arXiv:2404.17576v1 [stat.AP] 26 Apr 2024

Enhancing Longitudinal Clinical Trial Efficiency with Digital Twins and Prognostic Covariate-Adjusted Mixed Models for Repeated Measures (PROCOVA-MMRM)

Jessica L. Ross[1], Arman Sabbaghi[1], Run Zhuang[1], Daniele Bertolini[1],
for the ADCS[*], the Alzheimer's Disease Neuroimaging Initiative[†], the CPAD Database[‡], the European Prevention of Alzheimer's Disease (EPAD) Consortium[§], and the Pooled Resource Open-Access ALS Clinical Trials Consortium[¶]

[1]Unlearn.AI, Inc., San Francisco, CA

April 29, 2024

Abstract

Clinical trials are critical in advancing medical treatments but often suffer from immense time and financial burden. Advances in statistical methodologies and artificial intelligence (AI)

[*]Data used in preparation of this manuscript/publication/article were obtained from the University of California, San Diego Alzheimer's Disease Cooperative Study (https://www.adcs.org/)

[†]Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: https://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

[‡]Data used in the preparation of this article were obtained from the Critical Path for Alzheimer's Disease (CPAD) Database. In 2008, Critical Path Institute, in collaboration with the Engelberg Center for Health Care Reform at the Brookings Institution, formed the Coalition Against Major Diseases (rebranded to Critical Path for Alzheimer's Disease (CPAD) consortium in 2018). The consortium brings together patient groups, biopharmaceutical companies, and scientists from academia, the U.S. Food and Drug Administration (FDA), the European Medicines Agency (EMA), the National Institute of Neurological Disorders and Stroke (NINDS), and the National Institute on Aging (NIA). The Critical Path for Alzheimer's Disease (CPAD) consortium includes over 200 scientists from member and non-member organizations. The data available in the CPAD database has been volunteered by CPAD member companies and non-member organizations.

[§]Data used in preparation of this article were obtained from the Longitudinal Cohort Study (LCS), delivered by the European Prevention of Alzheimer's Disease (EPAD) Consortium. As such investigators within the EPAD LCS and EPAD Consortium contributed to the design and implementation of EPAD and/or provided data but did not participate in analysis or writing of this report. A complete list of EPAD Investigators can be found at: https://ep-ad.org/wp-content/uploads/2020/12/202010_List-of-epadistas.pdf

[¶]Data used in the preparation of this article were obtained from the Pooled Resource Open-Access ALS Clinical Trials (PRO-ACT) Database. As such, the following organizations and individuals within the PRO-ACT Consortium contributed to the design and implementation of the PRO-ACT Database and/or provided data, but did not participate in the analysis of the data or the writing of this report: Neurological Clinical Research Institute, MGH; Northeast ALS Consortium; Novartis; Prize4Life Israel; Regeneron Pharmaceuticals, Inc.; Sanofi; Teva Pharmaceutical Industries, Ltd.

present opportunities to address these inefficiencies. Here we introduce Prognostic Covariate-Adjusted Mixed Models for Repeated Measures (PROCOVA-MMRM) as an advantageous combination of prognostic covariate adjustment (PROCOVA) and Mixed Models for Repeated Measures (MMRM). PROCOVA-MMRM utilizes time-matched prognostic scores generated from AI models to enhance the precision of treatment effect estimators for longitudinal continuous outcomes, enabling reductions in sample size and enrollment times. We first provide a description of the background and implementation of PROCOVA-MMRM, followed by two case study reanalyses where we compare the performance of PROCOVA-MMRM versus the unadjusted MMRM. These reanalyses demonstrate significant improvements in statistical power and precision in clinical indications with unmet medical need, specifically Alzheimer's Disease (AD) and Amyotrophic Lateral Sclerosis (ALS). We also explore the potential for sample size reduction with the prospective implementation of PROCOVA-MMRM, finding that the same or better results could have been achieved with fewer participants in these historical trials if the enhanced precision provided by PROCOVA-MMRM had been prospectively leveraged. We also confirm the robustness of the statistical properties of PROCOVA-MMRM in a variety of realistic simulation scenarios. Altogether, PROCOVA-MMRM represents a rigorous method of incorporating advances in the prediction of time-matched prognostic scores generated by AI into longitudinal analysis, potentially reducing both the cost and time required to bring new treatments to patients while adhering to regulatory standards.

