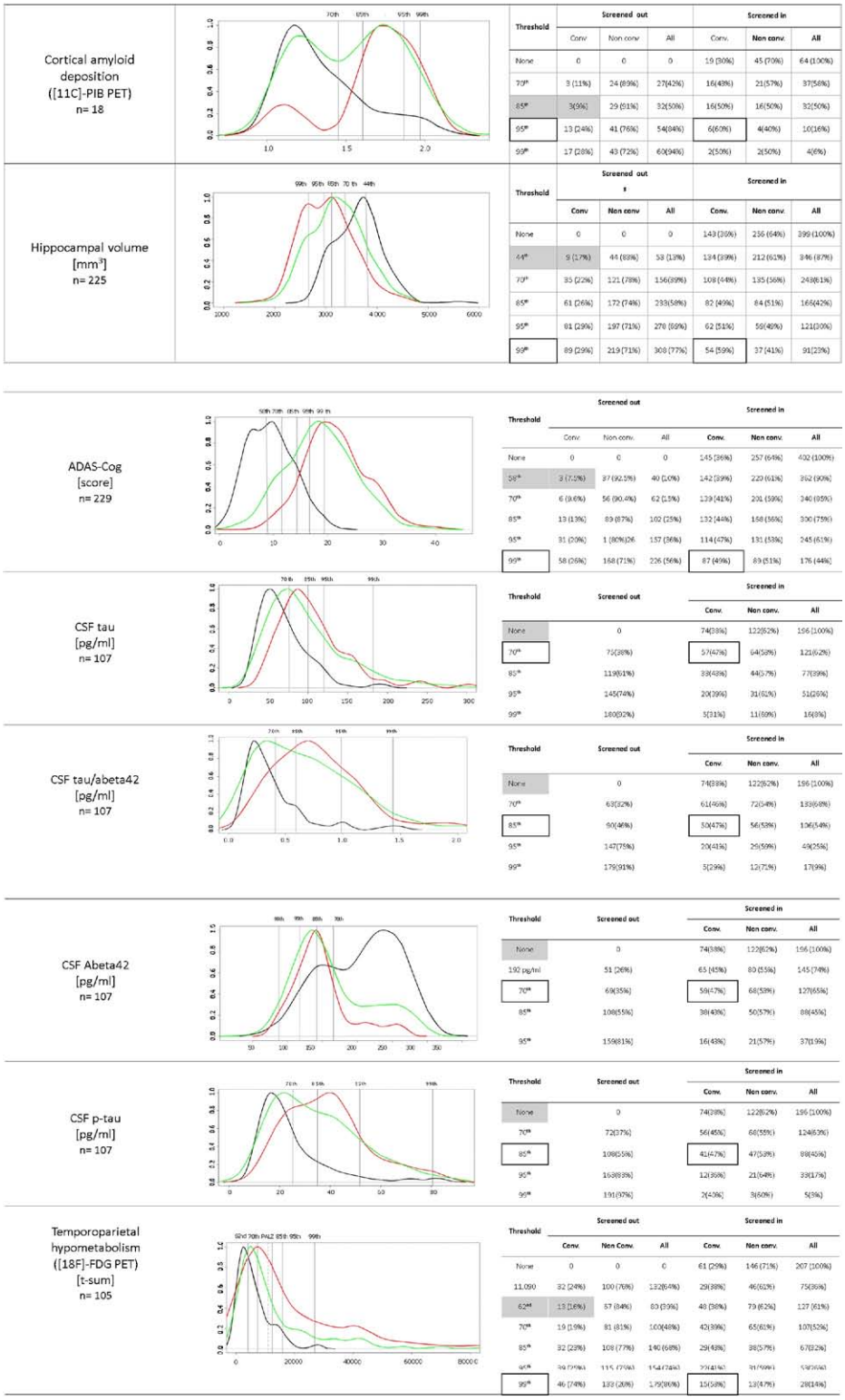
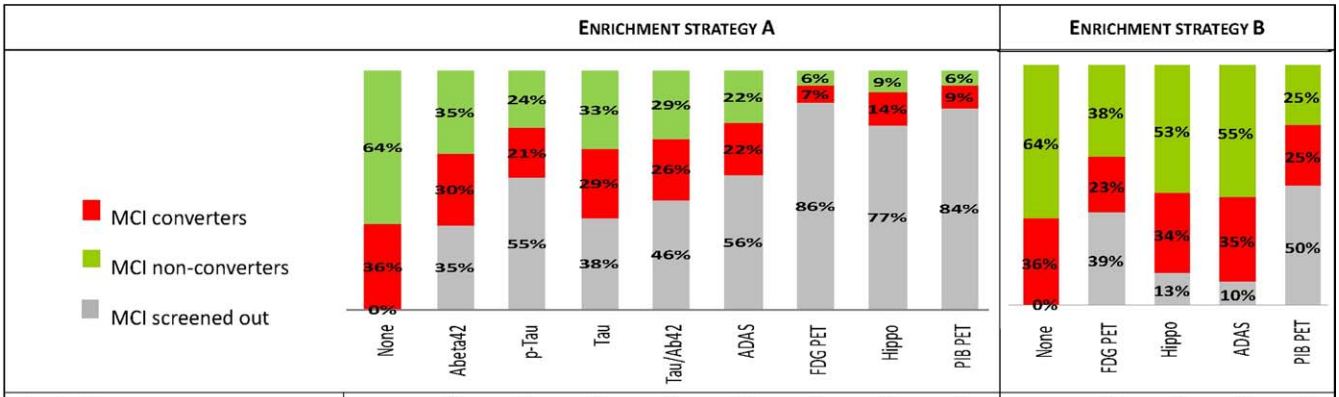


