

HIPPOCAMPAL NEUROIMAGING-INFORMED CLINICAL TRIAL ENRICHMENT TOOL FOR AMNESTIC MILD COGNITIVE IMPAIRMENT USING OPEN DATA

Daniela J. Conrado^{1*}, Jackson Burton², Derek Hill³, Brian Willis⁴, Vikram Sinha⁵, Julie Stone⁵, Neva Coello⁶, Wenping Wang⁷, Danny Chen⁸, Timothy Nicholas⁹, Michael Gold¹⁰, Emily Hartley², Volker D. Kern², Klaus Romero², for the Alzheimer's Disease Neuroimaging Initiative**, for the Critical Path for Alzheimer's Disease (CPAD)***

¹e-Quantify LLC, La Jolla, CA, USA;

²Critical Path Institute, Tucson, AZ, USA;

³Panoramic Digital Health, Grenoble, France; Consultant to the Critical Path Institute, Tucson, AZ, USA;

⁴Eli Lilly and Company, Indianapolis, IN, USA;

⁵Merck & Co. Inc., Philadelphia, PA, USA;

⁶Novartis Pharmaceuticals, Basel, Switzerland;

⁷Novartis Pharmaceuticals, Philadelphia, PA, USA;

⁸Pfizer Inc., Cambridge, MA, USA;

⁹Pfizer Inc., Groton, CT, USA;

¹⁰AbbVie, North Chicago, IL, USA

*Corresponding Author:

Daniela J Conrado, PhD

8860 Villa La Jolla Drive Unit 305, La Jolla, CA 92037

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1002/cpt.1766](https://doi.org/10.1002/cpt.1766)

This article is protected by copyright. All rights reserved

Cell: +1 (617) 800-5394

E-mail: DConrado@e-Quantify.com

Funding

**Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at:

http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

***Data used in the preparation of this article were obtained from the Critical Path for Alzheimer's Disease (CPAD) database. In 2008, Critical Path Institute, in collaboration with the Engelberg Center for Health Care Reform at the Brookings Institution, formed the Critical Path for Alzheimer's Disease (CPAD). The Consortium brings together patient groups, biopharmaceutical companies, and scientists from academia, the U.S. Food and Drug Administration (FDA), the European Medicines Agency (EMA), the National Institute of Neurological Disorders and Stroke (NINDS), and the National Institute on Aging (NIA). The Critical Path for Alzheimer's Disease (CPAD) includes over 200 scientists from member and non-member organizations. The data available in the CPAD database has been volunteered by CPAD member companies and non-member organizations.

Conflict of Interest Statement: The authors declared no competing interests for this work.

Journal: invitation from *Clinical Pharmacology & Therapeutics* journal – Impact factor of 6.336

Article Type: original research

Keywords: hippocampal volume, neuroimaging, clinical trial enrichment tool, amnesic mild cognitive impairment, open data

Number of Words: ~4200

Number of References: 13

Number of Figures/Tables: 4 tables and 3 figures

ABSTRACT

Our goal was to assess the enrichment utility of hippocampal volume (HV) as an enrichment biomarker in amnesic mild cognitive impairment (aMCI) clinical trials, and, hence, develop a HV neuroimaging-informed clinical trial enrichment tool. Modeling of integrated longitudinal patient-level data came from open-access natural history studies in patients diagnosed with aMCI – the Alzheimer’s disease Neuroimaging Initiative (ADNI)-1 and ADNI-2 – and indicated that a decrease of 1cm^3 with respect to the analysis dataset median baseline intracranial volume-adjusted HV (ICV-HV; $\sim 5\text{cm}^3$) is associated with more than 50% increase in disease progression rate as measured by the Clinical Dementia Rating Scale - Sum-of-Boxes (CDR-SB). Clinical trial simulations showed that the inclusion of aMCI subjects with baseline ICV-HV below the 84th or 50th percentile allowed an approximate reduction in trial size of at least 26% and 55%, respectively. This clinical trial enrichment tool can help design more efficient and informative clinical trials.

1 INTRODUCTION

There is emerging consensus that effective treatments to slow or delay the progress of Alzheimer disease (AD) have a higher probability of success when the prodementia stages of the disease continuum are targeted: (a) asymptomatic prodementia, also described as preclinical phase, and (b) symptomatic prodementia, also described as amnesic mild cognitive impairment (aMCI) (1). However, there is a heterogeneous rate of disease worsening in this population, which can be partially attributed to the diminished accuracy of diagnosis of AD using traditional clinical measurements. In fact, it is not uncommon for aMCI clinical trials to enroll patients who, later, do not undergo disease worsening over the course of the trial, leading to an increase of trial size.

Neuroimaging of hippocampal volume (HV) using structural magnetic resonance imaging (MRI) has been proposed as a noninvasive approach to identifying subjects with aMCI that are more likely to progress to dementia (1,2). While our focus is on enrichment instead of diagnostic biomarker, HV makes for an attractive enrichment biomarker for clinical trials targeting an aMCI population as it has the potential to identify aMCI patients who are more likely to undergo disease worsening during a clinical trial. To our knowledge, there has been no formal, quantitative study evaluating the utility of HV as an enrichment biomarker in a large population of subjects with early- and late-stage aMCI.

Our goal was to assess the utility of HV as an enrichment biomarker in aMCI clinical trials, and, hence, develop a HV neuroimaging-informed clinical trial enrichment tool. Enrichment utility is defined as the ability of the biomarker to increase clinical trial efficiency, with efficiency being a measurable feature such as sample size. We first built a disease progression model as a backbone – i.e., an algorithm that integrates information from the natural progression of the disease and individual patient characteristics that may be associated with differences in progression rate. As a second step, this model was used to develop a user-friendly web-based application which allows the simulation of different clinical trial designs and estimates statistical power. We used an approach similar to the one used before in the context of early-stage Parkinson disease demonstrating the utility of dopamine transporter (DAT) neuroimaging as enrichment biomarker for clinical trials (3–5). The analysis served as the basis for the European Medicines Agency (EMA) qualification of the DAT as an enrichment biomarker in that patient population (6). It constitutes a “model-informed biomarker qualification”, to our knowledge, presented for the first time in the literature (3).

In this work, individual patient characteristics evaluated as predictors of disease progression included not only HV neuroimaging, but also genetics, demographics, and cognitive measures. The last three of these have been previously shown to be associated with AD progression (7). Accounting for these predictors in a multivariable manner allows us to draw conclusions on the magnitude in which each patient characteristic can predict disease progression. This clinical trial enrichment tool is expected to assist in designing more efficient and informative clinical trials by enabling researchers to select patients who are more likely to experience disease progression during the course of a clinical trial.

Open-access clinical data was used to conduct this work. For the purpose of the modeling work presented here longitudinal patient-level data were derived from two natural history studies in patients diagnosed with aMCI: the Alzheimer's Disease Neuroimaging Initiative (ADNI)-1 and ADNI-2 (8). Disease progression was as measured by the Clinical Dementia Rating Scale - Sum of Boxes (CDR-SB). This work was carried out by the Critical Path for Alzheimer's Disease (CPAD) consortium. CPAD is a public-private partnership aimed at creating new tools and methods that can be applied to increase the efficiency of the drug development process leading to disease-modifying treatments for neurodegenerative diseases with shared characteristics as AD.

2 METHODS

2.1 Overview of the Analysis Dataset

2.1.1 Study Subjects

Longitudinal subject-level data were integrated from two studies in subjects diagnosed with aMCI: the ADNI-1 and ADNI-2 observational studies (8). These subjects presented clinical symptoms that aligned with the Stage 3 of the AD continuum, as defined by the recent FDA guidance on early AD (9). Subjects in ADNI-1 were 55 to 90 years of age with a Global CDR score of 0.5, MMSE scores of 24 to 30, a memory complaint, objective memory loss measured by education adjusted scores on Wechsler Memory Scale Logical Memory II (WMS-LMII), absence of significant levels of impairment in other cognitive domains, essentially preserved activities of daily living, and an absence of dementia. Inclusion criteria in ADNI-2 were analogous to ADNI-1 with subjects being categorized as

Early MCI and Late MCI based on the education adjusted WMS-LMII scores (Supplementary Information, Section S1.1.1).

Data standardization used the existing data standards published by the Clinical Data Interchange Standards Consortium (CDISC) (www.cdisc.org), appropriateness of pooling data from both studies was confirmed, and criteria for data exclusion were as described in the Supplementary Information, Sections S1.1.2 through S1.1.4.

2.1.2 *Dependent Variable, Time Metric and Covariates*

The dependent variable was the CDR-SB score (10). The CDR-SB score ranges from 0 to 18 and is obtained by summing ratings in each of six cognitive/functional domains or *boxes* including memory, orientation, judgment/problem solving, community affairs, home and hobbies, and personal care. Higher scores reflect higher global impairment.

The time metric was time in the study from screening to the 48-month visit. Because the *in-house* ADNI-2 data was mostly for up to 48 months, the analysis data set was cut-off at the 48-month visit (inclusive) for ADNI-1 and ADNI-2. Screening visits were included because CDR-SB was not assessed at time zero (i.e., baseline), and CDR-SB scores at screening were used as a surrogate for baseline when different subjects' baseline characteristics were compared.

Data permitting, the following subjects' characteristics were tested as predictors (i.e., covariates) of CDR-SB scores: sex, age at baseline, mini-mental state examination (MMSE) score at baseline, number of APOE- ϵ 4 alleles, amyloid beta neuroimaging at baseline (present *versus* absent), ICV-HV at baseline.

2.1.3 *Neuroimaging*

ICV-HV was calculated as the mean of the left and right HV, adjusted for ICV. HV measurements were derived from MRI at the 1.5 T field strength in ADNI-1 and 3T in ADNI-2 using two different image analysis algorithms (LEAP™ and FreeSurfer™). ICV is estimated using an automated scaling factor calculate by atlas-based head size normalization (11).

2.1.4 Missing Dependent Variable, Covariates and Dropout

Imputation of missing CDR-SB scores (dependent variable) was not performed. Subjects with missing scores were included in the development of the disease progression model as well as allowed to be sampled in the clinical trial simulations (Section 4.4), unless only screening assessments or no assessments at all were available. Herein, dropout denoted a situation in which subject outcome data was missing after a certain point until the 48-month visit, because the study participant abandoned the study for any reason, which included death. Dropout rates were not negligible for both studies, and a dropout model was developed. Details on the dropout analysis and assumptions of missing mechanism are presented in the Supplementary Information, Section S1.1.5.

Regarding the covariates included in the model, missing values were evidenced for ICV-HV. While imputation has not been performed, such subjects with missing ICV-HV values were still included in the development of the disease progression model and handled as explained in Section 2.4.

2.2 Assessment of the Correlation and Agreement between ICV-HV Image Analysis Algorithms

The two image analysis algorithms (LEAP™ and FreeSurfer™) used to obtain ICV-HV were compared using correlation and agreement analyses. To this end, the analysis dataset included two baseline ICV-HV values for each subject, one from LEAP™ and another from FreeSurfer™.

In the correlation analysis, the degree to which the ICV-HV values from the different algorithms were related was assessed by Pearson's and Spearman's correlation. The Pearson's correlation coefficient reflects the noise and direction of the linear relationship – not agreement – between the two algorithms. The Spearman's rank correlation coefficient assesses how well the relationship between the two algorithms can be described using a monotonic function.

The Bland-Altman method, the most popular method used in agreement research (12), was used to determine the agreement between the two algorithms (Supplementary Information, Section S1.2). Whether or not the algorithms showed agreement, the analysis steps presented in the following sections were to be executed for each algorithm to compare their enrichment utility and magnitude. An agreement in the Bland-Altman analysis was expected to translate to a comparable enrichment

utility/magnitude between the two algorithms. However, a lack of agreement was not expected to necessarily translate to a lack of interchangeability with respect to enrichment utility/magnitude.

2.3 Development of a Disease Progression Model

The model building process followed three main steps: (a) selection of the base model structure with incorporation of two levels of random effects (i.e., between-subject and residual variability); (b) building of covariate model using a “full model” approach; and (c) evaluation of model performance including simulation-based covariate-stratified model diagnostics. Modeling software and modeling selection criteria are described in the Supplementary Information, Section S1.3.1 and S1.3.5, respectively.