1 Introduction

Clinical trials are notoriously inefficient, requiring immense time and financial investment to introduce potentially life-saving treatments to the market (Fogel, 2018). In addition, 80% of trials fail to meet their initial enrollment targets (Brøgger-Mikkelsen et al., 2020), which can result in insufficient power. Sponsors can improve the precision and power of their trial by including more participants, but the associated increases in trial time and expense make alternative strategies for improving trial efficiency more appealing.

One such strategy is to utilize observed outcome information across all timepoints in a study (longitudinally) to estimate the average treatment effect at the final timepoint for the full trial population. The limitations of common Analysis of (Co)Variance (ANOVA/ANCOVA)-based methods in dealing with routine challenges in randomized clinical trials (RCTs), such as missing data, have become apparent, and in recent years, more sophisticated methods like Mixed Models for Repeated Measures (MMRM) and Generalized Estimating Equations (GEE) have grown in favor. While all statistical methods require some assumptions when considering the pattern of missing data/participant dropout, the assumptions required for ANOVA-based methods are more rigid than those required by MMRM and GEE, which can handle missing data without explicit imputation (Garcia and Marder, 2017; EMA, 2010). Data can be missing completely at random (MCAR), where there is no pattern to the missing data. More often, trials include data that are missing at random (MAR), which means that the missing data can be explained by observed variables in the study (Mallinckrodt et al., 2008), such as baseline covariates. For example, patients with worse functional measures at baseline might be more likely to drop out of a clinical trial. As their missingness would be partially explained by a variable measured in the study, this would be considered MAR. MMRM allows for strict type I error control and unbiased estimation of treatment effects for both MAR and

MCAR data (Mallinckrodt et al., 2001; Schuler, 2022), whereas the common practice of ANOVA with last observation carried forward (LOCF) and GEE can only be used safely with MCAR data (Mallinckrodt et al., 2001; Garcia and Marder, 2017). MMRM is effectively an extension of the linear regression model, in which the model is specified

- for all observed measurements on all participants,
- with different intercepts and treatment effects at distinct timepoints, and
- with a specified structure on the covariance matrix for the random error term in the model.

MMRM is suitable under current regulatory guidance (EMA, 2010) and is less sensitive to dropout bias and provides more power over standard analysis methods for continuous endpoints, contributing to its popularity for the analysis of phase 2 and 3 RCTs (Kabashima et al., 2020; Richeldi et al., 2022; Jastreboff et al., 2022).

Another strategy for increasing power in RCTs is through the use of covariate adjustment. Regulatory guidance recommends that only a small number of covariates be used for adjustment in the statistical analysis of clinical trials (FDA, 2023; EMA, 2015); but, determining the small set of covariates that can yield the greatest reduction in the variance of the treatment effect estimator can be challenging in the case of the many baseline measurements assessed in RCTs. Schuler et al. (2021) has shown that the optimal adjustment covariate in the estimation of a treatment effect would be the participant's expected outcome under control, conditional on the covariates. While a participant's expected control outcome cannot be known at study onset, the use of artificial intelligence (AI) can yield powerful predictions of expected control outcomes, or prognostic scores (Hansen, 2008), conditional on baseline covariates. AI models can handle complex, high-dimensional data and identify patterns that might not be apparent through conventional statistical models. As the availability of rich longitudinal participant data from historical clinical trials and observational studies has increased, so too has the predictive ability of AI models trained on this data. In fact, the European Medicines Agency (EMA) recently qualified a novel methodology, prognostic covariate adjustment (PROCOVA), which can enhance the efficiency of phase 2 and 3 RCTs without adding bias or risk of type I error in the case of continuous endpoint analysis. The power gain afforded by PROCOVA can also be anticipated during trial planning to prospectively reduce sample size over traditional analysis methods (EMA, 2022).

