Journal of Alzheimer's Disease 40 (2014) 967–979
DOI 10.3233/JAD-132090
IOS Press

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

Kaori Ito[a,*] and Matthew M. Hutmacher[b] for the Alzheimer's Disease Neuroimaging Initiative (ADNI)[1]
[a]*Pfizer Inc, Primary Care Business Unit, Groton, CT, USA*
[b]*Ann Arbor Pharmacometrics Group, Ann Arbor, MI, USA*

Accepted 31 December 2013

Abstract.
Background: Growing interest in treating Alzheimer's disease (AD) patients in the earliest stages requires new clinical endpoints. Currently, there is no established clinical endpoint or treatment duration for mild cognitive impairment (MCI) trials.
Objective: This analysis attempts to answer "how long the MCI clinical trial would be necessary" using the Clinical Dementia Rating Sum of Boxes (CDR-SB) as a clinical endpoint, where CDR-SB is an example of a suitable tool to assess both cognition and function as a single primary efficacy outcome.
Methods: A longitudinal model was developed to predict the CDR-SB time-profile. The CDR-SB is considered ideal to assess both cognition and function as a single primary endpoint in MCI trials. The median time for clinically "worsening", defined using several thresholds for change from baseline, was calculated using individual CDR-SB predictions. Covariates predictive of worsening were also evaluated.
Results: The median time to a 1-point change in CDR-SB was approximately 2 years in MCI patients. Higher baseline severity in disease, lower hippocampal volume, and ApoE4 carrier status were significant covariates predicting shorter times to worsening (faster progress). The results indicate that at least a 2-year trial would be necessary with 30% (or more) disease modifying drug with a sample size of $n = 350$ to detect the significant difference from placebo (80% power) and to achieve the target mean effect size of 0.5 point change in CDR-SB.
Conclusion: Predictions of CDR-SB changes from a longitudinal model are able to inform study design and possible enrichment strategies, based on covariate analyses, for prospective planning of clinical trials in MCI patients.

Keywords: ADNI, biomarkers, bounded outcome, Clinical Dementia Rating Sum of Boxes (CDR-SB), clinical trial, enrichment strategy, longitudinal data, median time to worsening, mild cognitive impairment

*Correspondence to: Kaori Ito, Pfizer Inc, Primary Care Business Unit, 445 Eastern Point Road, Groton, Connecticut 06340, USA. Tel.: +1 860 441 5743; E-mail: kaori.ito@pfizer.com.

[1]Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://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

ISSN 1387-2877/14/$27.50

INTRODUCTION

Recent advances in the understanding of the underlying pathophysiology of Alzheimer's disease (AD) have led to clinical testing of numerous new treatment modalities aimed at altering the disease early in its clinical progression, and growing interest in starting to treat patients in the earliest stages even before the disease manifests clinical symptoms. The Alzheimer's Disease Assessment Scale-cognitive subscale-11 items (ADAS-cog$_{11}$) has been the most commonly used outcome measure of cognitive function in anti-dementia clinical trials in mild to moderate AD patients. However, ADAS-cog$_{11}$ is not sensitive enough to detect changes in the pre-dementia stages of AD or mild cognitive impairment (MCI) [1], and there is no generally accepted, validated clinical endpoint for use in therapeutic trials. Therefore, tremendous efforts are currently ongoing to explore new, informative endpoints for clinical trials [2–8].

In addition, the FDA has recently released a draft guidance on treating early stages of AD [9], suggesting that the Clinical Dementia Rating Scale Sum of Boxes (CDR-SB) [10] or composite measurement tools may be acceptable for use in clinical trials. The guidance indicates specifically that CDR-SB would be appropriate to assess both cognition and function as a single primary efficacy outcome measure. CDR-SB is a widely used scale that has demonstrated validity and reliability in longitudinal assessment; therefore, CDR-SB may also be suitable for use in early AD [11, 12].

Early detection of AD is thought to offer the best opportunity for effective intervention; however, there is no clear answer for "how long the clinical study in MCI patients would be necessary" to demonstrate clinically meaningful changes. One possible clinical trial is to evaluate whether a new drug can delay the conversion from MCI to AD. However, it would require a long clinical study, possibly several years, since the estimated rate of progression from MCI to AD is up to 10% per year [13]. Another study indicates that about half of people who have visited a doctor for their MCI symptom develop dementia in three or four years [14]. This is problematic, because lengthy clinical trials are costly, which could discourage clinical development of novel compounds and delay beneficial treatments reaching patients. To address such issues, Eisen et al. [15] suggested that a 1 point change (or even 0.5 point) in CDR-SB signified a clinically meaningful "worsening", and that changes in CDR-SB could potentially be used to evaluate the clinically meaningful effect of a new drug in MCI clinical trials as a surrogate for conversion from MCI to AD.

In this analysis, the longitudinal CDR-SB data from an MCI population captured in the ADNI database was modeled, and covariates of interest were evaluated using method perhaps atypical to this indication [16]. The median time-to-worsening defined as a 0.5, 1, or 2 point change from baseline in CDR-SB were calculated and compared based on individual CDR-SB predictions. The median times to a 1 point change in CDR-SB were also calculated by disease severity, age, gender, hippocampal volume, and ApoE4 to demonstrate their effects on the disease progression. In addition, power was calculated as a function of sample size using a hypothetical disease modifying drug to show the utility of the approach.

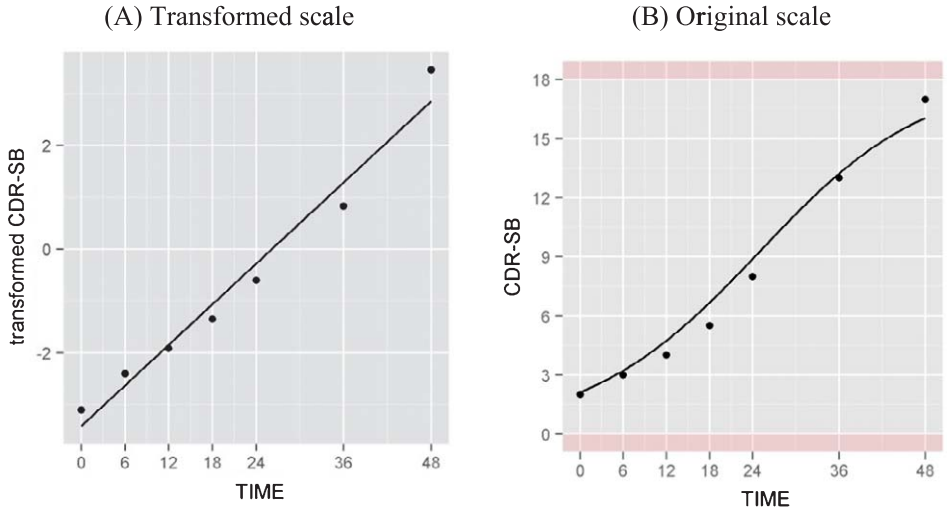
METHODS

Data

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.edu/). Analyses were performed on the data downloaded on June 10, 2013 from the ADNI web portal (http://adni.loni.usc.edu/data-samples/access-data/) using "adnimerge" package (R-version). All data points available for early and late MCI patients in the ADNI database were included in this analysis except the subjects confirmed withdrawal from the study with only baseline data. In addition, the data from all populations (normal, early MCI, late MCI, and AD) were used to visualize the whole trend of disease progression with CDR-SB (Supplementary Fig. 1). Detailed ADNI protocol information can be found at http://www.adni-info.org/.