2.3.1 Base Model (Fixed Effects)

Knowledge from the published CDR-SB longitudinal models were considered at this stage of model development. For the selection of the base model structure, a linear model was tested followed by non-linear models of increasing complexity. The base model structure was chosen using AIC_{mod} followed by other criteria cited in the Supplementary Information, Section S1.3.5. Details on previously published models and candidate base model structures are presented in the Supplementary Information, Section S1.3.2.

2.3.2 Random Effects

Between-subject random variability was incorporated for baseline scores and intrinsic rate of progression. Log-normally distributed between-subject variability was estimated for the baseline scores to prevent the prediction of nonsensical scores at the subject level. Normally distributed between-subject variability for intrinsic rate was estimated to allow for improvement, worsening or no change in scores over time. Residual variability was assumed to be beta distributed given the established bounded nature of the CDR-SB score (7) (Supplementary Information, Section S1.3.3).

2.3.3 Covariates

A covariate model approach with a focus on parameter estimation instead of stepwise hypothesis testing was implemented (well known as “full model approach”) (13). Covariates – as described in Section 4.1.2 (Dependent Variable, Time Metric and Covariates) – were pre-specified based on prior knowledge and/or clinical interest. Details on the covariate model building are described in the Supplementary Information, Section S1.3.4.

2.4 Assessment of the Utility of ICV-HV as an Enrichment Biomarker

To assess the utility of ICV-HV as an enrichment biomarker, clinical trial simulations using the Monte Carlo technique were performed to compare the statistical power by sample size in trials with or without ICV-HV enrichment. The simulations were performed with the two developed disease progression models; i.e., using ICV-HV calculated by each of the two image analysis algorithms, LEAP™ and FreeSurfer™. Simulated trials had different sizes and a placebo-controlled parallel group design. Non-enriched trials included subjects sampled from the whole distribution of ICV-HV in the analysis dataset. Conversely, enriched trials sampled subjects from truncated ICV-HV distributions based on illustrative cut-off values. Subjects were sampled with their respective covariate values to preserve the correlation between covariates.

For each non-enriched or enriched scenario, 1,000 clinical trials were simulated, including a hypothetical drug effect of 50% reduction in the disease progression rate, CDR-SB assessments of every three months and the herein developed dropout model. For a statistical power of 80% (type II error or $\beta = 0.20$) and 1,000 simulated trials, the Monte Carlo error was estimated to be 1.3% (Equation 1).

$$\text{Monte Carlo Error} = \sqrt{\frac{\text{Power} \times (1 - \text{Power})}{\text{Number of simulated trials}}} \quad (\text{Equation 1})$$

The statistical power for detecting the hypothetical drug effect was calculated as the proportion of trials for which the effect of treatment on progression rate was beneficial with a two-tailed P -value

less than 0.05. Details on the calculation of the statistical power are presented in the Supplementary Information, Section S1.4.

Clinical trial simulations with enrichment based on MMSE and *APOE* genotype were also performed to compare the enrichment utility/magnitude between ICV-HV, MMSE and *APOE* genotype.

3 RESULTS

3.1 Data Summary

The analysis data set included a total of 702 aMCI subjects, with a total of 3708 CDR-SB assessments in the screening-to-48 months interval. Table 1 shows the subjects characteristics by study. There were 381 subjects from ADNI-1 and 321 subjects from ADNI-2. A total of 32 subjects – 15 from ADNI-1 and 17 from ADNI-2 – were removed because they had only screening CDR-SB assessments; 4 ADNI-1 subjects were removed for having global CDR different from 0.5 (i.e., not meeting the criteria for aMCI). Two groups were specified by ADNI: the ‘early MCI’ (ADNI-2) and ‘late MCI’ (ADNI-1 and ADNI-2). The ‘early MCI’ and ‘late MCI’ groups represented about 24% (166 subjects) and 76% (535 subjects) of the entire analysis data set respectively. There was no classification for one subject in ADNI-1. Additional details on the analysis dataset are presented in the Supplementary Information, Section S2.1.

The time course of CDR-SB scores stratified by intracranial volume-adjusted hippocampal volume (ICV-HV) values is presented in Figure 1. The mean (95% confidence interval, CI) CDR-SB time course separates between the two baseline ICV-HV illustrative cut-off values (i.e., low with $ICV-HV \leq \text{median}$ and high with $ICV-HV > \text{median}$ ICV-HV of the analysis dataset) ICV-HV groups. The group with low ICV-HV shows faster progression than that with high ICV-HV. For ADNI-2, the high ICV-HV group shows a minimal increase of the mean CDR-SB scores over time (i.e., approximately 0.5 point).

Dropouts in ADNI-1 and ADNI-2 up to the 48-month visit represented 0.42 (95% CI = 0.37, 0.47) and 0.23 (95% CI = 0.18, 0.28), respectively. A dropout model was built to support the assumption of missing data mechanism and to account for dropout during the clinical trial simulations. The log normal base model performed better, and study, baseline age, and baseline mini-mental state

examination (MMSE) were identified as predictors (Supplementary Information, Section S2.1.1). Lower baseline MMSE and higher baseline age were associated to increased dropout.

3.2 Correlation and Agreement between HV Imaging Analysis Algorithms

The Learning Embeddings for Atlas Propagation (LEAP™) and FreeSurfer™ are two common HV analysis algorithms. LEAP™ and FreeSurfer™ ICV-HV values were dissimilar, with the FreeSurfer™ ICV-HV distribution shifted to higher values, and having wider range (i.e., lower boundary subtracted from the upper boundary) (Supplementary Information, Section S2.2.1). This suggests that a 1-cm³ ICV-HV increment in the LEAP™ and in the FreeSurfer™ 'scale' are not equivalent. Standardization of LEAP™ and FreeSurfer™ ICV-HV values was conducted to address the issue of dissimilar scales, and, hence, facilitate comparison (Supplementary Information, Section S2.2.1).

LEAP™ and FreeSurfer™ ICV-HV values, raw and standardized, were highly correlated (Pearson's and Spearman's correlation coefficients ≥ 0.75) (Supplementary Information, Figure S10). In the Bland-Altman plots for standardized LEAP™ and FreeSurfer™ ICV-HV values in ADNI-1 plus ADNI-2 (Supplementary Information, Figure S11), the mean (i.e., estimated bias) and standard deviation (SD) of the differences between standardized ICV-HV values for the two algorithms were 1.0 and 16, respectively. The 95% limits of agreement were -31 to 33, meaning that 95% of the differences are expected to lie between these limits under an assumption of normal distribution. Considering that the standardized LEAP™ and FreeSurfer™ ICV-HV values correspond to percentiles of the ADNI-1 control normal subjects ICV-HV distribution, the aforementioned 95% limits of agreement are meaningfully wide. This suggests limited agreement between the standardized LEAP™ and FreeSurfer™ ICV-HV values. Two disease progression models were developed using ICV-HV values from each analysis algorithm.

3.3 Base Disease Progression Model

The Richards model, a generalized logistic model, was the most appropriate to describe the time course of CDR-SB scores (Equation 2). It allowed for an asymmetric, inverted, concave relationship between disease progression rate and severity. The inflection point was estimated to occur at an

CDR-SB score of approximately 11.6 (Equation 3, Figure 2). Details on selection of the base model structure are presented in the Supplementary Information, Section S2.3.

$$\frac{d\text{Score}_i}{dt} = r_i \times \text{Score}_i \times \left[1 - \left(\frac{\text{Score}_i}{\max(\text{Score}_i)} \right)^\beta \right] \quad (\text{Equation 2})$$

$$\text{Inflection Point} = \left(\frac{1}{1 + \beta} \right)^{1/\beta} \times 18 \quad (\text{Equation 3})$$

Where:

- t denotes time
- r_i denotes the intrinsic rate of disease progression for subject i
- Score_i denotes the CDR-SB score for subject i
- β denotes the shape factor of the Richards' model, which was estimated as 3.3 in the frequentist LEAP™ covariate model (Section 2.4; Final Disease Progression Model)
- $\max(\text{Score}_i)$ or 18 denotes the upper boundary of the CDR-SB scale (i.e., the maximum possible observable score)

3.4 Final Disease Progression Model

The Richards model was used as the base model structure for the covariate model building. Final parameter estimates of the base model, covariate model using ICV-HV measured by LEAP™, and covariate model using ICV-HV measured by FreeSurfer™ are presented in Table 2. Parameter values for both covariate models were also estimated under a No-U-Turn Sampler (NUTS) Bayesian statistical framework. For the Bayesian approach, the respective final parameter estimates from the frequentist approach were used as uninformative priors. Noteworthy:

- Predictors of disease progression rate were sex, Apolipoprotein E (APOE) genotype, baseline ICV-HV, baseline MMSE and baseline age.
- The MMSE score at baseline was used as a measure of disease severity at baseline. The reason being that MMSE is the most commonly used scale to assess cognition at screening

Accepted Article

for trial subjects. In addition, the correlation between CDR-SB and baseline MMSE scores was lower than 30% (Supplementary Information, Figure S4).

- Approximately 20% of the subjects had missing ICV-HV values at baseline and were used for development of the disease progression model. Imputation of ICV-HV values was not performed, and the average covariate coefficient of the group of subjects with missing values was estimated.
- Incorporation of amyloid beta imaging status as a predictor of progression rate was attempted and led to an unsuccessful convergence of the minimization routine. Note that virtually none the ADNI-1 subjects underwent amyloid beta imaging; from the 321 subjects in ADNI-2, approximately 189 were amyloid beta positive. There was a total of 152 subjects who were amyloid beta positive and had ICV-HV information. In addition, there was some degree of association between amyloid imaging status, presence of *APOE-ε4* allele, and ICV-HV values.
- The relationship between progression rate and ICV-HV values was better described by a linear than a power relationship. Extrapolation to ICV-HV values outside the ICV-HV ranges in this dataset is not supported by this model.

Parameter estimates were comparable among the four covariate models explored, except for the age and ICV-HV effect on progression rate, and shape factor of the Richards' model. Plausible explanations for the different estimates are:

- The effect of age on rate is uncertain as reflected by the wide confidence intervals. Consequently, the different models settled in different point estimates. Given the known effect of age on the progression rate, this covariate was kept in the model. Moreover, sensitivity analyses of the frequentist LEAP™ and FreeSurfer™ covariate models were conducted by removing age as a predictor of rate, and similar estimates for the remaining parameters were obtained.
- The dissimilar estimates for the ICV-HV effect on rate between LEAP™ and FreeSurfer™ may be attributed to the lack of equivalence between a 1 cm³ increment in the LEAP™ and FreeSurfer™ scale (Section 2.2; Correlation and Agreement between LEAP™ and FreeSurfer™).

- The credible interval for the shape factor parameter was particularly wide under the Bayesian approach. The Bayesian approach is likely to yield a more reliable estimation for uncertainty than the one calculated under a normality assumption and using the asymptotic standard errors provided by Nonlinear Mixed Effects Modeling (NONMEM). This high uncertainty is likely associated to the somewhat limited number of subjects with scores above the inflection point.

Interpretation of parameter values and parameters-covariates relationships are presented in Table 3. Details on the model performance are presented in the Supplementary Information, Section S2.4. The NONMEM code and output for the four covariate models are presented in the Supplementary Information.

3.5 ICV-HV Enrichment Utility and Magnitude: Clinical Trial Simulations and Statistical Power

Estimated required sample sizes for simulated ICV-HV-enriched and non-enriched placebo-controlled parallel group clinical trials – to detect a drug effect of 50% reduction in the progression rate with an 80% probability (type II error or $\beta = 0.20$) at an α of 0.05 – using the LEAP™ and FreeSurfer™ covariate models are presented in Figure 3A and Figure 3B (left panels). Table 4 summarizes the required sample sizes (95% CI) estimated by the two models along with relative sample size reduction (95% CI) of enriched trials compared to non-enriched trials. The percentage of sample size savings due to ICV-HV enrichment estimated by the LEAP™ covariate model subtracted from the FreeSurfer™ covariate model were 2.2 (95% CI = -1.6, 6.0) %, 5.4 (95% CI = 1.0, 9.7) % and 4.5 (95% CI = 3.0, 6.0) % in enriched clinical trials including only subjects with baseline ICV-HV lower than +2 standard deviations (SD), +1 SD, and the 50th percentile (median) of the ICV-HV distribution in the analysis dataset, respectively. The point estimates for the three scenarios above suggest that FreeSurfer™ yields marginally greater sample size savings (2.2% to 5.4% higher) than LEAP™. However, statistical significance was not obtained for the scenario with the lowest sample size saving. It is not expected that the reduction in required sample size yielded by ICV-HV enrichment is related to a reduction in the dropout rate. The identified predictors of dropout were older age and lower MMSE score, and there was a positive, although weak, correlation between

baseline MMSE and ICV-HV (0.29, Supplementary Figure S4) and a negative, although weak, correlation between baseline age and ICV-HV (-0.37, Supplementary Figure S4).

