

Original Research

Evaluating inclusion of continuous multivariable cognitive scores for operational enrichment of preclinical Alzheimer's disease trials: a retrospective emulation study

Henry Bayly^{a,1} , Lindsay Salvati^{a,1} , Steven Lenio^{b,c} , Jesse Mez^{b,c},
Michael L. Alosco^{b,c,d,e}, Yorghos Tripodis^{a,c,*}

^a Department of Biostatistics, Boston University School of Public Health, 801 Massachusetts Avenue, Boston, MA 02118, USA

^b Department of Neurology, Boston University Chobanian & Avedisian School of Medicine, 72 E Concord St, Boston, MA 02118, USA

^c Boston University Alzheimer's Disease Research Center and CTE Center, Boston University Chobanian & Avedisian School of Medicine, 72 E Concord St, Boston, MA 02118, USA

^d Department of Anatomy & Neurobiology, Boston University Chobanian & Avedisian School of Medicine, 72 E Concord St, Boston, MA 02118, USA

^e Department of Neurology, Boston Medical Center, 1 Boston Medical Center Pl, Boston, MA 02118, USA

ARTICLE INFO

Keywords:

Alzheimer's disease
Clinical trial enrollment strategies
Trial efficiency
Cognitive decline risk prediction
Trial engineering

ABSTRACT

Background: Prevention trials for Alzheimer's Disease face significant challenges due to the slow and uncertain rate of cognitive decline in asymptomatic, amyloid-positive individuals. Biomarker positivity alone does not guarantee clinically meaningful progression, often leaving studies underpowered.

Objectives: Evaluate model-based operational trial enrichment within an amyloid-positive, cognitively normal subgroup.

Methods: Trained on a small cohort from the National Alzheimer's Coordinating Center ($N = 113$), with external validation using baseline data of a similar small cohort from the Alzheimer's Disease Neuroimaging Initiative ($N = 161$). Gradient-boosted decision trees estimated progression risk and projected power under alternative enrollment strategies. Predictors included neuropsychological scores, demographics, medical history, and ApoE $\epsilon 4$ genotype. The primary outcome was clinical progression to cognitive impairment within 1- and 3-year follow-up windows.

Results: In the independent testing set, the model successfully increased the effective prevalence of the outcome. The Positive Predictive Value for the 1-year window was 0.27 (a 35% relative increase over the 0.20 baseline prevalence) and 0.43 for the 3-year window (a 23% relative increase over the 0.35 baseline). Power simulations for an emulated trial ($N = 500$) demonstrated that this enrichment strategy consistently increased statistical power, potentially reducing the required sample size to achieve 80% power by up to 32% for small treatment effect sizes.

Conclusions: This operational methodology functions as a screening filter to optimize event rates within a highly specific, preselected trial population.

1. Introduction

Despite major advances in the biological characterization of Alzheimer's disease (AD) as a neuropathological process, the design of adequately powered clinical trials for dementia prevention remains challenging. A primary obstacle is the operational difficulty of efficiently

identifying which individuals will experience short-term clinical progression among those who already have underlying AD pathology, but whose symptoms are absent or subtle and nonspecific [1,2]. This uncertainty contributes to high screen-failure rates and substantial recruitment burden in prevention trials [3], and increases the risk that trials are underpowered to detect clinically meaningful treatment effects

* Corresponding author.

E-mail addresses: baylyh@bu.edu (H. Bayly), lrsalv@bu.edu (L. Salvati), stlenio@bu.edu (S. Lenio), jessemez@bu.edu (J. Mez), malosco@bu.edu (M.L. Alosco), yorghos@bu.edu (Y. Tripodis).

¹ Contributed equally.

<https://doi.org/10.1016/j.jarlif.2026.100081>

Received 26 May 2026; Received in revised form 21 July 2026; Accepted 23 July 2026

Available online 8 August 2026

2534-773X/© 2026 The Authors. Published by Elsevier Masson SAS on behalf of SERDI Publisher. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

within feasible follow-up periods.

The amyloid hypothesis, under which AD pathology is characterized by the accumulation of amyloid-beta neuritic plaques [4,5], has shaped contemporary AD trial design by motivating enrollment strategies that prioritize biological evidence of the disease [6]. The use of the amyloid hypothesis has transformed screening pipelines, particularly in large, multi-site trials [7], where the main inclusion criteria is restricted to biomarker positive individuals. While amyloid-based eligibility is a pragmatically chosen design choice for prevention studies, amyloid positivity is not synonymous with imminent cognitive decline. A substantial proportion of amyloid positive individuals exhibit slow, heterogeneous, or negligible cognitive decline over several years [8–10]. Furthermore, amyloid positivity should not be equated with the biological mechanisms underlying therapeutic response, as the relationship between amyloid burden, disease progression, and treatment benefit remains complex and incompletely understood. From a trial design perspective, the fact that many of these individuals may not progress within the follow-up period introduces prognostic uncertainty in clinical trajectories, which results in reduced statistical power and inflated sample sizes. As a result, trials may fail not because an intervention is ineffective, but because too few clinical events occur during follow-up to reliably detect a treatment effect.

Rather than treating amyloid positivity as an absolute indicator of conversion, trial designs can integrate further prognostic enrichment by the addition of neuropsychological measures. These can offer a complementary source of information by capturing subtle, clinically meaningful manifestations of the disease. Brief cognitive tests are inexpensive, widely available, and already embedded in standard clinical and research workflows. Historically, trial eligibility has been based on a combination of cognitive testing, PET imaging, and clinician assessment to determine the presence of absence of impairment due to AD [11,12]. Many large scale prevention trials, such as A4 [13], have implemented binary cognitive score thresholds as inclusion/exclusion criteria. Since the main goal of primary prevention trials is to understand the effect of a therapeutic on slowing cognitive decline, specifically in an amyloid positive cognitively normal group of people. However, with the increasing availability of data from Alzheimer's Disease Research Centers (ADRCs), there has been considerable work to try and redefine enrollment protocols for AD clinical trials in such a way that utilizes these rich sources of data.

Predictive modeling has emerged as a potential solution to this question [14–16], by offering an operational mechanism to optimize enrollment workflows. In comparison to a crude binary cognitive score threshold, a continuous, multivariable risk score which predicts risk of conversion to AD in a specified time frame can utilize the full depth of the available data. Existing studies that have looked into applying predictive modeling to enrich a study cohort have either exclusively relied on prior clinical trial data and therefore not utilized ADRC data [14] or failed to fully implement inclusion/exclusion criteria that adequately emulates contemporary AD clinical trials [15]. In this paper, we performed predictive modeling through Machine Learning (ML), though it should be noted that this was a design choice and not the only viable implementation method. Aside from predictive modeling, other existing enrichment strategies for AD clinical trials have included trajectory analysis [17], fMRI precision medicine prediction [18], and computational trial phenomaps [19].