Modeling longitudinal CDR-SB scores

The CDR-SB was assessed at baseline and at subsequent in-clinic visits (every 6 months) during the study. One could use a time-to-event analysis and the observed CDR-SB data to estimate the time, for example, to 1 point change from baseline (the threshold). However, technical and conceptual issues arise. First, a subject is only known to have crossed the threshold sometime between the prior clinic visit at which the threshold was not achieved, and the current clinic visit, at which the threshold difference was achieved. Thus the data are interval censored, which requires nonstandard time-to-event techniques for analysis. Evaluation of the observed CDR-SB data in this manner includes



$$y \in (0,18) \mapsto z \in (0,1): \quad z = \frac{y-a}{b-a}$$

$$z(0,1) \mapsto x \in (-\infty,\infty): \quad x = h(y;\alpha) = \begin{cases} \log\left[\dfrac{(1-z)^{-\alpha}-1}{\alpha}\right] & \alpha \neq 0 \\ \log(-\log(1-z)) & \alpha = 0 \end{cases}$$

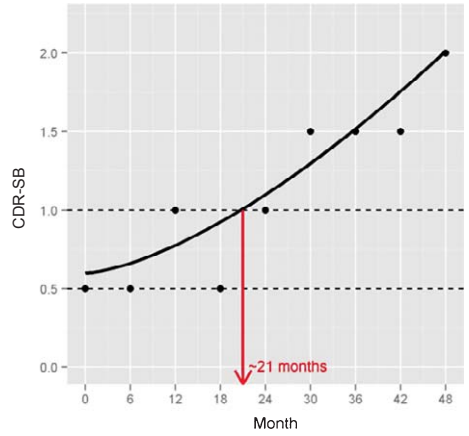


Fig. 1. Conceptual scheme for data transformation and model fitting. A) Transformed data were fitted using a linear model (Equation 1). B) Data and individual predictions by the model were transformed back to the original CDR-SB scale (range 0 to 18). Note the linear prediction on the transformed scale is not linear on the CDR-SB scale because the transformation is nonlinear. C) The dotted lines show the 0.5 point change from baseline in CSR-SB and the red arrow depicts the approximate time the model predicts a 0.5 increase relative to baseline. A standard time-to-event analysis ignores the underlying trend in the CDR-SB scores for that individual and would have used assumed that the worsening occurred between Month 6 and Month 12. The model however reflects a smooth underlying disease progression, and provides an estimate of when the subject crosses the threshold with respect to this smooth process (approximately 21 months using this conceptual example).

the day-to-day or visit-to-visit random variation within subject, which can induce unappealing bias. For example, a subject could achieve a 1 point change between Visit 2 and 3, yet this change could be lost between Visit 3 and 4, with a 1 point change observed sometime thereafter. Such a time-to-event analysis ignores the *underlying* trend in the CDR-SB scores for that individual. One would expect a smooth underlying disease progression, and it is desirable to know when the subject crosses the threshold with respect to this smooth

process (Fig. 1). Additionally, the time to event ignores the information contained in the data prior to and after the threshold, which can be helpful in predicting this clinical worsening precisely. Thus, we propose predicting the time to worsening from a longitudinal model. Efficient use of all the information is anticipated to reduce the sample size needed to show a difference.

The range of the CDR-SB is 0 to 18 by 0.5 point increments (bounded outcome scores). Such data can have non-normal and atypical distributions. These issues have been discussed and a method to handle bounded outcome score data was published by Hutmacher et al. [17]. This method was applied recently for modeling FAQ (Functional Assessment Questionnaire), which has a range of 0 to 30, from the ADNI database [16]. We used the same method to model longitudinal CDR-SB, and the methodology is described below.

A nonlinear mixed-effect likelihood-based approach was used, and the non-boundary data were scaled between 0 and 1 by dividing each CDR-SB value by 18. A transformation was applied to the data, which provides flexibility for handling the difficult data distribution shapes. The approach considers the boundary data as censored when formulating the likelihood. This is based on the assumption that the boundary data are reported as such because the measurement instrument lacks sufficient precision to differentiate the underlying true measurements from the boundaries. We consider the data sufficiently continuous for application of this approach.

The proposed model, conditional on the subject-specific random effects, on the transformed response scale is:

$$y = h\,(CDR\text{-}SB(t)) = \alpha_{INT} + \eta_{INT} + (\alpha_{SLP} + \eta_{SLP}) \cdot t + \varepsilon = \mu\,(t) + \varepsilon \qquad (1)$$

where y is the transformed CDR-SB value after applying the transformation $h()$, *CDR-SB(t)* is the CDR-SB score at time t. α_{INT} is a parameter predicting the baseline status (i.e., the intercept), η_{INT} is the random effect for the intercept, and α_{SLP} is the slope rate or rate of change in transformed CDR-SB over time (representing the underlying progression of the disease), which as a random effect η_{SLP}. The random effects were assumed to be normally distributed, and different variances were expected for each patient population (early MCI and late MCI). ε is residual error, assumed to be normally distributed with variance σ^2. The transformation applied in this analysis has a nice feature in that the scale used to measure the progression of the disease does not need to influence the rate of disease progression to maintain feasible predictions. One does not need to bind the slope (α_{SLP}) to be positive, nor does one need to force α_{SLP} to become smaller as the disease progresses toward the upper boundary to avoid predicting scores outside of the 0–18 range (which would unnecessarily complicate the model). Figure 1 provides a conceptual scheme for the data transformation and model fitting for predictions of individual profiles. More technical details for the interested can be found elsewhere [16, 17].

