Research Report

ADSS: A Composite Score to Detect Disease Progression in Alzheimer's Disease

Guogen Shan[a,*], Xinlin Lu[a], Zhigang Li[a], Jessica Z.K. Caldwell[b],
Charles Bernick[b] and Jeffrey Cummings[c]
[a]*Department of Biostatistics, University of Florida, Gainesville, FL, USA*
[b]*Cleveland Clinic Lou Ruvo Center for Brain Health, Las Vegas, NV, USA*
[c]*Department of Brain Health, School of Integrated Health Sciences, Chambers-Grundy Center for Transformative
Neuroscience, University of Nevada Las Vegas (UNLV) Las Vegas, NV, USA*

Abstract.
Background: Composite scores have been increasingly used in trials for Alzheimer's disease (AD) to detect disease
progression, such as the AD Composite Score (ADCOMS) in the lecanemab trial.
Objective: To develop a new composite score to improve the prediction of outcome change.
Methods: We proposed to develop a new composite score based on the statistical model in the ADCOMS, by removing
duplicated sub-scales and adding the model selection in the partial least squares (PLS) regression.
Results: The new **AD** composite **S**core with variable **S**election (ADSS) includes 7 cognitive sub-scales. ADSS can increase
the sensitivity to detect disease progression as compared to the existing total scores, which leads to smaller sample sizes
using the ADSS in trial designs.
Conclusions: ADSS can be utilized in AD trials to improve the success rate of drug development with a high sensitivity to
detect disease progression in early stages.

Keywords: ADCOMS, Alzheimer's disease, cognitive outcome, composite score, mild cognitive impairment, model selection

INTRODUCTION

In Alzheimer's disease (AD) trials, the primary
endpoint is often chosen as change from baseline on
cognitive measures in the treatment group compared
to the placebo group at the end of the double-blind
period [1, 2]. Cognitive outcome measures are used to
assess the severity of dementia and disease progres-
sion for mild cognitive impairment (MCI) and AD
patients. Frequently used cognitive outcome mea-
sures include the Clinical Dementia Rating – Sum
of Boxes (CDR-SB), Alzheimer's Disease Assess-
ment Scale – Cognitive Subscale (ADAS-Cog), and
Mini-Mental State Examination (MMSE). In addition
to these individual cognitive and global assessment
measures, several composite scores were developed
to increase the sensitivity to detect score change
from baseline, such as the AD Composite Score

---

*Correspondence to: Guogen Shan, Department of Biostatis-
tics, University of Florida, Gainesville, FL 32610, USA. Tel.: +1
352 294 5934; E-mail: gshan@ufl.edu.

(ADCOMS), and the Integrated Alzheimer's Disease Rating Scale (iADRS) [3–5]. Such composite scores were calculated from the sub-scales or the total scores of instruments used in the trials. The ADCOMS was one of the secondary outcomes in the most recent FDA approved AD drug, lecanamab [6, 7]. The primary outcome of that phase 3 trial was the CDR-SB. With a large sample size in phase 3 trials, the pre-specified clinically meaningful change in the CDR-SB can be detected. In early phase trials (e.g., proof of concept trials), sample size is often small and an outcome measure that is sensitive to detect outcome change is needed to identify potentially effective treatments.

For a composite score developed from a repeated-measure study, linear mixed-effect models may be used to develop optimal composite scores [8]. Unfortunately, composite scores based on mixed-effect models may decrease the statistical power [8] for the following two reasons: 1) multicollinearity among sub-scales; and 2) positive or negative association between cognitive measures and disease progression. Several statistical models were developed to overcome these challenges, including the nonlinear mixed model with constraints on the range of the parameters, and the partial least squares (PLS) regression.

Variable selection methods are commonly used in data analysis to reduce the number of predictors in the final model. The three classical methods are forward selection, backward elimination, and stepwise selection. Traditionally, these methods depend on one criterion, such as the p-value of a model. However, in the variable selection procedure for the composite score development, two criteria are considered: the relationship between cognitive outcome measures and time since baseline, and the variable importance (VIP). In practice, we would expect a monotonic relationship between time since baseline and cognitive outcome measures. Variable importance is referred to be as the importance index of a variable in the model prediction. We built the **AD** composite **S**core with variable **S**election (ADSS) to account for many of these issues based on the statistical model in the ADCOMS [3].

Sample size calculation is a key aspect of AD trial planning. Reducing the sample size of a trial will decrease the time required for recruitment and accelerate the ability to assess putative therapies in trials. The subtraction method is traditionally used to calculate sample size based on the two-sample z test with the score change as the outcome [9]. This method has a closed formula which is relatively easy to implement [10–12], but it has several limitations [13]. This method could under- or over-estimate sample size when a new study's follow-up is longer or shorter than the pilot study's follow-up time [13]. In addition, the correlation between the cognitive outcome measure at the follow-up and that at baseline could be utilized to improve designing AD trials [14, 15].

