Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Biomarker associations and outcome performance in preclinical Alzheimer's disease trials

Michael C. Donohue
mdonohue@usc.edu
University of Southern California
Oliver Langford
University of Southern California
Philip S. Insel
University of California
Rema Raman
University of Southern California
Michael S. Rafii
University of Southern California
Reisa A. Sperling
Harvard Medical School
Paul S. Aisen
University of Southern California

Research Article

Keywords: preclinical Alzheimer's disease, Preclinical Alzheimer Cognitive Composite (PACC), plasma P-tau217, clinical trial simulation, endpoint selection, Clinical Dementia Rating (CDR)

Posted Date: November 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11208391/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: Competing interest reported. MCD, OL, RR, and PSA receive research support from the National Institutes of Health (NIH). PSA reported personal fees from Merck, Biogen, Roche, AbbVie, ImmunoBrain Checkpoint, Bristol Myers Squibb, and Neurimmune and grants from Eisai outside the submitted work. RAS reported consulting fees from AbbVie, AC Immune, Acumen, Alector, Apellis, Biohaven, Bristol Myers Squibb, Janssen, Nervgen, Oligomerix, Prothena, Roche, and Vaxxinity outside

the submitted work. MCD reported personal fees from Roche (consultant), Therini Bio (consultant), and Johnson & Johnson (spouse is full-time employee) outside the submitted work. RR reported grants from American Heart Association, Gates Ventures, Eisai, Alzheimer's Association. No other disclosures were reported.

Abstract

Background

Cognitive, functional, and progression-based outcomes are all used in preclinical Alzheimer's disease trials, but their relationships to underlying pathology and their comparative statistical performance have not been evaluated within a common framework. We compared associations of candidate trial outcomes with Alzheimer's disease pathology and evaluated their operating characteristics under biologically motivated simulated treatment effects.

Methods

We used data from the double-blind phase, through week 216, of the Anti-Amyloid Treatment in Asymptomatic Alzheimer's Disease (A4) Study, a randomised, multicentre, placebo-controlled trial of 791 cognitively unimpaired adults aged 65–80 years with amyloid PET greater than 40 centiloids. Placebo groups were generated by non-parametric resampling; active-treatment groups were generated by enrichment for milder plasma P-tau217 change, while preserving correlations among outcomes. We assessed associations of baseline P-tau217 with decline on the Preclinical Alzheimer Cognitive Composite (PACC) versus progression on the Clinical Dementia Rating (CDR) Global score; and estimated statistical power for the PACC, the Cognitive Function Instrument (CFI), the CDR Sum of Boxes (CDR-SB), and time to CDR Global progression.

Results

Among 791 participants (mean [SD] age, 71.4 [3.9] years; 60.8% female), higher baseline P-tau217 showed a larger association with PACC decline than with CDR Global progression (odds ratio per SD, 3.08 [95% CI, 2.47–3.90] versus 1.84 [95% CI, 1.55–2.21]). Under biomarker-enrichment scenarios, the PACC had the greatest statistical power of the four outcomes: 80% power required enrichment from the expected 50% of participants with milder P-tau217 change to approximately 84–86%, corresponding to a simulated PACC effect of 0.153–0.175 points. The CDR-SB, CFI, and time to CDR Global progression required greater enrichment to achieve comparable power.

Conclusions

In this preclinical Alzheimer's disease cohort, baseline P-tau217 was more strongly associated with PACC decline than with CDR Global progression, and the PACC showed favourable operating characteristics under biomarker-enrichment scenarios anchored to this biomarker. Because these findings derive from a single trial cohort and a modelling framework with untested assumptions, they

support further evaluation of the PACC as a sensitive preclinical outcome rather than establishing it as the superior trial endpoint.

Background

As anti-amyloid therapies move earlier into the preclinical stage of Alzheimer's disease, endpoint selection has become one of the most important determinants of trial success. Whether cognitive composites, functional measures, or clinical progression outcomes are most informative in this setting depends not only on statistical efficiency, but also on how closely each outcome reflects the underlying biology of early Alzheimer's disease. Composite cognitive measures such as the Preclinical Alzheimer Cognitive Composite (PACC)[1] have been adopted in prevention trials because they capture subtle objective multidomain changes that may emerge before functional impairment. Functional and progression outcomes, such as Clinical Dementia Rating (CDR) Global progression,[2] Clinical Dementia Rating Sum of Boxes (CDR-SB),[3] and the Cognitive Function Instrument (CFI),[4] may be more intuitively clinically interpretable, but they may occur later in the disease cascade and may be influenced by factors less directly related to Alzheimer's disease biology.

Candidate trial outcomes are commonly compared by calculating power, or the required sample size, separately for each outcome. For example, investigators may estimate the number of participants needed to achieve 80% power to detect a 25% or 30% treatment effect on each outcome, expressed as a percentage reduction in the expected change in the control group[5–7]. These calculations rely on pilot data to estimate the mean change and variability of each outcome, and the outcome requiring the fewest participants is then viewed as offering the greatest power to detect a treatment effect. If we take the outcome which provides the smallest sample size as the most powerful, this implicitly assumes that a given intervention would produce a comparable percentage effect across those outcomes. This assumption is rarely stated explicitly and cannot be tested without treatment data, even though it is built into the calculation. This framework ignores differences in how strongly each outcome reflects the underlying biological mechanism targeted by treatment, as well as the correlation among outcomes as disease progresses. It also does not naturally extend to outcomes of different types, such as time-to-event measures, without strong assumptions about the comparability of percentage improvement on a continuous outcome and changes in progression rates. In prior work, we used a joint modelling approach to simulate correlated outcomes, but it relied on complex multi-outcome parametric models and did not readily incorporate biologically anchored treatment effects[8]. In this paper, we propose and apply a non-parametric weighted resampling, or enrichment, approach to simulate treatment effects that are biologically plausible, correlated across outcomes, and agnostic to outcome type, providing a more robust and broad framework for comparing and selecting candidate trial outcomes.