Covariate evaluation

Based on previous findings [16, 18–22], baseline covariates were selected for evaluation: disease severity, age, education, ApoE4 genotype, gender, total cholesterol, and hippocampal volume (imaging biomarker). The same approach used previously [16] in which a composite score derived from baseline MMSE and FAQ ($SEV_b = ((30 - MMSE_b) + FAQ_b)$), cognitive and functional domains, respectively) was used to describe baseline severity. Approximately 20% of baseline hippocampal volume data were missing in the database, and these missing values were imputed by replacing with median values for those after matching by age group, gender, ApoE4 status, and diagnosis (early MCI or late MCI), since there are relatively high correlations among these covariates (Supplementary Fig. 2). During the model building process, parameter estimates with/without imputed covariate were compared to assess the sensitivity to the imputation. Cerebrospinal fluid biomarkers (such as $A\beta_{42}$, p-tau) were not tested in the model, because these data were missing for nearly half of the patients.

Covariates were added one by one in a forward stepwise manner [23], examining the change in minimum objective function values (OFV), and also the precision of the parameter estimates. A decrease of >6.6 in the OFV indicated that a proposed model with 1 additional parameter provided a better fit than the reduced reference model ($p < 0.01$ Chi-Square). Covariates were retained if the model was stable and its parameter estimates demonstrated acceptable precision.

Model fitting was performed using a population analysis approach (NONMEM Version VII, Level 1.2, ICON Development Solutions, Ellicott City, MA) with the Laplace Conditional Estimation method. Data handling/missing covariate imputation, diagnostic graphics, and post-processing of output were performed using R (version 2.15.3). SAS (version 9.3) was used for power and sample calculation.

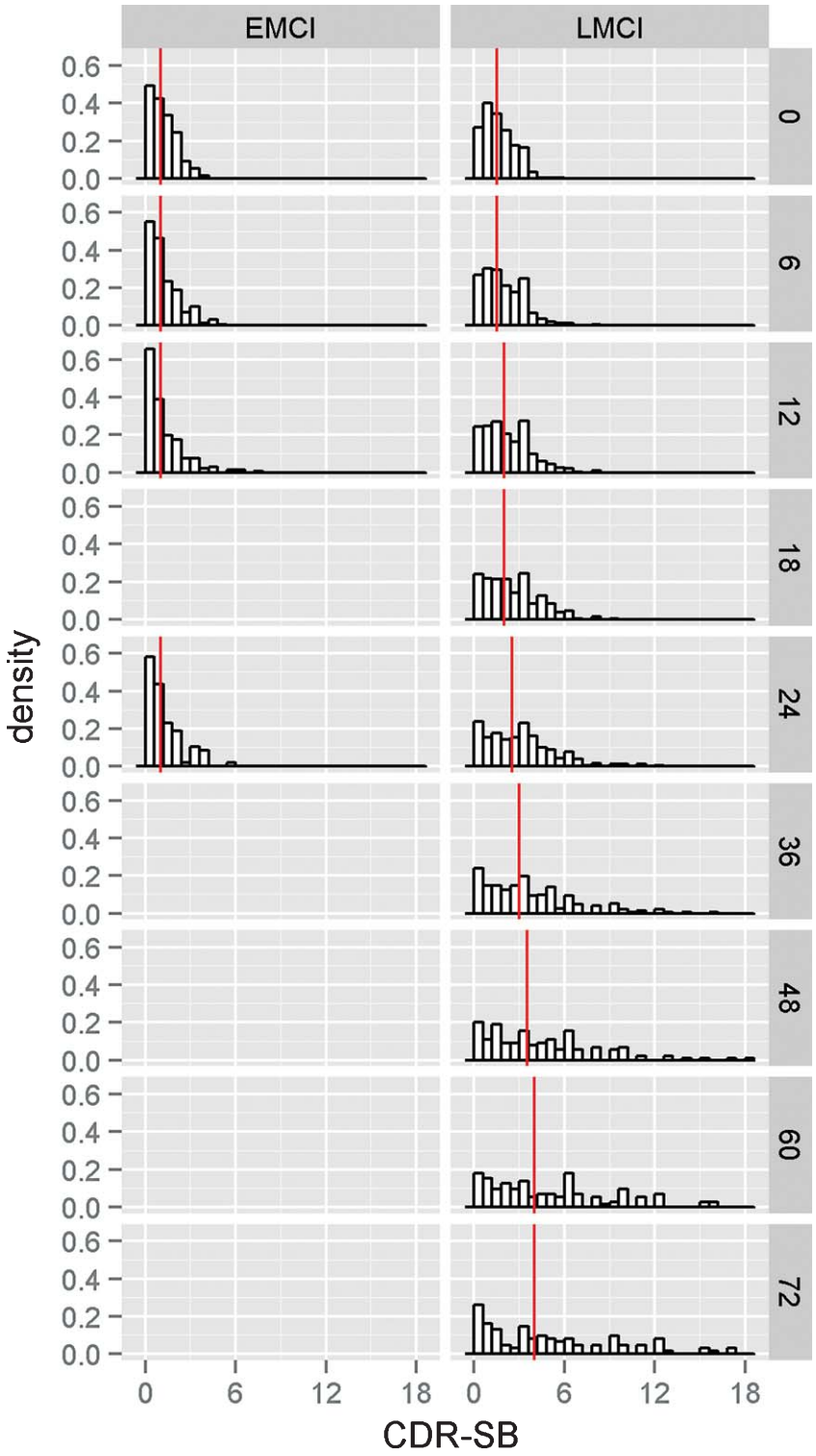


Fig. 2. Distribution of CDR-SB by each visit (Month) in early MCI (EMCI) and late MCI (LMCI) population. Vertical lines in red are median values at each visit.

Prediction of clinical worsening

Once the final longitudinal model was identified, CDR-SB was predicted for each individual patient by week and the median time to 1 point change from baseline was calculated, considered as clinical worsening for the patient. If 1 point change from baseline is not observed during the duration of the available data (either still following up, or dropped out from the study before it occurred), then the maximum time (the last observation time in the dataset) was recorded for the patient (and treated as censored in the summary discussed below). The same method was applied for 0.5 point change or 2 point change in CDR-SB.

Kaplan–Meier plots were generated to visualize the difference in 0.5, 1, and 2 point change in CDR-SB, and median time for each threshold was calculated using `survfit` function in R (version 2.15.3). The median time is considered as the time to 50% population (in the dataset) worsening to the specified thresholds (0.5, 1, and 2 points), therefore, the time to 20% population worsening was also calculated for comparison purpose which may help to reduce the duration of the clinical studies in the exploratory stage. In addition, the median times to a 1 point change were plotted by covariates (dichotomized into groups) to demonstrate their effect on disease progression and potential enrichment strategy.

Power as a function of sample size to detect significant differences from placebo in MCI clinical trials were computed using data generated from a hypothetical disease modifying drugs with varying effect magnitudes (10 to 50%) on the slope. The minimum disease modifying effect to achieve 0.5 point change from placebo at 24 and 36 months clinical trials was also determined.

