# Reduced sample sizes for atrophy outcomes in Alzheimer's disease trials: baseline adjustment

**J.M. Schott**[a,*], **J.W. Bartlett**[a,b], **J. Barnes**[a], **K.K. Leung**[a,c], **S. Ourselin**[a,c], **N.C. Fox**[a], and **The Alzheimer's Disease Neuroimaging Initiative investigators**

[a]Dementia Research Centre, Institute of Neurology, UCL, London WC1N 3BG, UK

[b]Medical Statistics Unit, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK

[c]Centre for Medical Image Computing (CMIC), Department of Medical Physics and Bioengineering, University College London, WC1E 6BT, UK

## Abstract

Cerebral atrophy rate is increasingly used as an outcome measure for Alzheimer's disease (AD) trials. We used the Alzheimer's disease Neuroimaging initiative (ADNI) dataset to assess if adjusting for baseline characteristics can reduce sample sizes. Controls (n = 199), patients with mild cognitive impairment (MCI) (n = 334) and AD (n = 144) had two MRI scans, 1-year apart; ~ 55% had baseline CSF tau, p-tau, and Aβ1-42. Whole brain (KN–BSI) and hippocampal (HMAPS-HBSI) atrophy rate, and ventricular expansion (VBSI) were calculated for each group; numbers required to power a placebo-controlled trial were estimated. Sample sizes per arm (80% power, 25% absolute rate reduction) for AD were (95% CI): brain atrophy = 81 (64,109), hippocampal atrophy = 88 (68,119), ventricular expansion = 118 (92,157); and for MCI: brain atrophy = 149 (122,188), hippocampal atrophy = 201 (160,262), ventricular expansion = 234 (191,295). To detect a 25% reduction relative to normal aging required increased sample sizes ~ 3-fold (AD), and ~ 5-fold (MCI). Disease severity and Aβ1-42 contributed significantly to atrophy rate variability. Adjusting for 11 predefined covariates reduced sample sizes by up to 30%. Treatment trials in AD should consider the effects of normal aging; adjusting for baseline characteristics can significantly reduce required sample sizes.

## Keywords

Alzheimer's disease; MRI; Clinical Trials; Biomarker; CSF

Alzheimer's disease (AD) is the commonest form of degenerative dementia, and is increasing in prevalence as the population ages (Ferri et al., 2005). Current treatments provide symptomatic benefits but have not been shown to alter the underlying progression of the disease. Rapid advances in our understanding of the underlying genetics and cellular biology of AD have led to the development of specific therapies targeting the pathological processes underlying AD. There is thus an urgent requirement to design trials that can distinguish

*Corresponding author at: Dementia Research Centre, UCL Institute of Neurology, London WC1N 3BG. Tel.: +44 20 7837 3611; fax: +44 20 7676 2066. jschott@dementia.ion.ucl.ac.uk. (J. M. Schott)..

symptomatic from disease-modifying effects. Ultimately, disease-modifying drugs should produce a sustained reduction in clinical decline and increase time to institutionalization or death; however trials aiming to show such effects are logistically difficult and lengthy. In disease modification trials there is therefore an interest in incorporating imaging or other biomarkers that can be measured repeatedly and ideally non-invasively, and can be used across the spectrum of disease severity (Cummings, 2009).

Pathological global and regional cerebral atrophy reflects neuronal cell loss and can be measured accurately from serially acquired MRI scans, (Fox and Schott, 2004; Jack et al., 2005). Atrophy rates have been shown to correlate with cognitive decline in AD, (Jack et al., 2009; Schott et al., 2008) and are increasingly used as an outcome measure in clinical trials of AD (Fox et al., 2005; Jack et al., 2003) and mild cognitive impairment (MCI) (Jack et al., 2008). Using change in cerebral volume as an outcome measure may also reduce the numbers of patients needed to show that a therapy has an effect on the pathological process.

Sample sizes are critically dependent on the variability of the measured outcome. In the case of rates of atrophy measured from serially acquired scans, reductions in within-subject variability and thereby sample size may be achieved in a number of ways, including: (1) improving acquisition stability; (2) using novel trials designs incorporating run-in periods, cross-over designs, or multitime point acquisition; or (3) using more sensitive and precise measures to detect change.

An alternative method to reduce sample sizes for trials is to decrease the between-subject variance, i.e. the heterogeneity of the study population. This can be achieved by limiting entry to the study, for example by only recruiting patients at similar disease stages; stratifying patients, for example on the basis of severity; or in the case of trials in MCI where patients with isolated memory impairment have a relatively high risk of converting to AD, (Gauthier et al., 2006) incorporating only those patients with additional genetic risk factors (i.e. possession of ApoE4 genotype). This approach however potentially limits the wider applicability of any subsequent findings, and the pool of eligible patients. Another possible approach is to adjust for such variables in the statistical analysis of atrophy measures. This has the potential advantage of allowing a wider range of patients to enter a study, while limiting sample size requirement by controlling for between-subject variability.

In this study, we used the publicly available Alzheimer's disease neuroimaging initiative (ADNI) dataset to establish: (1) the potential reduction in sample size that can be gained in treatment trials of AD, and MCI using measures of brain volume reduction; ventricular expansion; and an automated measure of hippocampal atrophy by adjusting for predefined baseline characteristics; (2) confidence intervals for these sample sizes; and (3) the numbers needed to power such studies with and without accounting for normal aging.

## 1. Methods

### 1.1. Subjects

All subjects were drawn from ADNI, which 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 5-year public-private partnership. The aims of ADNI included assessing the ability of imaging and other biomarkers to measure the progression of MCI and early AD.

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, 400 people with MCI, and 200 people with early AD. For up-to-date information see www.adni-info.org. Written informed consent was obtained for participation in these studies, as approved by the Institutional Review Board at each of the participating centers. We downloaded data from LONI (www.loni.ucla.edu/ADNI/) on 29 September, 2009, and included all subjects (controls, MCI or AD) at baseline that had usable 1.5 T MRI imaging at baseline and 1 year; scans were only rejected if nondisease related pathology potentially affecting measurement was seen. All subjects had a standardized cognitive assessment at baseline, which included: MMSE, CDR-SB, ADAS-Cog (13 point scale); and blood was drawn for ApoE4 genotyping. Approximately 60% of the ADNI cohort had a CSF examination at baseline, and measurement of CSF tau, p-tau, and $A\beta1$-42 was performed centrally, as previously described (Shaw et al., 2009).