This study fits within this broader enrichment literature but focuses specifically on whether routinely collected ADRC cognitive and clinical data can be translated into a practical, enrichment tool after biomarker based eligibility has already been established. Crucially, we are not aiming to advance the early detection of AD or provide a diagnostic innovation. Instead, our analysis was designed to evaluate the operational methodology of enrichment for short term symptomatic conversion – specifically transitioning from CDR of 0 to 0.5 or higher among an amyloid-positive, cognitively unimpaired cohort. The focus here is on a preclinical subpopulation that was cognitively intact at their

hypothetical time of enrollment but at possible elevated biological risk of progression, as assumed by the amyloid hypothesis and determined by amyloid positivity on PET or CSF measures. As such, our results should be interpreted cautiously, as they are derived from retrospective modeling rather than prospective validation and do not contribute to any biological literature of early disease progression. In practice, they are most appropriately viewed as supporting a two-stage framework applied within a population already meeting sponsor/FDA approved eligibility criteria.

2. Methods

2.1. Retrospective trial-enrichment emulation

Our analytic framework was designed to retrospectively emulate a prevention trial utilizing operational methodology in preclinical Alzheimer's disease, most closely resembling trials such as AHEAD [11] and Trailblazer-Alz 3 [20], in which cognitively unimpaired individuals with evidence of amyloid pathology are enrolled and followed longitudinally. Consistent with these designs, the target population for our emulated trial consisted of adults aged 55 years or older who were cognitively unimpaired at baseline (global Clinical Dementia Rating [CDR] = 0) and demonstrated amyloid positivity as assessed by positron emission tomography (PET) or cerebrospinal fluid (CSF) biomarkers. Amyloid status was based on study specific derived variables in each dataset. Individuals with any evidence of baseline cognitive impairment (CDR ≥ 0.5) were excluded.

The primary outcome for the emulated trial was short-term cognitive progression, defined as a transition from CDR 0 to CDR ≥ 0.5 at a subsequent annual visit, requiring sustained impairment in subsequent available study visits to minimize boundary fluctuation noise. This outcome was chosen to approximate the early symptomatic conversion events that drive statistical power in prevention trials with limited follow-up, and should not be conflated with the biological onset of Alzheimer's disease pathology. The estimand of interest was the cumulative incidence of cognitive impairment within a fixed follow-up window (1- or 3-year) under alternative enrollment strategies. Evaluating both a short and medium sized window allowed us to assess whether the relative performance of these strategies would be preserved across trial designs with different follow-up durations. To connect the estimand to a hypothetical prevention trial, we considered a two-arm randomized trial with equal allocation and a binary endpoint as defined above. Sample size projections were based on comparing two proportions, where the control-arm event rate was set to the observed progression rate among participants selected under each enrollment strategy. We note that these calculations are to serve as an illustrative emulation exercise and do not mirror complex analytic structures (e.g. time to event, repeated measures, composite outcomes) that active AD trials encounter. The treatment effect was expressed as an absolute risk reduction over the same window.

2.2. Data

We used the National Alzheimer's Coordinating Center (NACC) Uniform Data Set (UDS) Version 3, downloaded on November 5, 2025, to train our models. To evaluate our model in an independent cohort, we used the Alzheimer's Disease Neuroimaging Initiative (ADNI) data for our testing set due to its similarity to the NACC and its relatively large sample size.

2.3. Neuropsychological scores

Cognitive measures were harmonized across different UDS versions using the corresponding NACC definitions and crosswalk documentation [21]. The cognitive measures included: Category fluency tests (Animals, Vegetables), Trail Making Test parts A/B (TRAILA and TRAILB) and the

number of commission errors (TRAILARR and TRAILBRR), Montreal Cognitive Assessment (MoCA), Multilingual Naming Test (MNT), Craft Story 21 Immediate and Delayed Recall Paraphrase Scoring (CRAFTURS and CRAFTDRE), Number Span Test Forward (DIGFORCT), and Number Span Test Backward (DIGBACCT).

In March 2015, the Alzheimer's Disease Research Centers (ADRCs) adopted a new neuropsychological test battery which replaced the Mini Mental State Examination (MMSE) with the Montreal Cognitive Assessment (MoCA), the Boston Naming Test with the Multilingual Naming Test (MNT), the Logical Memory, Immediate and Delayed with the Craft Story 21 Recall, Immediate and Delayed, and the Digit Span, Forward and Backward with the Number Span, Forward and Backward to avoid licensing restrictions. Since both NACC and ADNI contained participants from before and after March 2015, standardization of the cognitive exams was required. Therefore, for all participants who received only the old version of the listed exams, we used the NACC UDS Crosswalk to map the observed old exam values to the new exam [21].

For the training set (NACC), we calculated features for the mean of cognitive scores. For individuals who ultimately progressed to cognitive impairment, only data from visits preceding the transition were used, ensuring that all predictors temporally preceded the outcome. Fig. 1 illustrates how we arrived at our final analytic cohort starting from all NACC participants. For the testing set (ADNI), we applied the same inclusion criteria to the baseline visit of each participant. To emulate real world trial enrollment difficulties, we only supplied the initial cognitive assessment from the baseline visit to the trained model to make predictions. This workflow demonstrates how longitudinal data from ADRCs can be utilized to make predictions for trial enrollment when only a single visit's worth of data is available. Finally, as a sensitivity

analysis, a 3-year transition window was evaluated which defined a positive transition if impairment occurred at any of the subsequent 3 visits.

2.4. Model construction

We constructed a series of models to predict those who would transition to cognitive impairment within a 1-year window. The model's performance was validated using a 1- and 3-year window in the independent testing set. Input for each model included cognitive measures and demographic characteristics, medical history variables, APOE $\epsilon 4$ genotype. All baseline demographics are described in Table 1. Amyloid status was not included as a predictor because all participants were selected conditional on amyloid positivity. We used gradient-boosted decision trees implemented via XGBoost as our primary mode of modeling, supplemented with SHAPley Additive Explanations (SHAP) for interpretability [22,23]. XGBoost forces the model to learn from mistakes made in earlier training rounds by sequentially fitting decision trees to the residual errors of the current ensemble [22]. Each iteration optimizes a regularized objective function using a second-order Taylor approximation of the loss function. We used 500 training rounds with a learning rate of 0.01. Confidence intervals were estimated using 100 bootstrap iterations. Because we had a relatively small training dataset ($N = 113$), we benchmarked XGBoost (a gradient boosted ensemble) against simpler models: Logistic Regression and Elastic Net (penalized regression). Further, we conducted classification threshold tuning over the range (0.3–0.7) to understand how sensitive each model was to changes in the decision threshold. The optimal model was chosen as the one which maximized PPV and balanced stability in terms of the number