RESULTS

The dataset used in this analysis included 301 early MCI patients (currently up to 24 months), and 550 late MCI patients (up to 72 months). The distribution of observed CDR-SB is displayed as histograms subset by patient group (early MCI, late MCI) and by visit (month) in Fig. 2. The red vertical line is the median CDR-SB value at each visit, and these lines shift toward the right (worsening) as the study progresses. The distributions are also widening (more variability) over time. Supplementary Fig. 1 also clearly indicates that the disease progresses over time in normal elderly, early MCI, late MCI, and AD patients in the ADNI database.

Table 1
Estimated median time to clinically worsening (1 points change from baseline in CDR-SB) by covariates

Covariates		Median time to worsening (Months)
Baseline severity		
≤6 (milder)	45.0	[33.0, 56.2]
>6 (severe)	14.5	[12.8, 16.0]
Hippocampal Volume (mm^3)		
≥6700	55.2	[39.0, NA]
<6700	18.5	[16.2, 21.2]
Baseline severity + Hippocampal Volume		
Group 1	NA	[56.2, NA]
Group 2	28.0	[23.2, 34.8]
Group 3	19.5	[13.8, 36.5]
Group 4	14.0	[11.8, 15.0]
Age (y)		
≤75	22.8	[20.5, 29.5]
>75	23.5	[19.5, 32.0]
Gender		
Male	29.5	[23.2, 35.2]
Female	19.5	[16.5, 22.0]
ApoE4		
Carrier	18.5	[15.8, 21.5]
Non-carrier	37.2	[29.5, 51.8]

Covariates are dichotomized at approximate median values in the late MCI population in ADNI database. Baseline severity is defined as bFAQ + (30-bMMSE). Since baseline severity and hippocampal volume are highly correlated, the median time to worsening for each combination was calculated: Group = 1 (mildest): milder in baseline severity (BSEV ≤6) and higher hippocampal volume (≥6700 mm^3), group = 2 (milder): milder in baseline severity and lower hippocampal volume (<6700 mm^3), group = 3 (more severe): more severe in baseline severity (BSEV >6) and higher hippocampal volume, group = 4 (most severe): more severe in baseline severity and lower hippocampal volume. [ ]: 95% CI.

Evaluation of covariates

Baseline severity, hippocampal volume, and ApoE4 (carrier versus non-carrier) genotype were significantly impacting the intercept (α_{INT}) and/or the rate of disease progression (α_{SLP}). ApoE4 was also tested as number of alleles (0, 1, or 2); however, there was no improvement in fit compared to using carrier (1 or 2) versus non-carrier (0). Age and gender were not found statistically significant; however, these covariates were kept in the final model as these are known to be clinically important covariates for understanding disease progression. Years of education and total cholesterol were not found to be significant. The parameter estimates from the base model and the final model are summarized in Supplementary Table 1. Overall, the final model parameters were well estimated with reasonable precision.