The Anti-Amyloid Treatment in Asymptomatic Alzheimer's Disease (A4) Study, a randomised, multicentre, placebo-controlled clinical trial, provides a unique opportunity to explore the operating characteristics of these endpoints in a well characterised cohort of cognitively unimpaired individuals with biomarker evidence of amyloid pathology.[9] The rich longitudinal data from A4 allows evaluation of

the empirical distributions and correlation structures among cognitive, functional, and clinical progression outcomes. These features are particularly important when designing simulation studies that aim to compare the statistical power of alternative endpoints and analysis approaches. Simulations that rely on simplified parametric assumptions may not fully reflect the complex structure of real world, multi-outcome, longitudinal data. We have also demonstrated the existence of latent classes of PACC decline in A4 data,[10] adding additional complexity. For these reasons, approaches that use nonparametric resampling are especially valuable because they preserve multidimensional relationships among outcomes and idiosyncrasies that may be difficult to accurately parameterize, while still enabling controlled assessment of hypothetical treatment effects.

Within a resampling-based simulation framework, placebo groups can be generated by resampling data from untreated or ineffectively treated participants. A simple parametric approach to simulating a treated group is to resample from the same population, and add, say, a linear treatment benefit over time. But if one is trying to compare outcomes, how do we determine comparable treatment benefit for each outcome? A treatment that meaningfully slows disease biology may not translate into the same percent benefit in cognition, function, and progression events, so power comparisons based on equal percent effects can be misleading.

In this paper we propose a non-parametric approach to simulating a treated group that preferentially samples, or enriches for, participants who have less Alzheimer’s pathology. This approach is motivated by the idea that a successful treatment would have a biological effect that reduces amyloid pathology, and those treated individuals might be expected to decline like individuals who naturally have less Alzheimer’s pathology. By varying the degree of enrichment, we can simulate treatment groups with varying magnitudes of simulated treatment effects. This allows us to compare the power of different endpoints under realistic assumptions about the link between the degree of Alzheimer’s pathology and clinical outcomes, and the correlation among outcomes over time. A critical prerequisite for this approach is whether the candidate outcome is itself meaningfully aligned with Alzheimer's disease pathology in the preclinical stage.

We selected longitudinal P-tau217 change as the primary enrichment variable because it reflects Alzheimer's disease-related pathological activity. Disease-modifying therapies are intended to alter the trajectory of pathology over time, making longitudinal biomarker change a relevant reference for simulated biologically motivated group differences. Although treatment-induced biomarker changes cannot be inferred from observational data, naturally occurring differences in longitudinal P-tau217 change may provide a useful framework for comparing endpoint behaviour under differing levels of disease progression. Baseline P-tau217, PACC progressor class, and CDR-G progressor class were also considered as an enrichment variables.

The present study addresses two related questions relevant to endpoint selection in prevention trials. First, we examine whether PACC decline is more closely related to Alzheimer's disease pathology than progression on the CDR Global score in the A4 population. Second, we evaluate whether outcomes that

are more closely aligned with Alzheimer's disease biology also provide greater power to detect simulated biologically plausible treatment effects in simulated trials. To address the second question, we compare four outcomes widely used or under consideration for Alzheimer's disease prevention trials: the PACC, CDR-SB, CFI total score, and time to CDR Global progression. By basing simulations on nonparametric resampling of A4 data and inducing simulaed treatment effects through enrichment for milder plasma P-tau217 change, we aim to provide a realistic framework for endpoint comparison that preserves the observed correlation among outcomes over time.

Statistical sensitivity, association with disease biomarkers, and clinical meaningfulness represent distinct properties of an outcome measure and should be evaluated separately. Therfore, we evaluate whether PACC change is associated with subsequent clinical progression. Specifically, we compare longitudinal PACC trajectories between participants who do and do not progress on the CDR Global score during follow-up. Together, these analyses are intended to clarify whether the PACC is both biologically sensitive and clinically relevant in preclinical Alzheimer's disease, and whether those properties translate into superior trial performance.

Methods

All analyses were conducted using A4 participants aged 65 to 80 years with baseline amyloid PET greater than 40 centiloids (CL). We performed three complementary analyses in this cohort: first, comparison of the association of baseline P-tau217 with PACC decline versus CDR Global progression; second, evaluation of the clinical meaningfulness of PACC decline using subsequent CDR Global progression; and third, simulation of randomised trials to compare the power of candidate endpoints under simulated biologically plausible treatment effects. Only data from the A4 double-blind phase up to the week 216 visit were included. Approval from an institutional review board was obtained at each study site, and all the participants provided written informed consent. The study was conducted in accordance with International Council for Harmonisation Good Clinical Practice guidelines.