## 1.2. MR imaging

MR Imaging was performed using a standardized protocol on 1.5-T MRI units from Siemens Medical Solutions, Phillips, and General Electric Healthcare. MR protocols included the acquisition of sagittal high-resolution volumetric T1-weighted, inversion recovery prepared, structural images (www.adni-info.org/images/stories//mritrainingmanualv1.pdf for more details). Before the MR images were uploaded to the central image repository, images underwent several preprocessing steps, as previously described (Evans et al., 2009). These included corrections for distortion due to gradient nonlinearity; for image intensity nonuniformity using N3; for B1 nonuniformity where required; and scalings based on phantom measures. Local image analysis was performed using the MIDAS software package (Freeborough et al., 1997).

## 1.3. Image postprocessing

Whole brain segmentation was performed using a semiautomated technique with manual editing as required. Baseline brain regions were propagated onto the follow-up MR datasets using affine and free-form deformation-based (FFD) nonrigid registration, as described by Evans (Evans et al., 2009). The ventricular system was outlined on baseline and follow-up scans registered to standard space, using a previously described semiautomated protocol with manual editing as required (Evans et al., 2009). Change (mL) in whole brain volume and ventricular size were obtained using the boundary shift integral (BSI) following a 9 degrees-of-freedom registration and differential bias correction of the follow-up to baseline scans. For whole brain changes, a recently validated enhancement to the BSI protocol which involves improved intensity normalization (KN–BSI) was used (Leung et al., 2009).

Hippocampal volume change was calculated using the automated hippocampal outlining measure hMAPS (hippocampal Multi-Atlas Propagation and Segmentation), which has previously been extensively validated using the ADNI dataset (Leung et al., 2010). In brief, baseline hippocampal regions were generated by registration of the eight best-matched hippocampi from a template library (Barnes et al., 2008) using FFD registration together with image intensity thresholding. These eight hippocampal regions were combined using STAPLE together with a Markov random field filter with a weighting of 0.2 (Warfield et al., 2004). Hippocampal volume change between the two time-points was given by calculating boundary shift integral (HBSI) using the baseline hippocampal regions.

## 1.4. Statistical analysis

**1.4.1. Sample sizes without covariate adjustment—**Separately for AD and MCI subjects, we estimated the number of patients needed for a randomized controlled trial, using either annualized whole brain atrophy (KN–BSI absolute loss), annualized ventricular

enlargement (vBSI absolute enlargement), or annualized hippocampal atrophy (hMAPS HBSI absolute loss) as outcome. We estimated the sample sizes required per arm, for 80% power and a 5% Type 1 error rate using the standard formula:

$$\text{Sample size per trial arm} = \frac{(0.84+1.96)^2\left(2\sigma^2\right)}{\Delta^2}$$

(1)

where $\sigma^2$ denotes the variance in outcome, estimated either in AD or MCI subjects. We calculated sample size estimates to detect a reduction in absolute rate equal to 25% of the rate in AD/MCI subjects, by setting $\Delta$ equal to 0.25 times the estimated mean AD/MCI rate. We also report estimated sample sizes which assume the maximal possible reduction in rate of atrophy/enlargement would be to reduce the AD/MCI rates to that seen in control subjects; this is equivalent to setting $\Delta$ equal to 0.25 times the estimated difference in means between AD/MCI subjects and controls.

**1.4.2. Reduction in sample size through covariate adjustment**—For each measure of atrophy and separately in each group (ADs and MCIs), we assessed the percentage reduction in sample size obtained by adjusting for each of 11 *a priori* selected baseline measures: age, baseline brain volume, baseline ventricle volume, baseline hippocampal volume, MMSE, CDR-Sum of boxes, ADAS-Cog, CSF tau, CSF Aβ1-42, CSF p-tau, or ApoE4 dose (0, 1 or 2) (see www.adni-info.org for details). Because the CSF variables were positively skewed, we used the logarithm of their values in our analyses. We also estimated the reduction in sample size which would be achieved if all of these variables were included as covariates.

The proportionate reduction in variance (and hence required sample sizes) accorded through adjustment for a single covariate is equal to the square of the population correlation coefficient, $\rho^2$ (Borm et al., 2007). It is well-known that the sample estimator $R^2$ is biased, and that the bias can be substantial when the number of covariates is large relative to the number of subjects used to fit the regression model (Lucke et al., 1984). Alf and Graf proposed a parametric marginal maximum likelihood estimator for $\rho,^2$ and showed that it has lower mean squared error compared with the sample $R^2$ estimator (Alf and Graf, 2002). Following this proposal, but without making parametric assumptions, we estimated $\rho^2$ by its empirical maximum likelihood estimate. We estimated the empirical likelihood function using the bootstrapping technique, as proposed by Pawitan (Pawitan, 2000), using 1 million bootstrap samples. This procedure also provides 95% bias-corrected and accelerated bootstrap confidence intervals for $\rho^2$. When adjusting for multiple covariates, sample sizes are reduced by the squared multiple correlation coefficient (Borm et al., 2007), which we again estimated using the above empirical likelihood procedure. Our calculations ignore the cost of estimating the covariate effects, which for a fixed number of covariates, tends to zero in randomized trials as sample sizes increase.

By design, not all participants in the ADNI study had CSF collected, leading to missing data for CSF tau, CSF Aβ1-42 and CSF p-tau. Complete-case analysis, i.e. using only those subjects who had CSF, is inefficient because it discards the observed information from those subjects who did not have CSF. Furthermore, such estimates may be biased if the decision as to whether a subject had CSF was not completely at random. For the estimates of sample size reduction using the CSF variables, and "all covariates", we therefore based estimation on the observed data likelihood function, assuming multivariate normality of the (logged) CSF variables, conditional on all others (other baseline variables and the three outcome variables), with an unstructured variance–covariance matrix (Little and Rubin, 2002). For those without CSF, this approach uses information on the fully observed baseline variables and atrophy measures to predict the missing CSF values, based on the relationship in those with CSF. Like the method of multiple imputation, it provides consistent estimates provided the decision to have CSF did

not depend on the unobserved CSF values, conditional on the other variables (the so-called "missing at random" assumption), and that the conditional normality assumption is valid.