METHODS

Data sets

Two data sets were used to develop and validate new composite scores to detect clinical progression in early stages of AD. First, in the development of the new ADSS, we included data of MCI patients from the ADNI database (downloaded on the date of November 29, 2022) [16, 17]. The total number of MCI patients was 1096 with $n = 396$ from ADNI-1, $n = 129$ from ADNI-Go, $n = 342$ from ADNI-2, and $n = 229$ from ADNI-3. Among them, there were over 800 participants having visits at the 6-month follow-up, 788 participants having year 1 visit, and 676 participants completed year 2 visit. Among 676 participants with year 2 follow-up, 624 participants had complete sub-scales for the outcomes of our interest, and they are MCI patients in building the model. Out of these 624 MCI patients, 173 patients progressed to dementia due to AD at the year 2 follow-up with the estimated conversion rate of 27.7%.

Second, in validating the ADSS, we utilized data of the MCI participants from the Alzheimer's Disease Cooperative Study (ADCS) to evaluate the efficacy of vitamin E and donepezil for the treatment of MCI in a randomized placebo-controlled three-arm study. That study enrolled a total of 769 MCI patients.

Baseline characteristics of these two cohorts are presented in Table 1. The proportion of female in the ADCS study is slightly higher than that in the ADNI MCI cohort. Other than that, these two studies have similar study populations.

Models for longitudinal data

Wang et al. [3] developed a statistical model to detect clinical decline with the outcome as the time from baseline. Let t be the time at the follow-up and t_0 be the time at baseline. For a study with K outcome measures (e.g., CDR-SB, ADAS-cog, their sub-scales), their proposed model is presented as

$$t - t_0 = \sum_{i=1}^{k} w_i \Delta x_i(t), \qquad (1)$$

Table 1
Baseline characteristics of the ADNI MCI cohort and the ADCS study

Measures	ADNI MCI ($n = 624$)	ADCS ($n = 769$)
Age, mean (SD)	73.04 (7.42)	72.98 (7.29)
Female, n (%)	254 (40.7%)	352 (45.77%)
CDR-SB, mean (SD)	1.51 (0.88)	1.82 (0.79)
ADAS-Cog-13, mean (SD)	16.50 (6.61)	17.69 (6.11)
MMSE, mean (SD)	27.63 (1.76)	27.27 (1.85)

ADNI, Alzheimer's Disease Neuroimaging Initiative; MCI, mild cognitive impairment; ADCS, Alzheimer's Disease Cooperative Study; CDR-SB, Clinical Dementia Rating – Sum of Boxes; ADAS-Cog, Alzheimer's Disease Assessment Scale – Cognitive Subscale; MMSE, Mini-Mental State Examination.

where $\Delta X_i(t)$ is the score change of the i-th cognitive score at time t from baseline (t_0), and w_i is its associated weight. The model was derived by using the linear item of the score change from the Taylor series [3]. In the ADNI data used to construct the ADSS, we used the exam date of the MMSE test to calculate the month t and t_0. The scheduled visit often does not occur exactly as scheduled. In the ADNI MCI patients, the average time for the 1-year follow-up was estimated as 13.73 months with the standard deviation of 1.5 months. For that reason, we used the exact months from baseline in this model to calculate $t - t_0$ [18–21]. For example, one participant has the visit time of $t = 13.2$ months. Then, $\Delta X_1(t = 13.2)$ is the difference of CDR-SB (the first cognitive test) at $t = 13.2$ and CDR-SB at baseline.

Outcome measures

We included the commonly used clinical trial total score measures: (I) CDR-SB, (II) the 13-item ADAS-cog, and (III) MMSE. Their sub-scales were included in the model. The CDR-SB has 6 domains: 1) memory, 2) orientation, 3) judgment and problem solving, 4) community, 5) home and hobbies, and 6) personal care. The ADAS-cog-13 include 13 domain areas: word recall, commands, constructional Praxis, delayed recall, naming, ideational praxis, orientation, word recognition, recall instructions, spoken language, word finding, comprehension, and number cancellation. The MMSE has 7 domain areas: orientation to time, orientation to place, language, attention and calculation, registration, recall, and constructional praxis [3, 22].

We also investigated the additional prediction power gain by adding other measures, such as Functional Activities Questionnaire (FAQ) and Neuropsychiatric Inventory (NPI) [23]. The FAQ includes 10 domain areas: manage finances, complete forms, shop, perform games of skill or hobbies, prepare hot beverages, prepare a balanced meal, follow current events, attend to television programs, books or magazines, remember appointments, and travel out of the neighborhood. Each domain has the score from 0 to 5, with the maximum total score of 50. The FAQ scores were used in predicting the disease progression [24]. NPI consists of 12 sub-scales to assess neuro-psychiatric symptoms [23].