Association of baseline P-tau217 with PACC decline and CDR Global progression

To assess whether change on PACC or CDR Global is more strongly associated with baseline P-tau217, we compared two logistic regression models applied to the same group of A4 participants. One model uses PACC latent class model estimated decline status as the binary outcome, and the other uses CDR Global progression status as the outcome. Both models include covariates for baseline P-tau217, age, treatment arm, APOE $\epsilon 4$ status, sex, and education.

Association of longitudinal PACC change with subsequent CDR Global progression

To evaluate the clinical meaningfulness of PACC changes, we fit a spline model of raw PACC scores over time that estimated mean PACC change for two groups: CDR Global progressors and CDR Global non-progressors. The model followed the same structure as the spline model of raw scores described above, except that CDR Global progression status replaced treatment arm, and we did not constrain the progression groups to share the same mean baseline score. The model included baseline amyloid PET SUVr, age, treatment arm, APOE $\epsilon 4$ status, sex, education, PACC version, time, CDR Global progression status, and the interaction between progression status and time. Time was modelled using natural cubic splines with two degrees of freedom.

Simulating trials

A control group was simulated by sampling, with replacement, from participants in the A4 Study. Active treatment groups were simulated by enriching for participants estimated to have less change in P-tau217 from baseline to 216 weeks than the median (more detail below and in Fig. 1). This results in a simulated treatment group with less decline than the aggregate population. The magnitude of the treatment effect can be varied by increasing the degree to which participants with less change in P-tau217 are oversampled. Using a biology based enrichment strategy, rather than a clinical based one, avoids favouring one clinical outcome over another, utilizes the observed correlation between outcomes without any parametric assumptions, and therefore makes power comparisons more meaningful. This strategy assumes that a treated participant whose pathology is pharmacologically reduced would come to resemble, prognostically, a participant who has always had milder pathology; we examine the plausibility of this assumption, and its likely direction of bias, in the Discussion.

Longitudinal P-tau217 change was chosen as the primary enrichment variable because it reflects differences in disease progression during follow-up rather than baseline disease severity alone. The objective was not to model the causal effect of treatment-induced P-tau217 lowering, but rather to construct biologically motivated population contrasts that preserve the observed relationships among biomarkers, cognition, function, and clinical progression within A4.

We estimated week-216 change in plasma P-tau217 using a linear mixed-effects model including fixed effects for baseline age, sex, APOE $\epsilon 4$ carrier status, time from baseline, and covariate-by-time interactions. Participant-level random intercepts and random slopes for time were included to account for within-participant correlation. This approach leverages all available longitudinal biomarker data and accommodates variability in visit timing relative to baseline.

A null-effect scenario was also simulated by sampling without enrichment for both treatment groups to evaluate Type I error. PACC, CDR SB, and CFI scores collected after CDR Global progression were excluded, to allow for open-label therapy. CDR Global progression was defined as the visit corresponding to the second consecutive CDR Global score greater than 0. Participants who progress at this point could be offered open label therapy.

Because the enrichment could yield simulated treatment groups with imbalanced baseline characteristics relative to simulated placebo groups, additional processing steps were applied to approximate the comparability expected under actual randomisation. Mean shifts were applied to baseline and post-baseline PACC, CDR-SB, and CFI scores to equalize baseline group means. No analogous adjustment was made to CDR Global progression data, which should favour this outcome in terms of simulated power. Covariates (age, sex, education, APOE $\epsilon 4$ carrier status, and amyloid PET SUVr) were independently permuted to minimize systematic differences between simulated treatment groups in these analysis covariates. Figure 1 demonstrates how the resampling approach induces a treatment effect and how the baseline means are adjusted.

To avoid duplicate observations in the simulated datasets, random "jitter" was added to PACC, CDR-SB, and CFI scores and to CDR-Global observation times. Jittering followed the default behaviour of the R jitter function, applying a uniform perturbation of $\pm 1/50$ of the smallest observed difference for each variable. This approach preserves the rank order of the original data while preventing identical values.

Additional enrichment strategies are described in Additional file 1, including enrichment for participants with (1) with baseline P-tau217 < 0.25 U/mL (roughly the median); (2) classified as PACC-stable based on a latent class model[11] (see Additional file 1, Supplementary Figure S2 for more details); or (3) who did not progress on CDR Global.

Approaches that explicitly favours one outcome over others are less suitable for objectively comparing outcome performance. In contrast, our primary approach is designed to approximate the biological effects of a successful anti-amyloid treatment without privileging a particular clinical measure.

Analyzing simulated data

We evaluated 4 outcomes: the Preclinical Alzheimer Cognitive Composite (PACC), Cognitive Function Instrument (CFI) total score, Clinical Dementia Rating Sum of Boxes (CDR-SB), and time to progression on the Clinical Dementia Rating (CDR) Global score. For the longitudinal outcomes (PACC, CFI, and CDR-SB), we fit models for (1) change from baseline with baseline score as a covariate and (2) raw scores over time with baseline included as part of the repeated-measures outcome vector. Within-participant residual covariance across visits was modelled as unstructured. All models adjusted for baseline covariates including age, sex, education, APOE $\epsilon 4$ status, and amyloid PET SUVr. The following models were evaluated:

MMRM of change from baseline. For PACC, we fit a mixed model for repeated measures with fixed effects for treatment arm, visit, and the treatment arm by visit interaction; baseline PACC and a baseline PACC by visit interaction; and the baseline covariates listed above[12]. The treatment effect was estimated as the between-arm contrast at week 216. Models for CFI and CDR-SB change from baseline had the same structure.