To investigate whether those subjects with CSF differed systematically from those without CSF, we compared the distributions of baseline characteristics between these two groups using two-sample *t*-tests with allowance for unequal variances and chi-squared tests. We also present estimates of sample size reductions found using data only from the subsets of the AD and MCI subjects who had CSF, for comparison with results based on using the available data from all subjects.

Analyses were performed in Stata 10 and R 2.10.1.

## 2. Results

Baseline characteristics are shown in Table 1. Controls, MCI subjects, and AD patients were well matched for age. MMSE was highest in controls; intermediate in MCI, and lowest in AD; ADAS-Cog was highest in AD; intermediate in MCI, and lowest in controls. Hippocampal and brain volumes were highest in controls, intermediate in MCI, and lowest in AD; and ventricular volumes were largest in AD, intermediate in MCI, and smallest in controls. CSF measures were available in 53.3% of controls; 51.8% of the MCI subjects; and 56.9% of the AD patients. Tau and P-tau levels were highest in AD, intermediate in MCI, and smallest in controls; and conversely $A\beta1\text{-}42$ levels were lowest in controls, intermediate in MCI, and highest in AD. 28.6% of the controls, 53.3% of the MCI subjects, and 66.6% of the AD patients had one or more ApoE4 alleles.

Rates of atrophy (mL/yr) are shown in Table 2. Mean whole brain and hippocampal volume loss, and ventricular expansion were highest in AD, intermediate in MCI, and smallest in controls, with statistically significant differences ($p > 0.05$) between the groups.

For each group, sample sizes were estimated to detect a 25% absolute reduction in rate of whole brain or hippocampal atrophy, or ventricular expansion. However, as the maximal reduction in rate of atrophy that can reasonably be expected is down to that seen in controls, we calculated the effective percentage reduction in these measures assuming a maximally efficacious treatment would reduce atrophy to the mean level in controls (Table 3). Using any of the three measures, in patients with established AD, a 25% absolute decline in rate of change is equivalent to a 36–43% reduction account for aging; and in MCI, this represents a 45–56% reduction.

Sample size estimates for a 25% mean reduction of the outcome (without allowing for normal aging), or 25% of the AD/MCI vs. control difference, are shown in Table 4 (AD) and 5 (MCI). All calculations were performed to provide 80% power with a 5% Type I error rate. The numbers required to power a trial to detect 25% mean reduction were ~ 3-fold higher in AD and ~ 5-fold higher in MCI when normal aging was accounted for.

The estimated reductions in sample size achieved by adjusting whole brain atrophy for baseline brain volume; or hippocampal atrophy for baseline hippocampal volume, were small (~ 1%). Larger (10% in AD, 16% in MCI) reductions were however achievable by adjusting ventricular enlargement for baseline ventricular volume. Adjusting for disease severity as measured by ADAS-Cog reduced estimated sample sizes by ~ 5% for AD; and ~ 8% in MCI. In AD subjects, accounting for CSF $A\beta1\text{-}42$ reduced sample sizes by 7–11%; and for MCI by 4–8%. Adjustment for all 11 covariates was estimated to reduce required sample sizes in AD using whole brain atrophy by ~ 16%; for ventricular enlargement by ~ 29%; and hippocampal atrophy by ~ 16% (see Table 4). In MCI, adjusting for all 11 covariates, sample sizes could be reduced

by ~ 18% using whole brain atrophy; by ~ 28% using ventricular enlargement; and by ~ 12% using hippocampal atrophy rates (Table 5).

There was no suggestion that subjects who had CSF differed compared with those who did not with respect to any of the baseline characteristics, in either the MCI group or AD group (see Supplementary Table 1a). Supplementary Tables 2a and 3 a show estimates of percentage reduction in sample sizes based on the subset of subjects for whom CSF data were available, which were similar to those found using data from all subjects (see Discussion).

## 3. Discussion

Rate of cerebral atrophy calculated from serially acquired MRI is increasingly used as an outcome measure for clinical trials in AD (Fox et al., 2005; Jack et al., 2003) and MCI (Jack et al., 2008). Attenuation of atrophy may provide a signal of a disease-modifying effect and sample size requirements may be much lower than those using traditional clinical outcome scores (Jack et al., 2003). Sample size calculations are proportional to the variance of the measure used, and such variability is a combination of within- and between-subject variability. Within-subject variability may arise because of measurement error and physiological variability over time, and numerous approaches to reducing these sources of error have been employed, including improving the stability of scan acquisition; employing multiple scanning time-points; (Schott et al., 2006) and developing novel and more accurate image analysis techniques, such as tensor-based morphometry (Hua et al., 2009).

Variation between individuals is likely to reflect several factors, including age, disease stage, differences in underlying pathological substrate (e.g. contribution from vascular disease and TDP-43 pathology (Josephs et al., 2008)), and other as yet unidentified epidemiological or genetic factors. Driving down these sources of variance, which have previously been estimated to contribute to over 50% of the variance in whole brain atrophy rate over 1 year in patients with established AD, (Schott et al., 2006) and are higher in MCI, is an alternative way to reduce sample sizes.

One method is to "enrich" trials by preselection of patients in an attempt to produce a more homogeneous group. This approach however potentially limits the wider applicability of the trial findings. An alternative approach is to include a broader range of individuals, but to predefine baseline characteristics that might be expected to explain inter-individual variation, and incorporate these into the analysis. Using this methodology, and incorporating baseline information routinely collected during the course of a clinical study, we have demonstrated that reduction of sample sizes of up to 15–30% in established AD and 10–30% in MCI may be achieved.

The raw sample size estimates we have produced to provide 80% power to show a 25% reduction in rate of change for a 1 year study of AD (i.e. ~ 80 per arm using the KN–BSI; ~ 120 per arm using the VBSI; and ~ 90 per arm using semiautomated hippocampal measures) are in line with those suggested by previous work (Barnes et al., 2008; Leung et al., 2009; Schott et al., 2005). In the context of patient recruitment, retention and cost, the 10–30% reduction in sample size potentially achievable by adjusting for baseline covariates, all of which are commonly measured, is not insignificant. The raw sample sizes required for an MCI trial are much larger (i.e. ~ 150 per arm using the KN–BSI; ~ 230 per arm using the VBSI; and ~ 200 per arm using hippocampal measures), but the percentage gains to be made by adjustment are similar, leading to sample sizes that are within the scope of Phase II studies. Few studies have reported confidence intervals on the "raw" sample sizes as we have done (Holland et al., 2009; Schott et al., 2006). Reporting such intervals for sample size estimates is essential, to indicate the precision with which they have been estimated.