Partial least square regression and model selection

The statistical model in Equation (1) was used in the PLS regression to derive the ADSS using the ADNI MCI data. PLS regression is used to find independent components from the existing variables. The first component was used as the composite score in this research. This component represents the eigenvector (e.g., the weights) for the largest eigenvalue of the covariance matrix between time ($t - t_0$) and cognitive scores. It is possible that the weight for each cognitive score may be negative although a positive correlation coefficient between time and each cognitive score is assumed. The backward variable selection approach was utilized to remove cognitive measures whose parameter estimates are negative in the fitted statistical model. The following steps were used to derive the final composite score based on the PLS and the backward variable selection approach.

- Step 1: Select the clinical scores that have a positive correlation with time $t - t_0$. These X_is are the initial set of the outcome measures which is a subset of $\{X_i : i = 1, 2, \ldots, K\}$.
- Step 2: Run a PLS regression with the identified set of outcome measures from Step 1.
- Step 3: The backward selection approach was used in the PLS regression model to delete outcome measures with negative parameters.
- Step 4: Among the remaining outcome measures, the ones with low VIP values were

then removed by using the backward selection approach. During this step, any cognitive outcome measure with a negative parameter estimate was removed first by using the method in Step 3.

This can be viewed as an iterative approach as Step 3 could be used multiple times. The backward selection approach was used in removing cognitive outcome measures with negative parameter estimates or low VIP values.

To avoid including similar orientation sub-scales from the three cognitive measures, we run this model selection approach three times by including the orientation sub-scale from one of the three measures in Step 1.

Comparison with other measures

We first compared the sensitivity of the developed ADSS with the existing measures by using the mean to standard deviation ratio (MSDR). We calculated the bias-corrected MSDR by using 10,000 bootstrap samples. The existing cognitive scores include three widely used total score outcome measures (CDR-SB, ADAS-Cog, and MMSE), and the composite score ADCOMS [3]). A larger MSDR value represents a larger effect size, which leads to greater sensitivity. In addition to the MCI population, we also compared the MSDR between the new composite scores and the existing scores in some enriched population (e.g., Apolipoprotein E ε4 (*APOE4*) carries and Amyloid-β positive patients assessed with positron emission tomography (PET)).

We then compared the sample sizes required for a parallel randomization study by using the ADNI MCI patients as the control group. The sample size for a randomized study was calculated by assuming a 25% slowing of cognitive decline as the observed benefit in a new treatment group as compared to the control group. In addition to the traditional subtraction method for sample size calculation, we considered the approach based on analysis of covariance (ANCOVA) where the baseline measure is the only covariate in the sample size calculation [15]. After analysis of the ADNI MCI data, we used the SMC cohort from the ADNI to validate the newly derived ADSS.

In a two-arm randomized trial, suppose μ_{G0} and μ_{G1} are the mean of the outcome measure at baseline and that at the follow-up for the group G, $G = c$ for the control group and $G = t$ for the treatment group. Then, $\mu_t = \mu_{t1} - \mu_{t0}$ (color blue in Fig. 1) and $\mu_c = \mu_{c1} -$
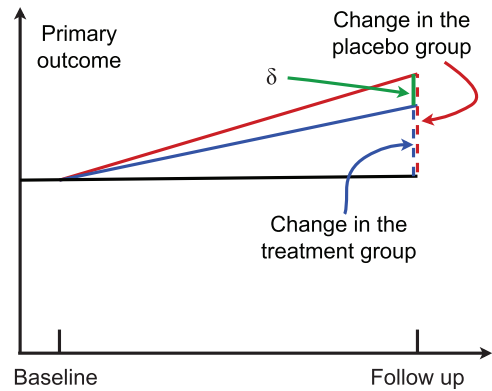

Fig. 1. A randomized two-arm study. Higher scores indicate worse cognitive performance.
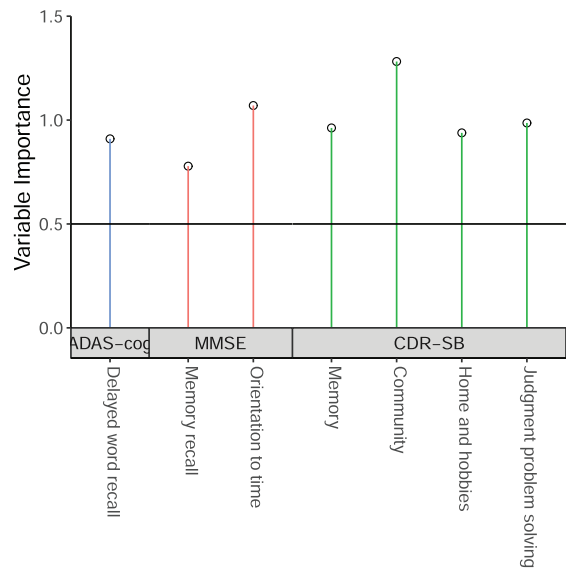

Fig. 2. The selected cognitive outcome measures in the new ADSS composite score.