Fig. 2. MCI nonconverters and MCI converters among screened in and screened out at increasingly restrictive thresholds of Alzheimer's disease markers. Red, green, and black lines denote the distributions of MCI converters, MCI nonconverters, and healthy elderly controls rescaled to the common range (0 to 1). Thresholds refer to the distribution of the marker values in healthy elders. The percentages of all screened out and all screened in refer to the whole group of MCI patients, while the percentages of converters and nonconverters refer to screened in and screened out. Cells with thick margins denote the threshold associated with the highest percentage of converters among screened in (enrichment strategy A) and gray cells those associated with the lowest percentage of converters among screened out (enrichment strategy B). These thresholds have been used to compute enrichment in Figure 3 and sample size in Table 2.



Threshold		---	70th	85th	70th	85th	99th	99th	99th	95th	---	62nd	44th	58th	85th
Ratio MCI converters/non-converters															
Point estimate		0.56	0.87	0.87	0.89	0.89	0.98	1.14	1.46	1.50	0.56	0.61	0,63	0.64	1
95% C.I.	Upper	0.64	1.11	1.17	1.15	1.17	1.20	1,97	1.95	3.69	0.64	0	0.78	0.52	1.97
	Lower	0.48	0.67	0.64	0.69	0.68	0.79	0,69	1.09	0.62	0.48	2	0.51	0.8	0.5
Signif. difference from:	PIB PET	<.05	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	---	n.s.	n.s.	n.s.	n.s.	---
	Hippo vol.	<.05	<.05	<.05	<.05	<.05	<.05	n.s.	---	n.s.	n.s.	n.s.	---	n.s.	n.s.
	ADAS-Cog	<.05	n.s.	n.s.	n.s.	n.s.	---	n.s.	n.s.	<.05	n.s.	n.s	n.s.	---	n.s.
	FDG-PET	<.05	n.s.	n.s.	n.s.	n.s.	n.s.	---	n.s.	<.05	n.s	---	n.s	n.s	n.s.
	Tau	<.05	n.s.	n.s.	---	n.s	n.s.	n.s.	n.s.	<.05	---	---	---	---	---
	Tau/Aβ42	<.05	n.s.	n.s.	n.s.	---	n.s.	n.s.	n.s.	<.05	---	---	---	---	---
	Aβ42	<.05	---	n.s.	n.s	n.s.	n.s.	n.s.	n.s.	<.05	---	---	---	---	---
	p-tau	<.05	n.s.	---.	n.s	n.s.	n.s.	n.s.	n.s.	<.05	---	---	---	---	---

