Received: 15 June 2020 | Accepted: 9 July 2020 | Published online: 4 September 2020

DOI: 10.1002/trc2.12071

Alzheimer's & Dementia
Translational Research & Clinical Interventions

RESEARCH ARTICLE

Enriching the design of Alzheimer's disease clinical trials: Application of the polygenic hazard score and composite outcome measures

Sarah J. Banks | Yuqi Qiu | Chun Chieh Fan | Anders M. Dale | Jingjing Zou | Brianna Askew | Howard H. Feldman | For the Alzheimer's Disease Neuroimaging Initiative

University of California San Diego, San Diego, California, USA

Correspondence
Sarah J Banks, University of California San Diego, 9500 Gilman Drive, m/c 0841, La Jolla CA 92093, USA.
E-mail: sbanks@health.ucsd.edu

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

Abstract

Introduction: Selecting individuals at high risk of Alzheimer's disease (AD) dementia and using the most sensitive outcome measures are important aspects of trial design.

Methods: We divided participants from Alzheimer's Disease Neuroimaging Initiative at the 50th percentile of the predicted absolute risk of the polygenic hazard score (PHS). Outcome measures were the Alzheimer's Disease Assessment Schedule-Cognitive Subscale (ADAS-Cog), ADNI-Mem, Clinical Dementia Rating-Sum of Boxes (CDR SB), and Cognitive Function Composite 2 (CFC2). In addition to modeling, we use a power analysis compare numbers needed with each technique.

Results: Data from 188 cognitively normal and 319 mild cognitively impaired (MCI) participants were analyzed. Using the ADAS-Cog to estimate sample sizes, without stratification over 24 months, would require 930 participants with MCI, while using the CFC2 and restricting participants to those in the upper 50th percentile would require only 284 participants.

Discussion: Combining stratification by PHS and selection of a sensitive combined outcome measure in a cohort of patients with MCI can allow trial design that is more efficient, potentially less burdensome on participants, and more cost effective.

KEYWORDS
clinical trial design, cognition, mild cognitive impairment, outcomes, polygenic hazard score

1 | INTRODUCTION

Pharmaceutical clinical trials in Alzheimer's disease (AD), particularly those that are directed at disease modification and prevention in those at risk or in the earliest stages of disease, are resource intensive, require long periods of intervention and follow-up, and have had a very high failure rate.[1,2] There are some encouraging results from several compounds currently in the experimental therapeutic pipeline[3] which will hopefully spur greater interest in clinical development at these stages. Trials may fail for drugs that lack efficacy but also for its participants not having the targeted disease pathology, or outcome measures being repeatedly selected despite awareness that they are insensitive

or tailored for the population under investigation.[4] To improve the likelihood of success developing high-quality therapeutics, further research attention to improving clinical trial design is warranted.

Older adults can have cognitive decline caused by both non-degenerative conditions including encephalopathies and vasculopathies, as well as with medical diseases and a spectrum of neurodegenerative diseases including AD, Parkinsonian dementias, frontotemporal degenerations, and limbic predominant age-related TDP 43 encephalopathy (LATE[5]). In clinical trials directed at the amyloidopathy of AD, 21.4% of all participants receiving bapineuzumab, a humanized monoclonal antibody directed at the n terminus of amyloid beta (Aβ), were classified on Pittsburgh compound B positron emission tomography (PIB PET) as being amyloid negative with 36.1 of non-carriers of the apolipoprotein E (*APOE*) ε4 allele being amyloid negative.[6] These findings within trials that include participants without underlying AD pathology[7] add significant noise to the analyses, potentially obscuring treatment effects and overall trial results. Enriching samples with those most likely to have underlying AD and thus the predicted trajectory of cognitive decline, with known underlying pathology, is an important research direction. Biomarkers including amyloid and tau pathology assessed via cerebrospinal fluid (CSF) or PET have gained popularity in clinical trial design; however, these tests are expensive and invasive, representing a large investment for both participants and investigators at the screening end of the study.[8]

Genetic tests for late onset AD have focused on the *APOE* ε4 allele, which has been used in trial ernichment.[3] More recently, polygenic risk and hazard scores have been developed, which take into account the smaller contributions of multiple single nucleotide polymorphisms (SNPs) to determine risk. In the case of hazard scores, age-specific absolute risk can be further derived based on the combination of baseline incidence proportion from epidemiological studies and the time-invariant hazard scores.[9,10] In the current study we use the Desikan polygenic hazard score[9] (PHS) to stratify a study sample to evaluate a low-cost, minimally invasive method to identify those most likely to decline during a typical trial period of 12 to 18 months. The PHS is derived from an analysis of a 31 SNP panel that has been demonstrated to reliably identify individuals at risk of AD dementia at any age, and provides a continuous measure of risk of AD dementia with higher scores indicating more risk at any given age.[9] The PHS extends beyond the risk conferred by the *APOE* ε4 genotype and enriches for both amyloid and tau pathology.[11]