Clinical trial simulations were also performed for other model covariates to compare the enrichment utility/magnitude between ICV-HV and other covariates (Figure 3A and Figure 3B, right panels). To have a fair comparison, attention was given to having a similarly estimated screening failure rate among the different enrichment scenarios. In this case, the chosen screening failure rate due to enrichment was approximately 50% (Supplementary Information, Section S2.5). Based on the simulations using the LEAP™ covariate model, the APOE-ε4 and MMSE-based enrichments translated to a reduction in trial size of approximately 39% and 46%, respectively. Similar reductions were obtained using the FreeSurfer™ covariate model: 41% and 51%, respectively.

2.5.1 Recommendations for a New ICV-HV Algorithm with respect to its Enrichment Utility

With technological advances, new ICV-HV algorithms will be introduced in the market. To determine whether the new algorithm provides greater or lower enrichment magnitude than LEAP™/FreeSurfer™ ('current image analysis algorithm'), one must analyze the new algorithm scores and subject-level clinical outcome data together. If a drug development sponsor does not have the resources/bandwidth to do such an analysis, a lower bound of the enrichment magnitude can be estimated based on the correlation between the ICV-HV values from the new and current algorithm (note that there was a linear relationship between ICV-HV values and progression rate). For the lower bound to be estimated, one must assume the worst-case scenario; i.e., the new algorithm is simply a noisy version of a current algorithm, where the noise is independent of the clinical outcome or the current algorithm. An algorithm that is noisier than the current algorithm would naturally have a reduced enrichment magnitude, in that an ICV-HV based-subject trial selection would be compromised. Under this assumption, new algorithms – where the ICV-HV values would correlate with those from LEAP™ ICV-HV by a Pearson's correlation coefficient of 0.9, 0.7, and 0.5, for instance – would require sample size increases of approximately 7.5%, 23% and 49%, respectively (Figure 3C). A detailed description of this statistical analysis is presented in the Supplementary Information, Section S2.5.1.

3.6 Clinical Trial Enrichment Tool

The disease progression model was used to develop a web-based simulator with a user-friendly interface to aid with clinical trial design. This tool simulates clinical trials based on user-defined subject characteristics at study entry and is available at <https://cpath.shinyapps.io/predemctegui/>.

4 DISCUSSION

This work supports the utility of ICV-HV as an enrichment biomarker in aMCI clinical trials. ICV-HV has the potential to significantly reduce trial size, with the enrichment magnitude being similar for the two commonly used image analysis algorithms, FreeSurfer™ and LEAP™. In light of the above, an HV neuroimaging-informed clinical trial enrichment tool was developed. The results suggest that the tool provides added value to optimize clinical trial design in aMCI. The backbone of the tool consists of a non-linear mixed-effects model of CDR-SB over time, which was developed using open-access patient-level data from ADNI-1 and ADNI-2. The model accounted for baseline ICV-HV, APOE-ε4 carrier status, baseline MMSE scores, baseline CDR-SB, baseline age and sex as relevant covariates. Such a clinical trial enrichment tool allows the user to perform simulations to estimate sample size and statistical power; enrichment strategies can be evaluated under different assumptions and trial design options. Together with the range of MMSE scores at baseline and the proportion of APOE-ε4 carriers, the most appropriate ICV-HV threshold can be selected to increase the likelihood of demonstrating drug effects in aMCI clinical trials.

Historical approaches for sample size estimation, based on literature metadata of the estimated standard deviation for the clinical endpoint and the expected effect size, do not typically account for: (a) individual differences in demographic, clinical and genetic characteristics of the enrolled trial population; (b) disease worsening profile over time and (c) the different levels of variability (e.g., between-subject, and residual variability). The HV neuroimaging-informed clinical trial enrichment tool accounts for the contribution of the aforementioned aspects and, conveniently, is presented as an open-access web-based simulator with a user-friendly graphical interface for a broader use. Its use is recommended for all clinical efficacy evaluation stages of drug development for aMCI, including early efficacy, proof-of-concept, dose-ranging, and registration studies.

Noteworthy is that: (a) the lack of clinical trial data precluded the development of a placebo effect model. It is understood that an unaccounted placebo effect, as well as other clinical trial components,

could have an impact on the estimated trial size. An expansion of the model with clinical trial data will allow the description of clinical trial components such as placebo response and dropout profile; (b) the limited number of patients who underwent amyloid beta imaging did not allow us to reach a conclusion on the role of amyloid beta imaging-based enrichment in comparison to ICV-HV and the other relevant predictors of disease progression. Our current recommendation is that ICV-HV not necessarily replace other currently used biomarkers but be considered as an alternative to or in combination with currently used biomarkers for aMCI trial enrichment. The determination on whether enrichment should be applied, and, if so, which biomarker(s) to use, must be made by the drug sponsor, who should weight potential advantages and disadvantages of the biomarkers in the context of the drug development program; for instance, i) hippocampal atrophy may occur at a later stage of the disease continuum than amyloid positivity, and ii) enrichment strategies in later stages of drug development would limit the understanding of drug effects in the broader aMCI population, and biomarker-based stratification may be considered as an alternative. There was a statistically significant negative correlation between the individual model estimated progression rate and the individual ICV-HV in subjects who were amyloid beta positive and had ICV-HV information (Supplementary Figure S18). The increase in progression rate with decrease of ICV-HV in amyloid beta positive subjects suggests that ICV-HV-based enrichment might have added value on the top of amyloid beta imaging-based enrichment. Clinical trial data in amnesic MCI patients is currently being pursued by the CPAD consortium; such data will allow for not only an external validation of the enrichment utility of hippocampal volume to be conducted, but also a better understanding of the potential value of ICV-HV-based enrichment in amyloid beta positive subjects.

This clinical trial simulation tool served as the basis for a Letter of Support issued by the European Medicines Agency (EMA) (14) to encourage: (a) the CPAD team to disseminate the tool; (b) the researchers that are actively designing clinical trials in aMCI to use the tool; and (c) the industry drug sponsors to share the patient-level data from completed phase 2 and 3 clinical trials in the target population. This will allow a continued improvement of the tool and, ultimately, trial design.

STUDY HIGHLIGHTS

1. What is the current knowledge on the topic?

Neuroimaging of hippocampal volume (HV) has been proposed as a noninvasive approach to identifying subjects with amnesic mild cognitive impairment (aMCI) that a more likely to progress to dementia. However, no formal, quantitative study demonstrating the enrichment utility of HV in a large population of subjects with early- and late-stage aMCI has been conducted.

2. What question did this study address?

Can HV help identify subjects with aMCI that a more likely to undergo disease worsening – as measured by the clinical dementia rating scale sum of boxes – over the course of a clinical trial?

3. What does this study add to our knowledge?

The developed hippocampal neuroimaging-informed clinical trial enrichment tool is expected to help design more efficient (e.g., reduced sample size) and informative (e.g., sufficiently powered) clinical trials for aMCI.

4. How might this change clinical pharmacology or translational science?

Optimization of clinical trial design for aMCI can increase the likelihood of finding an efficacious drug to treat Alzheimer disease at its prodementia stage.

AUTHOR CONTRIBUTIONS

D.J.C. wrote the manuscript; D.J.C., J.B., D.H., B.W., V.S., J.S., N.C., W.W., D.C., T.N., M.G., E.H., and V.D.K. designed the research; D.J.C., J.B., D.H., B.W., V.S., J.S., N.C., W.W., D.C., T.N., M.G., E.H., and V.D.K. performed research; D.J.C. analyzed the data.

REFERENCES

1. Jack CR, Albert MS, Knopman DS, McKhann GM, Sperling RA, Carrillo MC, et al. Introduction to the recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement J Alzheimers Assoc.* 2011 May;7(3):257–62.
2. Dubois B, Feldman HH, Jacova C, Dekosky ST, Barberger-Gateau P, Cummings J, et al. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. *Lancet Neurol.* 2007 Aug;6(8):734–46.
3. Conrado DJ, Timothy N, Kuenhi T, Macha Sreeraj, Sinha Vikram, Stone Julie, et al. Dopamine Transporter Neuroimaging as an Enrichment Biomarker in Early Parkinson's Disease Clinical Trials: A Disease Progression Modeling Analysis. *Clin Transl Sci.* 2018;11(1):63–70.
4. Romero K, Conrado D, Burton J, Nicholas T, Sinha V, Macha S, et al. Molecular Neuroimaging of the Dopamine Transporter as a Patient Enrichment Biomarker for Clinical Trials for Early Parkinson's Disease. *Clin Transl Sci.* 2019;12(3):240–6.
5. Stephenson D, Hill D, Cedarbaum JM, Tome M, Vamvakas S, Romero K, et al. The Qualification of an Enrichment Biomarker for Clinical Trials Targeting Early Stages of Parkinson's Disease. *J Park Dis.* 2019 Jul 8;
6. Qualification opinion on dopamine transporter imaging as an enrichment biomarker for Parkinson's disease clinical trials in patients with early Parkinsonian symptoms [Internet]. Available from: https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/qualification-opinion-dopamine-transporter-imaging-enrichment-biomarker-parkinsons-disease-clinical_en.pdf
7. Conrado DJ, Denney WS, Chen D, Ito K. An updated Alzheimer's disease progression model: incorporating non-linearity, beta regression, and a third-level random effect in NONMEM. *J Pharmacokinet Pharmacodyn.* 2014 Dec;41(6):581–98.

8. ADNI | Study Design [Internet]. [cited 2019 Jul 23]. Available from: <http://adni.loni.usc.edu/study-design/>
9. U.S. Department of Health and Human Services Food and Drug Administration. Early Alzheimer's Disease: Developing Drugs for Treatment Guidance for Industry [Internet]. 2018. Available from: <https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM596728.pdf>
10. Hughes CP, Berg L, Danziger WL, Coben LA, Martin RL. A new clinical scale for the staging of dementia. *Br J Psychiatry J Ment Sci.* 1982 Jun;140:566–72.
11. Buckner RL, Head D, Parker J, Fotenos AF, Marcus D, Morris JC, et al. A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: reliability and validation against manual measurement of total intracranial volume. *NeuroImage.* 2004 Oct;23(2):724–38.
12. Zaki R, Bulgiba A, Ismail R, Ismail NA. Statistical Methods Used to Test for Agreement of Medical Instruments Measuring Continuous Variables in Method Comparison Studies: A Systematic Review. Rapallo F, editor. *PLoS ONE.* 2012 May 25;7(5):e37908.
13. Gastonguay MR. Full Covariate Models as an Alternative to Methods Relying on Statistical Significance for Inferences about Covariate Effects: A Review of Methodology and 42 Case Studies. 2004 [cited 2016 Nov 29]; Available from: <http://metrumrg.com/assets/pubs/GastonguayPAGE2011.pdf>
14. Letter of support for Model-based CT enrichment tool for CTs in aMCI [Internet]. Available from: https://www.ema.europa.eu/en/documents/other/letter-support-model-based-ct-enrichment-tool-cts-amci_en.pdf
15. Conrado DJ, Chen D, Denney WS. Cardiovascular Safety Assessment in Early-Phase Clinical Studies: A Meta-Analytical Comparison of Exposure-Response Models. *CPT Pharmacomet Syst Pharmacol.* 2016 Jun;5(6):324–35.