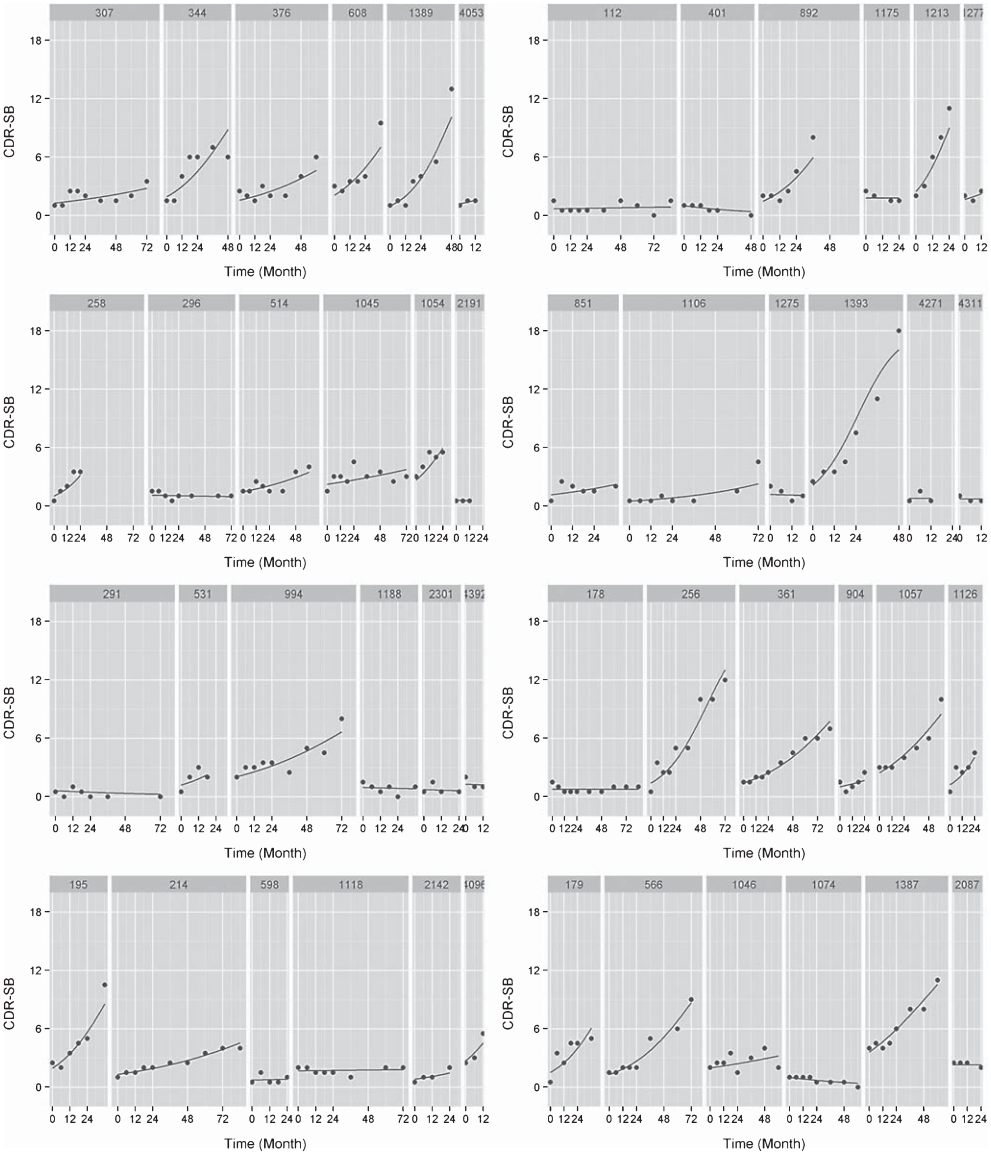


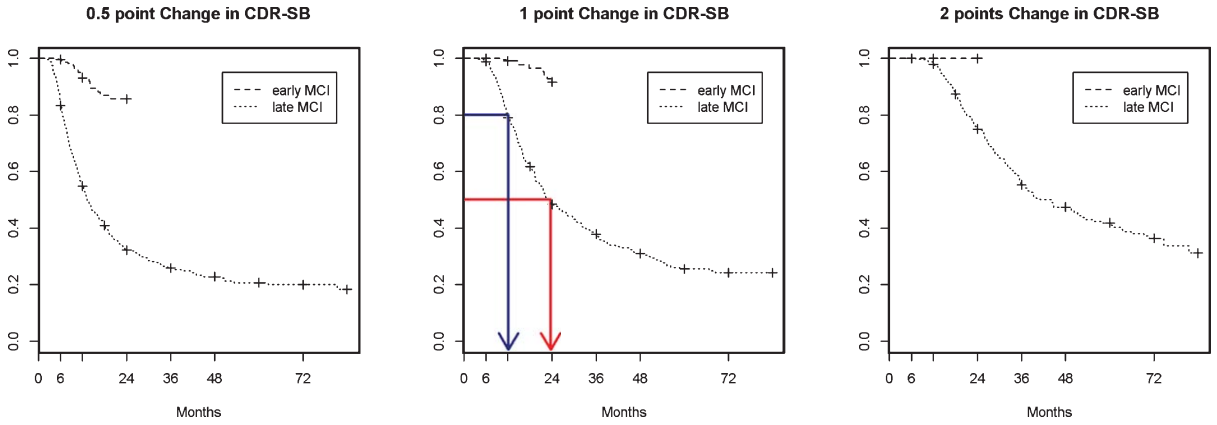
Fig. 3. Individual predicted CDR-SB overlaid with observed CDR-SB (few patients are shown here). Lines in black are individual predicted CDR-SB by every week, and circles are observed CDR-SB. Note that if the patient was dropped, or lost follow-up, or still ongoing, the predicted CDR-SB were obtained up to the maximum time observed for the patient.

Model evaluation

Figure 3 shows several predicted individual CDR-SB time-profiles from the final model, overlaid with observed data. These plots indicate that the model adequately characterizes the individual time courses. The performance of the final model was also assessed with regards to the population. This was achieved by calculating selected statistics of the observed data across the entire population and comparing these statistics to prediction intervals of these statistics computed by simulating data from the model. This model evaluation is provided as Supplementary Fig. 3. The observed data statistics are within the model-based predictions intervals indicating the model has captured the central tendency of the data over time as well as the variability in the population.

Comparison for the time to clinically "worsening"

Figure 4 shows the Kaplan-Meier plots for using different definitions of worsening (thresholds of 0.5,



(%)	Time to worsening (month) in late MCI population					
Population converted	0.5 point change		1 point change		2 points change	
20%	6.5	[6.3, 7.3]	12.0	[11.0, 13.5]	21.0	[19.5, 24.0]
50%	13.3	[12.3, 16.0]	23.3	[21.0, 28.3]	44.0	[36.0, 53.3]

Fig. 4. Kaplan–Meier plots for 0.5 point, 1 point, and 2 points change from baseline in CDR-SB and estimated time to worsening (for 20% and 50% population) in late MCI population. [ ]: 95% CI.

1, and 2 points change from baseline) as well as 20% or 50% populations achieving this threshold). Note that early MCI patients show little or no "worsening" with this dataset: therefore, clinical worsening was calculated only for late MCI patients. The results indicate that 50% of the population will achieve the 0.5 threshold in approximately 13.3 months (>1 year), or the 1.0 threshold in 23.3 month (∼2 years). Using the 20% population target, a 0.5 threshold was achieved at 6.5 months, or 1.0 threshold in 12 months.

Figure 5 shows the Kaplan–Meier plot for 1 point change from baseline in CDR-SB by covariates, and patients were dichotomized in to groups (above and below the approximate medians, i.e., age of 75 years, baseline severity of 6, and hippocampus volume of 6700 mm^3) in the late MCI population. Table 1 summarizes the median time to achieve a 1 point change for each group corresponding to Fig. 5. Baseline severity and baseline hippocampal volume are strong predictors for the median time-to-worsening, which stems from the strong significance found for these covariates during model building. For example, the median time to worsening was 45.0 months for "milder" severity versus 14.5 months for "more severe" patients in late MCI population. The estimate was 18.5 months for lower hippocampal volume patients and 55.2 months for higher hippocampal volume patients (Table 1). Since baseline severity and hippocampal volume are correlated, the median times were also calculated for each combination. Although there is some overlap (Fig. 5, top-right panel), the estimate was 14.0 months for the most severe patient (group 4: higher baseline severity and lower hippocampal volume) who are most likely to progress fast, and 19.5 months for the next severe group (group 3), and 28.0 months for the milder group (group 2). The median time was greater than 84 months for the mildest patients (group 1: lower baseline severity and higher hippocampal volume). The estimates for time-to-worsening were 22.8 and 23.5 months for ≤75 and >75 years of age, respectively with a threshold of 1.0 point change in CDR-SB (not a significant covariate). Gender was not a significant covariate either during the model building; however, the Kaplan–Meier plot shows some difference between male and female. This gender effect may due to correlations with other significant covariates (no further investigation was made in this analysis). The estimates were 29.5 and 19.5 months for male and female, respectively, which indicates, in general, female progresses faster than male. ApoE4 was a significant covariate, and the estimates were 18.5 and 37.2 months for carrier and non-carrier, respectively.

Power and sample size for disease modifying drugs in MCI trials

Figure 6A shows the power estimates to detect significant difference from placebo in 24-month and