Optimizing outcome measures in clinical trials involves identifying a sensitive measure to capture change over time in the population being measured, with the least variance, and with sampling of domains that predictably become impaired in the disease stage. The Alzheimer's Disease Assessment Schedule-Cognitive Subscale (ADAS-Cog[12]) has been frequently used in mild cognitive impairment (MCI) trials, despite its original development having been for mild to moderate dementia severity. It has been shown to have poor sensitivity to decline, especially in the earlier predementia stages,[13,14] which are now a frequent target of novel interventions. Furthermore, it has a high degree of variability,[15] further reducing its signal-to-noise ratio. ADAS-Cog has been preferred as the cognitive outcome measure in part due to its acceptance and familiarity by regulatory authorities alongside measures of functional impairment. However, the Food and Drug Administration (FDA) has recently updated its recommendations[16] and a wider array of cognitive and functional, or combined composite measures, are now acceptable as primary endpoints in trials. Less is known about such composite outcome measures, although improved power has been reported with measures such as the Clinical Dementia Rating Scale-Sum of Boxes (CDR-SB) scores[17,18] or the Cognitive Function Composite 2 (CFC2).[15] The Clinical Dementia Rating (CDR) scale[17] is a global assessment instrument combining assessment of cognitive elements with functional aspects of the disease. A step in the calculation is to provide box scores, whereby the score for each "box," that is, domain, is added up and not weighted as it is for the global score. This CDR-SB[19] score has been shown to be sensitive to change in MCI[20] and mild dementia.[18] The CFC2[15] composite was developed as a better

HIGHLIGHTS

- Combining genetic enrichment of participants with sensitive outcomes may improve trial design.
- Stratifying mild cognitive impairment with the polygenic hazard score appears effective.
- Using a combined cognitive-functional outcome with stratification reduces sample size needed.

RESEARCH IN CONTEXT

1. **Systematic review**: We used Google Scholar and PubMed to search the literature on trial enrichment and outcome measures in early Alzheimer's disease (AD). We found that lack of enrichment leads to many non-AD participants in trials aimed at the early stages of AD, and when used enrichment tends to use expensive and invasive techniques. Similarly, outcome measures used are often suboptimal.
2. **Interpretation**: The combination of lack of enrichment in recruitment and suboptimal outcome measures results in expensive and ineffective studies. We used new genetic methods to determine the polygenic hazard score (PHS) of participants, and modeled a hypothetical study, stratifying for risk by PHS score and comparing different outcome measures.
3. **Future directions**: This article outlines a proposed clinical trial design involving genetic enrichment using the PHS and combined cognitive-functional outcomes. We look forward to this method being tested in other cohorts and in prospective trials.

combined cognitive-functional endpoint with superior measurement properties than either the ADAS cog or the CDR-SB. It was empirically derived from the Alzheimer's Disease Neuroimaging Initiative (ADNI) data using linear mixed effects modeling, which can be used to derive composited which maximize change over time while minimizing variability. Measures combining multiple cognitive scores into a single cognitive composite, such as the ADNI-Mem,[21] are also being explored as potential primary outcome measures for MCI trials.

This study aimed to assess whether we could achieve a more efficient and effective trial design in predementia stages of AD including those at increased genetic risk while cognitively normal or with MCI by combining stratification based on PHS and testing across a range of the most frequently used composite outcome measures. We were interested in comparing change over time in these composite measures as outcome measures with determination of sample sizes necessary to achieve significant change at various follow-up periods.

2 | METHODS

2.1 | Participants

Data for these analyses were drawn from participants in the ADNI (adni.loni.usc.edu) who were classified as either being cognitively normal (CN) or having MCI at baseline entry to the cohort, had at least 12 months of follow-up data and had available PHS calculated by the Desikan Lab at University of California San Francisco (UCSF).

2.2 | Polygenic hazard score determination

The PHS[9] was developed using the AD Genetics Consortium data bases, which combine large-scale genomic studies. AD-associated single nucleotide polymorphisms (SNPs) and Cox proportional hazard models were used to develop the score, which was ultimately based on 31 SNPs in addition to the two *APOE* variants. Age-specific absolute risk for each individual was then estimated based on genotype scoring and age-specific incidence rates, with replication of the model on multiple other cohorts. For the current study, the PHS was calculated and the 50th percentile based on the age-specific absolute risks was used for stratification.

2.3 | Composite outcome measures

ADAS-Cog[22] is an 11-item measure with tests of verbal memory, praxis, orientation, and naming, and a scoring range of errors from 0 to 70. A higher score is worse.

The ADNI-Mem is an empirically derived memory measure that includes four memory tests (all verbal) available within the ADNI test battery including components of the ADAS-Cog, the Rey Auditory Verbal Learning Test (RAVLT),[23] Logical Memory from the Wechsler Memory Scale,[24] and the word list from the Mini-Mental State Examination.[25] It was derived using statistical modeling to enhance the psychometric properties of a composite in comparison with the individual tests. It benefits from adjusting scores to account for differences in RAVLT version, which are present in the ADNI dataset. It is centered around 0 with a variance of 1.

The CDR includes a rating of the individual domains of memory, orientation, judgment and problem solving, community affairs, home and hobbies, and personal care while a total score of the CDR-SB is derived by adding each domain score with a scoring range of 0 to 18, and with higher score being more impaired.[18]

The CFC2 is an empirically derived composite including the ADAS-3 (ie, the sum of word recall, delayed word recall, and orientation from the ADAS-Cog), the cognitive component of the CDR-SB, and the Functional Activities Questionnaire (range 0 to 30).[26] A higher score is more impaired.

2.4 | Data analysis

Demographic and test data were compared between the upper and lower PHS stratified risk groups using *t* tests or chi-squared tests, as appropriate. These analyses were completed independently within each of the CN and MCI groups. The analyses involved two steps. First, to compare change over time between high- and low-risk groups, two mixed models with repeated measures (MMRM) were fitted for each of the outcomes. Unstructured correlation matrix was assumed. The first mixed model controlled time and binary PHS levels (above 50% quantile as high risk group and below as low risk), and adjusted for sex, education, and age. The second model had the same covariates but additionally included an interaction term between time and the PHS levels. We compared the two models using the analysis of covariance (ANCOVA) to assess whether the PHS had a significant effect on the changes in the outcome measures over time by examining the significance levels of the effects of the added interaction terms in likelihood ratio test (LRT).