Fig. 3. Synopsis of the proportion of screened out, MCI converters, and MCI nonconverters and the ratio between MCI converters and nonconverters with the two different enrichment strategies. CI: confidence interval. Statistical significance of the difference between pairs of ratios was computed through simulations from the posterior distribution (see Methods).

out group size using both ADAS-cog (269 and sample size of 500 per arm) and CDR sum of boxes as outcome measure (157 cases and sample size of 291 per arm).

Table 2 also shows that with enrichment strategy B and ADAS-cog as outcome measure the lowest sample size per arm is associated with 18F-FDG PET (517 cases estimated

Table 2
Sample size and screened out per arm (95% CI) for a hypothetical 24-months trial in MCI patients of a disease modifying drug with 25% efficacy and 90% power with ADAS-COG and CDR sum of boxes as outcome measures. (A) Thresholds refer to the percentiles of the distribution of the marker values in healthy elders which maximize the number MCI converters among screened in (enrichment strategy A). [11C]-PIB PET estimates are missing due to inadequate group size of MCI PIB positive in the ADNI dataset at the threshold requested. (B) Thresholds refer to the percentiles in the distribution of the marker values in healthy elders which minimize the number of MCI converters among screened out (enrichment strategy B). In (A) and (B) columns of gray boxes denote similar sample size estimates

	Threshold	ADAS-Cog		CDR sum of boxes	
		Sample size	Screened out	Sample size	Screened out
A. Threshold maximizing % of MCI converters AMONG screened-in					
No enrichment	None	834 (631–1,154)	0	674 (524–900)	0
ADAS-Cog	99th	617 (430–959)	785 (547–1,220)	270 (213–354)	344 (271–450)
CSF tau	70th	531 (361–860)	325 (221–527)	310 (230–436)	190 (141–267)
CSF Abeta42	70th	500 (347–780)	269 (187–420)	291 (221–400)	157 (119–215)
CSF tau/abeta42	85th	453 (310–723)	386 (264–616)	264 (199–370)	225 (169–315)
Hippocampal volume	99th	434 (293–711)	1,452 (981–2,380)	191 (147–260)	639 (492–870)
CSF p-tau	85th	396 (269–643)	484 (329–786)	287 (207–424)	351 (253–518)
[18F]-FDG PET	99th	260 (151–553)	1,597 (927–3,397)	240 (139–506)	1,474 (854–3,108)
B. Threshold minimizing % of MCI converters among screened-out					
No enrichment	None	834 (631–1,154)	0	674 (524–900)	0
ADAS-Cog	58th	744 (566–1027)	82 (62–114)	509 (404–657)	56 (44–73)
[11C]-PIB PET	84th	740 (336–2779)	740 (336–2779)	351 (203–757)	351 (203–757)
Hippocampal volume	44th	676 (517–923)	101 (77–137)	566 (443–747)	84 (66–111)
[18F]-FDG PET	62th	517 (360–810)	330 (230–517)	464 (324–717)	296 (207–458)

from 127 patients and 330 screened out cases). By contrast, considering the CDR sum of boxes as outcome measure, 11C–PIB PET leads to the lowest sample size with 351 cases per arms and an equal number of screened out cases. AdAS-cog achieves the lowest number of screened out using both ADAS-cog (82 cases and a sample size of 744 cases per arm) and CDR sum of boxes (56 cases and sample size of 509) as an outcome measure.

Finally, we note that the use of CDR sum of boxes as an outcome measure brings about lower sample size compared with that associated with ADAS-cog.