16. Altman DG, Bland JM. Measurement in medicine: the analysis of method comparison studies. *The statistician*. 1983;307–317.
17. Bland JM, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. *The lancet*. 1986;327(8476):307–310.
18. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing [Internet]. Vienna, Austria; 2016. Available from: <https://www.R-project.org/>
19. Delor I, Charoin J-E, Gieschke R, Retout S, Jacqmin P. Modeling Alzheimer's Disease Progression Using Disease Onset Time and Disease Trajectory Concepts Applied to CDR-SOB Scores From ADNI. *CPT Pharmacomet Syst Pharmacol*. 2013;2:e78.
20. Ito K, Hutmacher MM. Predicting the time to clinically worsening in mild cognitive impairment patients and its utility in clinical trial design by modeling a longitudinal clinical dementia rating sum of boxes from the ADNI database. *J Alzheimers Dis JAD*. 2014;40(4):967–79.
21. Samtani M, Raghavan N, Novak G, Nandy P, Narayan VA. Disease progression model for Clinical Dementia Rating-Sum of Boxes in mild cognitive impairment and Alzheimer's subjects from the Alzheimer's Disease Neuroimaging Initiative. *Neuropsychiatr Dis Treat*. 2014 May;929.
22. Smithson M, Verkuilen J. A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychol Methods*. 2006 Mar;11(1):54–71.
23. Liu F, Eugenio EC. A review and comparison of Bayesian and likelihood-based inferences in beta regression and zero-or-one-inflated beta regression. *Stat Methods Med Res*. 2016 May 25;096228021665069.
24. Gibiansky L, Gibiansky E, Bauer R. Comparison of Nonmem 7.2 estimation methods and parallel processing efficiency on a target-mediated drug disposition model. *J Pharmacokinet Pharmacodyn*. 2012 Feb;39(1):17–35.
25. Hoffman MD, Gelman A. The No-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *J Mach Learn Res*. 2014;15(1):1593–1623.

- Accepted Article
26. William-Faltaos D, Chen Y, Wang Y, Gobburu J, Zhu H. Quantification of disease progression and dropout for Alzheimer's disease. *Int J Clin Pharmacol Ther.* 2013 Feb;51(2):120–31.
 27. Jackson CH. flexsurv: a platform for parametric survival modeling in R. *J Stat Softw.* 2016;70(8):1–33.
 28. Savic RM, Karlsson MO. Importance of Shrinkage in Empirical Bayes Estimates for Diagnostics: Problems and Solutions. *AAPS J.* 2009 Sep;11(3):558–69.

TABLES

Table 1 Study and subject characteristics (total sample size = 702 subjects)

Data set was cut-off at the 48-month visit (inclusive) for each study.

Characteristic	ADNI-1	ADNI-2
Sample size	381	321 * ³
Group names (%)	‘Late MCI’ (100)	‘Early MCI’ (52), ‘Late MCI’ (48)
Sex (%)	Female (36), Male (64)	Female (44), Male (56)
Baseline age in year, median (range)	75 (55, 89)	72 (55, 90)
Baseline body mass index in kg/m ² , median (range)	26 (18, 41)	27 (17, 51)
Number of (APOE)-ε4 alleles (%)	0 (46), 1 (42), 2 (12)	0 (49), 1 (40), 2 (12)
Baseline amyloid beta positivity (%) * ¹	No (2), Yes (2), Missing (96)	No (40), Yes (59), Missing (1)
Baseline LEAP TM ICV-HV in cm ³ , median (range)	5.1 (3.2, 7.7) [Missing for 88 subjects or 23%]	5.5 (3.1, 8.4) [Missing for 62 subjects or 19%]
Baseline Free Surfer TM ICV-HV in cm ³ , median (range)	7.2 (4.3, 12) [Missing for 88 subjects or 23%]	7.7 (2.0, 12) [Missing for 62 subjects or 19%]
Screening CDR-SB, median (range) * ²	1.5 (0.5, 5)	1.5 (0.5, 4.5)
Screening CDR-SB, mode * ²	1.0	0.5

Baseline MMSE, median (range)	27 (24, 30)	28 (24, 30)
Dropout by the 48- month visit (%) * ⁴	No (58), Yes (42)	No (77), Yes (23)
Subject follow-up duration in months, median (range)	36 (5.1, 58)	37 (4.7, 53)

Proportions not adding up to 100% are due to rounding. *¹ Amyloid beta positivity was determined by PET imaging; *² CDR-SB assessment were not performed at time zero or baseline. *³ There were 16 subjects who transitioned from ADNI-1 to ADNI-2 and were accounted for in ADNI-1 sample size but not in ADNI-2 sample size to prevent double counting. Their visits data in ADNI-2 were still included in the analysis dataset if they were within the dataset cut-off of up to 48-month visit (time zero is ADNI-1 start); *⁴ Subjects with CDR-SB scores at the 48-month visit were considered completers. Acronyms: ADNI = Alzheimer’s Disease Neuroimaging Initiative, CDR-SB = clinical dementia rating scale – sum of boxes, ICV-HV = intracranial volume-adjusted hippocampal volume, LEAPTM = Learning Embeddings Atlas Propagation, MCI = mild cognitive impairment, MMSE = mini-mental state examination.

Table 2 Base and final model parameter estimates

Estimates are for typical (or “average”) subjects: male, 73-year old, (*APOE*)- ϵ 4 carrier, 27.5-point MMSE, and 5.29 cm³ LEAPTM ICV-HV or 7.54 cm³ FreeSurferTM ICV-HV. The effect of a covariate change on a parameter estimate assumes that the other covariates are held constant and are at the values of a typical subject.

Parameter	Base model estimate	Covariate model estimate			
		Frequentist		Bayesian	
Statistical approach		Frequentist		Bayesian	
Estimation method		FOCE, Laplace		NUTS, Laplace	
Values format		Population mean (95% confidence interval*)		Median (2.5 th , 97.5 th)	
ICV-HV algorithm	Not applicable	LEAP TM	FreeSurfer TM	LEAP TM	FreeSurfer TM
CDR-SB Baseline (points)	0.082 (0.078, 0.086)	0.081 (0.077, 0.084)	0.081 (0.077, 0.084)	0.081 (0.078, 0.085)	0.081 (0.077, 0.085)
MMSE effect on baseline (centered at 27.5 points)	Not applicable	-2.2 (-2.8, -1.6)	-2.2 (-2.8, -1.6)	-2.2 (-2.8, -1.6)	-2.2 (-2.9, -1.6)
CDR-SB Intrinsic progression rate (year ⁻¹)	0.12 (0.098, 0.15)	0.13 (0.096, 0.16)	0.13 (0.1, 0.16)	0.12 (0.088, 0.15)	0.12 (0.089, 0.15)

Age effect on rate (centered at 73 years old)	Not applicable	1.5 (0.12, 2.8)	0.59 (-0.64, 1.8)	1.5 (0.046, 3)	0.69 (-0.84, 2.2)
Female sex effect on rate	Not applicable	1.3 (1, 1.6)	1.3 (1, 1.6)	1.3 (1, 1.7)	1.3 (0.98, 1.7)
MMSE effect on rate (centered at 27.5 points)	Not applicable	-3.2 (-5, -1.4)	-3.3 (-4.8, -1.8)	-3.1 (-5.1, -1.2)	-3.4 (-5.3, -1.4)
(<i>APOE</i>)- ϵ 4 non-carrier effect on rate	Not applicable	0.6 (0.41, 0.78)	0.6 (0.43, 0.77)	0.6 (0.41, 0.79)	0.61 (0.43, 0.78)
ICV-HV effect on rate (centered at 5.29 cm ³ for LEAP TM and 7.54 cm ³ for FreeSurfer TM)	Not applicable	-0.81 (-1.1, -0.48)	-0.52 (-0.7, -0.35)	-0.84 (-1.2, -0.49)	-0.56 (-0.77, -0.34)
Missing ICV-HV effect on rate	Not applicable	1.7 (1.2, 2.3)	1.7 (1.1, 2.2)	1.7 (1.1, 2.4)	1.7 (1.1, 2.4)
Shape factor of the Richards' model	3.6 (1.2, 6)	3.3 (1.5, 5.1)	3.1 (1.8, 4.3)	4.6 (-0.17, 9.5)	4.4 (-0.86, 9.7)
Variance of baseline random effects	0.27 (0.23, 0.3)	0.25 (0.21, 0.28)	0.25 (0.21, 0.28)	0.25 (0.21, 0.29)	0.25 (0.21, 0.29)
Covariance between baseline and rate random effects	0.071 (0.056, 0.086)	0.046 (0.033, 0.059)	0.045 (0.032, 0.058)	0.046 (0.033, 0.06)	0.046 (0.032, 0.059)

Variance of rate random effects	0.08 (0.065, 0.094)	0.062 (0.05, 0.073)	0.06 (0.049, 0.071)	0.063 (0.052, 0.075)	0.062 (0.05, 0.074)
Dispersion factor of the beta distribution	57 (54, 60)	57 (54, 60)	57 (53, 60)	57 (54, 60)	57 (54, 60)
Condition number	4.9	15	14	17	17

* Confidence intervals are calculated, under the normality assumption, using the asymptotic standard errors provided by NONMEM. Acronyms: *APOE* = Apolipoprotein E gene, CDR-SB = Clinical Dementia Rating Scale Sum of Boxes, FOCE = First-order conditional estimation, ICV-HV = intracranial volume-adjusted hippocampal volume, LEAP™ = Learning Embeddings Atlas Propagation, MMSE = mini-mental state examination, NUTS = No-U-Turn sampler.

Table 3 Interpretation of parameter values, covariates effects, and their relationships^{1,2}

Parameter	Frequentist FOCE Population Estimate	Transformation or parameter-covariate relationship	Interpretation for the population estimate at the original scale	Bayesian NUTS Median estimate (2.5 th , 97.5 th)
Baseline' (points)	0.081	$\text{Baseline}' \times 18$ Where Baseline' denotes the estimated transformed typical baseline CDR-SB of 0.081 (within the (0, 1) interval); the 18 points denotes the highest possible observed CDR-SB score.	The estimated baseline CDR-SB score is 1.5 points	0.081 (0.078, 0.085)
MMSE effect on baseline (centered at 27.5 points)	-2.2	$\text{Baseline} \times \left(\frac{\text{MMSE}}{27.5}\right)^{-2.2}$	A decrease in baseline MMSE score from 27.5 to 26.5 is associated to approximately 8% increase in baseline CDR-SB score	-2.2 (-2.8, -1.6)

Intrinsic progression rate (year ⁻¹)	0.13	$\frac{d\text{Score}_i}{dt} = r_i \times \text{Score}_i \times \left[1 - \left(\frac{\text{Score}_i}{\max(\text{Score}_i)} \right)^\beta \right] \times 18$ $0.13 \times 0.081 \times (1 - 0.081^{3.3}) \times 18$	The estimated typical rate of change in CDR-SB score is 0.2 point/year	0.12 (0.088, 0.15)
Age effect on rate (centered at 73 years old)	1.5	Rate of change $\times \left(\frac{\text{Age}}{73} \right)^{1.5}$	An increase in age from 73 to 74 years old is associated to approximately 2% increase in CDR-SB progression rate	1.5 (0.046, 3)
Female sex effect on rate	1.3	Rate of change $\times 1.3$	Females have approximately 30% higher CDR-SB progression rate than males	1.3 (1, 1.7)
MMSE effect on rate (centered at 27.5 points)	-3.2	Rate of change $\times \left(\frac{\text{MMSE}}{27.5} \right)^{-3.2}$	A decrease in baseline MMSE score from 27.5 to 26.5 is associated to approximately 12% increase in CDR-SB progression rate	-3.1 (-5.1, -1.2)

<i>APOE</i> -ε4 non-carrier effect on rate	0.60	Rate of change × 0.6	<i>APOE</i> -ε4 non-carriers have approximately 40% lower CDR-SB progression rate than carriers	0.60 (0.41, 0.79)
ICV-HV effect on rate ³ (centered at 5.29 cm ³)	-0.81	Rate of change × [1 – 0.81 × (ICVHV – 5.29)] Rate of change × [1 – 0.52 × (ICVHV – 7.54)]	A 1-cm ³ decrease in baseline ICV-HV is associated to approximately 81% increase in CDR-SB progression rate	-0.84 (-1.2, -0.49)
Missing ICV-HV effect on rate	1.7	Rate of change × 1.7	On average, the group of subjects with missing ICV-HV have approximately 70% higher CDR-SB progression rate than the subjects with a ICV-HV of 5.29 cm ³	1.7 (1.1, 2.4)
Shape factor of the Richards' model	3.3	Inflection point = $\left(\frac{1}{1 + \text{shape}}\right)^{1/\text{shape}} \times 18$	The inflection point of the rate of change in CDR-SB is	4.6 (-0.17, 9.5)

			estimated to occur at a score of approximately 11.6 points	
Variance of baseline random effects	0.25	<p>Coefficient of variation = $\sqrt{e^{0.25} - 1} \times 100$</p> <p>Where log-normally distributed between-subject variability was estimated for the baseline scores to prevent the prediction of nonsensical scores at the subject level.</p>	The coefficient of variation for the baseline CDR-SB scores is approximately 53%	0.25 (0.21, 0.29)
Covariance between baseline and rate random effects	0.046	<p>Correlation coefficient</p> $= \frac{0.046}{\sqrt{\text{variance of baseline}} \times \sqrt{\text{variance of rate}}}$	The correlation coefficient between baseline and rate random effects is 0.37	0.046 (0.033, 0.06)
Variance of rate random effects	0.062	<p>Coefficient of variation</p> $= \frac{\sqrt{0.062}}{\text{intrinsic progression rate}} \times 100$	The coefficient of variation for the CDR-SB intrinsic progression rate is approximately 196%	0.063 (0.052, 0.075)