μ_{c0} (color red in Fig. 1) are the change of the outcome at the follow-up from baseline in the treatment group and the control group, respectively. Their difference $\delta = \mu_c - \mu_t$ is the parameter of interest (green line in Fig. 1). The statistical hypothesis is:

$$H_0 : \delta = 0 \quad VS \quad H_a : \delta \neq 0.$$

The traditional subtraction method calculates the sample size by using the estimated value of $\hat{\delta}$ and its variance estimate [9, 15]. Suppose S_0^2 and S_1^2 are the variance estimate of outcome at baseline and that at the follow-up. Then, the variance estimate of the score change (S_d^2) is calculated as $S_d^2 = S_0^2 + S_1^2 - 2\rho S_0 S_1$,

Table 2
Weights for the new ADSS

Cognitive measure	Sub-scales	Weights
ADAS-Cog	Delayed word recall	0.0601
ADAS-Cog	Word finding	0.2037
MMSE	Orientation to time	0.1134
CDR-SB	Personal care	0.2542
CDR-SB	Community	0.4787
CDR-SB	Home and hobbies	0.2510
CDR-SB	Judgment and problem solving	0.2853

ADSS, AD composite Score with variable Selection; ADAS-Cog, Alzheimer's Disease Assessment Scale – Cognitive Subscale; MMSE, Mini-Mental State Examination; CDR-SB, Clinical Dementia Rating – Sum of Boxes.

where ρ is the correlation coefficient of outcome at baseline and outcome at the follow-up. The traditional approach is the one based on the two-sample z test [25, 26]. Suppose δ_d is the clinically meaningful improvement of the treatment as compared to the control group (e.g., 25% benefit in the treatment as compared to the control group with $\mu_c - \mu_t = 0.25\mu_c$). The required sample size per arm for a two-arm study based on the subtraction method is

$$n_s = \frac{2(z_{\alpha/2} + z_\beta)s_d^2}{\delta_d^2} \qquad (2)$$

where z_c is the $100(1 - c)$th percentile of the standard normal distribution. The subtraction method is easy to implement with a closed formula, but it has several limitations including the assumption of population homogeneity and consequent under- or over-estimate of sample sizes [13]. When the baseline outcome measure affects the disease progression, the subtraction method may lose statistical power substantially [27]. When the baseline cognitive measure is considered as a covariate in the ANCOVA sample size calculation, the required sample size per arm is presented as [14, 15]

$$n_c = (n_s + 1)(1 - \rho^2), \qquad (3)$$

where n_s is the sample size using the subtraction method in Equation (2), and one additional participant is added to maintain the statistical power level [14]. The sample size n_c goes down as the correlation coefficient increases. In addition to n_c, another exact approach based on the ratio of mean squares could be utilized, but the exact approach does not have a closed formula. For a study with a large sample size, their results are very close to each other [15].

RESULTS

We first used the sub-scales from CDR-SB, MMSE, ADAS-cog-13 to derive the ADSS by using the method described above, by using the ADNI MCI patients having baseline, 6 month, 1-year, and 2-year follow-up data. The threshold value of VIP could affect the model performance. For that reason, we run the model with VIP threshold value from 0.3 to 0.8 by 0.05, and found that the model with the threshold value of 0.75 with MMSE orientation sub-scales has a high MSDR. That final model includes 7 sub-scales: 4 sub-scales from CDR-SB (memory, community, home and hobbies, and judgment and problem solving), 1 sub-scale from ADAS-cog (delayed word recall), and 2 sub-scales from MMSE (orientation to time, and memory recall). Their weights are presented in Table 2. The VIP results are shown in Fig. 2.

We compared the MSDR of the new ADSS with the existing cognitive scores including the composite score ADCOMS in Table 3. In the MSDR calculation, the score change at 2-year follow-up from baseline was used to calculate the mean and the standard deviation of score change in 2 years. The estimated MSDR for the new composite score was 0.5726, and the bias-corrected MSDR estimated using the bootstrap approach was 0.5508. The range of MSDR is from 0.4026 to 0.5726 for these five scores. ADCOMS has similar MSDR as the proposed new composite score (0.5535 versus 0.5726). Orientation sub-scales from the three cognitive tests were all included in the ADCOMS. The duplicated effects from orientation sub-scales may increase the difficulty in interpretation. The proposed ADSS has a larger MSDR than the commonly used total score CDR-SB (0.4556). The MSDR ratio between the new ADSS and the existing scores is at least 1.03 as compared to the commonly used scores including ADCOMS. In the subgroups of *APOE4* carries or amyloid-β positive patients, the

Table 3
Sensitivity comparison between the new ADSS with the existing scores by using the MSDR
in MCI group and two enriched subgroups

Measures	Mean	SD	MSDR		
	MCI		MCI	MCI subgroups	
				APOE4 carriers	Amyloid-β positive
CDR-SB	0.3942	1.1113	0.3548	0.4688	0.3403
ADAS-cog	0.6464	5.2423	0.1233	0.2015	0.0639
MMSE	0.5654	2.2682	0.2493	0.3580	0.3218
ADCOMS	0.0509	0.1191	0.4277	0.5071	0.4248
new ADSS	0.1534	0.3716	0.4129	0.4856	0.3979

ADSS, AD composite Score with variable Selection; MSDR, mean to standard deviation ratio; MCI, mild cognitive impairment; CDR-SB, Clinical Dementia Rating – Sum of Boxes; ADAS-Cog, Alzheimer's Disease Assessment Scale – Cognitive Subscale; MMSE, Mini-Mental State Examination; ADCOMS, AD Composite Score.