3. Discussion

We have shown that the screening procedure with imaging and biological markers can lead to a significant enrichment of groups of MCI patients enrolled in clinical trials of AD drugs with "true AD cases",, i.e. patients who will convert in the following months. In this study, two different strategies enabled respectively to (a) enrich the screened in group with true AD patients (MCI converters) and (b) control the number of screened out cases, reducing the loss of MCI converters. The unenriched ratio between converters and nonconverters of 0.56, i.e. almost 1 : 2, can be reversed with strategy A to 1.46, i.e. about 3 : 2. This enrichment comes to the price of a sometimes relevant proportion of screened out MCI patients falling below threshold, that increase with increasing enrichment and can amount to as much as 84% of all MCIs. This percentage is reduced with strategy B, varying between 10% and 50%, and comes with the advantage of a reduced number of true converters lost (between 7.5% and 17% of the whole MCI population), although the screened-in populations are characterized by a lower ratio of converters to nonconverters, albeit significantly higher than the unenrichment scenario. Interestingly, CSF biomarkers did not exhibit a consistent threshold minimizing the number of excluded converters, reflecting high specificity. In fact, the curve describing the proportion of the excluded converters was monotonously increasing with decreasing threshold values, i.e. the proportion of MCI nonconverters excluded was increasing when closer to nonpathological values, preventing to define an optimal value for strategy B.

Thresholds resulting from strategy A often led to an unrealistic high percentage of screened out patients (up to 86%) as well as markers values lying in the pathological range, such as for the 11C–PIB whole brain uptake close to the value of 2 or ADAS-Cog score of 19.4. This inconvenient is mitigated by the adoption of strategy B, where marker values are closer to the thresholds employed for diagnostic purposes (Supp Table). This result is achieved at the expense of a less favorable proportion of MCI converters in the screened in population.

A key point emerging from the current study is the role of the markers thresholds chosen for the screening procedure, and the impact of their use in the resulting clinical practice.

In a hypothetical clinical trial, the balance between enrichment of screened in and loss of screened out patients should be viewed in the light of the gain of power and the relative decreased costs brought about by enrichment and the increased costs brought about by the exclusion of screened out patients.

A large number of studies have recently shown that MCI patients positive to one or more AD biological and imaging markers have greater chance to convert to AD (De Leon et al., 2007; Hampel et al., 2008). Some (Ferris, 2002; Risacher et al., 2009) have suggested that markers may help identify MCI individuals at increased risk of conversion to AD, thus assisting researchers striving to enrich clinical trial populations with people with latent AD, but to the best of our knowledge no study has so far estimated the extent of enrichment as well as the inevitable costs in terms of screened out. In 2002, Ferris argued that "one approach to reducing the cost would be to recruit 'enriched' samples of subjects who are at greater risk of developing AD during the trial" and underlined that the major effort required to screen and recruit large numbers of subjects for such trials would contribute to the cost. While acknowledging that research to develop more efficient assessment methods is needed, he suggested that data acquisition over the Internet might be an efficient and practical tool. Thanks to the recent availability of the public ADNI dataset, we showed that hippocampal volumetry might represent an efficient strategy for enrichment for the minimization of screened in patients for clinical trials. By contrast, a criterion optimizing the overall number of patients required for the trial (screened in + screened out), would lead to different conclusions. In fact, the results emerging from this study show that, despite the lowest sample size required for subjects to be included in the trial based on hippocampal volume, the number of screened out subjects is greater in comparison with other markers. For example, if we consider CDR sum of boxes as outcome measure under the strategy A, CSF Abeta42 guarantees similar sample size to hippocampal volume but limiting the number of screened out subjects, thus reducing the total amount of subjects required for the trial. The opportunity to resort to alternative strategies with lower enrichment power such as FDG PET and PIB PET markers should be judged in the context of the lower costs for screening and the biological mechanism of the drug under trial.