Here, we introduce PROCOVA-MMRM, an extension of PROCOVA intended for phase 2 and 3 RCTs with longitudinal continuous endpoints. Whereas PROCOVA is used for single timepoint analyses, PROCOVA-MMRM integrates prognostic covariate adjustment into the longitudinal analysis of continuous endpoints measured at multiple timepoints. Through the inclusion of time-matched prognostic scores, generated for each trial participant from only their baseline characteristics, as adjustment covariates, PROCOVA-MMRM can yield the following desirable results:

> Improved precision of treatment effect inferences compared to traditional MMRM analyses for a fixed sample size that either have no adjustment or that adjust solely for baseline covariates, and
>
> Reductions in RCT enrollment time and control arm sample size without sacrificing trial power.

Improvements in this framework for longitudinal outcomes have mirrored advances in prognostic modeling. The Conditional Restricted Boltzmann Machines (CRBMs) (Fisher et al., 2019;

Bertolini et al., 2020) that were used to support PROCOVA predicted at a 3 month cadence, so their utility with longitudinal outcomes was limited to studies with the same visit structure. Lang et al. (2023) and Unlearn.AI (2024) recently introduced the Neural Boltzmann Machine (NBM) architecture as an extension of CRBMs that allows for a more flexible representation of the underlying data. The NBM model can handle missing data and generate prognostic scores at arbitrary timepoints (rather than fixed timepoints), making it well-equipped to be used in trials that longitudinally collect continuous outcome measurements. Here, we demonstrate the efficiency gain achievable over traditional methods through the use of time-matched prognostic scores (i.e., predictions of control outcomes for an individual trial participant at planned observation timepoints) generated by NBMs for covariate adjustment in PROCOVA-MMRM. We begin with a description of the general methodology and step-by-step implementation of PROCOVA-MMRM, followed by case studies utilizing NBMs for Alzheimer's disease (AD) and amyotrophic lateral sclerosis (ALS) to demonstrate the practical implementation of this method, either for power gain or prospective sample size reduction, in clinical indications with unmet medical needs. Finally, we perform a series of simulations to demonstrate the robustness of the procedure across a variety of scenarios not encountered in our case studies.

2 The PROCOVA-MMRM Model

A Notations and Assumptions

We consider an RCT with N participants who are randomized to either the active treatment arm (denoted by $w_i = 1$ for participant i) or the control arm (denoted by $w_i = 0$). The vector of time-matched prognostic scores for each participant i is denoted by $\mathbf{x}_i$, with entry $x_{i,t} \in \mathbb{R}$ for time t determined from observations measured prior to treatment assignment. The continuous endpoint $y_{i,t} \in \mathbb{R}$ is measured at each time point $t = 1, \ldots, T_i$ for participant i, where $T_i \in \{1, \ldots, T\}$ is the final time point for which participant i has an observed outcome. Note that T_i is specified because of the potential for missing data, thus the last observed timepoint for participant i could differ from the last observed timepoint for another participant. We let $\mathbf{y}_i = (y_{i,1}, \ldots, y_{i,T_i})'$ denote the vector of observed outcomes for participant i. The trial dataset is the set of tuples $\{w_i, \mathbf{x}_i, \mathbf{y}_i : i = 1, \ldots, N\}$.