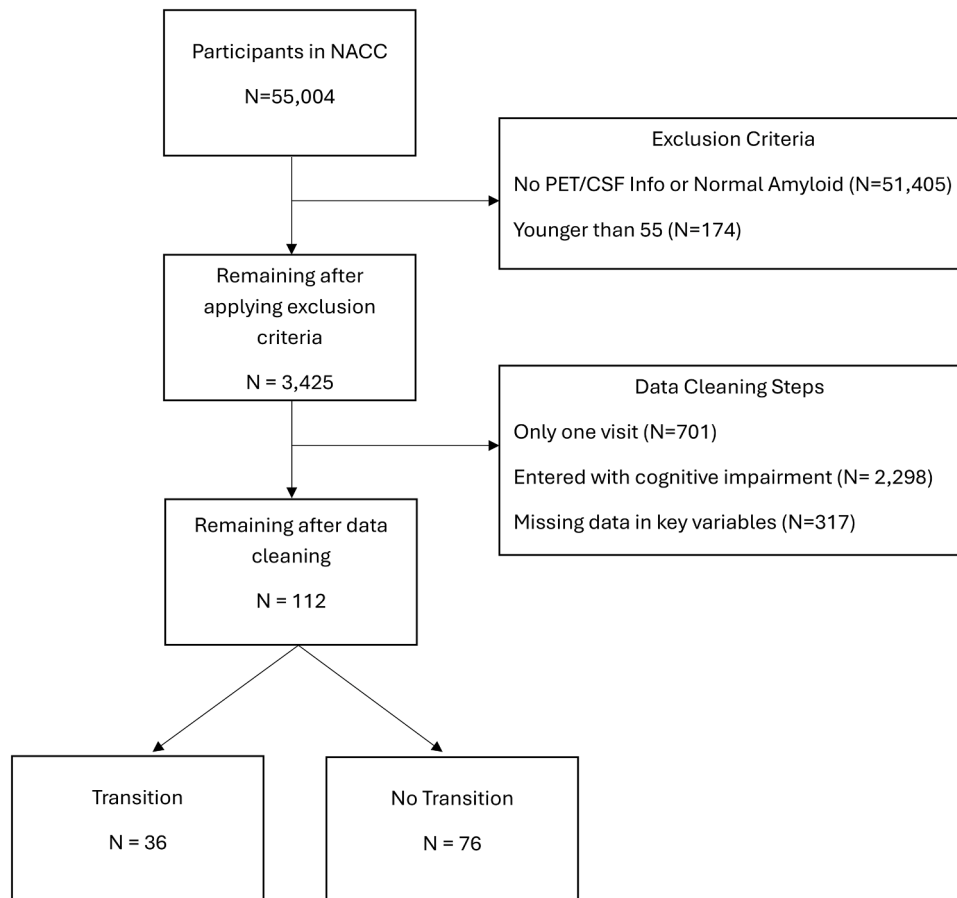


Fig. 1. Participant selection flowchart for the NACC training and ADNI validation cohorts. NACC = National Alzheimer's Coordinating Center; ADNI = Alzheimer's Disease Neuroimaging Initiative.

Table 1

Baseline Demographic and Clinical Characteristics of Amyloid-Positive Cognitively Unimpaired Participants in the NACC and ADNI Cohorts.

	NACC		ADNI	
	No Transition N = 76	Transition N = 37	No Transition N = 129	Transition N = 32
Number of Visits [N (SD)]	2.33 (1.30)	1.92 (1.57)	–	–
Female [N (%)]	43 (56.6)	19 of 37	65 (50.4)	18 of 32
Race [N (%)]				
White	67 (88.2)	35 of 37	124 (96.1)	30 of 32
Black / African American	6 (7.9)	2 of 37	5 (3.9)	1 of 32
American Indian or Alaska Native	1 (1.3)	0 of 37	0 (0.0)	0 of 32
Native Hawaiian or Other Pacific Islander	0 (0.0)	0 of 37	0 (0.0)	0 of 32
Asian	2 (2.6)	0 of 37	0 (0.0)	1 of 32
Family History of Dementia [N (%)]	60 (78.9)	23 of 37	120 (93.0)	31 of 32
Smoker [N (%)]	19 (25.0)	15 of 37	50 (38.8)	15 of 32
APOE4 Alleles [N (%)]				
0	24 (31.6)	19 of 37	70 (54.3)	19 of 32
1	45 (59.2)	12 of 37	53 (41.1)	12 of 32
2	7 (9.2)	6 of 37	6 (4.7)	1 of 32
Diabetes [N (%)]	8 (10.5)	2 of 37	5 (3.9)	1 of 32
Hypertension [N (%)]	42 (55.3)	22 of 37	58 (45.0)	17 of 32
B12 Deficiency [N (%)]	4 (5.3)	3 of 37	14 (10.9)	3 of 32
Years of Education [mean (SD)]	17.14 (2.10)	16.57 (2.74)	16.33 (2.54)	15.28 (2.68)
Geriatric Depression Score [mean (SD)]	1.30 (1.53)	1.54 (1.97)	0.97 (1.57)	1.28 (1.46)
Age [mean (SD)]	71.51 (6.84)	72.70 (7.29)	73.45 (6.10)	75.46 (5.61)
BMI [mean (SD)]	28.32 (4.86)	27.05 (5.70)	26.99 (5.07)	26.35 (5.23)

Note: In accordance with the *Journal of Aging Research & Lifestyle* instructions for authors, groups with 50 or fewer subjects (NACC Transition and ADNI Transition) are reported as text ratios ("X of Y") rather than percentages to ensure statistical clarity in small sample sizes.

of positives predicted. This criteria was chosen to emphasize the transportability of the produced model.

Our primary modeling objective was to maximize the Positive Predictive Value (PPV), also known as precision (defined as the proportion of positive predictions that are actually positive), of predicting individuals who would subsequently transition to cognitive impairment. In the context of trial enrollment, PPV directly determines the event rate in the placebo arm and therefore influences the attainable statistical power. We used the PPV to conduct power calculations for a hypothetical trial. We evaluated (i) the achievable power for a fixed sample size of 500 (250 per arm), approximating Phase 2 trials [24,25], and (ii) the sample size required to achieve 80% and 90% power across a range of treatment effect sizes (0.05 to 0.75).

Two enrollment strategies were compared: random selection among amyloid-positive individuals (baseline) and enrichment based on predictive modeling with cognitive measures combined with demographic and genetic variables. The baseline strategy was generated using the prevalence of conversion from the testing dataset; this in practice would mimic randomization in an amyloid positive sample. The second strategy used the PPV from the model; this would mimic randomizing *after* the model had predicted a person as a likely progressor. In this setting, the PPV values should be interpreted as the observed progression rate among the selected individuals within a given cohort and follow-up window. Since PPV depends on the underlying event prevalence in

each screened population, it cannot be generalized to different settings. These PPV estimates should only be interpreted as comparative measures of enrichment within biomarker-positive populations similar to those studied here.