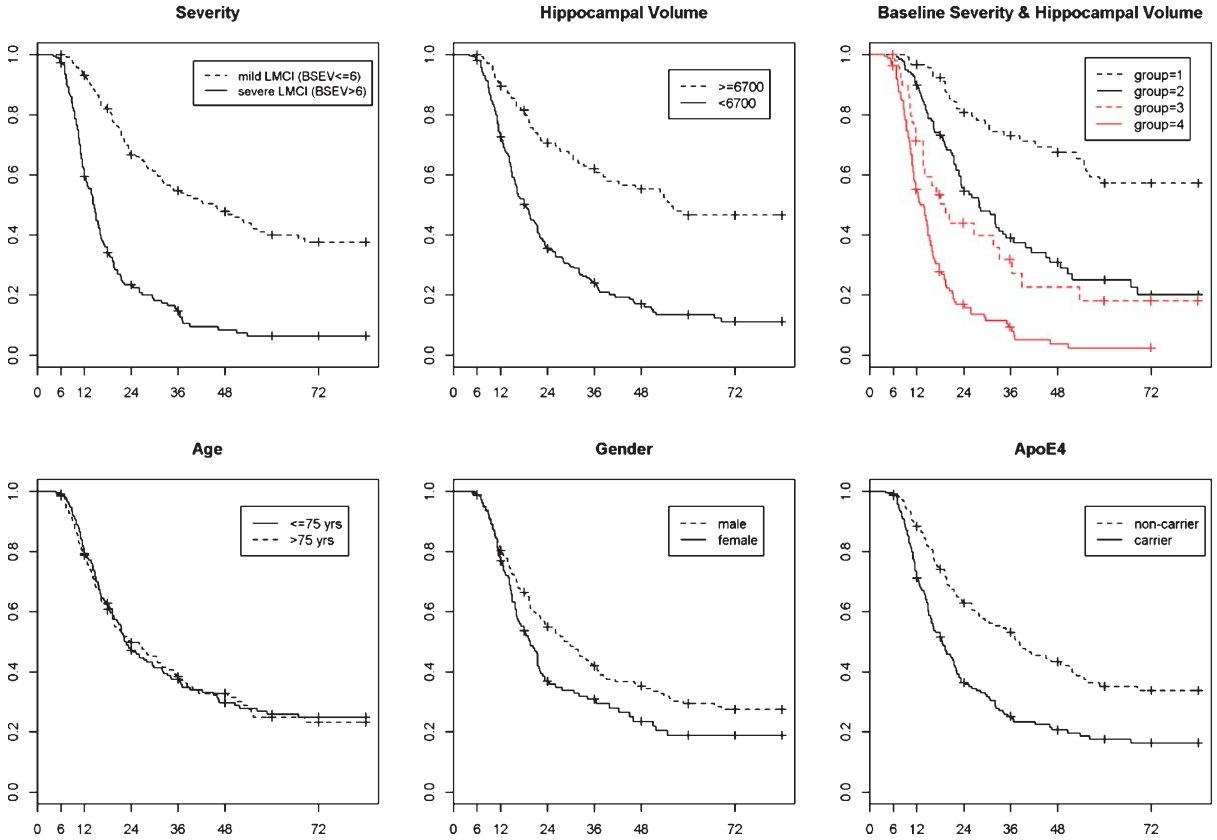


Fig. 5. Kaplan–Meier plots for 1 point change from baseline in CDR-SB by covariates. Covariates are dichotomized at approximate median values in the late MCI population in ADNI database. Baseline severity is defined as bFAQ + (30-bMMSE). Since baseline severity and hippocampal volume are highly correlated, interaction plot was shown in top-right panel. Group = 1 (mildest): milder in baseline severity (BSEV $\leq$6) and higher hippocampal volume ($\geq$6700 mm^3), group = 2 (milder): milder in baseline severity and lower hippocampal volume (<6700 mm^3), group = 3 (more severe): more severe in baseline severity (BSEV>6) and higher hippocampal volume, group = 4 (most severe): more severe in baseline severity and lower hippocampal volume. The x-axis in month.

36-month MCI trials. 80% power would be achieved with approximately $n = 350$ (per arm) for 24 month trial and $n = 200$ (per arm) for 36 month trial if the disease modifying effect were 30%. The results indicate that a 20% (or less) disease modifying drug has less power, and requires a greater sample size (e.g., $n = 800$ for 24-month trials with 20% disease modifying drug). Figure 6B shows the power estimates versus sample size with 24 month clinical study, with and without enrichment strategy (ApoE4 carrier or lower hippocampal volume (<6700 mm^3) patient groups versus all late MCI patients). The required sample size is reduced approximately 25% (i.e., $n = 350$ to $n = 250$ to achieve 80% power with disease modifying effect 30%, and $n = 700$ to $n = 550$ with disease modifying effect 20%). Figure 6C summarizes the predicted effect size (mean change from placebo) for different hypothetical disease modifying effects, and the disease modifying effect requires 35% and 20% for 24 and 36 month trials, respectively, to achieve a mean effect size of 0.5 points (difference from placebo) in CDR-SB.

DISCUSSION

A disease progression model was developed to describe longitudinal CDR-SB scores in MCI patient from the ADNI study. The novel approach of this analysis is the application of the mathematical model to predict the underlying trend in the CDR-SB scores for each individual patient and calculate the time to "clinically worsening" based on certain specified thresholds. In general, the use of a time-to-event survival analysis approach (e.g., time to a diagnosis of dementia) is a particularly appealing primary efficacy measure in clinical trials in early AD [9], but it would be challenging to conduct such studies considering the cost and length of the clinical trials.