In this study, we have analyzed volume loss (or ventricular enlargement) in mLs/yr, rather than as a percentage change. The approximate percentage changes we found are in keeping with prior studies (e.g. in AD ~ 1.5% whole brain atrophy/yr; ~ 5% hippocampal atrophy/yr). The whole brain atrophy rates were slightly smaller than in some previous studies (Fox et al., 2005; Schott et al., 2006), possibly reflecting either that the ADNI cohort were slightly older or had slightly milder disease than these other studies.

We found that while adjusting for baseline ventricular volume significantly reduced variability of VBSI, there was relatively little effect of adjusting KN–BSI or HMAPSHBSI for baseline brain or hippocampal volumes respectively. Thus while those with greater baseline ventricular volume tended to have greater subsequent ventricular enlargement, there was no evidence that baseline whole brain or hippocampal volumes were associated with subsequent atrophy in the same region.

Our results suggest that certain core features that contribute to the observed variance in atrophy rates; and when adjusted for, can significantly reduce the required sample sizes. Thus across all measures and in both AD and MCI, disease severity as measured using the ADAS-Cog is consistent in explaining some between-subject variance. Our results suggest that, for all three measures in AD and MCI subjects, CSF $A\beta1$–42 explains a moderate amount of variability in outcomes, with lower $A\beta1$–42 being associated with increased rates of atrophy; by contrast differences in baseline phosphorylated or total tau explained little variability. These results, seen in both the MCI and AD groups are perhaps surprising, as reduction of CSF $A\beta1$–42 reflecting deposition of fibrillar amyloid deposition within plaques is an early feature of AD and one that may begin to plateau in established disease (Jack et al., 2010). By contrast, elevation of CSF tau is thought to reflect ongoing neuronal degeneration, and thus might be expected to be a more sensitive measure of change throughout the course of the disease. Previous studies assessing the influence of CSF biomarkers on measures of atrophy have shown conflicting results. De Leon et al. (2006) and Schuff et al., (2009) (the latter analyzing the ADNI dataset) reported higher rates of hippocampal loss in MCI to be associated with lower levels of $A\beta1$–42. Several studies found increased hippocampal rates to be associated with higher levels of p-tau in MCI (de Leon et al., 2006; Hampel et al., 2005; Henneman et al., 2009), while in established AD, a weak association between baseline p-tau and whole brain atrophy has been reported (Sluimer et al., 2008). In interpreting the results for individual covariates in explaining atrophy rates, it is important to note that the confidence intervals for the estimated reductions in sample sizes are wide. Furthermore, we did not attempt to find the "optimal" subset of covariates to adjust, for two reasons. First, the optimal subset is likely to vary depending on the particular population studied. Second, defining the meaning of such an optimal subset, and finding it, is highly challenging from both a statistical and substantive perspective, given that all covariates provide some predictive value and that the "cost" of obtaining them often differs between variables (e.g. age v. CSF). Thus while our results suggest that disease severity and CSF $A\beta1$–42 may explain a relatively large proportion of between-subject differences in rate of atrophy, a degree of caution must be used when attempting to estimate the extent of influence of any one measure. The covariates found to be most predictive in these data-set, while biologically plausible, should not automatically be assumed to exert the same effect in all other AD/MCI studies.

Adjustment for baseline covariates can be performed by fitting a regression model for the outcome, with treatment group and the baseline covariates as "independent variables". If an adjusted analysis is to be used as the primary analysis of a trial, it is generally deemed as essential to prespecify in the trial's protocol the regression model which is to be used and which covariates will be adjusted for, although recently methods have been proposed which allow covariates to be selected using the trial data itself in such a way which does not lead to overestimated treatment effects (Tsiatis et al., 2008).

For continuous outcomes analyzed by linear regression models, the increase in statistical efficiency afforded through covariate adjustment depends on the strength of the associations between the covariates and outcome, and the size of the study (Cox and McCullagh, 1982). In large randomized studies, adjustment for a small number of baseline covariates incurs a negligible cost in degrees of freedom, because treatment group is independent of baseline covariates (a consequence of randomization). In smaller trials, where this cost is nonnegligible, the benefit of covariate adjustment in efficiency will be less, and may even be detrimental. The decision as to how many covariates are adjusted for should therefore been made in light of the size of the trial and the presumed strength of the associations between covariates and the outcome. In moderate to large trials, covariate adjustment is expected (approximately) to increase efficiency if the number of covariates is no more than $\rho^2$ (the population squared multiple correlation coefficient) times the number of subjects (Cox and McCullagh, 1982).

It is likely that maximum gain from neuroprotective agents will be achieved if these are given as early as possible in the disease process, and ideally at an asymptomatic stage even before fulfilment of criteria for MCI (Petersen, 2009). However if clinical trials are to be powered appropriately, it is critically important that the effect of normal aging is not ignored. It is unlikely that any neuroprotective agent will slow the rate of atrophy to below that seen in normal aging, and as rates of atrophy in MCI are smaller than in AD and consequently closer to normal aging, studies of MCI that do not acknowledge normal aging as a floor effect are in danger of being underpowered. This is demonstrated in this study, where an absolute 25% reduction in atrophy rate equates to a relative reduction accounting for the effects of normal aging of ~ 35% in AD; but as much as 50% in MCI, with consequent large increases in required sample sizes when normal aging is taken into account. Simply comparing sample sizes which do not take into account normal aging disadvantages outcomes that have little aging effects (e.g. some cognitive measures), and flatters those with relatively large changes in normal aging (e.g. atrophy).

This study suggests that simply in terms of study power, using standard placebo/control designs, preliminary studies of disease modifying drugs are more likely to show an effect when tested in patients with established AD. This conclusion however does not acknowledge that different disease processes may peak at different stages of the disease; that it may be more difficult to halt a wide-spread and advanced pathological process; and that there is more brain and cognition to be saved in early disease. Advances in accurate, early diagnosis of AD, and novel trial designs, incorporating multiple scanning time-points, run-in periods (Frost et al., 2008) or cross-over designs (Cummings, 2009), may however be able to reduce within-subject variability still further and make early treatment studies more viable.