3. Results

Table 1 shows the baseline demographic and clinical characteristics of the analytic samples. For both the 1- and 3-year windows, the optimal model was XGBoost with a 0.5 classification threshold (Supplement Table 1). Although some alternative model/threshold combinations produced higher PPVs, they identified significantly fewer positive cases and exhibited wide CIs around PPV. The full set of predictive performance metrics is shown in Table 2. We also evaluated practical operating characteristics such as sensitivity, specificity, and expected calibration error (ECE). The final model demonstrated good calibration and an appropriate balance between sensitivity and specificity at the 0.5 threshold (Supplement Table 1). To create a more parsimonious model, we ran a secondary reduced model using demographics, health history, and the top four most predictive exams as identified by SHAP. Because performance metrics were similar between the full model and the reduced model, we only present the results from the reduced model for clarity and translational relevance. The most predictive cognitive measures were the Craft Story 21 Immediate Recall, Craft Story 21 Delayed Recall, Montreal Cognitive Assessment (MoCA), and the Multilingual Naming Test (MNT).

The baseline prevalence of the outcome in our training set (NACC) was approximately 0.33, meaning that about 33% of the sample would transition to cognitive impairment at their subsequent visit. The predictive model, when applied within a held-out validation set from NACC, achieved a PPV of approximately 0.57 (95% Confidence Interval (CI): [0.54,0.64]). We saw similar results when we applied the model to the fully held-out testing set (ADNI). The 1- and 3-year prevalence of the outcome in ADNI were approximately 0.20 and 0.35, respectively. The PPVs from the modeling in the 1- and 3-year tasks were 0.27 (95% CI: [0.20,0.36]) and 0.43 (95% CI: [0.35,0.52]). These represent about a 35% and 23% relative increase in the effective prevalence of the outcome.

Figs. 2 and 3 translate predictive performance into trial-relevant quantities by illustrating achievable power and estimated required sample size under alternative enrollment strategies. Fig. 2 shows that, in simulation, an enrollment strategy based on predictive modeling uniformly increases power in a 1-year and 3-year event horizon. Fig. 3 quantifies the potential percent reduction in sample size needed to achieve 80% and 90% power. For a small effect size, for example 0.10, the predictive model can potentially reduce the sample size needed to achieve 80% power in a 1-year horizon by 32%. Importantly, the values presented do not account for attrition during follow-up and therefore represent optimistic projections of trial efficiency. Differential dropout across enrollment strategies could attenuate the projected gains, especially if selected high risk individuals are also more likely to withdraw from the study. Therefore, these results should be interpreted as

Table 2

Predictive Performance Metrics for the NACC Training and ADNI Validation Cohorts.

	NACC (Training Results)	ADNI (Testing Results)	
	Mean (95% CI)	1-year window	3-year window
Accuracy	0.69 [0.68,0.71]	0.54 [0.27,0.71]	0.56 [0.37,0.7]
Sensitivity	0.40 [0.38,0.43]	0.71 [0.36,0.99]	0.66 [0.37,0.97]
Specificity	0.84 [0.83,0.86]	0.50 [0.11,0.76]	0.51 [0.07,0.81]
PPV	0.57 [0.54,0.64]	0.27 [0.20,0.36]	0.43 [0.35,0.52]

*Core modeling results. NACC results represent the mean and 95% confidence intervals based on a 70/30 train/test split and 100 bootstrap iterations. ADNI results represent external validation for 1-year and 3-year conversion windows.

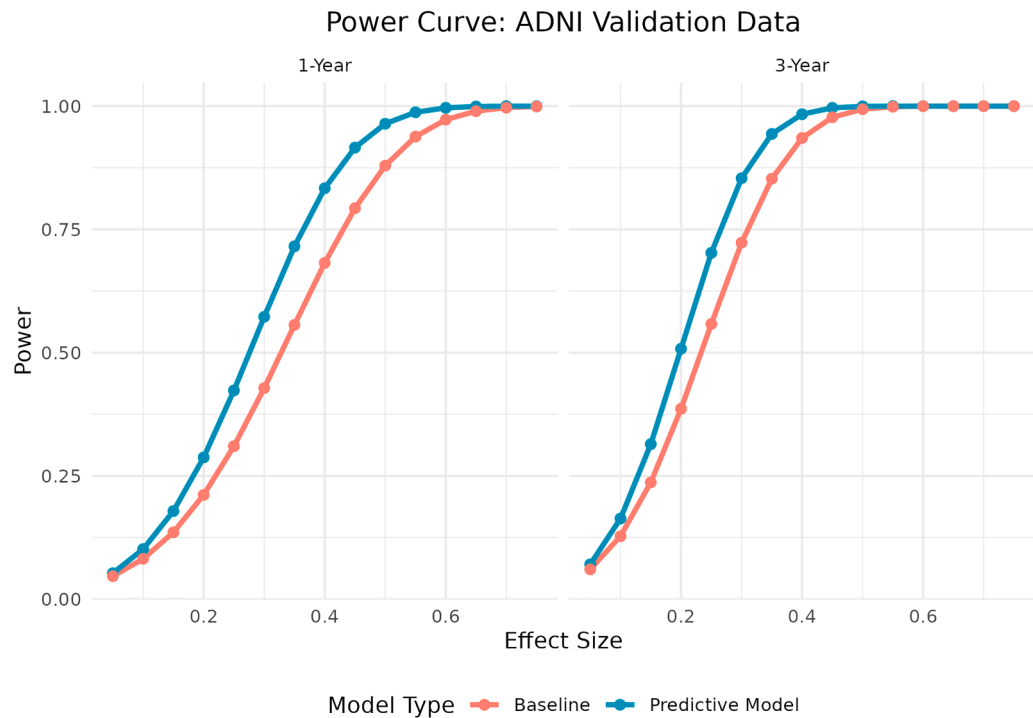


Fig. 2. Statistical power curves by treatment effect size. Comparison of projected statistical power between the baseline clinical trial enrollment strategy (biomarker-positive only) and the machine learning-based two-stage enrichment framework across an emulated trial size of 500 participants.