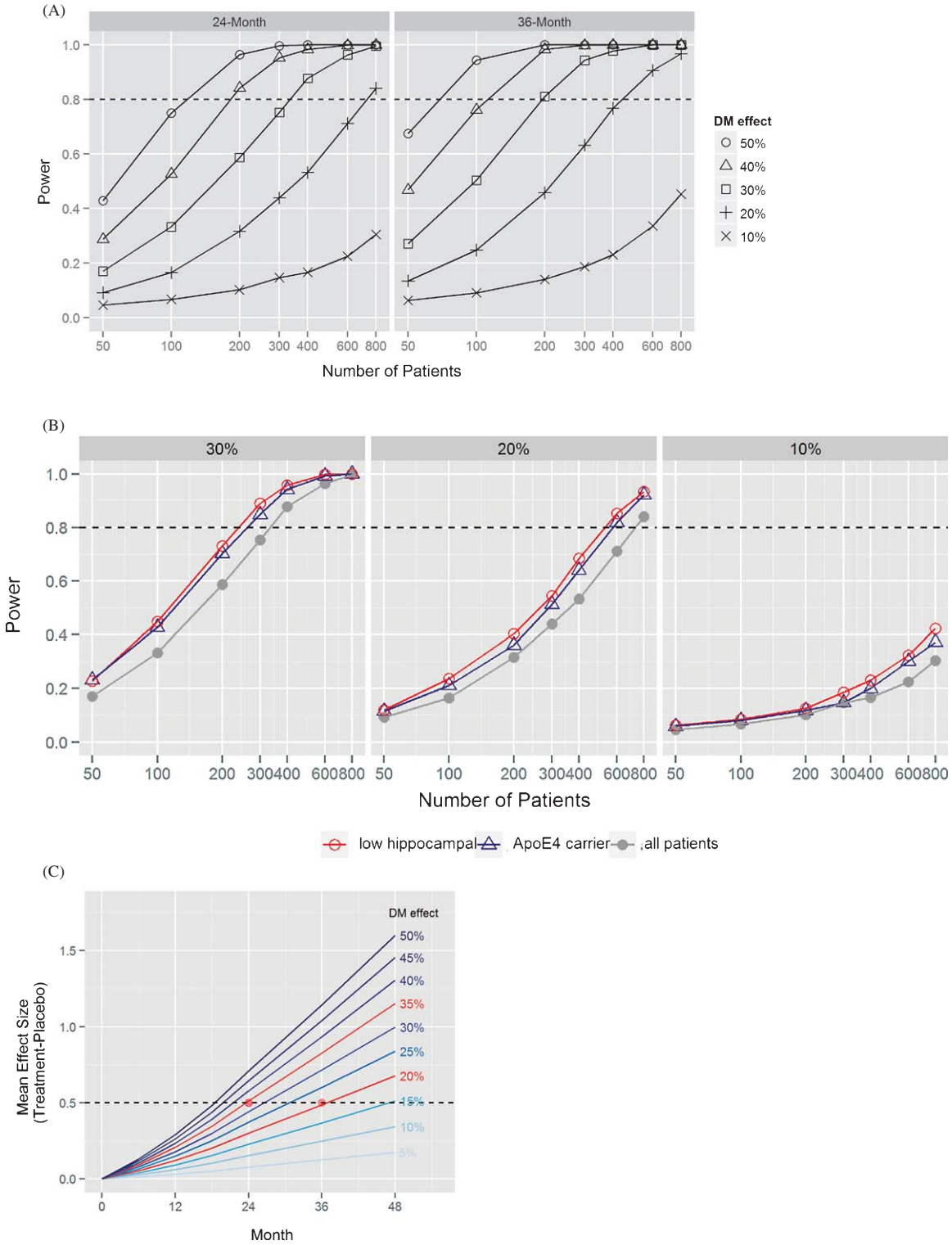


Fig. 6. Power and sample size to detect significant difference in CDR-SB in late MCI trials. A) 24 and 36 month clinical trials for hypothetical disease modifying drugs. B) 24 month trial for hypothetical disease modifying drugs (30, 20, 10%) with enrichment strategy. C) Predicted effect size (mean change from placebo) over time for hypothetical disease modifying drugs.

Ultimately, this analysis attempts to answer "how long the MCI clinical trial would be necessary using CDR-SB as a clinical endpoint" by testing different threshold as a "clinical worsening", where CDR-SB is an example of a suitable tool to assess both cognition and function as a single primary efficacy outcome measure [9]. The results indicate that the underlying disease progression takes approximately 1 year for 50% of the population to experience "worsening", defined by a 0.5 point change from baseline, and it takes 2 years for a 1 point change and 4 years for 2 point change from baseline (Fig. 4). If we consider in the time needed for 20% of the population to become worse instead of 50% of the population, the length of the clinical trial becomes shorter, i.e., approximately half year for 0.5 point change and 1 year for 1 point change.

The length of a clinical study is also dependent on the effect size and sample size. The hypothetical disease modifying effects were simulated from the model, and the results indicate that at least 2-year trial would be necessary with 30% (or more) disease modifying drug with a reasonable sample size ($n = 350$ per arm) to detect the significant difference from placebo (80% power) and to achieve the target mean effect size (treatment-placebo) of 0.5 point change in CDR-SB (Fig. 6).

Baseline severity, hippocampal volume, and ApoE4 were significant covariates and have a large impact on disease progression; the median time to worsening (1 point change from baseline) was dramatically shortened (progress faster) with severe MCI patients compared to milder patients (14.5 versus 45.0 months), lower hippocampal volume versus higher (18.5 versus 55.2 months), and ApoE4 carrier versus non-carrier (18.5 versus 37.2 months) (Fig. 5, Table 1). This evaluation would be expected to improve the patient selection by predicting which patients would likely progress to AD dementia within a time range appropriate for a therapeutic clinical trial.

The power and sample size calculations for lower hippocampal volume and ApoE4 carrier indicate that the sample size could be reduced approximately 25% with these groups when 80% power is maintained (Fig. 6B). This enrichment strategy will enable enrollment of subjects who are more likely to benefit from a treatment, and thus will enable a trial sponsor to reduce subject numbers and increase power.

In particular, the findings of lower hippocampal volume patients with faster disease progression in this analysis align with the newly proposed diagnosis of AD by Dubois et al. [24, 25], in which the use of biomarkers is emphasized and one or more abnormal biomarker is considered among structural neuroimaging with MRI, molecular neuroimaging with PET, and cerebrospinal fluid analysis of Aβ or tau proteins for diagnosis of MCI. However, no consensus on the quantitative thresholds for these biomarkers has been reached to define prodromal AD due to the variability between assay/measurement methods. Harmonization and validation of these tools are currently needed in AD research. To address these challenges, the C-Path Institute (http://c-path.org/programs/camd/camd-overview/) has formed two biomarker working groups: imaging biomarker and cerebrospinal fluid biomarker, including contributors from regulatory agencies, academia, and pharmaceutical companies. These working groups were established to facilitate the discussion about, harmonization of, and validation of the biomarkers being used to define AD pathology, and to establish these biomarkers as qualified tools to be used in drug development. In our analysis, we used an approximate median hippocampal volume ($6700\,mm^3$) as a cut-off to select the lower hippocampal volume population. Therefore, we need to predict the time of achieving a certain point change and calculate the power and sample size for prodromal AD trials once the biomarker threshold(s) for prodromal AD is defined and agreed among researchers. In addition, we intend to re-visit this analysis and conduct further evaluate the validation of the model when more data from ADNI-2 becomes available.

In conclusion, the approach to model the longitudinal CDR-SB first, then to calculate the median time to threshold for clinically "worsening", was able to estimate the duration of the clinical studies in MCI population with different scenarios, and possible enrichment strategy. For future research, a drop out model will be incorporated to simulate realistic clinical trial outcomes, and to identify designs and inclusion/exclusion criteria that may lead to more sensitive trials.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Raymond Miller (Daiichi Sankyo Pharma Development) for his invaluable input and support in the preparation of this article. The authors would also like to acknowledge members for the AD disease development team at Pfizer (S. Duvvuri, J. Werth, M. Bednar, T. Fullerton, and many others) who have provided valuable input.

Authors' disclosures available online (http://www.j-alz.com/disclosures/view.php?id=2088).

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Alzheimer's Association; Alzheimer's Drug Discovery Foundation; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; GE Healthcare; Innogenetics, N.V.; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of California, Los Angeles. This research was also supported by NIH grants P30 AG010129 and K01 AG030514.

SUPPLEMENTARY MATERIAL

Supplementary table and figures are available in the electronic version of this article: http://dx.doi.org/10.3233/JAD-132090.

REFERENCES

[1] Black R, Greenberg B, Ryan JM, Posner H, Seeburger J, Amatniek J, Resnick M, Mohs R, Miller DS, Saumier D, Carrillo MC, Stern Y (2009) Scales as outcome measures for Alzheimer's disease. *Alzheimers Dement* **5**, 324-339.

[2] ADNI Private Partner Scientific Board (PPSB) meeting (2013), http://www.alz.org/research/funding/partnerships/2013_meeting/25-ppsb-perspective.pdf, Accessed on December 22, 2013.