Table 4
Sample size comparison between the new ADSS with the existing scores by using the MSDR in MCI
group and two enriched subgroups, to detect a 25% benefit in a new treatment as compared to the
control group. Two methods were used for sample size calculation: the subtraction
method and the ANCOVA method

Measures	MCI		MCI subgroups			
			APOE4 carries		Amyloid-β positive	
	n_s: Sample size using the subtraction approach					
	n_s	Ratio	n_s	Ratio	n_s	Ratio
CDR-SB	1996	1.35	1143	1.07	2169	1.37
ADAS-cog	16521	11.21	6186	5.80	61512	38.76
MMSE	4042	2.74	1960	1.84	2426	1.53
ADCOMS	1374	0.93	977	0.92	1392	0.88
new ADSS	1474		1066		1587	
	n_c: ANCOVA Sample size					
CDR-SB	1210	1.54	677	1.30	1206	1.56
ADAS-cog	6041	7.70	2253	4.34	24695	31.95
MMSE	2897	3.69	1437	2.77	1760	2.28
ADCOMS	719	0.92	439	0.85	716	0.80
new ADSS	785		519		773	

ADSS, AD composite Score with variable Selection; MSDR, mean to standard deviation ratio; MCI, mild cognitive impairment; CDR-SB, Clinical Dementia Rating – Sum of Boxes; ADAS-Cog, Alzheimer's Disease Assessment Scale – Cognitive Subscale; MMSE, Mini-Mental State Examination; ADCOMS, AD Composite Score.

ADSS has a larger MSDR than the existing scores, and the difference is similar to that in the overall MCI population.

In Table 4, we presented sample sizes for a study to detect a 25% reduction in clinical decline at 2-year follow-up from baseline by using the existing 4 scores and the new ADSS. When a study was designed by using the subtraction method, the required sample size is a function of the MSDR. For that reason, the new composite score requires the smaller sample size as compared to the existing total scores as shown in Table 3 comparing MSDR values. The ratio of the required sample sizes between the existing cognitive measure and the new composite score is the square of the MSDR ratio between them. ADCOMS requires

similar sample sizes as the ADSS. The individual total scores require at least 59% more participants than the ADSS in MCI patients.

When the ANCOVA approach is used in sample size calculation with baseline cognitive score being adjusted in the outcome, the sample size formula in Equation (3) can be used to calculate the required trial sample sizes. As compared to the sample size n_s based on the subtraction method, the sample size n_c is often much smaller, especially when the correlation ρ is high. For these measures, the range of ρ is from 0.45 to 0.78. ADAS-cog has the highest correlation, followed by ADSS, and ADCOMS. The new score reduces sample sizes at least 77% as compared to CDR-SB.

Validation

The developed ADSS was validated by using MCI patients' data from the ADCS [28]. We identified 536 MCI participants having complete CDR-SB, ADAS-cog, MMSE, and their sub-scales at 18 months. At the 18-month follow-up visit, the donepezil treatment group ($n = 168$) had the lowest MCI-to-AD conversion rate of 14% as compared to the AD onset rate of 24% and 22% in the vitamin E arm ($n = 184$) and the placebo arm ($n = 184$), respectively. When combining data from all the three arms, the new ADSS has the highest MSDR of 0.405, followed by CDR-SB (0.383), MMSE (0.270), and ADAS-cog (0.231). In this example, the ADCOMS has a slightly higher MSDR than the ADSS. Within each arm, the new ADSS and the CDR-SB often have larger MSDR values than others. The ADSS has similar MSDR values as the CDR-SB in the donepezil arm and the vitamin E arm, but the ADSS has 15% higher MSDR than CDR-SB in the placebo arm.

We calculated the predictive validity of the ADSS by using the 624 MCI patients from the ADNI [29]. They can be separated into two subgroups: 173 patients progressed to dementia due to AD at the year 2 follow-up visit, and the remaining 451 patients who remained MCI status. In these 2 years, MCI patients who remained stable in that 2 years had very slight decline of the ADSS (mean change of 0.35 with SD of 1.41), while the ADSS was increased by 3.56 (SD = 2.07) in the 173 patients progressed to dementia due to AD. These results indicate high predictive validity of the proposed ADSS.

DISCUSSION

Our goal in developing the ADSS was to create a valid score that will allow sample size reductions for clinical trials. The slowest aspect of clinical trial conduct is recruitment of patients; the recruitment time usually exceeds the period of drug exposure in the trial [30]. Reducing sample size can decrease the recruitment time and accelerate assessment of candidate treatments in trials. To overcome the challenge of multicollinearity, we first calculated the Pearson correlation coefficients between each cognitive total score or sub-scale and the outcome. We included only the cognitive outcome measures with outcomes measures in the expected direction of correlation in the statistical model to derive a new composite score. Different cognitive tests may have similar domains (e.g., orientation) [10], leading to highly correlated

measures [31, 32]. The PLS regression was used in this project to compute independent components.