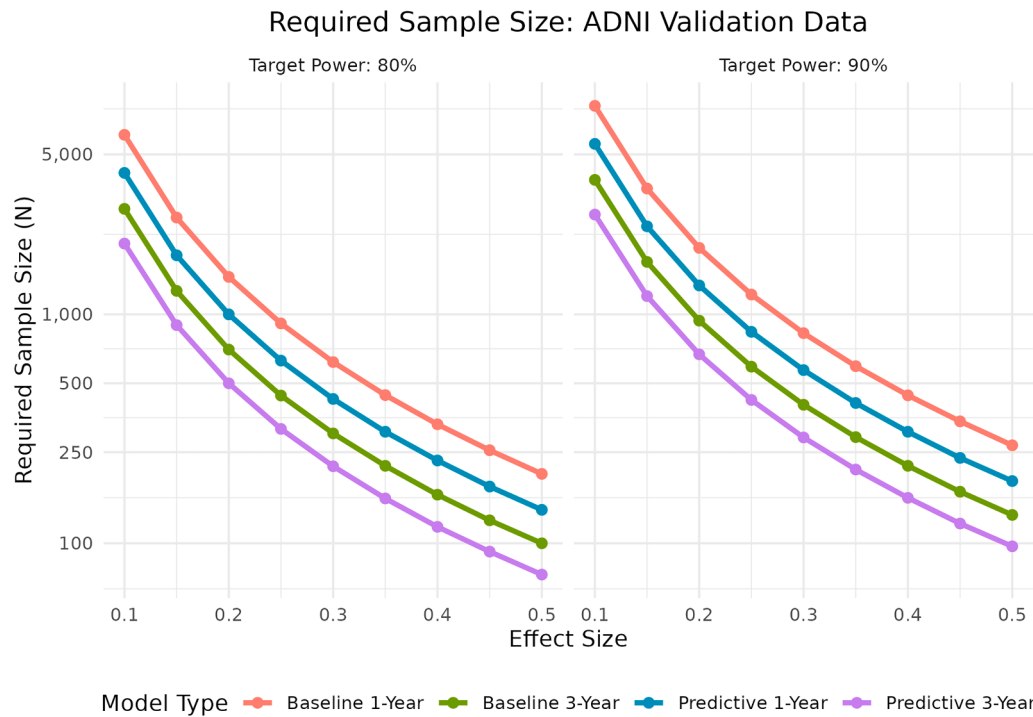


Fig. 3. Required sample size projections to achieve 80% statistical power. Total sample size requirements across varying treatment effect sizes under standard enrollment versus the predictive model enrichment framework.

illustrating the potential design impact of enrichment on event rates rather than as fully operational sample size projections under any retention assumptions. Nevertheless, the relative differences between enrollment strategies are expected to remain stable under reasonable assumptions about the dropout mechanism.

4. Discussion

In this study we examined the extent to which predictive modeling with ADRC data can be used to enhance clinical trial efficiency for AD through operational methodology and optimized enrollment. In contemporary AD trials, study enrollment is guided by predetermined

and agreed upon inclusion/exclusion criteria. Following the application of these criteria, randomization into the treatment and placebo arm occurs. Our results suggest that adding a modeling-based enrollment approach using continuous cognitive scores on top of standard inclusion/exclusion criteria might yield more optimally powered trials that can detect smaller treatment effects. In both a short (1-year) and medium (3-year) term event window, our models were able to meaningfully increase the effective event rate. As shown through power simulations, the achievable power and required sample sizes to achieve target powers consistently increased and decreased, respectively. It is well known that selective enrollment of individuals more likely to experience endpoint-relevant events leads to increased power and reduced sample sizes. Our proposal is therefore focused on the emphasis of specifically the pragmatic two-stage enrollment framework in which predictive modeling is layered into existing eligibility criteria to selectively enrich for individuals most likely to experience short-term clinical progression.

An important consideration in applying a modeling-based enrollment method is that the screen failure rate increases in the second stage of the approach. After participants satisfy the initial inclusion/exclusion criteria and complete standard screening procedures, an additional model-based filter is applied where participants with low likelihood of clinically progressing are excluded despite otherwise meeting a trial's eligibility. Operationally, this introduces a clear tradeoff: a larger number of screen participants must be evaluated to achieve a given randomized sample size while the actual enrolled sample is significantly enriched for the event of interest. For example, under our 1-year model, randomizing 250 amyloid positive cognitively normal participants would require screening approximately 213 additional candidates (463 total) to reach the same n , vs 250 under unenriched selection. From a design perspective, the key question is whether the gains in statistical power outweigh the added screening complexity. Whether a modeling based enrichment strategy such as the one proposed in this paper will ultimately be advantageous will depend on practical considerations such as site capacity, and recruitment rates. In this sense, the increased screening burden represents an operational investment to obtain a more informative trial population rather than a limitation of the strategy itself.

Our fundamental contribution is methodological and operational rather than diagnostic, focusing on enrollment enrichment and trial efficiency rather than disease classification or individual-level prognosis. The model does not identify the early presence of disease, as all participants are already known to be amyloid-positive at baseline. Instead, it attempts to answer the question: given an amyloid-positive, cognitively unimpaired cohort that already meets biological eligibility criteria, how can an investigator forecast short-term symptomatic transition to protect a trial from underpowered readouts? Thus, the risk score functions not as a diagnostic biomarker, but as a clinical trial filter designed to manage operational risk. Previous work has been done in attempting to use predictive modeling to enhance AD clinical trials either through historical clinical trials data [14] or through ADRC data [15,16]. While these approaches did demonstrate improvements in trial efficiency, they were often tied to specific datasets or narrowly defined modeling pipelines, which limit generalizability. Sample size projections should be treated as illustrative gains possible under model-based enrichment, rather as inputs to a real trial's power analysis, and thus, real-world prospective efficiency gains could be substantially lower. In contrast, we aimed to develop a reproducible framework in which trial researchers can develop their trial specific eligibility criteria and then utilize ADRC data to implement a model-based enrichment approach to enroll a study cohort. By training our models on longitudinal ADRC data and testing them on baseline only data, we are able to mimic the real-world constraint of single-visit screening while still leveraging the breadth of longitudinal data.

This study has several limitations. First, NACC and ADNI utilize different data collection and cognitive assessments protocols, requiring

the assumption that these measures share an underlying relationship with the outcome across cohorts. However, even under similar inclusion/exclusion criteria, the absolute predictive performance and PPV of this approach may vary across independent datasets because of differences in cohort composition, testing protocols, follow-up structure, and outcome prevalence. That said, the external validation in ADNI suggests that the relative ranking of enrollment strategies will be preserved. Second, NACC included longitudinally followed participants with repeated exposure to the same cognitive instruments. As a result, model development may partly reflect practice effects. This may reduce transportability to prevention trials which recruit participants assessed for the first time at screening. Further, as we focused on optimizing tightly for short-term converters, this could preferentially select individuals with lower cognitive reserve or specific baseline phenotypes. This makes the resulting cohort operationally efficient but biologically unrepresentative of the broader preclinical AD population. We view this as an inherent consequence of any prognostic enrichment strategy rather than a flaw of the approach itself, and investigators should weigh the improved event rate against any reduction in representativeness based on the scientific objectives of the trial. We do not propose utilizing our specific model, but rather incorporating the general concept, framed for the specific population and timing of interest of the user's trial. We also recognize the training data is exceptionally small for a complex ensemble method like XGBoost and may introduce the risk of model instability. While we sought to mitigate this concern through external validation in an independent cohort, the model's performance and generalizability may nevertheless be limited when applied to other populations or study settings that differ from those represented in the training and validation data. Finally, the use of the CDR as an outcome variable introduces potential noise in the modeling process since it is common for CDR levels to fluctuate between 0 and 0.5 at subsequent visits. To address this limitation, we considered a transition to be progression to CDR > 0.5 and maintaining a 0.5 or higher in future visits. However, we recognize the model still may be enriching for ascertainment tendencies, boundary effects, or general frailty rather than for disease-relevant progression specifically.