Dispersion factor of the beta distribution	57	<p>Standard deviation = $\frac{\text{Score}' \times (1 - \text{Score}')}{57 + 1}$</p> <p>Where Score' denotes the expected CDR-SB score of the beta distribution within the (0, 1) interval.</p>	<p>At the typical baseline CDR-SB score, the standard deviation of the beta distributed residual variability is 0.036 points</p>	57 (54, 60)
Condition number	15	Not applicable	<p>Condition number is the ratio of the largest to the smallest eigenvalue of the covariance matrix and measures ill-conditioning. There is no consensus in the literature of what constitutes a large condition number. In the field of Pharmacometrics, it is commonly accepted that a condition number exceeding 1,000 is indicative of severe ill-conditioning.</p>	17

¹ CDR-SB scores were constrained to an open unit interval (0, 1) for implementation of the beta regression.

² Unless otherwise specified, estimates are for typical (or “average”) subjects: male, 73-year old, *APOE*- ϵ 4 carrier, 27.5-point MMSE, and 5.29 cm³ LEAPTM ICV-HV. The effect of a covariate change on a parameter estimate assumes that the other covariates are held constant and are at the values of a typical subject.

³ Determined by the LEAPTM imaging algorithm.

Acronyms: *APOE* = Apolipoprotein E, CDR-SB = Clinical Dementia Rating-Sum of Boxes, FOCE = First-Order Conditional Estimation, ICV-HV = Intracranial volume-adjusted hippocampal volume, LEAPTM = Learning Embeddings Atlas Propagation, MMSE = Mini-Mental State Examination, NUTS = No-U-Turn Sampler Bayesian estimation.

Table 4 Sample sizes to achieve 80% power in simulated placebo-controlled parallel group with FreeSurfer™ or LEAP™ ICV-HV-enriched and non-enriched clinical trials

Thresholds for enrichment are illustrative and are with respect to the median baseline ICV-HV value of the analysis dataset (~ 5 cm³). The simulations used: (a) the frequentist LEAP™ or FreeSurfer™ covariate models; (b) a hypothetical drug effect of 50% reduction in the disease progression rate; (c) the developed dropout model. Number of simulations was 1,000 for each non-enriched or enriched scenario.

Clinical trials with:	Algorithm	Sample size for 80% power (95% CI*)	Sample size reduction of enriched <i>versus</i> non-enriched trials (%) (95% CI)
No enrichment	LEAP™	474 (468, 481)	Reference
Only ICV-HV < 97.7 th (+2SD) subjects	LEAP™	469 (459, 479)	1 (-1, 4)
Only ICV-HV < 84.1 th (+1SD) subjects	LEAP™	353 (338, 363)	26 (23, 28)
Only ICV-HV < 50 th (median) subjects	LEAP™	214 (210, 218)	55 (54, 56)
No enrichment	FreeSurfer™	456 (446, 465)	Reference
Only ICV-HV < 97.7 th (+2SD) subjects	FreeSurfer™	440 (431, 448)	3 (1, 6)
Only ICV-HV < 84.1 th (+1SD) subjects	FreeSurfer™	315 (300, 325)	31 (28, 34)
Only ICV-HV < 50 th (median) subjects	FreeSurfer™	186 (183, 188)	59 (58, 60)

* Assumes independence. CI = confidence intervals, ICV-HV = intracranial volume-adjusted hippocampal volume, LEAP™ = Learning Embeddings Atlas Propagation, SD = standard deviation.

FIGURES

Figure 1 CDR-SB observed scores *versus* years from study baseline stratified by ICV-HV determined by (A) LEAP™ and (B) FreeSurfer™

Thresholds used for ICV-HV stratification are illustrative and correspond to the respective median values of the dataset. Acronyms: ADNI = Alzheimer's Disease Neuroimaging Initiative, CDR-SB = clinical dementia rating scale – sum of boxes, LEAP™ = Learning Embeddings Atlas Propagation, MCI = mild cognitive impairment.

Figure 2 Non-linearity in CDR-SB progression estimated with the Richards model

Data is from subject 888 in the analysis dataset. Open circles are observed scores; solid lines are the frequentist LEAP™ covariate model predictions (Section 7.5, Final Model); dashed line is the estimated inflection point of 11.6. Acronyms: CDR-SB = clinical dementia rating scale – sum of boxes.

Figure 3 Statistical power *versus* sample size for simulated placebo-controlled parallel group enriched and non-enriched clinical trials

(A, B, C) ICV-HV thresholds for enrichment are illustrative. The simulations used: the frequentist LEAP™ or FreeSurfer™ covariate models as applicable; a hypothetical drug effect of 50% reduction in the disease progression rate; the developed dropout model. Number of simulations was 1,000 for each non-enriched or enriched scenario. (C) Enrichment scenarios are for LEAP™ ICV-HV, and hypothetical new ICV-HV algorithms whose ICV-HV values are correlated with LEAP™ ICV-HV [Pearson's correlation coefficient, $R(\text{Pearson})$, of 0.5, 0.7 or 0.9]. It is assumed that the novel algorithm equals the original algorithm plus noise. For each simulation, the noise (k) was randomly sampled, $k \sim N(0, \sigma_k)$, where σ_k was calculated as in Supplementary Information, Equation S16.

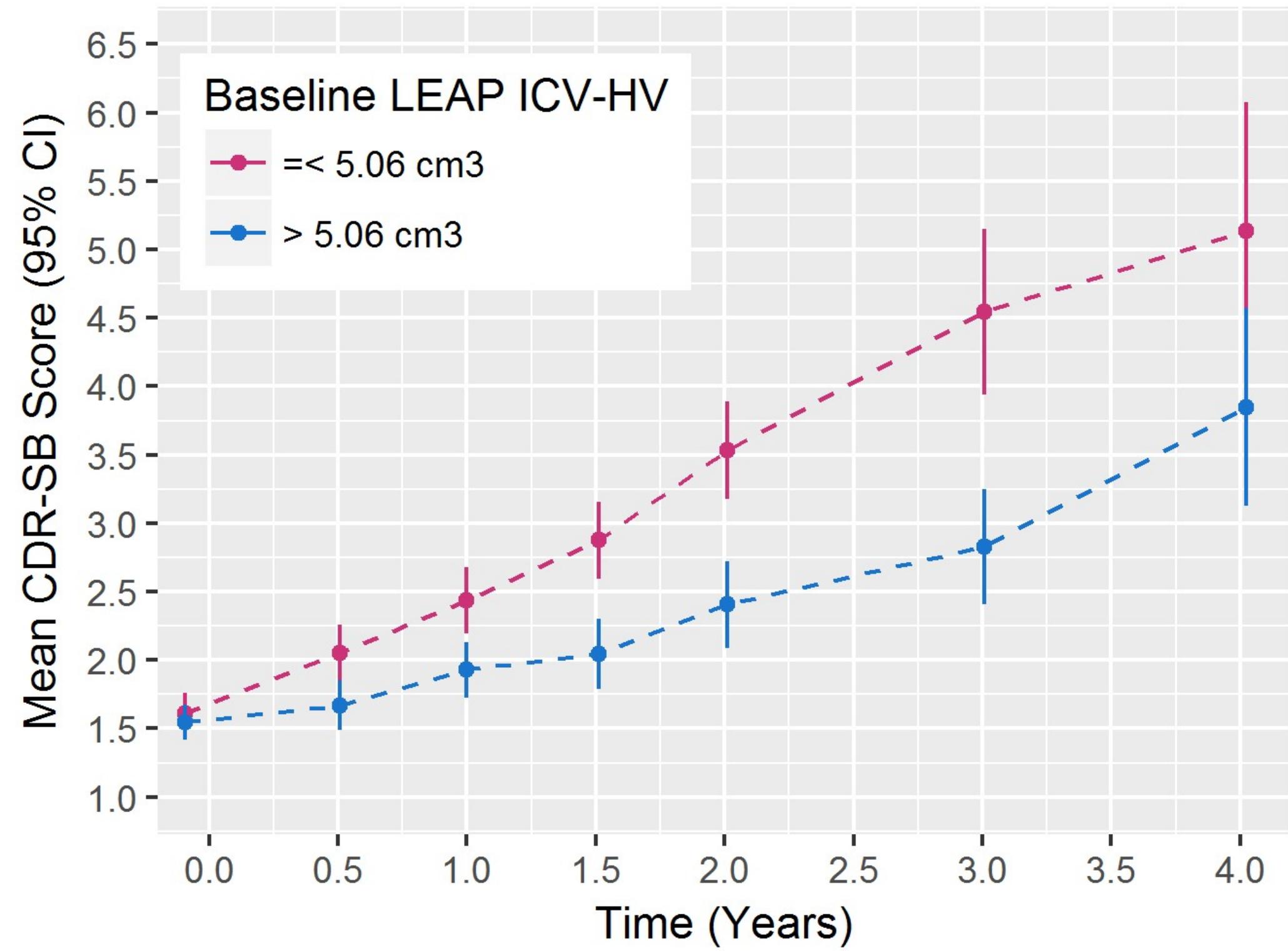
Acronyms: *APOE* = Apolipoprotein E gene, ICV-HV = intracranial volume-adjusted hippocampal volume, LEAP™ = Learning Embeddings Atlas Propagation, MMSE = mini-mental state examination, SD = standard deviation.

Supplementary Information

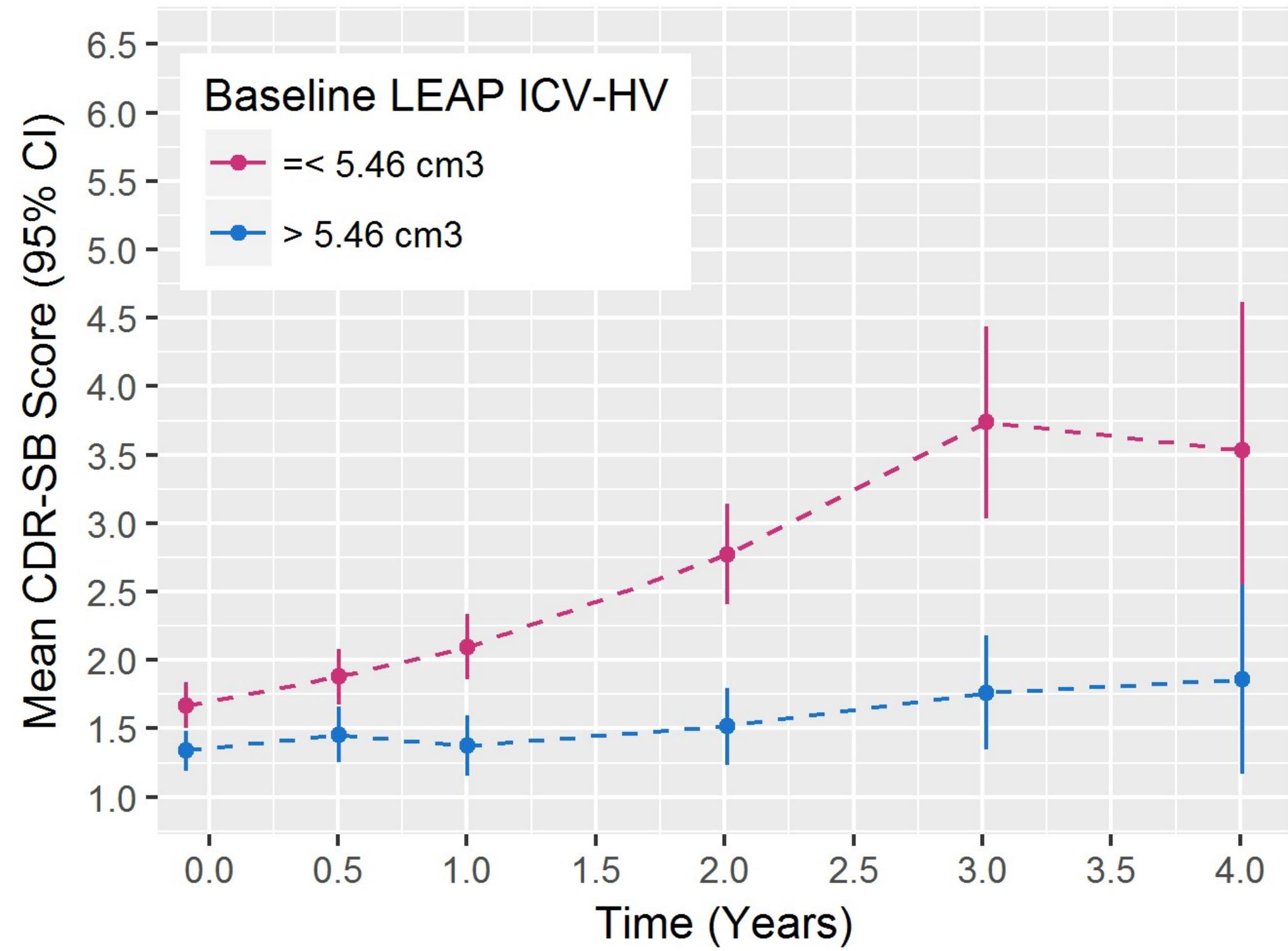
(Supplemental Material)