PROCOVA-MMRM is a form of the MMRM-II working model (with interactions) described by Wang and Du (2023). This model is

$$[\mathbf{y}_i \mid w_i, \mathbf{x}_i] \sim \text{Normal}_{T_i}\left(\begin{pmatrix}\beta_{0,1}\\ \vdots \\ \beta_{0,T_i}\end{pmatrix} + w_i \begin{pmatrix}\beta_{w,1}\\ \vdots \\ \beta_{w,T_i}\end{pmatrix} + \mathbf{B}_x\mathbf{x}_i, \mathbf{R}_i\right) \tag{1}$$

where $\mathbf{B}_x = \begin{pmatrix}\boldsymbol{\beta}'_{x,1}\\ \boldsymbol{\beta}'_{x,2}\\ \cdots \\ \boldsymbol{\beta}'_{x,T_i}\end{pmatrix}$ is the $T_i \times T_i$ matrix of unknown model coefficients in which each row consists of a distinct set of parameters indexed by time for the T_i-dimensional vector of time-matched prognostic scores and $\mathbf{R}_i$ is the $T_i \times T_i$ positive definite covariance matrix for the random error term of participant i. The predictor vectors in this model are effectively defined in terms of the indicators for the timepoints $\mathbb{I}_t$, the products of the timepoint indicators with the treatment indicator $w_i\mathbb{I}_t$,

and the products of the timepoint indicators with the prognostic scores $x_{i,t}\mathbb{I}_t$. Here, the indicator vector $\mathbb{I}_t$ is a $T \times 1$ vector that has a 1 in entry t, and zeros for all other entries.

It is important to recognize that, in this model, the time-matched prognostic scores are not a function of treatment nor of any intermediate outcomes, and are purely a transformation of the baseline covariates via a prognostic model that can be calculated and observed at baseline. Consequently, the prognostic scores are not time-varying covariates (Diggle et al., 2013), but are equivalent to a vector of baseline covariates.

PROCOVA-MMRM proceeds by fitting an MMRM model with the adjustment for time-matched prognostic scores $x_{i,t}$ (including interactions with the time indicator) to the data using Restricted Maximum Likelihood (REML) with an unstructured covariance matrix. Robust standard errors are utilized for the treatment effect estimator. Once the model has been fitted, the point estimate, p-value, and confidence interval for $\beta_{w,T}$ (which is the estimand of interest) are calculated.

B Sample Size Estimation

PROCOVA-MMRM yields unbiased treatment effect estimators with desired interval coverage rates and controlled type I error rates in the large sample setting. Closed-form prospective sample size reduction formulae for the MMRM-II working model do not exist. Using simulation studies, efficiency gains (i.e., power boost and improved precision for treatment effect estimation) using PROCOVA-MMRM can be quantified with respect to the traditional MMRM analyses that either do not adjust for any covariates or that adjust solely for baseline covariates. This comparison using specific trial characteristics enables the fullest benefit of PROCOVA-MMRM for sample size reduction to be quantified.

As an alternative, one may use the PROCOVA formula for sample size reduction (EMA, 2022). PROCOVA-MMRM will always achieve a lower variance of the treatment effect estimator over complete-case PROCOVA due to the additional longitudinal information incorporated into the model (Appendix B). Thus, the PROCOVA sample size estimation formula can also be utilized as a conservative estimate for sample size in PROCOVA-MMRM. This implementation of PROCOVA-MMRM is described in the next section.

3 Implementation for Prospective Trial Planning

A Step 1: Validate the Prognostic Score for Use in the Target Trial

The purpose of Step 1 is to validate the prognostic score generated by a prognostic model for use in a particular planned trial, which we will call the Target Trial. In the case that one plans to use the PROCOVA sample size estimation formula as a conservative estimate, the validation will require estimating the Pearson correlation coefficient R between the end-of-study prognostic score and the actual outcomes obtained from a separate dataset that was not used to train the prognostic model, which contains data from subjects whose baseline characteristics are similar to those in the Target Trial.