We then used two-sample *t* tests to calculate required sample size for each outcome based on hypothetical, two-arm studies over 12, 24, and 36 months to achieve a 25% effect size with a power of $\geq$ 80%. Effect size was calculated as Cohen's *d* based on estimated mean change from baseline and residual standard deviation from regression models (MMRM). This practical analysis allows us to understand the ramifications of using particular outcome measures, and how this might be improved by risk stratification. All analyses were completed with R version 3.6.1. R package "pwr" was used for power calculation.

3 | RESULTS

3.1 | Participants

Demographic, disease severity, and cognitive data are displayed in Table 1.

TABLE 1 Demographic and disease severity data for the sample

	CN		MCI	
	Low risk (n = 113)	High risk (n = 75)	Low risk (n = 113)	High risk (n = 206)
Age at initial assessment	75.91 (5.10)	76.11 (4.60)	76.14 (7.45)*	74.47 (7.11)*
Years of education	16.10 (2.73)	16.37 (2.76)	15.65 (2.96)	15.64 (3.12)
% women	43.4	45.3	38.9	33.0
% white	100	100	100	99.5
ADNI-Mem	0.94 (0.50)	1.07 (0.60)	0.08 (0.61)***	−0.15 (0.57)**
CDR-SB	0.03 (0.12)	0.02 (0.10)	1.50 (0.79)	1.64 (0.89)
ADAS-Cog 11	6.32 (2.77)	5.94 (3.05)	10.37 (4.07)*	11.66 (4.09)*
CFC2	6.21 (2.90)	5.68 (2.79)	14.55 (6.03)**	17.02 (6.65)**

Abbreviations: ADAS-Cog, Alzheimer's Disease Assessment Schedule-Cognitive Subscale; ADNI, Alzheimer's Disease Neuroimaging Initiative; CN, cognitively normal; CDR-SB, Clinical Dementia Rating-Sum of Boxes; CFC2, Cognitive Function Composite 2; MCI, mild cognitive impairment.
*$P < .05$.
**$P < .005$.
***$P < .0005$ referring to significant differences between high and low risk within the cohort (CN or MCI). Numbers in parentheses are the standard deviation.

TABLE 2 Model results for CN

Test	Findings				
CFC2	Effect @ 12 months for low risk only	Effect @ 24 months for low risk only	Effect @ 36 months for low risk only	Main effect of high risk at baseline	LRT (comparing Models 1 and 2)
	0.03	−0.10	0.88*	−0.46	3.23 ($P = .36$)
	Additional effect for high risk group	Additional effect for high risk group	Additional effect for high risk group		
	0.46	0.68	0.88		
ADAS-Cog 11	Effect @ 12 months for low risk only	Effect @ 24 months for low risk only	Effect @ 36 months for low risk only	Main effect of High Risk at baseline	LRT (comparing Models 1 and 2)
	−0.65*	−0.50	−0.60	−0.34	1.42 ($P = .70$)
	Additional effect for high risk group	Additional effect for high risk group	Additional effect for high risk group		
	0.03	0.49	0.27		
CDR-SB	Effect @ 12 months for low risk only	Effect @ 24 months for low risk only	Effect @ 36 months for low risk only	Main effect of High Risk at baseline	LRT (comparing Models 1 and 2)
	0.07*	0.11*	0.18*	−0.01	1.13($P = .77$)
	Additional effect for high risk group	Additional effect for high risk group	Additional effect for high risk group		
	−0.01	0.04	0.09		
ADNI-Mem	Effect @ 12 months for low risk only	Effect @ 24 months for low risk only	Effect @ 36 months for low risk only	Main effect of High Risk at baseline	LRT (comparing Models 1 and 2)
	0.05	0.07	0.09	0.11	2.91 ($P = .41$)
	Additional effect for high risk group	Additional effect for high risk group	Additional effect for high risk group		
	−0.06	0.00	−0.09		

Notes: For each outcome, results of model 2 are displayed, including: low risk effects of time (mean changes in outcomes at each time point comparing to the baseline for low risk [reference group]), main effects of risk (difference in outcomes between the two risk categories at the baseline). Results of the analysis of covariance comparing the two models are also reported, and the interaction effects for risk (PHS) by time at each time point are displayed as the addition effects of high risk group.
Abbreviations: ADAS-Cog, Alzheimer's Disease Assessment Schedule-Cognitive Subscale; ADNI, Alzheimer's Disease Neuroimaging Initiative; CN, cognitively normal; CDR-SB, Clinical Dementia Rating-Sum of Boxes; CFC2, Cognitive Function Composite 2; MCI, mild cognitive impairment; PHS, polygenic hazard score.
*$P < .05$, **$P < .005$, ***$P < .0005$.