Supplementary Information

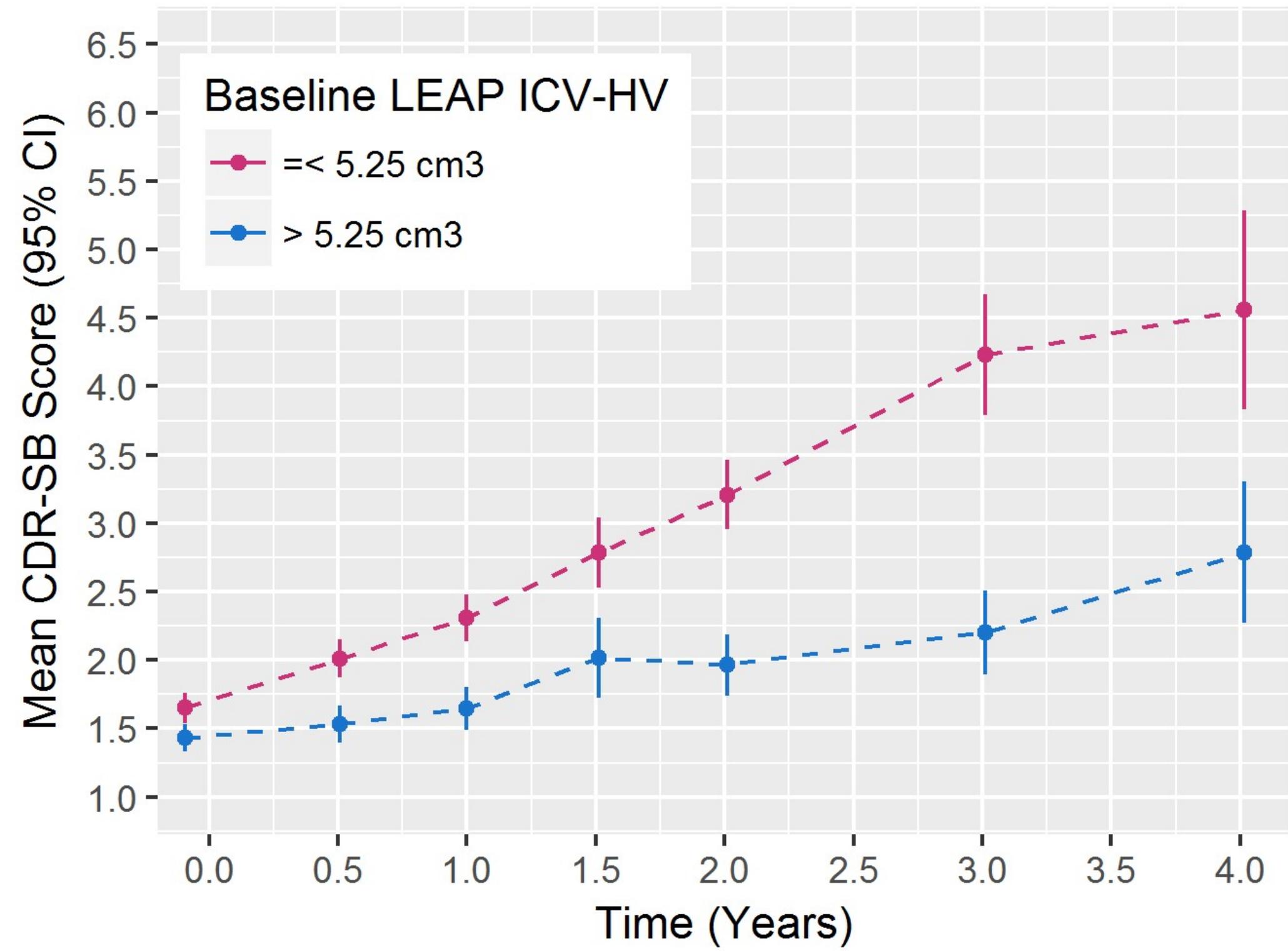
MCI ADNI-1



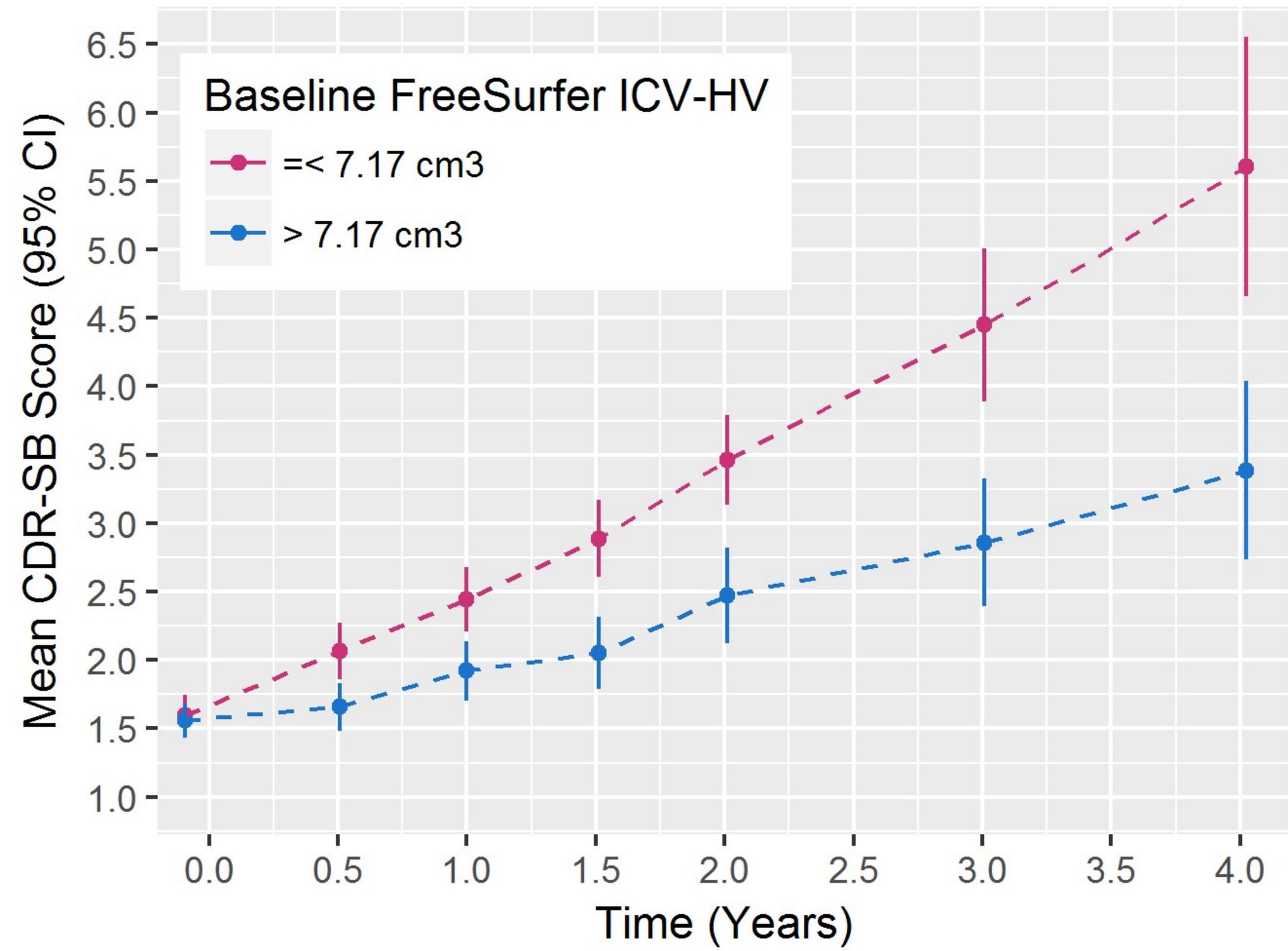
MCI ADNI-2



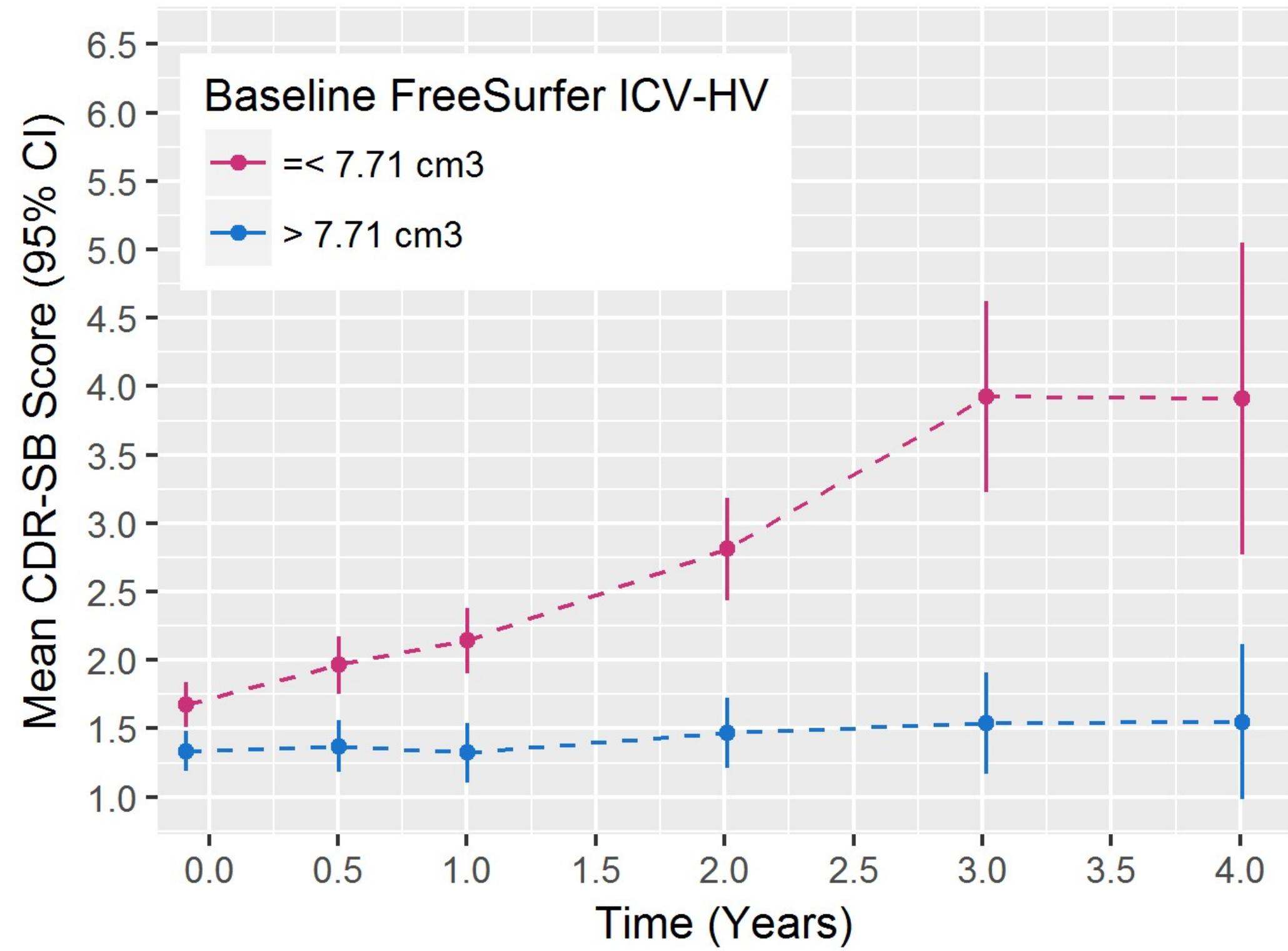