B Step 2: Estimate Sample Size and Plan the Target Trial Taking the Prognostic Score into Account

B.1 Step 2a: Estimate Sample Size

Gather the standard inputs needed to compute a sample size for a given power (i.e., the target effect size, the standard deviation of the outcome, the proportion of subjects to be randomized to the intervention, the expected dropout rate, and the alpha level), and define the inflation and deflation factors used to avoid undue optimism in the PROCOVA sample size calculation, i.e., λ (the deflation factor for the correlation coefficient R) and γ (the inflation factor for the standard deviation). Use the rules outlined in the PROCOVA Handbook for the Target Trial Statistician to select λ and γ (EMA, 2022). With all of the parameters now defined, compute the power over a range of sample sizes to obtain a power curve, and choose the minimum sample size to achieve the desired power for the Target Trial. In general, the PROCOVA sample size formula does not require a 1:1 randomization and allows for the λ and γ values to differ between the control and treatment groups. Here, we provide an example of a simplified formula where a common γ and a common λ are used for two equally sized treatment groups. We also assume that the outcomes have a common variance σ^2; that n represents the trial sample size without dropouts; and that d represents the proportion of subjects who drop out.

- Estimate the standard error ν of the treatment effect via $\sqrt{\{n(1-d)\}^{-1}(2\gamma\sigma)^2\left\{1-(\lambda R)^2\right\}}$.
- Compute the power as $\Phi(\Phi^{-1}(\alpha/2)+\beta/\nu)+\Phi(\Phi^{-1}(\alpha/2)-\beta/\nu)$, where $\Phi(\cdot)$ is the cumulative distribution function of the standard Normal distribution, α is the type I error rate, and β is the target treatment effect.
- For a range of n (e.g., 100-1000) compute the power and plot the corresponding curve.

If the Target Trial utilizes co-primary endpoints, repeat the bulleted steps above for each endpoint. Select the sample size that would be sufficient to simultaneously address the null hypothesis for both endpoints. Note that repeating the process for multiple endpoints necessitates having a prognostic score for each endpoint, and that, in turn, necessitates either a machine learning model that predicts multivariate results or multiple individual models for each endpoint of interest.

B.2 Step 2b: Plan the Target Trial and Pre-Specify its Characteristics

To apply PROCOVA-MMRM, a validated prognostic model used to generate the time-matched prognostic scores for the Target Trial must be locked and prespecified in the Target Trial protocol and/or statistical analysis plan (SAP) as appropriate, prior to implementation, to prevent data-driven model selection. This documentation should also include model weights and any hyperparameters associated with the sampling necessary to compute the prognostic scores.

- Consider if individual baseline covariates will be included (along with the time-matched prognostic scores) in the primary analysis and/or sensitivity analysis. When PROCOVA-MMRM is applied to trials utilizing stratified randomization, the strata should be included as covariates in the primary analysis (note that the prognostic scores are not designed to be used for stratification).

- Choose the strategy for handling dropout by defining the appropriate estimand. For example, the estimand could be treatment policy or hypothetical depending on how missing data are handled, even though both methods can use PROCOVA-MMRM.

- Select and pre-specify sensitivity analyses that probe the assumptions of PROCOVA-MMRM, as you would with MMRM alone. To evaluate the ability of PROCOVA-MMRM to address missing data, methods such as MICE (White et al., 2011) or tipping point analyses (Liublinska and Rubin, 2014) can be applied. Additional sensitivity analyses for PROCOVA-MMRM may include principal stratification (EMA, 2020), unadjusted MMRM, and/or MMRM with baseline covariates, dependent on Trial Statistician preference.

- Although PROCOVA-MMRM uses an unstructured covariance matrix, select and pre-specify at least two backup options for the structure of the covariance matrix to account for the possibility that the unstructured covariance matrix fails to converge. We recommend using a Toeplitz structure with an additional backup of a compound symmetry covariance structure in the case that both the unstructured and Toeplitz matrices fail to converge.

- Document all analysis choices described in Step 2b in the Target Trial protocol and/or SAP, as appropriate, prior to implementation.

C Step 3: Analyze Target Trial Results Using MMRM That Adjusts for the Time-Matched Prognostic Scores.

Step 3a: When clean baseline data on all participants randomized into the Target Trial are available, provide them to the model developers for the generation of trial-specific prognostic scores. The model developers must remain blinded to the randomization code until after the final trial-specific prognostic score is delivered to the Target Trial Statistician and applied in treatment effect estimation as described below.