In conclusion, our findings suggest that cognitive measures provide complementary prognostic information beyond binary biomarker positivity alone, which does not guarantee clinically meaningful progression within a trial's duration. Importantly, the cognitive measures used here are familiar to clinicians and trialists, facilitating interpretability and operational adoption relative to opaque risk scores. AD trials already screen on cognition by using cutoffs. Our contribution demonstrates the potential value of replacing binary thresholds with a continuous, multivariable risk score derived from the same routine measures. Rather than serving as a singular screening gate for cognitive intactness, cognitive measures can function as quantitative enrichment tools to help assemble high-risk cohorts in which treatment effects are most likely to be detected within feasible trial timelines. Our results suggest that incorporating operational methodology into enrollment design may represent a pragmatic strategy to improve trial efficiency.

The present study contains many practical limitations and, therefore, the most critical future work arising from this study concerns real-world prospective validation of the modeling-based enrollment framework within an active or planned AD trial. Here we used predictive ML as an example of an approach that illustrates the potential benefit of model-based enrichment. Future methodological work might consider trying to bulk leverage multi-site and multi-modal data through methods like transfer learning and domain adaptation to make robust generalizable conversion forecast models. While our results paint an optimistic picture for the future of AD clinical trials, careful attention to regulatory, ethical, and equitable considerations will remain essential to ensure that these algorithmic based enrichment strategies do not inadvertently limit access or reduce diversity around trial populations.

Declaration Of generative AI and AI-assisted technologies

AI and AI-assisted technologies was not used in the writing process.

Funding

This work was supported by National Institute of Aging Boston University AD Research Center, P30AG072978. The sponsors had no role in the design and conduct of the study; in the collection, analysis, and interpretation of data; in the preparation of the manuscript; or in the review or approval of the manuscript.

Data statement

The data used in this study are publicly available to qualified investigators through approved applications to the National Alzheimer's Coordinating Center (NACC) and the Alzheimer's Disease Neuroimaging Initiative (ADNI). Because these data are subject to data-use agreements, the authors are not permitted to redistribute the datasets or provide direct links to the analytic files used in this study. Investigators interested in accessing the data should consult the NACC and ADNI websites for instructions on obtaining access: NACC: <https://www.naccdata.org>; ADNI: <https://adni.loni.usc.edu>.

CRedit authorship contribution statement

Henry Bayly: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Lindsay Salvati:** Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Steven Lenio:** Writing – review & editing. **Jesse Mez:** Writing – review & editing. **Michael L. Alosco:** Writing – review & editing. **Yorghos Tripodis:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Yorghos Tripodis reports financial support was provided by National Institute of Aging. Yorghos Tripodis reports a relationship with American Medical Association that includes: consulting or advisory. Jesse Mez reports a relationship with National Institutes of Health that includes: funding grants. Jesse Mez reports a relationship with US Department of Defense that includes: funding grants. Michael Alosco reports a relationship with National Institutes of Health that includes: funding grants. Michael Alosco reports a relationship with US Department of Defense that includes: funding grants. Michael Alosco reports a relationship with from Life Molecular Imaging Inc Research Support that includes: funding grants. Michael Alosco reports a relationship with Michael J Fox Foundation Inc that includes: consulting or advisory. Michael Alosco reports a relationship with Oxford University Press Inc that includes: consulting or advisory. Michael Alosco reports a relationship with Concussion Legacy Foundation that includes: consulting or advisory. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

NACC Data.

The NACC UDS is a standardized clinical data set collected longitudinally to better characterize Alzheimer's Disease Research Center (ADRC) participants with mild Alzheimer's dementia and mild cognitive impairment compared to cognitively normal controls [26]. The data

contains over 900 variables from participants, including demographics, features of symptom onset and course, personal medical history, concurrent medications, family history of dementia, and performance measures from neurological and neuropsychological examinations, that assess the major cognitive domains that are affected by AD.

The NACC database is funded by NIA/NIH Grant U24 AG072122. NACC data are contributed by the NIA-funded ADRCs: P30 AG062429 (PI James Brewer, MD, PhD), P30 AG066468 (PI Oscar Lopez, MD), P30 AG062421 (PI Bradley Hyman, MD, PhD), P30 AG066509 (PI Thomas Grabowski, MD), P30 AG066514 (PI Mary Sano, PhD), P30 AG066530 (PI Helena Chui, MD), P30 AG066507 (PI Marilyn Albert, PhD), P30 AG066444 (PI David Holtzman, MD), P30 AG066518 (PI Lisa Silbert, MD, MCR), P30 AG066512 (PI Thomas Wisniewski, MD), P30 AG066462 (PI Scott Small, MD), P30 AG072979 (PI David Wolk, MD), P30 AG072972 (PI Charles DeCarli, MD), P30 AG072976 (PI Andrew Saykin, PsyD), P30 AG072975 (PI Julie A. Schneider, MD, MS), P30 AG072978 (PI Ann McKee, MD), P30 AG072977 (PI Robert Vassar, PhD), P30 AG066519 (PI Frank LaFerla, PhD), P30 AG062677 (PI Ronald Petersen, MD, PhD), P30 AG079280 (PI Jessica Langbaum, PhD), P30 AG062422 (PI Gil Rabinovici, MD), P30 AG066511 (PI Allan Levey, MD, PhD), P30 AG072946 (PI Linda Van Eldik, PhD), P30 AG062715 (PI Sanjay Asthana, MD, FRCP), P30 AG072973 (PI Russell Swerdlow, MD), P30 AG066506 (PI Glenn Smith, PhD, ABPP), P30 AG066508 (PI Stephen Strittmatter, MD, PhD), P30 AG066515 (PI Victor Henderson, MD, MS), P30 AG072947 (PI Suzanne Craft, PhD), P30 AG072931 (PI Henry Paulson, MD, PhD), P30 AG066546 (PI Sudha Seshadri, MD), P30 AG086401 (PI Erik Roberson, MD, PhD), P30 AG086404 (PI Gary Rosenberg, MD), P20 AG068082 (PI Angela Jefferson, PhD), P30 AG072958 (PI Heather Whitson, MD), P30 AG072959 (PI James Leverenz, MD).