Most disease modifiers presently in Phase II and III clinical trials are targeting beta amyloid and an enrichment strategy aiming to select MCI patients with brain amyloidosis might be appropriate. We have shown that the use of CSF Abeta42 to select MCI patients to enroll in a trial has significantly lower effectiveness at enrichment with fast converters than hippocampal volume. Although it might be contended that CSF Abeta42 has high sensitivity and specificity to recruit patients with brain amyloidosis (Jagust et

al., 2009), it should also be acknowledged that some of these patients might convert significantly later than the 24 months of a clinical trial (Jack et al., 2010). Thus, a judgment should be made over which criterion should be followed for enrichment, i.e. efficiency or biological plausibility. Unfortunately, the low group size of MCI patients for whom 11C–PIB PET is available prevents accurate estimates of the effectiveness of this enrichment strategy. Future studies with larger group sizes will allow us to answer this question.

This study is a technical exercise that should be translated into practice with some caveats. The enrichment strategy in a clinical trial of drug "x" that will prove effective in slowing disease progression should be viewed in light of the intended licence. Showing the effectiveness of a drug in a specific subpopulation positive to a biomarker (e.g. MCI patients with small hippocampi), might exclude from the benefit of prescription the proportion of negative patients, that for some biomarker might be much larger than the proportion of the biomarker positive. However, cases such as the one above are not unprecedented in medicine: tamoxifen is currently used for the treatment of estrogen receptor positive but not estrogen receptor negative breast cancer (Jordan, 1993).

This study has several limitations. First, the proportion of converters enrolled in the ADNI is going to change as MCI patients are followed for longer periods and more will convert to Alzheimer's dementia. Some studies of MCI patients enrolled in clinical settings are presently available with long follow-up (Busse et al., 2006; Ganguli et al., 2004; Mitchell and Shiri-Feshki, 2008; Tyas et al., 2007) showing that the vast majority of conversions occur in the first 5 years after first assessment. Because the mean follow-up of the patients of this study is 11 months, it seems likely that a sizable proportion of converters will show up in future years, and the present estimates of the ratio between MCI converters and nonconverters will need to be updated. Second, the ratio estimates for some markers are poorly accurate for the small number of patients and healthy controls in whom the marker has been collected. Future expansion of the ADNI dataset will allow an increase in the accuracy of the estimate. Further improvements of the present analysis can be addressed for a future expansion.

In the present study, a single marker enrichment strategy was assessed. It might be of great interest to explore the effect of combinations of markers to improve enrichment quality and efficiency. Moreover, the same strategies adopted here could be applied to different sets of markers, based on their relevance to discriminating disease pathology at the early stages. As in the present study, potential measures relevant for discriminative purposes, such as neuropsychological measures (CDR-SOB or FAQ) or even imaging measurements external to the ADNI Database (e.g. fMRI connectivity in key regions), could be tested and compared with the results shown here.

Disclosure statement

The authors report no actual or potential conflicts of interest.

Acknowledgements

The Foundation for the National Institutes of Health (www.fnih.org) coordinates the private sector participation of the US$60mn ADNI public-private partnership that was begun by the National Institute on Aging and supported by the National Institutes of Health. To date, more than US$27mn has been provided to the Foundation for NIH by Abbott, AstraZeneca AB, Bayer Schering Pharma AG, Bristol-Myers Squibb, Eisai Global Clinical Development, Elan Corporation, Genentech, GE Healthcare, GlaxoSmithKline, Innogenetics, Johnson and Johnson, Eli Lilly, and Co., Merck & Co., Inc., Novartis AG, Pfizer, Inc., F. Hoffmann-La Roche, Schering-Plough, Synarc, Inc., and Wyeth, as well as nonprofit partners the Alzheimer's Association and the Institute for the Study of Aging. The project is funded through the European Community's 'Seventh Framework' Programme (FP7/2007-2013) for an innovative scheme, the Innovative Medicines Initiate (IMI), under Grant Agreement n°115009. IMI is a young and unique public-private partnership, founded in 2008 by the pharmaceutical industry (represented by the European Federation of Pharmaceutical Industries and Associations), EFPIA and the European Communities (represented by the European Commission).

References

Albert, J., 2009. Bayesian Computation with R. New York: Springer; pp 22-26. ISBN 978-0-387-92297-3.