The threshold value of VIP in the model selection was chosen to be as 0.75. The model prediction could be affected by the VIP threshold value [33]. When the VIP threshold value is too low, many cognitive outcome measures will be selected, which could increase the complexity of model interpretation. Meanwhile, a high VIP threshold value reduces the number of predictors in the final model. In this case, the model prediction may be affected. The commonly used threshold value of 0.8 for VIP [34] did not perform as well as the lower threshold values. The threshold value of 0.75 was used to select enough cognitive outcome measures in the ADSS.

We investigated the improvement of the final composite score by adding additional predictors from the FAQ for measuring impairment in instrumental activities of daily living [24] and the NPI for assessing neuropsychiatric symptoms [23]. The estimated MSDR of change from baseline can be increased when the following 7 sub-scales from the FAQ were added to the final composite score: finances, paperwork, preparing a meal, events, travel, game, and shopping. But not all AD trials collect FAQ data which limits the usage of the model with FAQ sub-scales. For the NPI sub-scales, given the significant amount of missing data, we did not observe improvement in the prediction power.

We only utilized the natural history data (the ADNI study) in developing the model, which could be considered as a limitation as compared to the ADCOMS developed by using the ADNI and the control group data from AD trials. ADSS was developed to avoid including multiple highly correlated orientation sub-scales. As ADSS was developed by using the statistical model in the ADCOMS, we would expect them have similar prediction performance in some applications. In general, the ADSS has similar performance as the ADCOMS. It should be noted that the ADNI data is an observational dataset [35]. MCI patients in the ADNI study may have different disease progression rates as compared to the MCI patients who are assigned to the control group in AD trials.

In addition to the commonly used cognitive outcome measures (e.g., ADAS-cog), more sensitive outcome measures are in great need to detect cognitive change. One example would be the digital cognitive testing, which can be assessed remotely [36]. The PLS method provides the parameter estimates that can be used directly in computing the composite score. The statistical model is used to

select the sub-scales, which may not lead to better interpretation of clinical meaningfulness of the composite scores. Alternatively, machine learning methods that can provide parameter estimates may be considered in the future to further improve the prediction of disease progression. We consider utilizing machine learning methods to develop new composite scores as future work.

AUTHOR CONTRIBUTIONS

Guogen Shan (Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Project administration; Resources; Software; Supervision; Validation; Writing – original draft; Writing – review & editing); Xinlin Lu (Methodology; Software; Validation; Writing – original draft); Zhigang Li (Methodology; Writing – original draft); Jessica Z.K. Caldwell (Methodology; Resources; Writing – original draft); Charles Bernick (Conceptualization; Methodology; Writing – original draft); Jeffrey Cummings (Conceptualization; Data curation; Methodology; Writing – original draft).

ACKNOWLEDGMENTS

FUNDING

CONFLICT OF INTEREST

JC has provided consultation to Acadia, Actinogen, Alkahest, AlphaCognition, Aprinoia, AriBio, Biogen, BioVie, Cassava, Cerecin, Corium, Cortexyme, Diadem, EIP Pharma, Eisai, GemVax, Genentech, Green Valley, GAP Innovations, Grifols, Janssen, Karuna, Lilly, Lundbeck, LSP, Merck, NervGen, Novo Nordisk, Oligomerix, Optoceutics, Ono, Otsuka, PRODEO, Prothena, ReMYND, Resverlogix, Roche, Sage Therapeutics, Signant Health, Simcere, Sunbird Bio, Suven, SynapseBio, TrueBinding, and Vaxxinity pharmaceutical, assessment, and investment companies.

All other authors have nothing to disclose.

DATA AVAILABILITY

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

The vitamin E and donepezil trial was supported by NIH (U19AG010483), Pfizer and Eisai, and DSM nutritional products (vitamin E donation). The complete list of institutions and persons participated in the ADCS can be found their website.

REFERENCES

[1] Salloway S, Farlow M, McDade E, Clifford DB, Wang G, Llibre-Guerra JJ, Hitchcock JM, Mills SL, Santacruz AM, Aschenbrenner AJ, Hassenstab J, Benzinger TLS, Gordon BA, Fagan AM, Coalier KA, Cruchaga C, Goate AA, Perrin RJ, Xiong C, Li Y, Morris JC, Snider BJ, Mummery C, Surti GM, Hannequin D, Wallon D, Berman SB, Lah JJ, Jimenez-Velazquez IZ, Roberson ED, van Dyck CH, Honig LS, Sánchez-Valle R, Brooks WS, Gauthier S, Galasko DR, Masters CL, Brosch JR, Hsiung G-YR, Jayadev S, Formaglio M, Masellis M, Clarnette R, Pariente J, Dubois B, Pasquier F, Jack CR, Koeppe R, Snyder PJ, Aisen PS, Thomas RG, Berry SM, Wendelberger BA, Andersen SW, Holdridge KC, Mintun MA, Yaari R, Sims JR, Baudler M, Delmar P, Doody RS, Fontoura P, Giacobino C, Kerchner GA, Bateman RJ, Formaglio M, Mills SL, Pariente J, van Dyck CH, the Dominantly Inherited Alzheimer Network–Trials U (2021) A trial of gantenerumab or solanezumab in dominantly inherited Alzheimer's disease. *Nat Med* **27**, 1187-1196.