Spline model of raw scores. We modelled raw scores over time, treating baseline as an observed outcome and modelling time as a continuous variable[13]. The model included fixed effects for time, the treatment arm by time interaction, PACC version (alternate forms across visits), and the baseline covariates listed above. The model excluded a main effect of treatment, constraining baseline means to be equal across simulated treatment groups. Time was represented using natural cubic splines with 2 degrees of freedom. The treatment effect was estimated as the between-arm contrast at week 216. CFI and CDR-SB models were analogous but did not include a version term.

Cox proportional hazards model for time to CDR Global progression. We fit a Cox proportional hazards model with treatment arm as the primary effect and adjustment for the baseline covariates listed above[14]. The treatment effect was summarised using the hazard ratio.

Results

Pilot data characteristics

A total of $N = 791$ A4 participants were identified with age 65 to 80 years and amyloid PET greater than 40 centiloids. The mean (standard deviation [SD]) age at baseline was 71.42 (3.94) years, and 60.8% were female. The median follow-up time (prior to any CDR Global progression) in the double-blind phase up to the week 216 visit was 4.2 years (interquartile range 4.1 to 4.3). Baseline characteristics stratified by P-tau217 week-216 change status (< 0.07 U/ml vs ≥ 0.07 U/ml) are presented in Table 1. These baseline relationships suggested that participants with milder biomarker burden also tended to have better cognitive and clinical outcomes, prompting formal comparison of the association of P-tau217 with PACC decline versus CDR Global progression.

Table 1: Baseline Demographic, Clinical, and Biomarker Characteristics by P-tau217 change status.

Baseline characteristics of participants classified by P-tau217 change status (< 0.07 U/ml vs ≥ 0.07 U/ml). PACC is presented as the average of four components, rather than the sum.

Continuous variables are presented as mean (SD); categorical variables as No. (%). Abbreviations: APOE = apolipoprotein E; PET = positron emission tomography; P-tau217 = plasma phosphorylated tau 217; SUVr = standardized uptake value ratio; U/mL = arbitrary units proportional to assay signal; Am. = American; PI = Pacific Islander; CDR = Clinical Dementia Rating; Prog. = Progressor. Footnotes: Amyloid PET and tau PET SUVr values represent mean cortical uptake relative to cerebellar reference region. Hippocampal atrophy are residualized for intracranial volume and z-scored. Education reported in years of formal schooling. CDR Progressors were observed to have a CDR Global score greater than zero at two consecutive visits, or their last visit.

	Below median P-tau217 change (N = 395)	Above median P-tau217 change (N = 396)	Total (N = 791)	p value
Group				0.005
- Placebo	177 (44.8%)	217 (54.8%)	394 (49.8%)	
- Solanezumab	218 (55.2%)	179 (45.2%)	397 (50.2%)	
Age, mean (SD), y	71.08 (3.94)	71.76 (3.91)	71.42 (3.94)	0.015
Sex, No. (%)				0.008
- Male	173 (43.8%)	137 (34.6%)	310 (39.2%)	
- Female	222 (56.2%)	259 (65.4%)	481 (60.8%)	
Race, No. (%)				0.028
- N missing	2	3	5	
- Am. Indian or Alaska Native	1 (0.3%)	1 (0.3%)	2 (0.3%)	
- Asian	9 (2.3%)	4 (1.0%)	13 (1.7%)	
- Native Hawaiian or Other PI	0 (0.0%)	0 (0.0%)	0 (0.0%)	
- Black or African Am.	10 (2.5%)	4 (1.0%)	14 (1.8%)	
- White	367 (93.4%)	384 (97.7%)	751 (95.5%)	
- Unknown or Not Reported	6 (1.5%)	0 (0.0%)	6 (0.8%)	
Ethnicity, No. (%)				0.979
- Hispanic or Latino	11 (2.8%)	12 (3.0%)	23 (2.9%)	
- Not Hispanic or Latino	381 (96.5%)	381 (96.2%)	762 (96.3%)	
- Unknown or Not reported	3 (0.8%)	3 (0.8%)	6 (0.8%)	
Education, mean (SD), y	16.64 (2.75)	16.29 (2.84)	16.47 (2.80)	0.074
APOE ε4, No. (%)				0.306

- non-carrier	128 (32.4%)	142 (35.9%)	270 (34.1%)	
- carrier	267 (67.6%)	254 (64.1%)	521 (65.9%)	
PACC, mean (SD)	0.07 (0.63)	-0.07 (0.70)	-0.00 (0.67)	0.003
Amyloid PET, mean (SD), CL	70.54 (25.42)	81.92 (30.59)	76.24 (28.68)	< 0.001
P-tau217, mean (SD), U/ml	0.26 (0.14)	0.34 (0.18)	0.30 (0.17)	< 0.001
P-tau217 216-week change, mean (SD), U/ml	0.02 (0.05)	0.20 (0.19)	0.11 (0.17)	< 0.001
CDR Prog., No. (%)				< 0.001
- CDR Non-progressor	286 (72.4%)	236 (59.6%)	522 (66.0%)	
- CDR Progressor	109 (27.6%)	160 (40.4%)	269 (34.0%)	

Baseline P-tau217 was more strongly associated with PACC decline than with CDR Global progression