Ashburner, J., 2007. A fast diffeomorphic image registration algorithm. Neuroimage 38, 95–113.

Busse, A., Angermeyer, M.C., Riedel-Heller, S.G., 2006. Progression of mild cognitive impairment to dementia: a challenge to current thinking. Br J Psychiatry 189, 399–404.

De Leon, M.J., Mosconi, L., Blennow, K., DeSanti, S., Zinkowski, R., Mehta, P.D., Pratico, D., Tsui, W., Saint Louis, L.A., Sobanska, L., Brys, M., Li, Y., Rich, K., Rinne, J., Rusinek, H., 2007. Imaging and CSF studies in the preclinical diagnosis of Alzheimer's disease. Ann NY Acad Sci 1097, 114–145.

Dubois, B., Feldman, H.H., Jacova, C., Dekosky, S.T., Barberger-Gateau, P., Cummings, J., Delacourte, A., Galasko, D., Gauthier, S., Jicha, G., Meguro, K., O'Brien, J., Pasquier, F., Robert, P., Rossor, M., Salloway, S., Stern, Y., Visser, P.J., Scheltens, P., 2007. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. Lancet Neurol 6, 734–746.

Feldman, H.H., Ferris, S., Winblad, B., Sfikas, N., Mancione, L., He, Y., Tekin, S., Burns, A., Cummings, J., del Ser, T., Inzitari, D., Orgogozo, J.M., Sauer, H., Scheltens, P., Scarpini, E., Herrmann, N., Farlow, M., Potkin, S., Charles, H.C., Fox, N.C., Lane, R., 2007. Effect of rivastigmine on delay to diagnosis of Alzheimer's disease from mild cognitive impairment: the InDDEx study. Lancet Neurol 6, 501–512.

Ferris, S.H., 2002. Clinical trials in AD: are current formats and outcome measures adequate? Alzheimer Dis Assoc Disord suppl 1, S13–S17.

Ganguli, M., Dodge, H.H., Shen, C., deKosky, S.T., 2004. Mild cognitive impairment, amnestic type: an epidemiologic study. Neurology 63, 115–121.

Hampel, H., Bürger, K., Teipel, S.J., Bokde, A.L., Zetterberg, H., Blennow, K., 2008. Core candidate neurochemical and imaging biomarkers of Alzheimer's disease. Alzheimers Dement 4, 38–48.

Herholz, K., Salmon, E., Perani, D., Baron, J.C., Holthoff, V., Frölich, L., Schönknecht, P., Ito, K., Mielke, R., Kalbe, E., Zündorf, G., Delbeuck, X., Pelati, O., Anchisi, D., Fazio, F., Kerrouche, N., Desgranges, B., Eustache, F., Beuthien-Baumann, B., Menzel, C., Schröder, J., Kato, T., Arahata, Y., Henze, M., Heiss, W.D., 2002. Discrimination between Alzheimer dementia and controls by automated analysis of multicenter FDG PET. Neuroimage 17, 302–316.

Jack, C.R., Jr, Knopman, D.S., Jagust, W.J., Shaw, L.M., Aisen, P.S., Weiner, M.W., Petersen, R.C., Trojanowski, J.Q., 2010. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. Lancet Neurol 9, 119–128.

Jack, C.R., Theodore, W.H., Cook, M., McCarthy, G., 1995. MRI-based hippocampal volumetrics: data acquisition, normal ranges, and optimal protocol. Magn Reson Imaging 13, 1057–1064.

Jagust, W.J., Landau, S.M., Shaw, L.M., Trojanowski, J.Q., Koeppe, R.A., Reiman, E.M., Foster, N.L., Petersen, R.C., Weiner, M.W., Price, J.C., Mathis,C.A.. Alzheimer's Disease Neuroimaging Initiative, 2009. Relationships between biomarkers in aging and dementia. Neurology 73,1, 193–199.

Jordan, V.C., 1993. Fourteenth Gaddum Memorial Lecture. A current view of tamoxifen for the treatment and prevention of breast cancer. Br J Pharmacol 110, 507–517.