[2] Knopman DS, Jones DT, Greicius MD (2021) Failure to demonstrate efficacy of aducanumab: An analysis of the EMERGE and ENGAGE trials as reported by Biogen, December 2019. *Alzheimers Dement* **17**, 696-701.

[3] Wang J, Logovinsky V, Hendrix SB, Stanworth SH, Perdomo C, Xu L, Dhadda S, Do I, Rabe M, Luthman J, Cummings J, Satlin A (2016) ADCOMS: A composite clinical outcome for prodromal Alzheimer's disease trials. *J Neurol Neurosurg Psychiatry* **87**, 993-999.

[4] Wessels AM, Siemers ER, Yu P, Andersen SW, Holdridge KC, Sims JR, Sundell K, Stern Y, Rentz DM, Dubois B, Jones RW, Cummings J, Aisen PS (2015) A combined measure of cognition and function for clinical trials: The Integrated Alzheimer's Disease Rating Scale (iADRS). *J Prev Alzheimers Dis* **2**, 227-241.

[5] Jacobs DM, Thomas RG, Salmon DP, Jin S, Feldman HH, Cotman CW, Baker LD, Alzheimer's Disease Cooperative Study EXERT Study Group; Alzheimer's Disease Neuroimaging Initiative (2020) Development of a novel cognitive composite outcome to assess therapeutic effects of exercise in the EXERT trial for adults with MCI: The ADAS-Cog-Exec. *Alzheimers Dement* **6**, e12059.

[6] van Dyck CH, Swanson CJ, Aisen P, Bateman RJ, Chen C, Gee M, Kanekiyo M, Li D, Reyderman L, Cohen S, Froelich L, Katayama S, Sabbagh M, Vellas B, Watson D, Dhadda S, Irizarry M, Kramer LD, Iwatsubo T (2022) Lecanemab in early Alzheimer's disease. *N Engl J Med* **388**, 9-21.

[7] Swanson CJ, Zhang Y, Dhadda S, Wang J, Kaplow J, Lai RYK, Lannfelt L, Bradley H, Rabe M, Koyama A, Reyderman L, Berry DA, Berry S, Gordon R, Kramer LD, Cummings JL (2021) A randomized, double-blind, phase 2b proof-of-concept clinical trial in early Alzheimer's disease with lecanemab, an anti-Aβ protofibril antibody. *Alzheimers Res Ther* **13**, 80.

[8] Ard MC, Raghavan N, Edland SD (2015) Optimal composite scores for longitudinal clinical trials under the linear mixed effects model. *Pharm Stat* **14**, 418-426.

[9] Rosner B (2016) *Fundamentals of biostatistics*, 8th edition. Cengage Learning, Boston, MA.

[10] Schneider LS, Goldberg TE (2020) Composite cognitive and functional measures for early stage Alzheimer's disease trials. *Alzheimers Dement* **12**, e12017.

[11] Logovinsky V, Satlin A, Lai R, Swanson C, Kaplow J, Osswald G, Basun H, Lannfelt L (2016) Safety and tolerability of BAN2401 – a clinical study in Alzheimer's disease with a protofibril selective Aβ antibody. *Alzheimers Res Ther* **8**, 14.

[12] Cummings J, Lee G, Zhong K, Fonseca J, Taghva K (2021) Alzheimer's disease drug development pipeline: 2021. *Alzheimers Dement* **7**, e12179.

[13] Chen K, Guo X, Pan R, Xiong C, Harvey DJ, Chen Y, Yao L, Su Y, Reiman EM, Alzheimer's Disease Neuroimaging Initiative (2022) Limitations of clinical trial sample size estimate by subtraction of two measurements. *Stat Med* **41**, 1137-1147.

[14] Borm GF, Fransen J, Lemmens WA (2007) A simple sample size formula for analysis of covariance in randomized clinical trials. *J Clin Epidemiol* **60**, 1234-1238.

[15] Shan G, Ma C (2014) A comment on sample size calculation for analysis of covariance in parallel arm studies. *J Biom Biostat* **5**, 184.

[16] Weiner MW, Veitch DP, Aisen PS, Beckett LA, Cairns NJ, Cedarbaum J, Donohue MC, Green RC, Harvey D, Jack Jr. CR, Jagust W, Morris JC, Petersen RC, Saykin AJ, Shaw L, Thompson PM, Toga AW, Trojanowski JQ, Alzheimer's Disease Neuroimaging Initiative (2015) Impact of the Alzheimer's Disease Neuroimaging Initiative, 2004 to 2014. *Alzheimers Dement* **11**, 865-884.