The NACC database is funded by NIA/NIH Grant U24 AG072122. SCAN is a multi-institutional project that was funded as a U24 grant (AG067418) by the National Institute on Aging in May 2020. Data collected by SCAN and shared by NACC are contributed by the NIA-funded ADRCs as follows:

Arizona Alzheimer's Center - P30 AG072980 (PI: Eric Reiman, MD); R01 AG069453 (PI: Eric Reiman (contact), MD); P30 AG019610 (PI: Eric Reiman, MD); and the State of Arizona which provided additional funding supporting our center; Boston University - P30 AG013846 (PI Neil Kowall MD); Cleveland ADRC - P30 AG062428 (James Leverenz, MD); Cleveland Clinic, Las Vegas - P20AG068053; Columbia - P50 AG008702 (PI Scott Small MD); Duke/UNC ADRC - P30 AG072958; Emory University - P30AG066511 (PI Levey Allan, MD, PhD); Indiana University - R01 AG19771 (PI Andrew Saykin, PsyD); P30 AG10133 (PI Andrew Saykin, PsyD); P30 AG072976 (PI Andrew Saykin, PsyD); R01 AG061788 (PI Shannon Risacher, PhD); R01 AG053993 (PI Yu-Chien Wu, MD, PhD); U01 AG057195 (PI Liana Apostolova, MD); U19 AG063911 (PI Bradley Boeve, MD); and the Indiana University Department of Radiology and Imaging Sciences; Johns Hopkins - P30 AG066507 (PI Marilyn Albert, PhD.); Mayo Clinic - P50 AG016574 (PI Ronald Petersen MD PhD); Mount Sinai - P30 AG066514 (PI Mary Sano, PhD); R01 AG054110 (PI Trey Hedden, PhD); R01 AG053509 (PI Trey Hedden, PhD); New York University - P30AG066512-01S2 (PI Thomas Wisniewski, MD); R01AG056031 (PI Ricardo Osorio, MD); R01AG056531 (PIs Ricardo Osorio, MD; Girardin Jean-Louis, PhD); Northwestern University - P30 AG013854 (PI Robert Vassar PhD); R01 AG045571 (PI Emily Rogalski, PhD); R56 AG045571, (PI Emily Rogalski, PhD); R01 AG067781, (PI Emily Rogalski, PhD); U19 AG073153, (PI Emily Rogalski, PhD); R01 DC008552, (M.-Marsel Mesulam, MD); R01 AG077444, (PIs M.-Marsel Mesulam, MD, Emily Rogalski, PhD); R01 NS075075 (PI Emily Rogalski, PhD); R01 AG056258 (PI Emily Rogalski, PhD); Oregon Health & Science University - P30 AG066518 (PI Lisa Silbert, MD, MCR); Rush University - P30 AG010161 (PI David Bennett MD); Stanford - P30AG066515; P50 AG047366 (PI Victor Henderson MD MS); University of Alabama, Birmingham - P20; University of California, Davis - P30 AG10129 (PI Charles DeCarli, MD); P30 AG072972

(PI Charles DeCarli, MD); University of California, Irvine - P50 AG016573 (PI Frank LaFerla PhD); University of California, San Diego - P30AG062429 (PI James Brewer, MD, PhD); University of California, San Francisco - P30 AG062422 (Rabinovici, Gil D., MD); University of Kansas - P30 AG035982 (Russell Swerdlow, MD); University of Kentucky - P30 AG028283-15S1 (PIs Linda Van Eldik, PhD and Brian Gold, PhD); University of Michigan ADRC - P30AG053760 (PI Henry Paulson, MD, PhD) P30AG072931 (PI Henry Paulson, MD, PhD) Cure Alzheimer's Fund 200775 - (PI Henry Paulson, MD, PhD) U19 NS120384 (PI Charles DeCarli, MD, University of Michigan Site PI Henry Paulson, MD, PhD) R01 AG068338 (MPI Bruno Giordani, PhD, Carol Persad, PhD, Yi Murphey, PhD) S100D026738-01 (PI Douglas Noll, PhD) R01 AG058724 (PI Benjamin Hampstead, PhD) R35 AG072262 (PI Benjamin Hampstead, PhD) W81XWH2110743 (PI Benjamin Hampstead, PhD) R01 AG073235 (PI Nancy Chiaravalloti, University of Michigan Site PI Benjamin Hampstead, PhD) 1I01RX001534 (PI Benjamin Hampstead, PhD) IRX001381 (PI Benjamin Hampstead, PhD); University of New Mexico - P20 AG068077 (Gary Rosenberg, MD); University of Pennsylvania - State of PA project 2019NF4100087335 (PI David Wolk, MD); Rooney Family Research Fund (PI David Wolk, MD); R01 AG055005 (PI David Wolk, MD); University of Pittsburgh - P50 AG005133 (PI Oscar Lopez MD); University of Southern California - P50 AG005142 (PI Helena Chui MD); University of Washington - P50 AG005136 (PI Thomas Grabowski MD); University of Wisconsin - P50 AG033514 (PI Sanjay Asthana MD FRCP); Vanderbilt University - P20 AG068082; Wake Forest - P30AG072947 (PI Suzanne Craft, PhD); Washington University, St. Louis - P01 AG03991 (PI John Morris MD); P01 AG026276 (PI John Morris MD); P20 MH071616 (PI Dan Marcus); P30 AG066444 (PI John Morris MD); P30 NS098577 (PI Dan Marcus); R01 AG021910 (PI Randy Buckner); R01 AG043434 (PI Catherine Roe); R01 EB009352 (PI Dan Marcus); UL1 TR000448 (PI Brad Evanoff); U24 RR021382 (PI Bruce Rosen); Avid Radiopharmaceuticals / Eli Lilly; Yale - P50 AG047270 (PI Stephen Strittmatter MD PhD); R01AG052560 (MPI: Christopher van Dyck, MD; Richard Carson, PhD); R01AG062276 (PI: Christopher van Dyck, MD); 1Florida - P30AG066506-03 (PI Glenn Smith, PhD); P50 AG047266 (PI Todd Golde MD PhD)

ADNI Data:

The ADNI dataset is an extensive and frequently utilized repository of longitudinal data, designed to approximate an interventional clinical trial cohort, comprising clinical, imaging, genetic, and other biomarker information¹⁹. The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The original goal of ADNI was to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). The current goals include validating biomarkers for clinical trials, improving the generalizability of ADNI data by increasing diversity in the participant cohort, and to provide data concerning the diagnosis and progression of Alzheimer's disease to the scientific community.

Data collection and sharing for the Alzheimer's Disease Neuroimaging Initiative (ADNI) is funded by the National Institute on Aging (National Institutes of Health Grant U19AG024904). The grantee organization is the Northern California Institute for Research and Education. In the past, ADNI has also received funding from the National Institute of Biomedical Imaging and Bioengineering, the Canadian Institutes of Health Research, and private sector contributions through the Foundation for the National Institutes of Health (FNIH) including 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; Euro-Immun; 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

All data used in this study were fully de-identified prior to analysis. The study involved secondary analysis of existing de-identified datasets (NACC and ADNI), and no direct contact with human participants occurred. Therefore, informed consent was not required for this analysis.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.jarlif.2026.100081](https://doi.org/10.1016/j.jarlif.2026.100081).