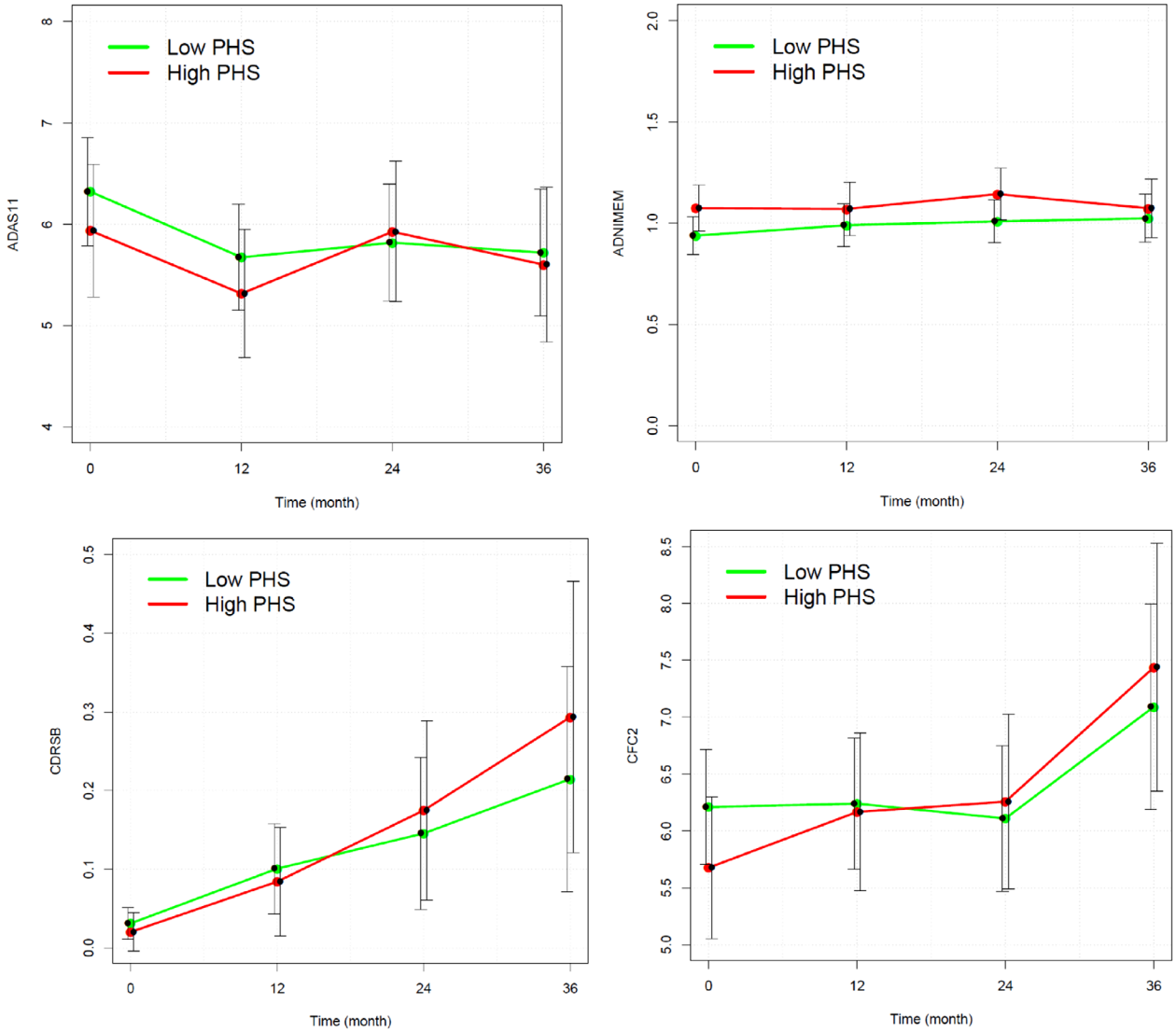

FIGURE 1 Change over time on each outcome metric by risk group in the cognitively normal cohort. PHS, polygenic hazard score

3.2 | Cognitively normal

There were no significant changes at 1 year, 2 years, or 3 years from baseline for ADNI-Mem (see Table 2 and Figure 1). CFC2 demonstrated worsening from baseline at 36 months, while ADAS-Cog 11 showed a significant time effect (ie, improvement) at 12 months, but no significant effects at later time points. Only CDR-SB showed a significant time effect at all three time points compared to the baseline. None of the four outcomes had significant differences between high- and low-risk groups either at the baseline or in their modeled slope. While there were several statistically significant differences, the actual differences were small and potentially not clinically detectable. There were significant main effects of sex and education for CFC2, ADAS, and ADNI-Mem; women and more highly educated people had better scores in those measures at baseline. To examine whether the sex had a synergistic effect when combining with time and risk, we fitted the same regression models with the three-way interaction terms of risk, time, and sex, but none of these terms were significant compared to the reduced models using LRT, implying that the added impact of stratification was not helpful for the CN group.

The MCI findings are displayed in Table 3 and visualized in Figure 2.

3.3 | Mild cognitive impairment

For the MCI cohort, the overall LRTs for the interaction terms between time and risk (PHS) were significant for all outcomes except ADAS-Cog 11 ($P = .10$), indicating that stratifying time with PHS made a significant contribution in explaining the changes in CFC2, CDR-SB, and ADNI-Mem.