The strengths of this study include the use of a large, well validated, publicly available dataset, consisting of representative patients acquired from multiple sites and different scanners (Petersen et al., 2010); robust statistical methodology; and a critical analysis of a range of different potential covariates in patients with MCI, AD and normal controls, using three different measures of structural change. We did not include PIB-PET measures (Jack et al., 2009) or other genetic haplotype data (Potkin et al., 2009) which might have been able to explain some of the large unexplained between-subject variability. Only ~ 55% of subjects had CSF results, potentially limiting the validity of our estimates for the benefit of adjusting for the CSF variables, as well as using all the covariates. To deal with the missing CSF values, we used a principled statistical technique for dealing with missing data. This approach uses the relationship between CSF variables and the other variables, estimated in those who had CSF, to (implicitly) predict the missing CSF variables in those who did not have lumbar puncture. The resulting estimates are consistent provided the decision to have CSF did not depend on the unobserved CSF values (conditional on observed variables), which seems reasonable, and provided the underlying statistical model is correctly specified. A comparison of the

distribution of fully observed baseline characteristics between those who had CSF and those who did not revealed no statistically significant differences. Furthermore, the estimates of percentage sample size reduction found using the subset of AD/MCI subjects for whom CSF was available were similar to those found using the available data from all subjects. Differences between the estimates may be due to several reasons (Sterne et al., 2009). First, estimates based on data from all subjects are, providing the modeling assumptions are valid, more precise than those based on the subset (~ 55% of each group) for whom CSF was available. Second, results may differ if the CSF data are not missing completely at random, although as noted there was little evidence against this assumption. We also note that in trials some outcome data are typically missing for some subjects, for a variety of potential reasons. Allowing for such missing values at both the design and analysis stage (e.g. through the use of linear mixed models or imputation methods) is essential.

The linear regression model used is based on a number of assumptions, including linearity of effects, no interactions, constant variance and normality of residuals. However, it has been shown that the covariate adjusted treatment effect estimates are (in large samples) unbiased without requiring these assumptions (Tsiatis et al., 2008). Using the standard sample size formula, we have assumed that in a future trial the variance of the atrophy/ventricular enlargement outcome would be the same in the two treatment arms, equal to that estimated using the ADNI data. Our estimates of sample sizes with covariate adjustment are valid with the additional assumption that the covariances of the covariates with the atrophy/ventricular enlargement outcomes would be the same in the two treatment arms. The extent to which covariates can explain variability in the outcome, and hence reduce sample sizes, depends critically on the variability of the covariate in the sample. Strictly speaking therefore, our estimates are applicable for future studies in which AD/MCI patients are recruited using the same criteria as that in the ADNI study. In particular, the covariates may explain a larger proportion of variability between patients in the wider AD/MCI populations, since the covariates are likely to have greater variability than in the ADNI study. However, the ADNI dataset has been shown to be representative of patients who might be recruited for therapeutic studies (Petersen et al., 2010).

In summary, we have shown that useful reductions in sample sizes may be achieved in AD and MCI trials using measures of cerebral volume change as an outcome measure if baseline characteristics are used as covariates. Required sample sizes are substantially higher in MCI trials than those carried out in patients with established AD, and the effect of accounting for normal aging as a floor threshold below which excess atrophy cannot fall implies significantly higher patient numbers will be needed for a given drug effect, particularly in MCI. It is critical that future trials of potentially disease-modifying therapies are appropriately powered so as not to miss a potential effect, and these data may help to inform such trial designs.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

# References

Alf EF Jr, Graf RJ. A new maximum likelihood estimator for the population squared multiple correlation. J Educ Behav Stat 2002;27:223–35.

Barnes J, Foster J, Boyes RG, Pepple T, Moore EK, Schott JM, Frost C, Scahill RI, Fox NC. A comparison of methods for the automated calculation of volumes and atrophy rates in the hippocampus. Neuroimage 2008;40:1655–71. [PubMed: 18353687]

Borm GF, Fransen J, Lemmens WA. A simple sample size formula for analysis of covariance in randomized clinical trials. J Clin Epidemiol 2007;60:1234–8. [PubMed: 17998077]

Cox DR, McCullagh P. Some aspects of analysis of covariance. Biometrics 1982;38:541–61. [PubMed: 7171689]

Cummings JL. Defining and labeling disease-modifying treatments for Alzheimer's disease. Alzheimers Dement 2009;5:406–18. [PubMed: 19751920]

de Leon MJ, DeSanti S, Zinkowski R, Mehta PD, Pratico D, Segal S, Rusinek H, Li J, Tsui W, Saint Louis LA, Clark CM, Tarshish C, Li Y, Lair L, Javier E, Rich K, Lesbre P, Mosconi L, Reisberg B, Sadowski M, deBernadis JF, Kerkman DJ, Hampel H, Wahlund LO, Davies P. Longitudinal CSF and MRI biomarkers improve the diagnosis of mild cognitive impairment. Neurobiol Aging 2006;27:394–401. [PubMed: 16125823]

Evans MC, Barnes J, Nielsen C, Kim LG, Clegg SL, Blair M, Leung KK, Douiri A, Boyes RG, Ourselin S, Fox NC. Volume changes in Alzheimer's disease and mild cognitive impairment: cognitive associations. Eur Radiol 2009;20:674–82. [PubMed: 19760240]

Ferri CP, Prince M, Brayne C, Brodaty H, Fratiglioni L, Ganguli M, Hall K, Hasegawa K, Hendrie H, Huang Y, Jorm A, Mathers C, Menezes PR, Rimmer E, Scazufca M. Global prevalence of dementia: a Delphi consensus study. Lancet 2005;366:2112–17. [PubMed: 16360788]

Fox NC, Black RS, Gilman S, Rossor MN, Griffith SG, Jenkins L, Koller M. Effects of Abeta immunization (AN1792) on MRI measures of cerebral volume in Alzheimer disease. Neurology 2005;64:1563–72. [PubMed: 15883317]

Fox NC, Schott JM. Imaging cerebral atrophy: normal ageing to Alzheimer's disease. Lancet 2004;363:392–4. [PubMed: 15074306]