Step 3b: Estimate the treatment effect with PROCOVA-MMRM using MMRM that adjusts for the time-matched prognostic scores. If specified in the protocol and SAP, adjust for additional covariates (including stratification factors in the case of trials with stratified randomization). In all cases, estimate the variance for the treatment effect using a heteroskedasticity-consistent (HC) covariance matrix. In SAS, this is the empirical option in PROC MIXED which computes the Huber-White covariance of the fixed parameters (Huber, 1967; White, 1980). To achieve similar results, the implementation in R uses the `mmrm` package with the Kenward-Roger procedure to adjust degrees of freedom. The treatment effect estimator corresponds to the estimator of $\beta_{w,T}$ from the fit of the PROCOVA-MMRM model to the entire data. The treatment effect estimate, as well as the standard error associated with the estimate, are obtained in SAS via the LSMEANS statement.

For this primary analysis, test the null hypothesis by computing a two-sided p-value based on a t-distribution using the covariate-adjusted treatment effect and the HC standard error.

4 Case Studies

Recruitment and enrollment are known challenges for Phase 2 and 3 RCTs in both AD and ALS. Furthermore, sufficient historical data (from RCTs and observational studies) exist that can be used to train prognostic models for each of these conditions. Thus, to illustrate the broad utility of PROCOVA-MMRM across therapeutic indications where highly predictive prognostic models are available, we reanalyze two previously completed phase 2/3 clinical trials in individuals with AD or ALS.

A General Methods

The AI models for AD and ALS that were used to generate time-matched prognostic scores for trial participants in these case studies utilized a newer model architecture that improved upon the previous CRBMs (Fisher et al., 2019; Bertolini et al., 2020), which were limited by their underlying representational power and restriction of prediction times. To overcome these limitations, we generalized CRBMs to Neural Boltzmann Machines (NBM), which promote parameters of a CRBM to full neural networks and allow for greater flexibility. A comprehensive discussion of the model architecture is outside the scope of this manuscript and we refer to Lang et al. (2023) for details. In brief, the main objective of the model is to predict a multivariate distribution of outcomes over time, when prompted with the measured baseline characteristics of a given trial participant. Two separate neural networks, which take as input baseline characteristics, predict the mode and the per-participant precision of the multivariate distribution at arbitrary followup times. Trajectories for a given trial participant (i.e., the samples forming their digital twin distribution) are then generated autoregressively, where predictions at a future time t_{future} are produced from baseline and from predictions at the previous time t_{current}. A third neural network, a linear auto-encoder, imputes missing baseline values and it is co-trained with the NBM. This new architecture enables greater flexibility in missing data handling and prediction cadence, making it particularly powerful for generating time-matched prognostic scores.

Rather than a single fixed 50% / 20% / 30% split into training, validation, and test datasets as was performed with previous versions of our model for AD progression (Fisher et al., 2019; Bertolini et al., 2020), the filtered training data was divided into five folds of similar size according to the following strategy: clinical trials were assigned to individual folds while observational studies were split across folds and stratified by baseline severity and other basic demographic variables. Assigning clinical trials to individual folds makes cross-validation more robust and improves estimation of how performance will generalize to future trials. On the other hand, for observational studies, which are typically larger, stratification ensures that the severity distribution in each fold roughly matches the distribution over the full population. For each run of a potential model, four of the folds were used for training and one was held out for evaluation of validation performance. This process was repeated until each fold had been held out for evaluation once, and the mean and standard deviation of these five evaluations of independent model training were used to represent the spread of possible performance of the prognostic model. In order to form a more robust prognostic model for real-world use cases, we use these five independently trained prognostic models to form an ensemble sampler, with each model contributing one fifth of the requested samples that form the digital twin distribution for each individual trial participant. This cross-validation procedure was performed for both the AD and ALS prognostic models.