The ADAS-Cog 11 showed a significant main effect of time at 24 months (2.19 higher than baseline) and 36 months (3.70 higher than baseline), and an additional effect of high PHS group at 36 months (2.35 higher than low PHS group). There was a negative main effect of

TABLE 3 Model results for MCI

Test	Findings				
CFC2	Effect @ 12 months for low risk only	Effect @ 24 months for low risk only	Effect @ 36 months for low risk only	Main effect of high risk at baseline	LRT (comparing Models 1 and 2)
	1.94***	4.98***	7.67***	2.40**	13.19** (*P* = .004)
	Additional effect for high risk group	Additional effect for high risk group	Additional effect for high risk group		
	1.69*	3.70**	5.28**		
ADAS-Cog 11	Effect @ 12 months for low risk only	Effect @ 24 months for low risk only	Effect @ 36 months for low risk only	Main effect of high risk at baseline	LRT (comparing Models 1 and 2)
	0.78	2.19***	3.70***	1.33*	6.21 (*P* = .10)
	Additional effect for high risk group	Additional effect for high risk group	Additional effect for high risk group		
	0.19	1.35	2.35*		
CDR-SB	Effect @ 12 months for low risk only	Effect @ 24 months for low risk only	Effect @ 36 months for low risk only	Main effect of High Risk at baseline	LRT (comparing Models 1 and 2)
	0. 42***	1.02***	1.35***	0.13	12.05* (*P* = .007)
	Additional effect for high risk group	Additional effect for high risk group	Additional effect for high risk group		
	0.25	0.59*	1.24***		
ADNI-Mem	Effect @ 12 months for low risk only	Effect @ 24 months for low risk only	Effect @ 36 months for low risk only	Main effect of High Risk at baseline	LRT (comparing Models 1 and 2)
	0.00	−0.07	−0.14*	−0.23**	14.26** (*P* = .003)
	Additional effect for high risk group	Additional effect for high risk group	Additional effect for high risk group		
	−0.09*	−0.22***	−0.27**		

Notes: For each outcome, results of model 2 are displayed, including: low risk effects of time (mean changes in outcomes at each time point comparing to the baseline for low risk [reference group]), main effects of risk (difference in outcomes between the two risk categories at the baseline). Results of the analysis of covariance comparing the two models are also reported, and the interaction effects for risk (PHS) by time at each time point are displayed as the addition effects of high-risk group.

Abbreviations: LRT, likelihood ratio test; MCI, mild cognitively impaired.

*$P < .05$.

**$P < .005$.

***$P < .0005$.

Abbreviations: ADAS-Cog, Alzheimer's Disease Assessment Schedule-Cognitive Subscale; ADNI, Alzheimer's Disease Neuroimaging Initiative; CN, cognitively normal; CDR-SB, Clinical Dementia Rating-Sum of Boxes; CFC2, Cognitive Function Composite 2; MCI, mild cognitive impairment.

education (the higher the education, the lower the ADAS-Cog). There were no significant main effects of age or sex.

The ADNI-Mem showed a significant effect of low-risk group at 36 months compared to baseline, and additional significant effects at each time point for the high-risk group. LRT suggested an overall significant interaction effect between high-risk group and time. There was a baseline difference between low- and high-risk groups. To note, though, the actual change was relatively small. There was a significantly positive main effect of education (the higher the education, the higher the ADNI-Mem) and female sex, but no main effect of age.

The CDR-SB showed significant increasingly main effects of time at each time point (changes of 0.42, 1.02, and 1.35, respectively) and additional significant effects of high PHS group at 24 and 36 months (additional 0.59 and 1.24). There was no main effect of sex, age, or education.

The CFC2 also showed significant main effects of time at 12, 24, and 36 months (1.94, 4.98, and 7.67, respectively). Furthermore, the high- and low-risk groups differing at baseline, and had additional significant effects at 12, 24, and 36 months. There was no main effect of sex, age, or education.

3.4 | Power analyses

Table 4 shows the sample size needed in each arm of a two-arm hypothetical trial with MCI participants, to show a 25% effect size with a minimum of 80% power. At each time point, two sample sizes are presented, one showing the number of participants needed without stratification, the other if only high PHS participants are included. The results are calculated separately for each outcome measure, and are calculated for the cases of full sample and high-risk sample only. Twelve months may not be an appropriate timeline for a MCI trial—all four outcomes required huge sample size to achieve an 80% power. At 24