MCI ADNI-1 and ADNI-2



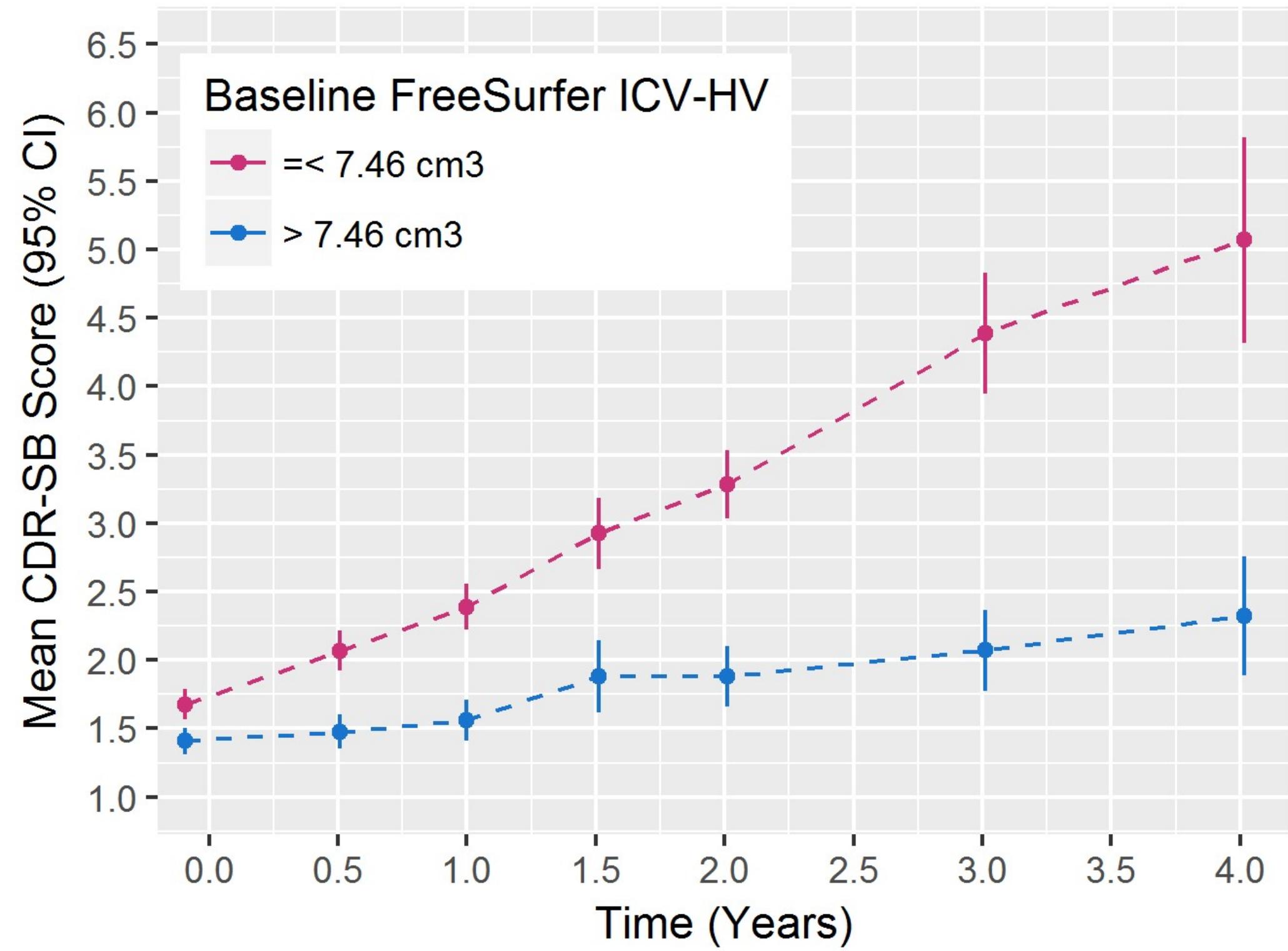
MCI ADNI-1

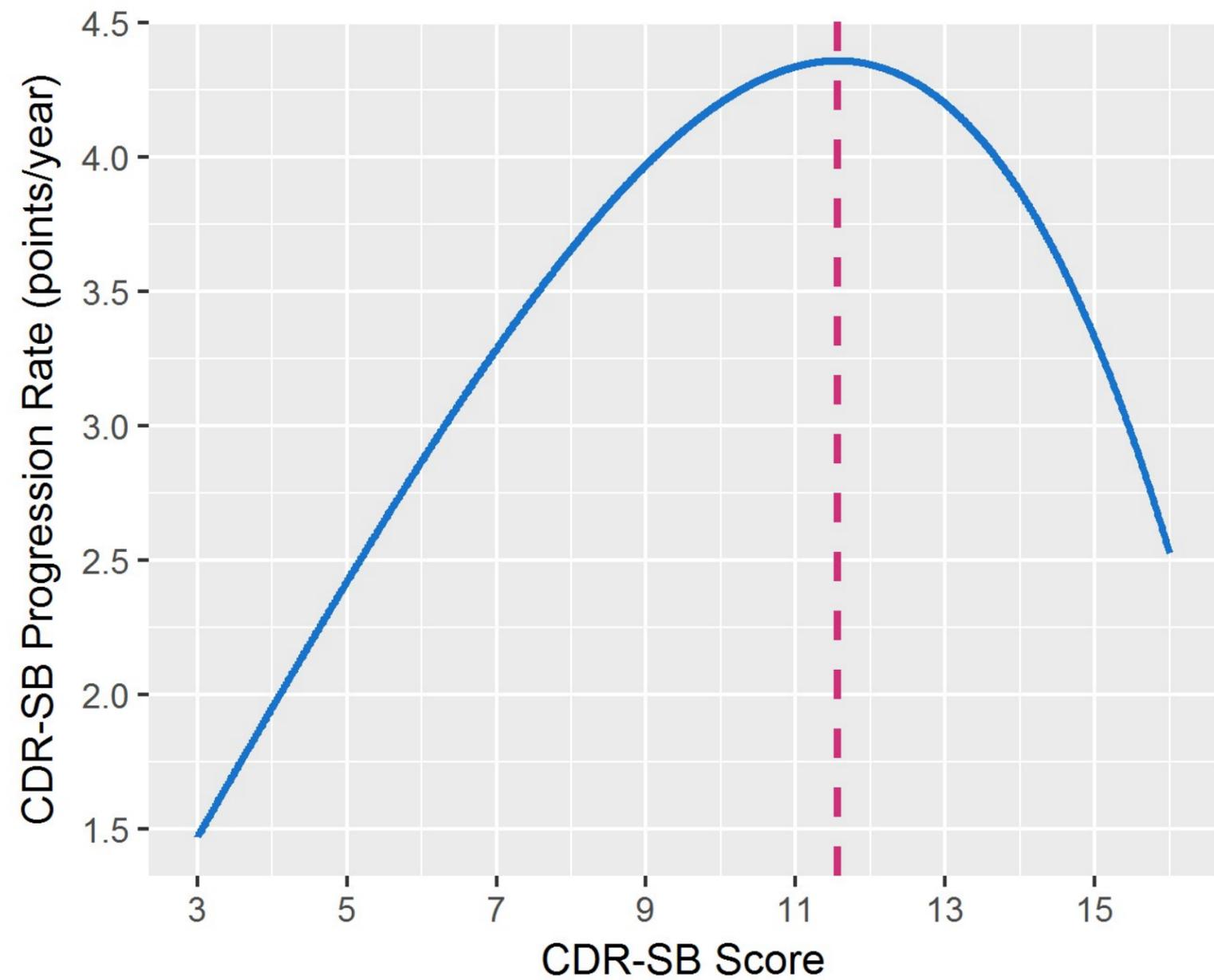
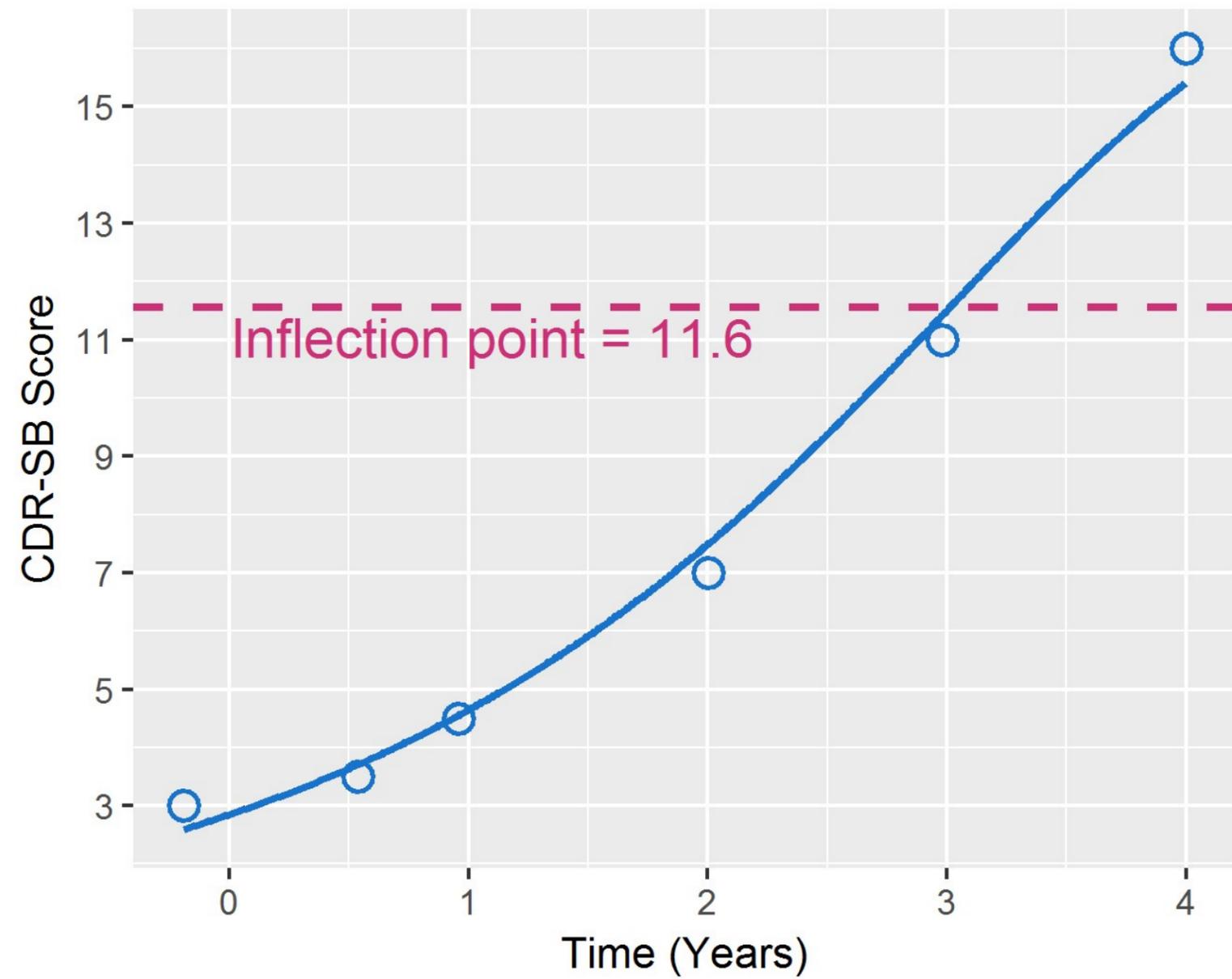