The first reanalysis was performed using a phase 3 study investigating the effect of docosahexaenoic acid (DHA) on cognitive function in patients with AD (Quinn et al., 2010), to be referred to hereafter as the AD Target Trial. The second reanalysis was conducted on a phase 2/3 study (Cudkowicz et al., 2014) investigating the effect of ceftriaxone on function and survival in patients with ALS, to be referred to hereafter as the ALS Target Trial (NCT00349622). Importantly, neither of these Target Trial data sets were used in model training. For each of the trials, we present the results of two experiments:

Experiment 1. Pre-specified primary analysis of phase 2 and 3 trials, to deliver higher power/confidence in the results compared to unadjusted analyses.

Experiment 2. Prospective design/sample size estimation for phase 2 and 3 trials, to attain the desired level of power/confidence with a smaller sample size compared to unadjusted designs. This follows the implementation procedure using PROCOVA sample size estimation as described in Section 3.

We show that PROCOVA-MMRM yields smaller estimated variances of the treatment effect estimators in each of the re-analyzed studies, regardless of whether a conservative reduction in sample size (via PROCOVA sample size estimation) is prospectively applied. The MMRM and PROCOVA-MMRM models were fit using PROC MIXED in SAS Version 9.4 for LIN64 (SAS Institute, Cary, NC). The robust heteroskedasticity-consistent variance estimator was used to perform statistical tests using $\alpha = 0.05$. Validity of the assumptions for the analyses were assessed via standard diagnostics.

B Case Study 1 - Alzheimer's Disease

Data from the AD Target Trial were obtained from the University of California, San Diego through the Alzheimer's Disease Cooperative Study (ADCS), a consortium of academic medical centers and private AD clinics funded by the National Institute on Aging to conduct clinical trials on AD (NCT00440050). In this trial, 238 participants were randomized to the active treatment arm, and 164 participants were randomized to placebo (3:2 randomization). The AD Target Trial measured multiple baseline characteristics including demographics (e.g., sex, age, weight, geographic region), lab tests (e.g., blood pressure, ApoE4 status) (Coon et al., 2007; Safieh et al., 2019), and component scores of cognitive tests. Our reanalysis was performed on both co-primary endpoints in this trial: the 18-month change in the Alzheimer's Disease Assessment Scale–Cognitive Subscale 11 (ADAS-Cog 11) score (Rosen et al., 1984) and the 18-month change in the Clinical Dementia Rating Scale Sum of Boxes (CDR-SB) score (Morris, 1993).

Along with the change in model architecture from CRBM to NBM, this version of the prognostic model included additional variables and was trained on a significantly larger historical dataset. A dataset of 25,212 participants, covering 76 background and longitudinal variables, was used to train the prognostic model to predict control outcomes of a trial participant given their baseline characteristics. These historical data were obtained from the National Alzheimer's Coordinating Center (NACC) (Besser et al., 2018), the Knight Alzheimer's Disease Research Center (Koenig et al., 2020), and three additional data sources. Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner,

MD. The primary goal of ADNI has been 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 AD (Petersen et al., 2010). Data used in the preparation of this article were obtained from the Critical Path for Alzheimer's Disease (CPAD) Database. As such, the investigators within CPAD contributed to the design and implementation of the CPAD database and/or provided data, but did not participate in the analysis of the data or the writing of this report (Romero et al., 2009; Neville et al., 2015). Data used in preparation of this article were obtained from the EPAD Longitudinal Cohort Study (LCS) dataset V.IMI, doi:10.34688/epadlcs_v.imi_20.10.30. The EPAD LCS was launched in 2015 as a public-private partnership, led by Chief Investigator Professor Craig Ritchie MB BS. The primary research goal of the EPAD LCS (NCT02804789) is to provide a well-phenotyped probability-spectrum population for developing and continuously improving disease models for Alzheimer's disease in individuals without dementia (Solomon et al., 2018). Importantly, the training dataset did not include data from the AD Target Trial.