[17] Jagust WJ, Bandy D, Chen K, Foster NL, Landau SM, Mathis CA, Price JC, Reiman EM, Skovronsky D, Koeppe RA, Alzheimer's Disease Neuroimaging Initiative (2010) The Alzheimer's Disease Neuroimaging Initiative positron emission tomography core. *Alzheimers Dement* **6**, 221-229.

[18] Shan G, Banks S, Miller JB, Ritter A, Bernick C, Lombardo J, Cummings JL (2018) Statistical advances in clinical trials and clinical research. *Alzheimers Dement* **4**, 366-371.

[19] Cummings JL, Fulkerson N (2018) Neurodegeneration research: Advances in clinical translational neuroscience infrastructure and methods. *Alzheimers Dement* **4**, 326-329.

[20] Caldwell JZK, Berg JL, Shan G, Cummings JL, Banks SJ (2018) Sex moderates the impact of diagnosis and amyloid PET positivity on hippocampal subfield volume. *J Alzheimers Dis* **64**, 79-89.

[21] Miller JB, Shan G, Lombardo J, Jimenez-Maggoria G (2018) Biomedical informatics applications for precision management of neurodegenerative diseases. *Alzheimers Dement* **4**, 357-365.

[22] Schultz-Larsen K, Kreiner S, Lomholt RK (2007) Mini-Mental Status Examination: Mixed Rasch model item analysis derived two different cognitive dimensions of the MMSE. *J Clin Epidemiol* **60**, 268-279.

[23] Cummings JL, Mega M, Gray K, Rosenberg-Thompson S, Carusi DA, Gornbein J (1994) The Neuropsychiatric Inventory. *Neurology* **44**, 2308-2308.

[24] Ito K, Hutmacher MM, Corrigan BW (2012) Modeling of Functional Assessment Questionnaire (FAQ) as continuous bounded data from the ADNI database. *J Pharmacokinet Pharmacodyn* **39**, 601-618.

[25] Shan G (2016) Exact confidence intervals for randomized response strategies. *J Appl Stat* **43**, 1279-1290.

[26] Shan G, Gerstenberger S (2017) Fisher's exact approach for post hoc analysis of a chi-squared test. *PLoS One* **12**, e0188709.

[27] Hernández AV, Steyerberg EW, Habbema JDF (2004) Covariate adjustment in randomized controlled trials with dichotomous outcomes increases statistical power and reduces sample size requirements. *J Clin Epidemiol* **57**, 454-460.

[28] Petersen RC, Thomas RG, Grundman M, Bennett D, Doody R, Ferris S, Galasko D, Jin S, Kaye J, Levey A, Pfeiffer E, Sano M, van Dyck CH, Thal LJ (2005) Vitamin E and donepezil for the treatment of mild cognitive impairment. *N Engl J Med* **352**, 2379-2388.

[29] Jutten RJ, Papp KV, Hendrix S, Ellison N, Langbaum JB, Donohue MC, Hassenstab J, Maruff P, Rentz DM, Harrison J, Cummings J, Scheltens P, Sikkes SAM (2023) Why a clinical trial is as good as its outcome measure: A framework for the selection and use of cognitive outcome measures for clinical trials of Alzheimer's disease. *Alzheimers Dement* **19**, 708-720.

[30] Cummings J, Lee G, Nahed P, Kambar MEZN, Zhong K, Fonseca J, Taghva K (2022) Alzheimer's disease drug development pipeline: 2022. *Alzheimers Dement* **8**, e12295.

[31] Hura Ahmad M, Adnan R, Adnan N (2006) A comparative study on some methods for handling multicollinearity problems. *MATEMATIKA* **22**, 109-119.

[32] Shan G, Wilding GE, Hutson AD (2017) Computationally intensive two-stage designs for clinical trials. In *Wiley StatsRef: Statistics Reference Online*, Balakrishnan N, Colton T, Everitt B, Piegorsch W, Ruggeri F, Teugels JL, eds. pp. 1-7.

[33] Chong I-G, Jun C-H (2005) Performance of some variable selection methods when multicollinearity is present. *Chemometrics Intellig Lab Syst* **78**, 103-112.

[34] Wold H (1966) Estimation of principal components and related models by iterative least squares. *Multivariate Anal-*

ysis, Krishnajah PR, ed. Academic Press, New York, pp. 391-420.

[35] Lu X, Shan G (2023) Two-stage response adaptive randomization designs for multi-arm trials with binary outcome. *J Biopharm Stat*, doi: 10.1080/10543406.2023.2234028.

[36] Öhman F, Hassenstab J, Berron D, Schöll M, Papp KV (2021) Current advances in digital cognitive assessment for preclinical Alzheimer's disease. *Alzheimers Dement (Amst)* **13**, e12217.