Freeborough PA, Fox NC, Kitney RI. Interactive algorithms for the segmentation and quantitation of 3-D MRI brain scans. Comput Methods Programs Biomed 1997;53:15–25. [PubMed: 9113464]

Frost C, Kenward MG, Fox NC. Optimizing the design of clinical trials where the outcome is a rate. Can estimating a baseline rate in a run-in period increase efficiency? Stat Med 2008;27:3717–31. [PubMed: 18484598]

Gauthier S, Reisberg B, Zaudig M, Petersen RC, Ritchie K, Broich K, Belleville S, Brodaty H, Bennett D, Chertkow H, Cummings JL, de Leon M, Feldman H, Ganguli M, Hampel H, Scheltens P, Tierney MC, Whitehouse P, Winblad B. Mild cognitive impairment. Lancet 2006;367:1262–70. [PubMed: 16631882]

Hampel H, Burger K, Pruessner JC, Zinkowski R, deBernardis J, Kerkman D, Leinsinger G, Evans AC, Davies P, Moller HJ, Teipel SJ. Correlation of cerebrospinal fluid levels of tau protein phosphorylated at threonine two hundred and thirty-one with rates of hippocampal atrophy in Alzheimer disease. Arch Neurol 2005;62:770–3. [PubMed: 15883264]

Henneman WJ, Vrenken H, Barnes J, Sluimer IC, Verwey NA, Blankenstein MA, Klein M, Fox NC, Scheltens P, Barkhof F, van der Flier WM. Baseline CSF p-tau levels independently predict progression of hippocampal atrophy in Alzheimer disease. Neurology 2009;73:935–40. [PubMed: 19770469]

Holland D, Brewer JB, Hagler DJ, Fenema-Notestine C, Dale AM. Subregional neuroanatomical change as a biomarker for Alzheimer's disease. Proc. Natl. Acad. Sci. U S A 2009;106:20954–59. [PubMed: 19996185]

Hua X, Lee S, Yanovsky I, Leow AD, Chou YY, Ho AJ, Gutman B, Toga AW, Jack CR Jr, Bernstein MA, Reiman EM, Harvey DJ, Kornak J, Schuff N, Alexander GE, Weiner MW, Thompson PM. Optimizing power to track brain degeneration in Alzheimer's disease and mild cognitive impairment with tensor-based morphometry: an ADNI study of 515 subjects. Neuroimage 2009;48:668–81. [PubMed: 19615450]

Jack CR Jr, Knopman DS, Jagust WJ, Shaw LM, Aisen PS, Weiner MW, Petersen RC, Trojanowski JQ. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. Lancet Neurol 2010;9:119–28. [PubMed: 20083042]

Jack CR Jr, Lowe VJ, Weigand SD, Wiste HJ, Senjem ML, Knopman DS, Shiung MM, Gunter JL, Boeve BF, Kemp BJ, Weiner M, Petersen RC. Serial PIB and MRI in normal, mild cognitive impairment and Alzheimer's disease: implications for sequence of pathological events in Alzheimer's disease. Brain 2009;132:1355–65. [PubMed: 19339253]

Jack CR Jr, Petersen RC, Grundman M, Jin S, Gamst A, Ward CP, Sencakova D, Doody RS, Thal LJ. Longitudinal MRI findings from the vitamin E and donepezil treatment study for MCI. Neurobiol Aging 2008;29:1285–95. [PubMed: 17452062]

Jack CR Jr, Shiung MM, Weigand SD, O'Brien PC, Gunter JL, Boeve BF, Knopman DS, Smith GE, Ivnik RJ, Tangalos EG, Petersen RC. Brain atrophy rates predict subsequent clinical conversion in normal elderly and amnestic MCI. Neurology 2005;65:1227–31. [PubMed: 16247049]

Jack CR Jr, Slomkowski M, Gracon S, Hoover TM, Felmlee JP, Stewart K, Xu Y, Shiung M, O'Brien PC, Cha R, Knopman D, Petersen RC. MRI as a biomarker of disease progression in a therapeutic trial of milameline for AD. Neurology 2003;60:253–60. [PubMed: 12552040]

Josephs KA, Whitwell JL, Knopman DS, Hu WT, Stroh DA, Baker M, Rademakers R, Boeve BF, Parisi JE, Smith GE, Ivnik RJ, Petersen RC, Jack CR Jr, Dickson DW. Abnormal TDP-forty-three immunoreactivity in AD modifies clinicopathologic and radiologic phenotype. Neurology 2008;70:1850–7. [PubMed: 18401022]

Leung KK, Barnes J, Ridgway GR, Bartlett JW, Clarkson MJ, Macdonald K, Schuff N, Fox NC, Ourselin S. Automated cross-sectional and longitudinal hippocampal volume measurement in mild cognitive impairment and Alzheimer's disease. Neuroimage 2010;51:1345–59. [PubMed: 20230901]

Leung KK, Clarkson MJ, Bartlett JW, Clegg S, Jack CR Jr, Weiner MW, Fox NC, Ourselin S. Robust atrophy rate measurement in Alzheimer's disease using multi-site serial MRI: Tissue-specific intensity normalization and parameter selection. Neuroimage 2009;50:516–23. [PubMed: 20034579]

Little, RJA.; Rubin, DB. Statistical Analysis With Missing Data. John Wiley & Sons; 2002.

Lucke JF, Embretson, Whitely S. The biases and mean squared errors of estimators of multinormal squared multiple correlation. J Ed Stat 1984;9:183–92.

Pawitan Y. Computing empirical likelihood from the bootstrap. Stat Probability Lett 2000;47:337–45.