[3] Raghavan N, Samtani MN, Farnum M, Yang E, Lobanov V, Novak G, Narayan V, DiBernardo A (2011) Optimizing the ADAS-Cog for MCI and early AD. *FDA-Industry Statistics Workshop*, September 19–21, 2011, Washington, DC.

[4] Raghavan N, Samtani MN, Farnum M, Yang E, Lobanov V, Novak G, Narayan V, DiBernardo A (2011) Optimizing the ADAS-Cog for MCI and early AD, *Clinical Trials on Alzheimer's Disease*, November 3–5, 2011, San Diego, CA.

[5] Hendrix S, Logovinsky V, Perdomo C, Wang J, Satlin A (2012) Introducing a new tool for optimizing responsiveness to decline in early AD, *Alzheimer's Association International Conference*, July 14–19, 2012, Vancouver, Canada.

[6] Hendrix S, Logovinsky V, Perdomo C, Wang J, Satlin A (2012) A new tool for optimizing responsiveness to decline in early AD, *Clinical Trials on Alzheimer's Disease*, October 29–31, 2012, Monte Carlo, Monaco.

[7] Raghavan N, Samtani MN, Farnum M, Yang E, Novak G, Grundman M, Narayan V, Dibernardo, Alzheimer's Disease Neuroimaging, Initiative (2013) The ADAS-Cog revisited: Novel composite scales based on ADAS-Cog to improve efficiency in MCI and early AD trials. *Alzheimers Dement* **9** (1 Suppl), S21-S31.

[8] Logovinsky V, Hendrix, S, Perdomo C, Wang J, Satlin A (2013) New composite score demonstrates sensitivity to disease progression and treatment effects, *Alzheimer's and Parkinson's Disease Congress*, March 6–10, 2013, Florence, Italy.

[9] FDA Guidance for Industry: Alzheimer's Disease: Developing Drugs for the Treatment of Early Stage Disease, February 2013, http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM338287.pdf.

[10] Hughes CP, Berg L, Danziger WL, Coben LA, Martin RL (1982) A new clinical scale for the staging of dementia. *Br J Psychiatry* **140**, 566-572.

[11] O'Bryant SE, Lacritz LH, Hall J, Waring SC, Chan W, Khodr ZG, Massman PJ, Hobson V, Cullum CM (2010) Validation of the new interpretive guidelines for the clinical dementia rating scale sum of boxes score in the national Alzheimer's coordinating center database. *Arch Neurol* **67**, 746-749.

[12] Cedarbaum JM, Jaros M, Hernandez C, Coley N, Andrieu S, Grundman M, Vellas B, the Alzheimer's Disease Neuroimaging, Initiative (2013) Rationale for use of the Clinical Dementia Rating Sum of Boxes as a primary outcome measure for Alzheimer's disease clinical trials. *Alzheimers Dement* **9** (1 Suppl), S45-S55.

[13] Manly JJ, Tang MX, Schupf N, Stern Y, Vonsattel JP, Mayeux R (2008) Frequency and course of mild cognitive impairment in a multiethnic community. *Ann Neurol* **63**, 494-506.

[14] Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E (1999) Mild cognitive impairment: Clinical characterization and outcome. *Arch Neurol* **56**, 303-308.

[15] Aisen PS, Andrieu S, Sampaio C, Carrillo M, Khachaturian ZS, Dubois B, Feldman HH, Petersen RC, Siemers E, Doody RS, Hendrix SB, Grundman M, Schneider LS, Schindler RJ, Salmon E, Potter WZ, Thomas RG, Salmon D, Donohue M, Bednar MM, Touchon J, Vellas B (2011) Report of the task force on designing clinical trials in early (predementia) AD. *Neurology* **76**, 280-286.

[16] Ito K, Hutmacher MM, Corrigan WB (2012) Modeling of functional assessment questionnaire (FAQ) as continuous bounded data from the ADNI database. *J Pharmacokinet Pharmacodyn* **39**, 601-618.

[17] Hutmacher MM, French JL, Krishnaswami S, Menon S (2011) Estimating transformations for repeated measures modeling of continuous bounded outcome data. *Stat Med* **30**, 935-949.

[18] Ito K, Ahadieh S, Corrigan B, French J, Fullerton T, Tensfeldt T, Alzheimer's Disease Working, Group (2010) A disease

progression meta-analysis model in Alzheimer's disease. *Alzheimers Dement* **6**, 39-53.

[19] Ito K, Corrigan B, Zhao Q, French J, Miller R et al. (2011) Disease progression model for cognitive deterioration from ADNI database. *Alzheimers Dement* **7**, 151-160.

[20] William-Faltaos D, Chen Y, Wang Y, Gobburu J, Zhu H (2013) Quantification of disease progression and drop-out for Alzheimer's disease. *Int J Clin Pharmacol Ther* **51**, 120-131.

[21] Samtani MN, Farnum M, Lobanov V, Yang E, Raghavan N, DiBernardo A, Narayan V, Alzheimer's Disease Neuroimaging initiative (2012) An improved model for disease progression in patients from the Alzheimer's Disease Neuroimaging Initiative. *J Clin Pharmacol* **52**, 629-644.

[22] Rogers JA, Polhamus D, Gillespie WR, Ito K, Romero K, Qiu R, Stephenson D, Gastonguay MR, Corrigan B (2012) Combining patient-level and summary-level data for Alzheimer's disease modeling and simulation: A β regression meta-analysis. *J Pharmacokinet Pharmacodyn* **39**, 479-498.

[23] Wählby U, Jonsson EN, Karlsson MO (2002) Comparison of stepwise covariate model building strategies in population pharmacokinetic-pharmacodynamic analysis. *AAPS PharmSci* **4**, 68-79.

[24] Dubois B, Feldman HH, Jacova C, Dekosky ST, Barberger-Gateau P, Cummings J, Delacourte A, Galasko D, Gauthier S, Jicha G, Meguro K, O'brien J, Pasquier F, Robert P, Rossor M, Salloway S, Stern Y, Visser PJ, Scheltens P (2007) Research criteria for the diagnosis of Alzheimer's disease: Revising the NINCDS-ADRDA criteria. *Lancet Neurol* **6**, 734-746.

[25] Dubois B, Feldman HH, Jacova C, Cummings JL, Dekosky ST, Barberger-Gateau P, Delacourte A, Frisoni G, Fox NC, Galasko D, Gauthier S, Hampel H, Jicha GA, Meguro K, O'Brien J, Pasquier F, Robert P, Rossor M, Salloway S, Sarazin M, de Souza LC, Stern Y, Visser PJ, Scheltens P (2010) Revising the definition of Alzheimer's disease: A new lexicon. *Lancet Neurol* **9**, 1118-1127.