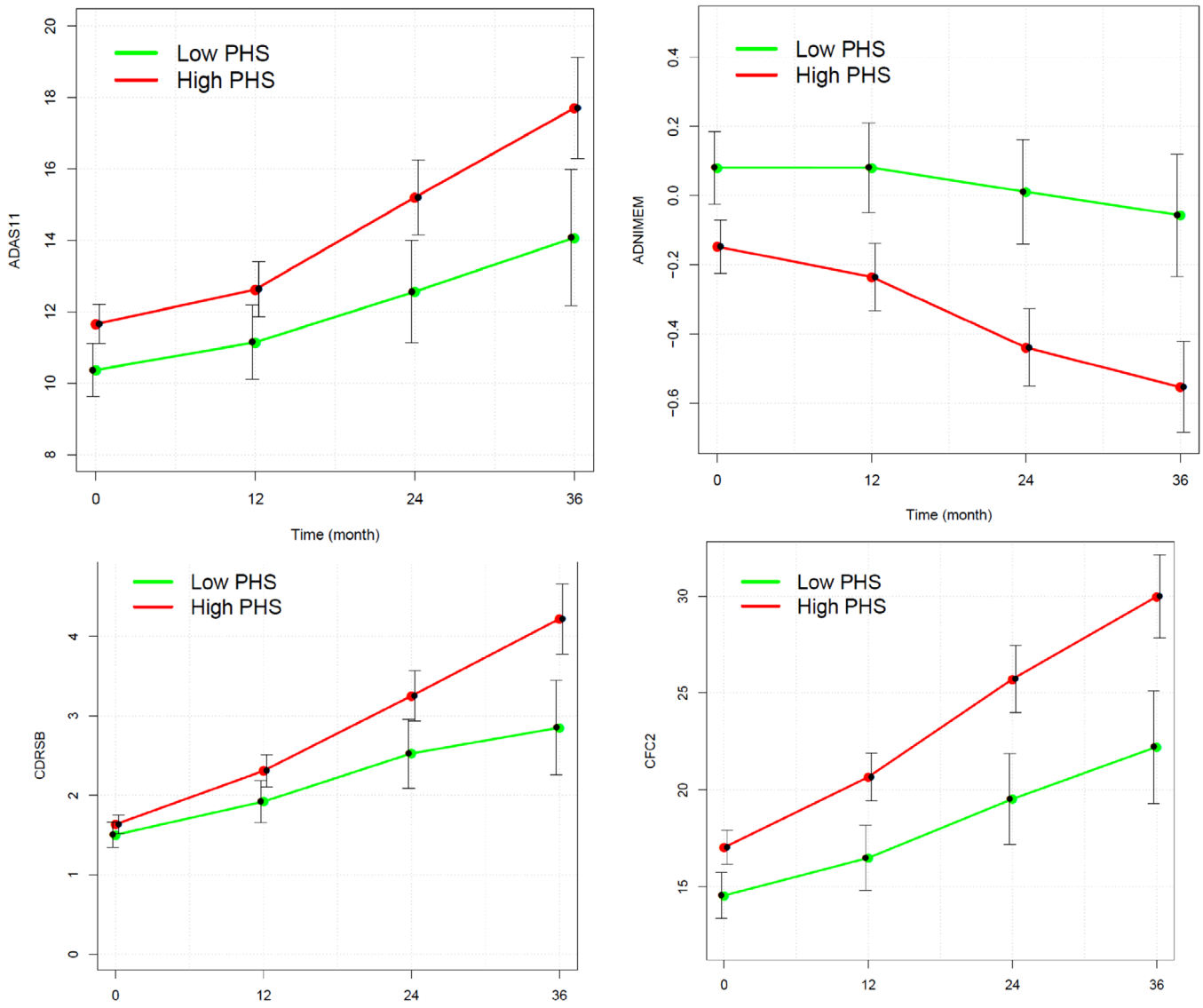

FIGURE 2 Change over time on each outcome metric by risk group in the mild cognitive impairment cohort. PHS, polygenic hazard score

TABLE 4 Sample size needed in the hypothetical clinical trial using a regression model

		Number needed in each group		
Test	MCI	12 months	24 months	36 months
ADAS-Cog	Full sample	5470	930	559
	High risk	5161	638	472
CDR-SB	Full sample	957	508	419
	High risk	647	409	320
CFC2	Full sample	722	360	272
	High risk	538	284	230
ADNI-Mem	Full sample	9713	1315	1032
	High risk	3938	742	667

Abbreviations: ADAS-Cog, Alzheimer's Disease Assessment Schedule-Cognitive Subscale; ADNI, Alzheimer's Disease Neuroimaging Initiative; CDR-SB, Clinical Dementia Rating-Sum of Boxes; CFC2, Cognitive Function Composite 2; MCI, mild cognitive impairment.

months, the full sample group required 45.8% more samples in ADAS-Cog, 24.2% more samples in CDR-SB, 26.8% more samples in CFC2, and 77.2% more samples in ADNI-Mem than the high PHS group. All four outcome measures further reduced the required sample size at 36 months when the full sample group required 30.9% more samples in CDR-SB (419 vs 320) and 18.3% more samples in CFC2 (272 vs 230) than the high PHS group. CFC2 required the smallest sample size comparing any other outcome at each time point for either full sample or high-risk group.

4 | DISCUSSION

Our study analyses were undertaken to address the utility of enriching clinical trial designs in populations of participants who might be enrolled while cognitively normal or with MCI while identified to be above median age adjusted genetic risk according to the Desikan PHR.

In the CN cohort, there was little meaningful change over time seen on these composite outcomes within the stratified sample above and below PHS median risk. While a slight improvement (potentially a practice effect) was captured at the 12-month visit in the ADAS-Cog, ADNI-Mem, which is designed to avoid this phenomenon, did not demonstrate this change. There was statistically significant decline on the CDR-SB and CFC2; however, the actual difference was numerically small, below the threshold of being clinically discernible. This is consistent with earlier reports that tests sensitive to subclinical levels of change are elusive,[27] or alternatively that there is simply insufficient decline in the time period studied to capture and that much longer periods of follow-up potentially to 6 years may be necessary.[28]

PHS stratification around a median split was not helpful in the CN cohort, although from the graph (Figure 1) it appears that there is some divergence between groups beginning at 36 months on the combined cognitive–functional measures.

Given known differences in sex on verbal memory, better scores in women[29,30] on the ADAS-Cog was not entirely unexpected, as this test relies heavily on verbal memory. This is an important feature to take into account in the design of trials in preclinical AD that women are likely to be more resilient to decline on these measures despite increasing levels of AD pathology. Using nonverbal memory measures (not available in the ADNI dataset) due to their indifference to sex differences[31] or using sex-specific norms[32] are two practical ways to counter this effect.

The potential utility of the median split PHS stratification in MCI is informative. Cognitive measures demonstrated change at 24 and 36 months for the low-risk group for the ADAS-Cog, and 36 months only on the ADNI-Mem. Adding risk stratification demonstrated more change over time in the high-risk category at all time points for ADNI-Mem, and only for 36 months on ADAS-Cog. There was an advantage to women for the ADNI-Mem but not on any other outcome measure, perhaps reflective of the loss of reserve seen during this stage.[29,30] The power analysis suggested important savings with stratification in terms of numbers needed with each of the cognitive only measures, and an apparent advantage of extending a trial to 36 months.