Liu, G., Liang, K.Y., 1997. Sample size calculations for studies with correlated observations. Biometrics 53, 937–947.

Loy, C., Schneider, L., 2006. Galantamine for Alzheimer's disease and mild cognitive impairment. Cochrane Database Syst Rev 1, CD001747.

Mitchell, A.J., Shiri-Feshki, M., 2008. Temporal trends in the long term risk of progression of mild cognitive impairment: a pooled analysis. J Neurol Neurosurg Psychiatry 79, 1386–1391.

Mueller, S.G., Weiner, M.W., Thal, L.J., Petersen, R.C., Jack, C., Jagust, W., Trojanowski, J.Q., Toga, A.W., Beckett, L., 2005. The Alzheimer's disease neuroimaging initiative. Neuroimaging Clin N Am 15, 869–877.

Petersen, R.C., Thomas, R.G., Grundman, M., Bennett, D., Doody, R., Ferris, S., Galasko, D., Jin, S., Kaye, J., Levey, A., Pfeiffer, E., Sano, M., van Dyck, C.H., Thal, L.J.; Alzheimer's Disease Cooperative Study group, 2005. Vitamin E and donepezil for the treatment of mild cognitive impairment. N Engl J Med 352, 2379–2388.

R Development Core Team, 2009. A language and environment for statistical computing. In R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Raschetti, R., Albanese, E., Vanacore, N., Maggini, M., 2007. Cholinesterase inhibitors in mild cognitive impairment: a systematic review of randomised trials. PLoS Med. 4, e338.

Risacher, S.L., Saykin, A.J., West, J.D., Shen, L., Firpi, H.A., McDonald, B.C., Alzheimer's Disease Neuroimaging Initiative (ADNI), 2009. Baseline MRI predictors of conversion from MCI to probable AD in the ADNI cohort. Curr Alzheimer Res 6, 347–361.

Salloway, S., Ferris, S., Kluger, A., Goldman, R., Griesing, T., Kumar, D., Richardson, S.; Donepezil 401 Study Group, 2004. Efficacy of donepezil in mild cognitive impairment: a randomized placebo-controlled trial. Neurology 63, 651–657.

Shaw, L.M., Vanderstichele, H., Knapik-Czajka, M., Clark, C.M., Aisen, P.S., Petersen, R.C., Blennow, K., Soares, H., Simon, A., Lewczuk, P., Dean, R., Siemers, E., Potter, W., Lee, V.M., Trojanowski, J.Q., Alzheimer's Disease Neuroimaging Initiative, 2009. Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. Ann Neurol 65, 403–413.

Tyas, S.L., Salazar, J.C., Snowdon, D.A., 2007. Transitions to mild cognitive impairments, dementia, and death: findings from the Nun Study. Am J Epidemiol 165, 1231.

Visser, P.J., Scheltens, P., Verhey, F.R., 2005. Do MCI criteria in drug trials accurately identify subjects with predementia Alzheimer's disease? J Neurol Neurosurg Psychiatry 76, 1348–1354.

Wilson, E.B., 1927. Probable Inference, the Law of Succession, and Statistical Inference. J Am Stat Assoc 22, 209–212.

Supplementary Table
Marker thresholds associated with the percentiles of the distributions identified by enrichment strategies (A) and (B)

		Enrichment strategy A	Enrichment strategy B
Cortical amyloid deposition ([11C]-PIB PET)	cm^3	1.86 (95^{th})	1.58 (84^{th})
Hippocampal volume	mm^3	2,819 (99^{th})	3,806 (44^{th})
ADAS-Cog	score	19.4 (99^{th})	10.3 (58^{th})
CSF tau	pg/ml	76.2 (70^{th})	/
CSF tau/abeta42	—	0.59 (85^{th})	/
CSF Abeta42	pg/ml	165.8 (70^{th})	/
Temporoparietal hypometabolism ([18F]-FDG PET)	t-sum	26,848 (99^{th})	6,078 (62^{nd})