Baseline P-tau217 showed a substantially stronger association with PACC decline than with CDR Global progression (Fig. 2, Supplementary Table S1). In logistic regression models adjusted for age, treatment arm, APOE ε4 status, sex, and education, each standard deviation increase in baseline P-tau217 was associated with more than a threefold increase in the odds of belonging to the PACC decline class (OR, 3.08; 95% CI, 2.47 to 3.90; $p < 0.001$), whereas the corresponding association with CDR Global progression was appreciably smaller (OR, 1.84; 95% CI, 1.55 to 2.21; $p < 0.001$). A similar pattern was observed for baseline amyloid PET, which was also more strongly associated with PACC decline than with CDR Global progression, supporting the robustness of this finding across Alzheimer's disease biomarkers (Supplementary Table S2). These findings indicate that cognitive decline as measured by the PACC is more closely aligned with underlying Alzheimer's disease pathology than progression on the CDR Global score in this preclinical population.

APOE ε4 carrier status showed a similar pattern. APOE ε4 was significantly associated with increased odds of PACC decline (OR, 2.22; 95% CI, 1.48 to 3.39; $p < 0.001$) but was not significantly associated with CDR Global progression in the same participants (OR, 1.29; 95% CI, 0.92 to 1.81; $p = 0.137$). This pattern is consistent with the known relationship between APOE genotype and amyloid-related cognitive change and further supports the interpretation that the PACC is more sensitive than CDR Global progression to Alzheimer's disease-related decline during the preclinical stage.[15]

Sex showed the opposite pattern. Female sex was significantly associated with a lower likelihood of CDR Global progression (OR, 0.63; 95% CI, 0.46 to 0.87; p = 0.005) but was not significantly associated with PACC decline (OR, 1.33; 95% CI, 0.91 to 1.94; p = 0.141). Because this sex effect does not appear to align with Alzheimer's disease biology, it may reflect measurement, informant, or household-structure factors that influence progression-based outcomes.

Taken together, these findings indicate that both baseline P-tau217 and APOE $\epsilon 4$ status are more strongly linked to PACC decline than to CDR Global progression, providing further support for the use of the PACC as a sensitive outcome for tracking early Alzheimer's disease-related change. The observed sex effect on CDR Global progression appears less biologically grounded and may reflect informant related or household structure differences rather than true differences in clinical decline. To determine whether this biologic sensitivity also remained clinically interpretable, we next examined longitudinal PACC trajectories according to subsequent CDR Global progression status.

Longitudinal PACC decline tracked subsequent CDR Global progression

In a spline model of PACC scores over time that adjusted for baseline amyloid PET SUVr, age, treatment arm, APOE ε4 status, sex, education, and PACC version, participants who later progressed on the CDR Global score had lower PACC scores at baseline and showed earlier and steeper decline than participants who did not progress. Modeled mean trajectories (Supplementary Figure S3) diverged over follow-up, with progressively lower PACC scores among progressors and relatively stable scores among nonprogressors. These findings indicate that PACC differences are detectable before CDR Global progression and that longitudinal PACC decline tracks subsequent clinically recognizable worsening. Given these observed relationships between PACC, Alzheimer's disease pathology, and subsequent clinical progression, we next evaluated whether the PACC also offered greater power to detect simulated biologically plausible treatment effects in simulated prevention trials.

Simulated power results

We then evaluated whether the greater biologic sensitivity of the PACC translated into improved operating characteristics in trial simulations. Figure 3 and Table 2 summarize the main simulation results under the primary scenario in which active treatment groups were simulated by enriching for participants with lower week-216 P-tau217 change. Under this biologically grounded enrichment strategy, the PACC was consistently the most powerful outcome for detecting simulated treatment effects, followed by CDR-SB, CDR Global progression, and CFI. The PACC required the least enrichment to attain 80% power, indicating that it was the most sensitive to shifts in the underlying biology represented by milder P-tau217 change. Achieving 80% power required enrichment from the expected 50% of participants with milder P-tau217 change in the control group to approximately 84% to 86% in the biomarker-enriched group, corresponding to a simulated PACC effect of 0.153 to 0.175 points. In

contrast, CDR-SB required approximately 95% enrichment in the active arm, and time to CDR Global progression required approximately 97% enrichment to achieve comparable power.

Results were similar when treatment effects were simulated by enriching for individuals with lower baseline P-tau217, and type I error remained close to 5% in the null scenario with no enrichment. In alternative simulation approaches that explicitly favoured a particular outcome, that same outcome tended to appear most powerful (Supplementary Figure S1), underscoring the importance of biologically grounded, outcome-agnostic enrichment when comparing endpoint performance.

Table 2: Simulated effect required to attain 80% power. Based on the results shown in Fig. 3, we can linearly interpolate the simulated treatment effect required to achieve 80% power for each outcome and analysis approach. The table below shows the estimated treatment effect (on the original scale of the outcome), the residual standard deviation at the final visit, and the proportion of milder individuals in the active arm for each approach at the point where power is approximately 80%. For the Cox model of CDR Global progression, the treatment effect is presented as a hazard ratio. For the PACC, the treatment effect and residual SD are rescaled to represent the average of four components rather than the sum.