The combined cognitive–functional measures were sensitive to change at each time point in MCI, with enhanced decline noted in the higher risk group at 24 and 36 months for CDR-SB and all time points for CFC2. In addition, these measures were robust to sex, and required the fewest participants for a trial. The best performer in the power analysis was arguably the CFC2 in the high-risk group at 24 months. There was minimal benefit in terms of number of participants needed when extending to 36 months. A trial with this stratification measure/outcome measure combination requires 70% fewer participants to one using ADAS-Cog without stratification represents a potentially large gain in efficiency in trial design that would translate into very large savings in participant recruitment, opportunity cost, and time to completion.

Limitations of this study include its lack of generalizability to those outside of white European ancestry. The PHS was developed primarily on available large-scale genetics studies, which have had a strong bias toward this ethnicity.[33] We were also limited by the test battery given that this was retrospective data. Future studies may include more sensitive measures of decline in the preclinical phase.

One aspect of stratification that should be taken into account when considering this as an option is issues around the disclosure of PHS risk status. Until there is more known about its performance in those of non-white and diverse populations with evidence that its findings are robust and clinically important, it seems premature to be deploying this within a clinical setting. In the future, careful education and counseling will be recommended before disclosure potentially perhaps following strategies used successfully with ApoE disclosure.[34]

Our results emphasize the importance of selecting the right patients and the right outcome measures for clinical trials.[35] We further propose a relatively cost efficient way of doing this, using the PHS as a stratification factor and using a combined cognitive–functional, empirically derived composite measure as outcome. This combination strategy could significantly reduce the number of participants needed, and the time frame required for a trial in MCI.

ACKNOWLEDGMENTS

Funding for Dr. Banks and Feldman from the UCSD Alzheimer's Research Center 1P30AG062429-01. Funding for Dr. Banks from 1R01AG066088-01. Dr. Feldman received support from the CIHR for this work 201901CNA-417847-CAN-ABPI-32054.

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI; National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie; Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC; Johnson & Johnson Pharmaceutical Research & Development LLC; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. 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 Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

We thank Dr. Karen Messer for her guidance on the statistical plan.

CONFLICTS OF INTEREST

Dr. Banks has no relevant disclosures for this paper. She discloses the following consulting relationship outside this paper: Boston University, on the DIAGNOSE CTE project (U01NS093334); and research funding from the NIA/NIH, Alzheimer's Association, and California Department of Public Health. Dr. Feldman has no relevant disclosures for this paper. He discloses the following relationships outside of this paper: UCSD service agreements for consulting with Axon Neuroscience, Banner Health/Roche, Genentech, Samus Therapeutics, Tau Consortium, and Novo Nordisk; ADCS Clinical Trials grant funding from Annovis (Posiphen), Biohaven (BH 4157), Vivoryon (PQ 912), AC Immune (ACI-24-1301), and LuMind (ADC-059-LIFE-DSR); and research funding from the NIA/NIH (U19 AG010483, R01 AG061146-01, R01 AG051618), CIHR (137794, 254450, 201901CNA-417847-CAN-ABPI-32054), Brain Canada (4469), Alzheimer's Association (SG-20-690388-PEACE AD). Dr. Dale reports that he was a founder of and holds equity in CorTechs Labs, Inc., and serves on its Scientific Advisory Board. He is a member of the Scientific Advisory Boards of Human Longevity, Inc., and the Mohn Medical Imaging and Visualization Centre. He receives funding through research grants from GE Healthcare to UCSD. Dr. Fan is a consultant of CorTechs Labs, in addition to his research appointment at the University of California, San Diego. Drs. Qiu and Zou and Ms. Askew report no conflicts.