References

- [1] Sperling R, Aisen P, Beckett L, Bennett D, Craft S, Fagan A, et al. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* 2011;7:280-92.
- [2] Sperling R, Rentz D, Johnson K, Karlawish J, Donohue M, Salmon D, et al. The A4 study: stopping AD before symptoms begin? *Sci Transl Med* 2014;6:228fs13.
- [3] Grill J, Karlawish J. Addressing the challenges to successful recruitment and retention in Alzheimer's disease clinical trials. *Alzheimers Res Ther* 2010;2:34.
- [4] Braak H, Braak E. Neuropathological staging of Alzheimer-related changes. *Acta Neuropathol* 1991;82:239-59.
- [5] Iqbal K, Liu F, Gong C-X. Tau and neurodegenerative disease: the story so far. *Nat Rev Neurol* 2016;12:15-27.
- [6] Zammit A, Capuano A, Barnes L, Schneider J, Sperling R, Bennett D, et al. Cognitive decline across five cognitive batteries: sample size implications for clinical trials. *Alzheimer 19s Dement* 2025;11:e70137.
- [7] Cummings J, Zhou Y, Lee G, Zhong K, Fonseca J, Leisgang-Osse A, et al. Alzheimer's disease drug development pipeline: 2025. *Alzheimer 19s Dement* 2025;11:e70098.
- [8] Vos S, Xiong C, Visser P, Jasielec M, Hassenstab J, Grant E, et al. Preclinical Alzheimer's disease and its outcome: a longitudinal cohort study. *Lancet Neurol* 2013;12:957-65.
- [9] Insel P, Hansson O, Mackin R, Weiner M, Mattsson N. Alzheimer's Disease Neuroimaging Initiative. Amyloid pathology in the progression to mild cognitive impairment. *Neurobiol Aging* 2018;64:76-84.
- [10] Donohue M, Sperling R, Petersen R, Sun C, Weiner M, Aisen P, et al. Association between elevated brain amyloid and subsequent cognitive decline among cognitively normal persons. *JAMA* 2017;317:2305.
- [11] Eisai Inc. AHEAD 3-45 study: a placebo-controlled, double-blind, parallel-treatment arm, 216 week study with an extension phase to evaluate efficacy and safety of treatment with BAN2401 in subjects with preclinical Alzheimer's disease and elevated Amyloid (A45 Trial) and in subjects with early preclinical Alzheimer's disease and intermediate amyloid (A3 Trial). Clinical trial registration NCT04468659. [clinicaltrials.gov](https://clinicaltrials.gov/study/NCT04468659); 2024. <https://clinicaltrials.gov/study/NCT04468659>, accessed 7 October 2025.
- [12] Petersen R, Graf A, Brady C, De Santi S, Florian H, Landen J, et al. Operationalizing selection criteria for clinical trials in Alzheimer's disease: biomarker and clinical considerations: proceedings from the Alzheimer's Association Research Roundtable (AARR) Fall 2021 meeting. *Alzheimers Dement* 2025;11:e70038.
- [13] Sperling R, Donohue M, Raman R, Rafi M, Johnson K, Masters C, et al. Trial of Solanezumab in preclinical Alzheimer's disease. *N Engl J Med* 2023;389:1096-107.
- [14] Ezzati A, Davatzikos C, Wolk D, Hall C, Habeck C, Lipton RB. Application of predictive models in boosting power of Alzheimer's disease clinical trials: a post hoc analysis of phase 3 solanezumab trials. *Alzheimers Dement* 2022;8:e12223.
- [15] Prediction of cognitive decline for enrichment of Alzheimer's Disease clinical trials. *J Prev Alzheimer 19s Dis* 2022;9:400-9.
- [16] Tam A, Laurent C, Colina A. de L., Dansereau C. AI-driven patient selection can drastically reduce sample sizes in Alzheimer's disease clinical trials. DOI: 10.1002/alz70859.100413.
- [17] Younes K, Johns E, Young C, Kennedy G, Mukherjee S, Vossler H, et al. Amyloid PET predicts longitudinal functional and cognitive trajectories in a heterogeneous cohort. *Alzheimer 19s Dement* 2025;21:e70075.
- [18] Biel D, Luan Y, Brendel M, Hager P, Dewenter A, Moscoso A, et al. Combining tau-PET and fMRI meta-analyses for patient-centered prediction of cognitive decline in Alzheimer's disease. *Alzheimers Res Ther* 2022;14:166.
- [19] Oikonomou E, Thangaraj P, Bhatt D, Ross J, Young L, Krumholz H, et al. An explainable machine learning-based phenomapping strategy for adaptive predictive enrichment in randomized clinical trials. *npj Digit Med* 2023;6:217.
- [20] Eli Lilly and Company. A study of Donanemab versus Placebo in participants at risk for cognitive and functional decline of Alzheimer's disease. Clinical trial registration NCT05026866. [clinicaltrials.gov](https://clinicaltrials.gov/study/NCT05026866); 2026. <https://clinicaltrials.gov/study/NCT05026866>, accessed 12 February 2026.

- [21] Monsell S, Dodge H, Zhou X, Bu Y, Besser L, Mock C, et al. Results from the NACC uniform data set neuropsychological battery crosswalk study. *Alzheimer Dis Assoc Disord* 2016;30:134–9.
- [22] Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–94.
- [23] Lundberg S, Lee S-I. A unified approach to interpreting model predictions. Epub ahead of print 25 November 2017. DOI: 10.48550/arXiv.1705.07874.
- [24] Swanson C, Zhang Y, Dhadda S, Wang J, Kaplow J, Lai R, et al. A randomized, double-blind, phase 2b proof-of-concept clinical trial in early Alzheimer's disease with lecanemab, an anti- $\text{A}\beta$ protofibril antibody. *Alzheimers Res Ther* 2021;13:80.
- [25] Company EL and. Lilly's donanemab slowed Alzheimer's disease progression in phase 2 trial: full data presented at AD/PD™ 2021 and published in NEJM, <https://www.prnewswire.com/news-releases/lillys-donanemab-slowed-alzheimers-disease-progression-in-phase-2-trial-full-data-presented-at-adpd-2021-and-published-in-nejm-301246745.html> (accessed 5 November 2025).
- [26] Beekly D, Ramos E, Lee W, Deitrich W, Jacka M, Wu J, et al. The National Alzheimer's Coordinating Center (NACC) database: the uniform data set. *Alzheimer Dis Assoc Disord* 2007;21:249–58.