Approach	Average simulated treatment effect	Average final visit residual SD	Average proportion mild in active arm (%)
PACC Spline	0.153	1.291	84.8
PACC MMRM	0.175	0.987	86.5
CDR-SB MMRM	-0.203	1.156	95.0
CDR-SB Spline	-0.191	1.174	95.4
CDR-G Cox	0.738		97.3

Discussion

In this cohort of cognitively unimpaired individuals with elevated amyloid, baseline P-tau217 showed substantially stronger association with latent-class-defined PACC decline than with CDR Global progression. Comparable findings were observed for baseline amyloid PET. Participants who subsequently progressed on the CDR Global score also exhibited earlier and steeper PACC decline. Together, these findings suggest that the PACC captures clinically relevant variation associated with underlying Alzheimer's disease pathology during the preclinical stage.

The simulation analyses provide an extension of these observations. Under biomarker-based enrichment scenarios designed to preserve the observed correlation structure among outcomes, the PACC demonstrated favourable operating characteristics relative to the alternative outcomes examined. These

findings should be interpreted as characteristics of the observed A4 data structure under specified enrichment assumptions rather than predictions of treatment effects or trial outcomes.

A specific concern about this approach deserves direct comment. Our enrichment strategy assumes that a treated participant whose amyloid-related pathology is pharmacologically reduced would come to resemble, prognostically, an untreated participant who has always had milder pathology. This is unlikely to hold exactly: participants enriched for milder natural P-tau217 change have, by definition, not experienced the downstream consequences — reduced cognitive reserve, accumulated neurodegeneration, and neuroinflammation — of a sustained period of higher amyloid and tau burden, a period that any actually treated participant in this cohort (all with amyloid PET > 40 centiloids at entry) would already have experienced before a disease-modifying therapy could act. If these prior downstream effects are only partially reversible, our enrichment-based comparisons likely provide an optimistic, best-case approximation of achievable treatment effects rather than an unbiased estimate, and the true benefit of amyloid-lowering therapy on each outcome may be smaller than these simulations suggest; observational data alone cannot quantify how much smaller. Because this optimism should apply similarly across the outcomes compared — all are evaluated within the same enriched contrast — we believe the relative ranking of outcomes is more robust to this concern than the absolute power estimates, but this should be tested directly once trials with confirmed target engagement and long-term follow-up become available.

CDR-SB and CFI, while potentially more intuitively clinically meaningful, appeared less sensitive than the PACC in the simulated anti-amyloid context, particularly when enrichment targeted biologic rather than clinical stability. This pattern is consistent with the conceptual framework of disease-modifying therapies, which are expected to exert early effects on cognition before functional decline becomes readily measurable. It also aligns with the observed A4 data, in which baseline P-tau217 and APOE ε4 carrier status were more strongly associated with PACC decline than with CDR Global progression. A similar pattern was observed for baseline amyloid PET, which was also more strongly associated with PACC decline than with CDR Global progression, supporting the robustness of this finding across Alzheimer's disease biomarkers. Together, these findings suggest that the PACC is more sensitive to Alzheimer's disease-related processes during the earliest clinical stages. This pattern should be interpreted specifically in relation to P-tau217 — one of several relevant Alzheimer’s disease biomarkers — rather than as evidence that the PACC captures the whole of Alzheimer’s disease biology.

Differences between outcomes also highlighted potential nonbiological influences on progression-based measures. Female sex is significantly associated with CDR Global progression but not with PACC decline. This association is unlikely to reflect true biological differences and may instead arise from factors beyond Alzheimer's disease biology such as social or informant related factors. For example, the sex of the spouse or partner available to provide informant ratings could influence the timing of CDR based progression in ways unrelated to underlying disease. Although exploratory, this finding underscores the importance of considering both biological and nonbiological influences on clinical outcome measures in preclinical populations.

An important strength of the present study is that the simulation design was anchored to an observed biologic relationship, namely the stronger association of lower P-tau217 burden with better cognitive outcomes, while still avoiding explicit favouritism toward any single clinical endpoint. This design links the observational and simulation components of the study and allows endpoint comparisons under a common biologically grounded treatment framework. Under this framework, differences in outcome performance reflect intrinsic sensitivity to treatment-associated changes rather than artifacts of trial construction. Alternative enrichment strategies that explicitly favoured stability on a particular outcome yielded less informative comparisons and emphasised the value of outcome-agnostic enrichment when evaluating relative endpoint performance.

We also explored the impact of longitudinal model specification on treatment effect detection. Across outcomes, spline-based models generally provided greater statistical efficiency than categorical time MMRM. This result is consistent with our prior parametric based work[13] and suggests that flexible time modelling may improve sensitivity in longer prevention trials.

Several limitations should be considered. First, the simulations rely on assumptions derived from A4 data and from modelled P-tau217 trajectories, which may not fully capture the complexity of biomarker and clinical responses in future cohorts that may be observed in response to potent amyloid lowering therapeutics. Notably, data from the A4 study cannot be used to predict the treatment response of more effective amyloid lowering therapy. The primary simulation strategy was based on naturally observed longitudinal P-tau217 trajectories. Consequently, these analyses do not imply that treatment-induced P-tau217 changes would produce the same clinical trajectories as naturally occurring differences in P-tau217 change. Rather, the enrichment strategy was designed to generate biologically motivated differences in disease progression while preserving the observed correlation structure among outcomes. Second, baseline mean alignment and covariate permutation were used to approximate randomised comparability; although necessary for the simulation structure, these adjustments may reduce realism. Third, the exclusion of post-progression data reflects anticipated trial procedures but increases missingness and could result in bias if the missing at random assumption is not met. Fourth, these findings derive from a single trial cohort (A4) evaluating a single mechanism of action (anti-amyloid antibody treatment); the relative performance of these outcomes may differ for other drug classes, populations, or trial durations, and independent replication is needed. Fifth, PACC decline in the A4 cohort was slower and more heterogeneous than originally anticipated when the trial was designed, an outcome-specific limitation that tempers, without negating, the case for the PACC's sensitivity.