REFERENCES

1. Cummings JL, Morstorf T, Zhong K. Alzheimer's disease drug-development pipeline: few candidates, frequent failures. *Alzheimer's Res Ther*. 2014;6(4):37.
2. Mehta D, Jackson R, Paul G, Shi J, Sabbagh M. Why do trials for Alzheimer's disease drugs keep failing? A discontinued drug perspective for 2010-2015. *Expert Opin Investig Drugs*. 2017;26(6):735-739.
3. Cummings J, Lee G, Ritter A, Zhong K. Alzheimer's disease drug development pipeline: 2018. *Alzheimer's Dement Transl Res Clin Interv*. 2018;4:195-214.
4. Mangialasche F, Solomon A, Winblad B, Mecocci P, Kivipelto M. Alzheimer's disease: clinical trials and drug development. *Lancet Neurol*. 2010;9(7):702-716.
5. Nelson PT, Dickson DW, Trojanowski JQ, et al. Limbic-predominant age-related TDP-43 encephalopathy (LATE): consensus working group report. *Brain*. 2019;142(6):1503-1527.
6. Salloway S, Sperling R, Fox NC, et al. Two phase 3 trials of Bapineuzumab in mild-to-moderate Alzheimer's disease. *N Engl J Med*. 2014;370(4):322-333.
7. Karran E, Hardy J. Antiamyloid Therapy for Alzheimer's Disease—Are We on the Right Road?. *N Engl J Med*. 2014;370(4):377-378.
8. Lorenzi M, Donohue M, Paternicò D, et al. Enrichment through biomarkers in clinical trials of Alzheimer's drugs in patients with mild cognitive impairment. *Neurobiol Aging*. 2010;31:1443-1451.
9. Desikan RS, Fan CC, Wang Y, et al. Genetic assessment of age-associated Alzheimer disease risk: development and validation of a polygenic hazard score. *PLoS Med*. 2017;14(3):e1002258.
10. Tan CH, Bonham LW, Fan CC, et al. Polygenic hazard score, amyloid deposition and Alzheimer's neurodegeneration. *Brain*. 2019;14(2):460-470.
11. Tan CH, Fan CC, Mormino EC, et al. Polygenic hazard score: an enrichment marker for Alzheimer's associated amyloid and tau deposition. *Acta Neuropathol*. 2018;135(1):85-93.
12. Rosen WG, Mohs RC, Davis KL. A new rating scale for Alzheimer's disease. *Am J Psychiatry*. 1984;141(11):1356-1364.
13. Sano M, Raman R, Emond J, et al. Adding delayed recall to the Alzheimer disease assessment scale is useful in studies of mild cognitive impairment but not alzheimer disease. *Alzheimer Dis Assoc Disord*. 2011;25(2):122-127.
14. Kueper JK, Speechley M, Montero-Odasso M. The Alzheimer's disease assessment scale-cognitive subscale (ADAS-Cog): modifications and responsiveness in pre-dementia populations. A narrative review. *J Alzheimer's Dis*. 2018;63(2):423-444.
15. Raghavan N, Samtani MN, Farnum M, et al. The ADAS-Cog revisited: novel composite scales based on ADAS-Cog to improve efficiency in MCI and early AD trials. *Alzheimer's Dement*. 2013(1 Suppl):S21-31.
16. Sabbagh MN, Hendrix SB, Harrison JE. FDA position statement "Early Alzheimer's disease: developing drugs for treatment, guidance for industry. *Alzheimer's Dement Transl Res Clin Interv*. 2019;5:13-19.
17. Hughes CP, Berg L, Danziger WL, Coben LA, Martin RL. A new clinical scale for the staging of dementia. *Br J Psychiatry*. 1982;140:566-572.
18. Lynch CA, Walsh C, Blanco A, et al. The clinical dementia rating sum of box score in mild dementia. *Dement Geriatr Cogn Disord*. 2006;21(1):40-43.
19. Berg L, Hughes CP, Coben LA, Danziger WL, Martin RL, Knesevich J. Mild senile dementia of Alzheimer type: research diagnostic criteria, recruitment, and description of a study population. *J Neurol Neurosurg Psychiatry*. 1982;45(11):962-968.
20. Samtani MN, 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;10:929-952.
21. Crane PK, Carle A, Gibbons LE, et al. Development and assessment of a composite score for memory in the Alzheimer's disease neuroimaging initiative (ADNI). *Brain Imaging Behav*. 2012;6(4):502-516.
22. Mohs RC. *Administration and Scoring Manual for the Alzheimer's Disease Assessment Scale 1994 Revised Edition*. New York: The Mount Sinai School of Medicine; 1994.
23. Rey A, *L'examen clinique en psychologie*. Presses Universitaries De France [The clinical examination in psychology]: Presses Universitaries De France. 1958.
24. Wechsler D. *Wechsler Memory Scale-Revised*. San Antonio, Texas: Psychological Corporation; 1987.
25. Folstein MF, Folstein SE, McHugh PR. "Mini-mental State": a practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. 1975;12:189-198.
26. Pfeffer RI, Kurosaki TT, Harrah CH, Chance JM, Filos S. Measurement of functional activities in older adults in the community. *J Gerontol*. 1982;37(3):323-329.
27. Goldman WP, Price JL, Storandt M, et al. Absence of cognitive impairment or decline in preclinical Alzheimer's disease. *Neurology*. 2001;56(3):361-367.
28. Insel PS, Weiner M, Scott MacKin R, et al. Determining clinically meaningful decline in preclinical Alzheimer disease. *Neurology*. 2019;93(4):e322-e333.
29. Sundermann EE, Maki PM, Rubin LH, Lipton RB, Landau S, Biegon A. Female advantage in verbal memory: evidence of sex-specific cognitive reserve. *Neurology*. 2016;87(18):1916-1924.
30. Caldwell JZK, Berg J-L, Cummings JL, Banks SJ. Moderating effects of sex on the impact of diagnosis and amyloid positivity on verbal memory and hippocampal volume. *Alzheimers Res Ther*. 2017;9(1):72.
31. Brunet HE, Caldwell JZK, Brandt J, Miller JB. Influence of sex differences in interpreting learning and memory within a clinical sample of older adults. *Aging, Neuropsychol Cogn*. 2019;27(1).

32. Sundermann EE, Maki P, Biegon A, et al. Sex-specific norms for verbal memory tests may improve diagnostic accuracy of amnestic MCI. *Neurology*. 2019;93(20).
33. Clyde D. Making the case for more inclusive GWAS. *Nat Rev Genet*. 2019;20(9):500-501.
34. Green RC, Roberts JS, Cupples LA, et al. Disclosure of APOE genotype for risk of Alzheimer's disease. *N Engl J Med*. 2009;361(3):245-254.
35. Cummings J, Feldman HH, Scheltens P. The "rights" of precision drug development for Alzheimer's disease. *Alzheimer's Res Ther*. 2019.

How to cite this article: Banks SJ, Qiu Y, Fan CC, et al. Enriching the design of Alzheimer's disease clinical trials: Application of the polygenic hazard score and composite outcome measures. *Alzheimer's Dement*. 2020;6:e12071. https://doi.org/10.1002/trc2.12071