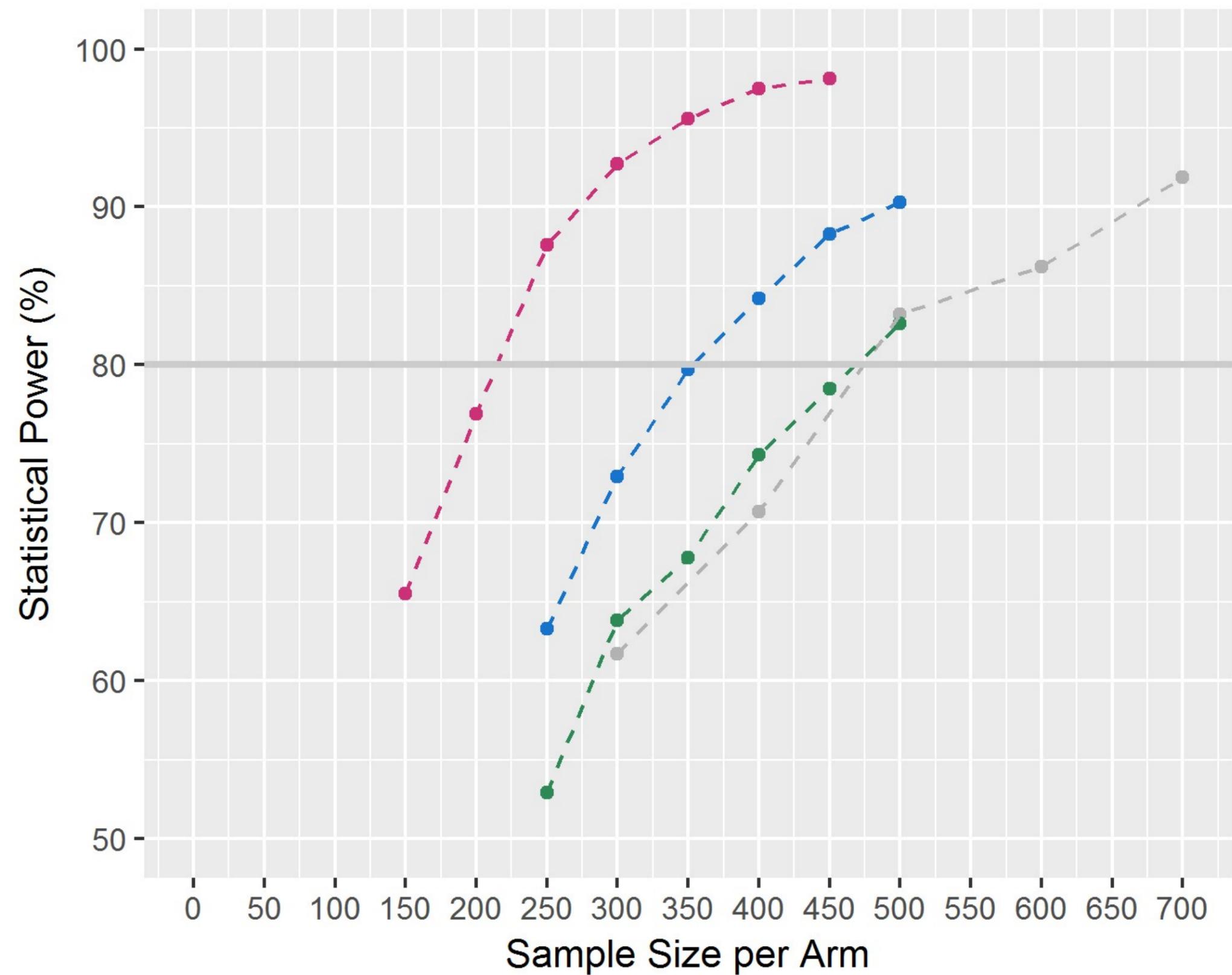
MCI ADNI-2



MCI ADNI-1 and ADNI-2

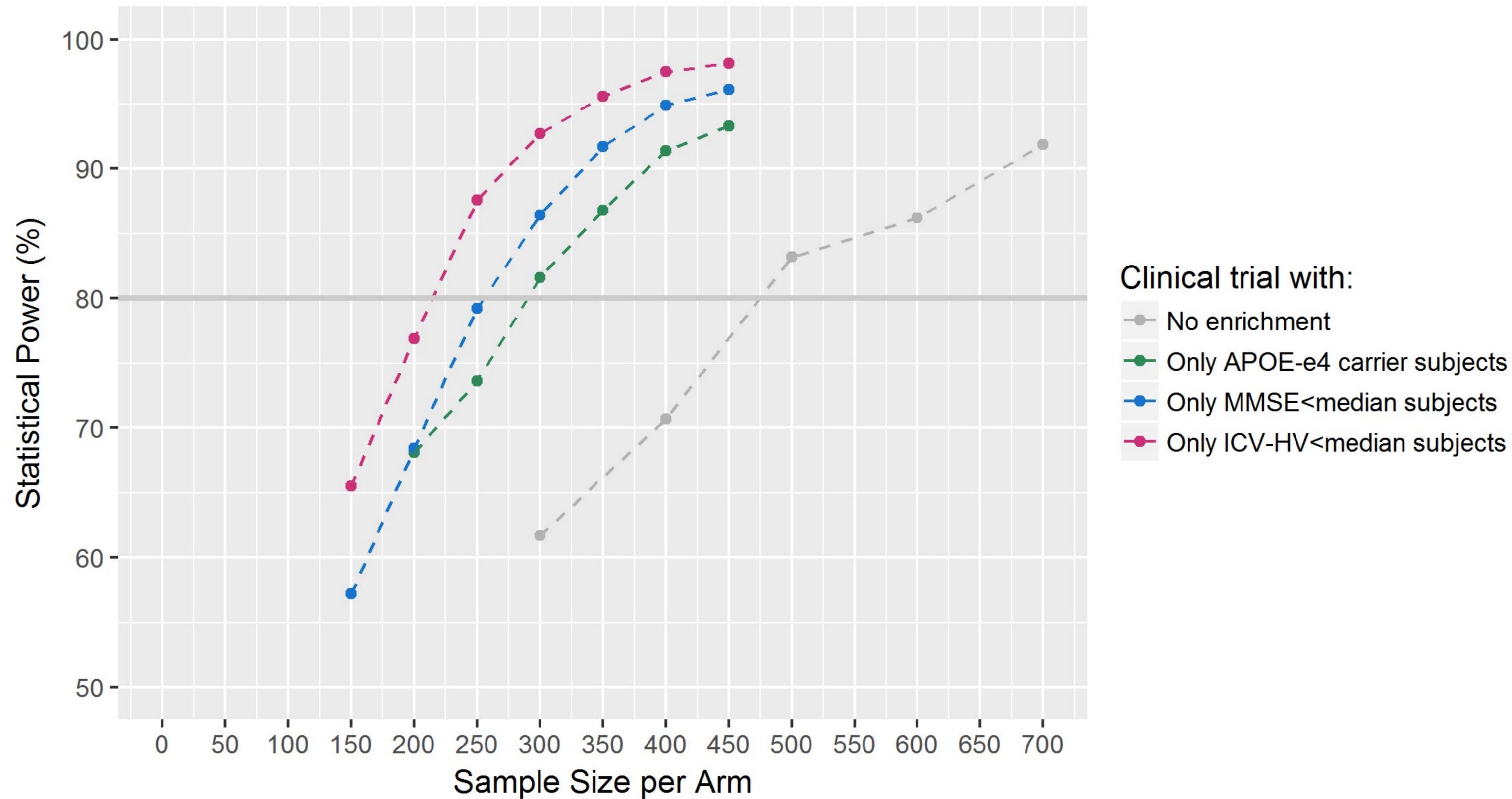


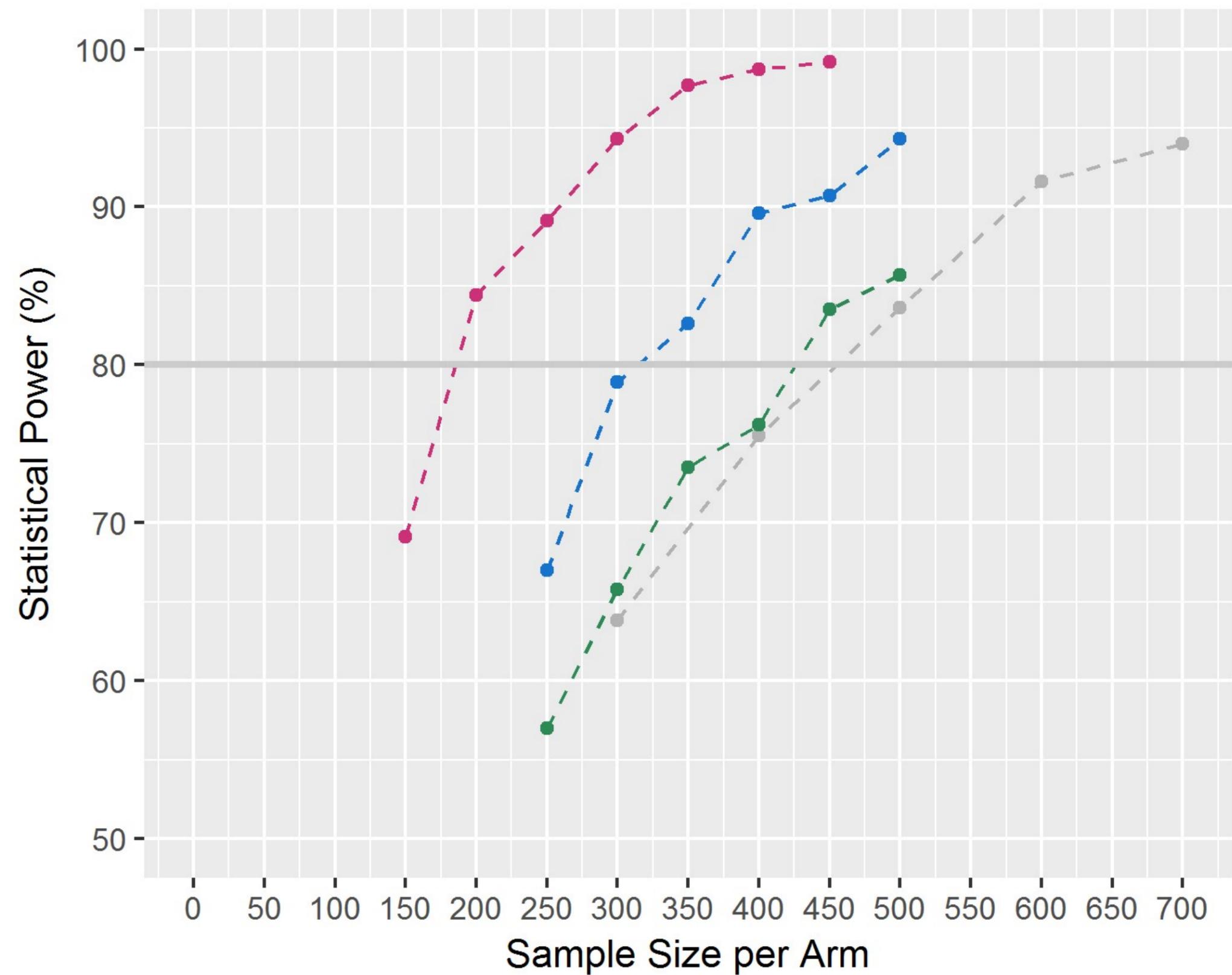




Clinical trial with:

- No enrichment
- Only ICSV-HV < 97.7th (+2SD) subjects
- Only ICSV-HV < 84.1th (+1SD) subjects
- Only ICSV-HV < 50th (median) subjects





Clinical trial with:

- No enrichment
- Only ICSV-HV < 97.7th (+2SD) subjects
- Only ICSV-HV < 84.1th (+1SD) subjects
- Only ICSV-HV < 50th (median) subjects