Several conclusions should not be drawn from these analyses. The findings do not establish that treatment-induced biomarker change would produce the same outcome trajectories as naturally occurring biomarker variation. They do not establish clinically meaningful change thresholds for any outcome. Nor do they predict the relative performance of outcomes in a specific clinical trial. Rather, they describe observed relationships within the A4 cohort and operating characteristics under a defined set of biomarker-based enrichment scenarios.

Despite these limitations, the results provide convergent, though not definitive, evidence that the PACC may be well suited for detecting early treatment effects of anti-amyloid therapies in enriched populations, a conclusion that warrants replication beyond this single trial cohort. The modest gains observed with spline-based longitudinal models suggest that flexible modelling may offer incremental benefits, particularly as follow-up durations increase. Future work should extend these simulations to alternative enrichment strategies and to other classes of disease-modifying therapies, as well as explore implications for trial duration and sample size planning.

Conclusions

Overall, these findings suggest that the PACC occupies an important middle ground for preclinical Alzheimer's disease trials: it is more closely aligned with Alzheimer's disease biology than progression-based outcomes in this population, yet it also tracks subsequent clinical worsening and performs favourably in biologically grounded trial simulations. Endpoint selection in prevention trials should therefore consider not only clinical intuitiveness, but also the extent to which an outcome reflects the underlying disease process and can detect early treatment-related change. This apparent advantage is specific to the biomarkers examined here (P-tau217 and APOE ε4 status) and to this trial cohort, and should not be read as a general claim that the PACC captures Alzheimer’s disease biology more broadly than other candidate outcomes.

Declarations

Ethics approval and consent to participate

The A4 Study protocol was approved by the University of Southern California Institutional Review Board for the coordinating center, and by the institutional review board at each participating site; all participants provided written informed consent, as previously described [9]. This manuscript reports a secondary analysis of de-identified A4 Study data and did not involve additional data collection, participant contact, or site-level ethics review beyond the original trial approvals.

Competing Interests

MCD, OL, RR, and PSA receive research support from the National Institutes of Health (NIH). PSA reported personal fees from Merck, Biogen, Roche, AbbVie, ImmunoBrain Checkpoint, Bristol Myers Squibb, and Neurimmune and grants from Eisai outside the submitted work. RAS reported consulting fees from AbbVie, AC Immune, Acumen, Alector, Apellis, Biohaven, Bristol Myers Squibb, Janssen, Nervgen, Oligomerix, Prothena, Roche, and Vaxxinity outside the submitted work. MCD reported personal fees from Roche (consultant), Therini Bio (consultant), and Johnson & Johnson (spouse is full-time employee) outside the submitted work. RR reported grants from American Heart Association, Gates Ventures, Eisai, Alzheimer’s Association. No other disclosures were reported.

Funding

The A4 and LEARN Studies were supported by a public-private-philanthropic partnership which included funding from the National Institute of Aging of the National Institutes of Health (R01 AG063689, U19AG010483 and U24AG057437), Eli Lilly (also the supplier of active medication and placebo), the Alzheimer's Association, the Accelerating Medicines Partnership through the Foundation for the National Institutes of Health, the GHR Foundation, the Davis Alzheimer Prevention Program, the Yugilbar Foundation, an anonymous foundation, and additional private donors to Brigham and Women's Hospital, with in-kind support from Avid Radiopharmaceuticals, Cogstate, Albert Einstein College of Medicine and the Foundation for Neurologic Diseases.

Other support was provided by grants from the Epstein Family Foundation.

Author Contribution

MCD, OL, PSI, PSA, RAS, and RR conceived and designed the study. MCD, OL, and PSI performed the statistical analyses. MCD drafted the manuscript. OL, PSI, RR, MSR, RAS, and PSA contributed to interpretation of the results and critically revised the manuscript for important intellectual content. All authors read and approved the final manuscript.

Acknowledgement

The authors would like to thank the A4 Study Team and site principal investigators and staff. Special gratitude to the A4 and LEARN participants and their study partners, without whom these studies would not be possible.

Data Availability

Data from the A4 Study are available to qualified researchers, subject to a data use agreement, through the study's established data-sharing process via https://www.a4studydata.org.

References

1. Donohue MC, Sperling RA, Salmon DP, et al. The preclinical alzheimer cognitive composite: Measuring amyloid-related decline. JAMA Neurol. 2014;71(8):961–70.
2. Morris JC. The Clinical Dementia Rating (CDR) current version and scoring rules. Neurology. 1993;43(11):2412–2412.
3. Berg L, Miller JP, Storandt M, et al. Mild senile dementia of the alzheimer type: 2. Longitudinal assessment. Ann Neurol. 1988;23(5):477–84.