Petersen RC. Early diagnosis of Alzheimer's disease: is MCI too late? Curr Alzheimer Res 2009;6:324–30. [PubMed: 19689230]

Petersen RC, Aisen PS, Beckett LA, Donohue MC, Gamst AC, Harvey DJ, Jack CR Jr, Jagust WJ, Shaw LM, Toga AW, Trojanowski JQ, Weiner MW. Alzheimer's Disease Neuroimaging Initiative (ADNI): clinical characterization. Neurology 2010;74:201–9. [PubMed: 20042704]

Potkin SG, Guffanti G, Lakatos A, Turner JA, Kruggel F, Fallon JH, Saykin AJ, Orro A, Lupoli S, Salvi E, Weiner M, Macciardi F. Hippocampal atrophy as a quantitative trait in a genome-wide association

study identifying novel susceptibility genes for Alzheimer's disease. PLoS ONE 2009;4:e6501. [PubMed: 19668339]

Schott JM, Crutch SJ, Frost C, Warrington EK, Rossor MN, Fox NC. Neuropsychological correlates of whole brain atrophy in Alzheimer's disease. Neuropsychologia 2008;46:1732–7. [PubMed: 18395233]

Schott JM, Frost C, Whitwell JL, MacManus DG, Boyes RG, Rossor MN, Fox NC. Combining short interval MRI in Alzheimer's disease: Implications for therapeutic trials. J Neurol 2006;253:1147–53. [PubMed: 16998650]

Schott JM, Price SL, Frost C, Whitwell JL, Rossor MN, Fox NC. Measuring atrophy in Alzheimer disease: a serial MRI study over 6 and 12 months. Neurology 2005;65:119–24. [PubMed: 16009896]

Schuff N, Woerner N, Boreta L, Kornfield T, Shaw LM, Trojanowski JQ, Thompson PM, Jack CR Jr, Weiner MW. MRI of hippocampal volume loss in early Alzheimer's disease in relation to ApoE genotype and biomarkers. Brain 2009;132:1067–77. [PubMed: 19251758]

Shaw LM, Vanderstichele H, Knapik-Czajka M, Clark CM, Aisen PS, Petersen RC, Blennow K, Soares H, Simon A, Lewczuk P, Dean R, Siemers E, Potter W, Lee VM, Trojanowski JQ. Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. Ann Neurol 2009;65:403–13. [PubMed: 19296504]

Sluimer JD, Bouwman FH, Vrenken H, Blankenstein MA, Barkhof F, van der Flier WM, Scheltens P. Whole-brain atrophy rate and CSF biomarker levels in MCI and AD: A longitudinal study. Neurobiol Aging 2010;31(5):758–64. [PubMed: 18692273]

Sterne JA, C. White IR, Carlin JB, Spratt M, Royston P, Kenward MG, Wood AM, Carpenter JR. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ 2009;338:b2393. [PubMed: 19564179]

Tsiatis AA, Davidian M, Zhang M, Lu X. Covariate adjustment for two-sample treatment comparisons in randomized clinical trials: a principled yet flexible approach. Stat Med 2008;27:4658–77. [PubMed: 17960577]

Warfield SK, Zou KH, Wells WM. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Trans Med Imaging 2004;23:903–21. [PubMed: 15250643]

**Table 1**

Baseline characteristics of subject groups. Values shown are mean (SD) unless stated otherwise.

| Baseline variable | Control (n = 199) | MCI (n = 334) | AD (n = 144) |
|---|---|---|---|
| Brain volume (mL) | 1058.7 (102.6) | 1050.2 (115.7) | 1007.0 (113.6) |
| Vent, volume (mL) | 38.2 (20.7) | 47.4 (24.7) | 54.1 (28.1) |
| hMAPS hippocampal volume (mL) | 5.16 (0.73) | 4.45 (0.87) | 3.93 (0.91) |
| Male (n,%) | 106 (53.3) | 212 (63.5) | 75 (52.1) |
| Age (yrs) | 76.0 (5.1) | 74.9 (7.2) | 75.2 (7.3) |
| MMSE (/30) | 29.1 (1.0) | 27.0 (1.8) | 23.5 (1.9) |
| CDR-sum | 0.03 (0.11) | 1.58 (0.89) | 4.19 (1.55) |
| ADAS-cog | 9.3 (4.1) | 18.5 (6.6) | 28.2 (8.3) |
| Tau (pg/mL) | 68.8 (28.2)[a] | 100.4 (52.9)[c] | 122.6 (56.7)[b] |
| Aβ 1–42 (pg/mL) | 205.3 (54.0)[a] | 163.0 (54.2)[c] | 141.0 (40.5)[b] |
| P-tau (pg/mL) | 24.8 (13.8)[a] | 35.0 (16.7)[c] | 41.6 (19.4)[b] |
| ApoE4 dose (n,%) | | | |
| 0 | 142 (71.4) | 156 (46.7) | 48 (33.3) |
| 1 | 52 (26.1) | 138 (41.3) | 67 (46.5) |
| 2 | 5 (2.5) | 40 (12.0) | 29 (20.1) |

[a] n = 106

[b] n = 82

[c] n = 173

**Table 2**

Rates of change over 1 year by subject group. Values shown are mean (SD) [95% CI for mean].

| Atrophy rate (mL/yr) | Control (n = 199) | MCI (n = 334) | AD (n = 144) |
| --- | --- | --- | --- |
| Whole Brain (KN-BSI) | 6.27 (6.15) [5.41, 7.13] | 11.39 (8.78) [10.45, 12.34] | 15.19 (8.64) [13.77, 16.62] |
| Ventricular expansion (VBSI) | 1.43 (1.63) [1.21, 1.66] | 2.85 (2.75) [2.56, 3.15] | 4.52 (3.10) [4.00, 5.03] |
| Hippocampi (hMAPS-HBSI) | 0.052 (0.089) [0.040, 0.065] | 0.117 (0.105) [0.106, 0.129] | 0.175 (0.104) [0.158, 0.192] |

**Table 3**

Effective percentage reduction in atrophy rates accounting for normal aging, based on absolute percentage reduction (not accounting for aging) = 25%

|  | MCI | AD |
|---|---|---|
| Whole Brain (KN-BSI) | 55.6% | 42.6% |
| Ventricular expansion (VBSI) | 50.2% | 36.6% |
| Hippocampi (hMAPS-HBSI) | 45.0% | 35.6% |

**Table 4**

Sample size estimates required in each arm of a placebo-controlled AD trial (80% Power) to demonstrate 25% absolute reduction in atrophy; and 25% reduction in atrophy relative to normal aging.

| Variables adjusted for | Whole Brain atrophy (KN-BSI) | | | Ventricular expansion (VBSI) | | | Hippocampal loss (hMAPS-HBSI) | | |
|---|---|---|---|---|---|---|---|---|---|
| | Sample size (per arm) (95% CI) | | % reduction in sample size (95% CI) | Sample size (per arm) (95% CI) | | % reduction in sample size (95% CI) | Sample size (per arm) (95% CI) | | % reduction in sample size (95% CI) |
| | with aging | without aging | | with aging | without aging | | with aging | without aging | |
| No adjustments | 235 (164, 373) | 81 (64, 109) | — | 254 (186, 375) | 118 (92, 157) | — | 179 (125, 275) | 88 (68, 119) | — |
| Brain volume | 230 | 79 | 2.32 (0.02, 11.05) | 246 | 114 | 3.12 (0.09, 11.07) | 178 | 27 | 0.82 (<0.01, 6.01) |
| Vent volume | 228 | 79 | 3.05 (0.06, 11.18) | 229 | 106 | 9.82 (2.25, 22.67) | 175 | 86 | 2.06 (0.02, 8.12) |
| Hippo volume | 232 | 80 | 1.07 (<0.01, 7.55) | 252 | 117 | 0.62 (<0.01, 6.32) | 179 | 88 | 0.02 (<0.01, 1.61) |
| Age | 227 | 78 | 3.28 (0.09, 11.59) | 249 | 116 | 1.96 (0.02, 9.13) | 178 | 87 | 0.73 (<0.01, 7.01) |
| MMSE | 225 | 77 | 4.45 (0.22, 14.41) | 237 | 110 | 6.85 (0.98, 17.79) | 176 | 86 | 1.81 (0.01, 8.89) |
| CDR-sum | 235 | 81 | 0.02 (<0.01, 1.54) | 246 | 114 | 3.03 (0.02, 12.26) | 179 | 88 | 0.05 (<0.01, 1.93) |
| ADAS-cog (13) | 223 | 77 | 5.10 (0.83, 12.81) | 238 | 110 | 6.49 (1.13, 16.04) | 171 | 84 | 4.49 (0.11, 13.38) |
| CSF Tau[a] | 235 | 81 | 0.06 (<0.01, 5.58) | 252 | 117 | 0.82 (<0.01, 12.38) | 179 | 88 | <0.01 (<0.01, 0.01) |
| CSF Aβ1-42[a] | 218 | 75 | 7.18 (0.22, 21.26) | 231 | 107 | 9.20 (0.80, 22.02) | 160 | 79 | 10.79 (1.69, 24.30) |
| CSF P-tau[a] | 232 | 80 | 1.39 (<0.01, 13.16) | 254 | 118 | <0.01 (<0.01, 0.58) | 179 | 88 | 0.08 (<0.01, 4.39) |
| ApoE4 dose | 229 | 79 | 2.47 (0.09, 10.31) | 244 | 114 | 3.80 (0.24, 12.57) | 176 | 86 | 1.88 (0.07, 8.72) |
| All the above | 197 | 68 | 16.06 (8.22, 28.07) | 181 | 84 | 28.74 (17.22, 42.24) | 150 | 74 | 16.23 (6.80, 31.69) |

[a]Natural logarithm of CSF variables used. See statistical methods

**Table 5**

Sample size estimates required in each arm of a placebo-controlled MCI trial (80% Power) to demonstrate 25% absolute reduction in atrophy; and 25% reduction in atrophy relative to normal aging

| Variables adjusted for | Whole Brain atrophy (KN-BSI) | | | Ventricular expansion (VBSI) | | | Hippocampal loss (hMAPS-HBSI) | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | Sample size (per arm) (95% CI) | | | Sample size (per arm) (95% CI) | | | Sample size (per arm) (95% CI) | | |
| | with aging | without aging | % reduction in sample size (95% CI) | with aging | without aging | % reduction in sample size (95% CI) | with aging | without aging | % reduction in sample size (95% CI) |
| No adjustments | 739 (474, 1334) | 149 (122, 188) | — | 944 (621, 1681) | 234 (191, 295) | — | 648 (402, 1197) | 201 (160, 262) | — |
| Brain volume | 734 | 148 | 0.74 (<0.01, 3.55) | 929 | 230 | 1.61 (0.05, 5.26) | 646 | 200 | 0.37 (<0.01, 2.57) |
| Vent volume | 695 | 140 | 5.94 (1.63, 12.28) | 789 | 196 | 16.39 (8.09, 26.36) | 647 | 201 | 0.15 (<0.01, 2.00) |
| Hippo volume | 739 | 149 | 0.05 (<0.01, 1.06) | 940 | 233 | 0.43 (<0.01, 2.64) | 647 | 201 | 0.18 (<0.01, 1.81) |
| Age | 735 | 148 | 0.52 (<0.01, 3.45) | 931 | 231 | 1.36 (0.02, 5.69) | 645 | 200 | 0.45 (<0.01, 3.03) |
| MMSE | 706 | 142 | 4.47 (1.21, 9.85) | 912 | 226 | 3.35 (0.80, 7.48) | 635 | 197 | 2.03 (0.15, 6.17) |
| CDR-sum | 732 | 148 | 1.00 (0.03, 3.61) | 930 | 231 | 1.46 (0.07, 5.20) | 643 | 199 | 0.79 (<0.01, 3.66) |
| ADAS-cog (13) | 680 | 137 | 7.95 (3.40, 14.16) | 870 | 216 | 7.86 (3.17, 14.49) | 589 | 183 | 9.08 (4.38, 15.11) |
| CSF Tau[a] | 721 | 145 | 2.46 (0.13, 8.05) | 940 | 233 | 0.44 (<0.01, 3.59) | 614 | 190 | 5.26 (1.11, 12.95) |
| CSF Aβ1-42[a] | 678 | 137 | 8.30 (3.09, 15.81) | 890 | 221 | 5.67 (2.06, 10.91) | 625 | 194 | 5.26 (1.11, 12.95) |
| CSF P-tau[a] | 707 | 143 | 4.29 (0.62, 11.01) | 933 | 231 | 1.24 (<0.01, 5.66) | 608 | 188 | 6.25 (1.61, 14.26) |
| ApoE4 dose | 623 | 146 | 2.22 (0.25, 6.55) | 925 | 229 | 2.00 (0.23, 6.21) | 623 | 193 | 3.84 (0.95, 8.85) |
| **All the above** | **605** | **122** | **18.18 (11.72, 26.16)** | **675** | **167** | **28.49 (20.19, 37.57)** | **572** | **178** | **11.66 (6.74, 18.02)** |

[a]Natural logarithm of CSF variables used. See statistical methods