4. Li C, Neugroschl J, Luo X, et al. The utility of the cognitive function instrument (CFI) to detect cognitive decline in non-demented older adults. J Alzheimer's Disease. 2017;60(2):427–37.
5. Cullen NC, Zetterberg H, Insel PS, et al. Comparing progression biomarkers in clinical trials of early alzheimer's disease. Ann Clin Transl Neurol. 2020;7(9):1661–73.
6. Harvey DJ, Tosun D, Jack CR Jr, Weiner M, Beckett LA, Initiative ADN. Standardized statistical framework for comparison of biomarkers: Techniques from ADNI. Alzheimer's Dement. 2024;20(10):6834–43.
7. Holland D, Brewer JB, Hagler DJ et al. Subregional neuroanatomical change as a biomarker for alzheimer's disease. *Proceedings of the National Academy of Sciences*. 2009;106(49):20954–20959.
8. Li D, Iddi S, Aisen PS, et al. The relative efficiency of time-to-progression and continuous measures of cognition in presymptomatic alzheimer's disease. Alzheimer's Dementia: Translational Res Clin Interventions. 2019;5:308–18.
9. Sperling RA, Donohue MC, Raman R, et al. Trial of Solanezumab in Preclinical Alzheimer's Disease. N Engl J Med. 2023;389(12):1096–107. 10.1056/NEJMoa2305032.
10. Li R, Langford O, Insel PS et al. Divergent patterns of cognitive decline in preclinical alzheimer's disease: Implications for secondary prevention trials. *Alzheimer's & Dementia*. Published online in press. https://doi.org/10.64898/2025.12.15.25342305
11. Proust-Lima C, Philipps V, Liquet B. Estimation of extended mixed models using latent classes and latent processes: the R package lcmm. J Stat Softw. 2017;78:1–56.
12. Mallinckrodt CH, Clark WS, David SR. Accounting for dropout bias using mixed-effects models. J Biopharm Stat. 2001;11(1–2):9–21.
13. Donohue MC, Langford O, Insel PS, et al. Natural cubic splines for the analysis of alzheimer's clinical trials. Pharm Stat. 2023;22(3):508–19.
14. Cox DR. Regression models and life-tables. J Roy Stat Soc: Ser B (Methodol). 1972;34(2):187–202.
15. Donohue MC, Sperling RA, Petersen R, et al. Association between elevated brain amyloid and subsequent cognitive decline among cognitively normal persons. JAMA. 2017;317(22):2305–16.

Figures

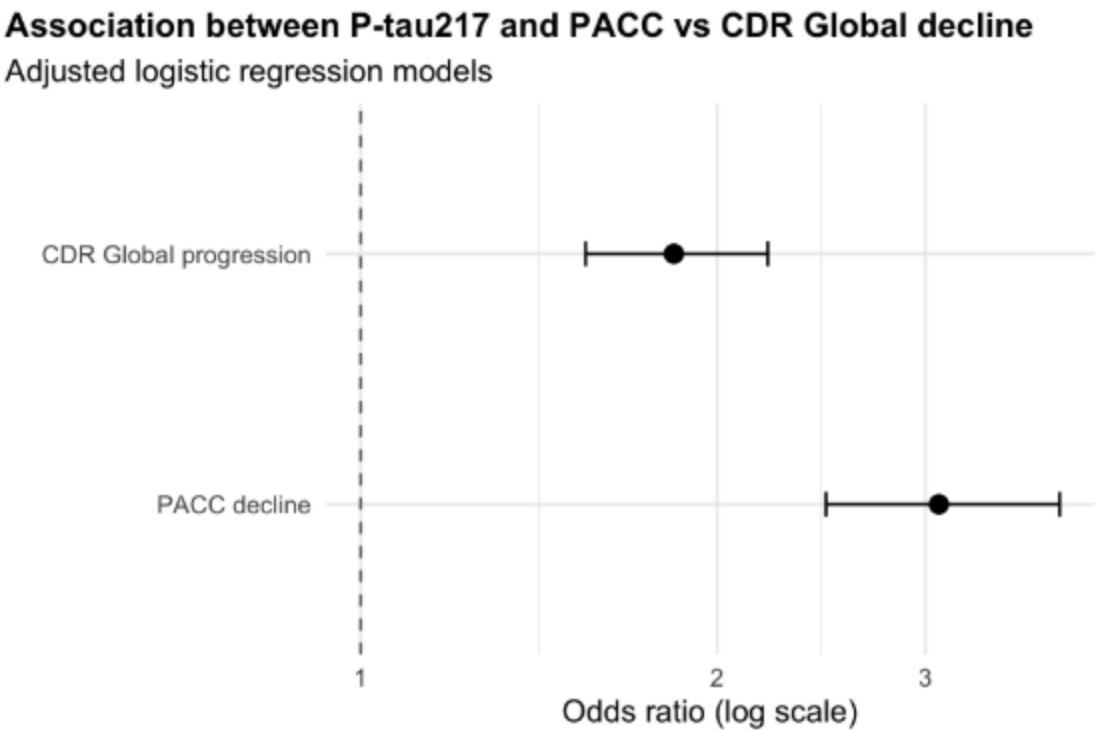


Figure 2

Adjusted association of baseline P-tau217 with PACC decline and CDR Global progression. Odds ratios and 95% confidence intervals for baseline P-tau217 are shown from two logistic regression models fit to the same A4 participants. One model used latent-class-defined PACC decline as the outcome and the other used CDR Global progression. Both models adjusted for age, treatment arm, APOE ε4 status, sex, and education. Baseline P-tau217 was more strongly associated with PACC decline than with CDR Global progression, supporting greater biologic sensitivity of the PACC in preclinical Alzheimer disease.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- evalpadoutcomesARTtrackedchanges20260928suppl.docx