B.1 AD Experiment 1

After fitting the prognostic model and obtaining time-matched prognostic scores for each study visit, we analyzed the results from the AD Target Trial using four approaches: MMRM; MMRM with covariate adjustment for baseline (BL) ADAS-Cog-11 or CDR-SB score; PROCOVA-MMRM; and PROCOVA-MMRM with adjustment for BL ADAS-Cog-11 or CDR-SB score as an additional covariate. These reanalyses used the same number of participants and randomization ratio as in the original analysis reported by Quinn et al. (2010). The point estimates and variance of the treatment effect for ADAS-Cog11 (Table 1) and CDR-SB (Table 2) at 18 months were obtained in each of the four analysis conditions. All variance data presented are heteroskedasticity consistent and were computed using the Huber-White estimator.

Table 1 Estimated treatment effects and variances for ADAS-Cog11 at month 18 in the reanalysis of the AD Target Trial.

Analysis Method	Treatment Effect Estimate	Estimated Variance
Unadjusted MMRM	-0.035	1.024
MMRM + BL ADAS-Cog11	-0.028	0.955
PROCOVA-MMRM	0.158	0.907
PROCOVA-MMRM + BL ADAS-Cog11	0.117	0.857

Table 2 Estimated treatment effects and variances for CDR-SB at month 18 in the reanalysis of the AD Target Trial.

Analysis Method	Treatment Effect Estimate	Estimated Variance
Unadjusted MMRM	-0.092	0.113
MMRM + BL CDR-SB	-0.022	0.108
PROCOVA-MMRM	-0.116	0.099
PROCOVA-MMRM + BL CDR-SB	-0.038	0.093

We also evaluated how well the outcomes predicted by our prognostic model correlated with

the subsequent study observations. For ADAS-Cog11, the correlation between the time-matched prognostic scores generated from baseline data and the study visit observations was 0.36, 0.34, and 0.38 at 6, 12, and 18 months, respectively. For CDR-SB, the correlation between the time-matched prognostic scores predicted from baseline data only and the actual outcomes was 0.34, 0.37, and 0.37 at 6, 12, and 18 months.

B.2 AD Experiment 2

In this second experiment, we analyze the performance of PROCOVA-MMRM vs MMRM when following the prospective implementation procedure described in Section III. To use the PROCOVA formula as a conservative estimate for PROCOVA-MMRM, we first needed to understand the Pearson correlation between the AD prognostic model predictions and actual outcomes in historical data that has similar characteristics to the AD Target Trial. Thus, a subset of the modeling data with baseline MMSE [14, 26] and age greater than 50 was used to estimate Pearson correlation coefficients of 0.267 and 0.361, for ADAS-Cog/11 and CDR-SB, respectively. This translates to an estimated sample size reduction of 7.1% and 13.0% using the PROCOVA formula with no deflation factor applied and assuming a gamma (inflation factor for the standard deviation) value of 1. This sample size reduction was then applied by randomly removing participants from both treatment and control arm of the AD Target Trial for PROCOVA-MMRM analysis. This process was repeated 1000 times and the average of the treatment effect variances with PROCOVA-MMRM was calculated. The estimated potential sample size reduction was accurate if the calculated treatment effect variance with PROCOVA-MMRM was equal to or less than the unadjusted MMRM estimate.

Table 3 Comparison of sample size, estimated treatment effect, and variance for ADAS-Cog11 at month 18 in reanalysis of the AD Target Trial.

Analysis Method	Sample Size	Treatment Effect Estimate	Estimated Variance
Unadjusted MMRM	402	-0.035	1.024
PROCOVA-MMRM	334	0.160	0.979

Table 4 Comparison of sample size, estimated treatment effect, and variance for CDR-SB at month 18 in reanalysis of the AD Target Trial.

Analysis Method	Sample Size	Treatment Effect Estimate	Estimated Variance
Unadjusted MMRM	402	-0.092	0.113
PROCOVA-MMRM	321	-0.115	0.114

Results in Table 3 shows that the treatment effect variances for PROCOVA-MMRM was lower than the unadjusted MMRM, indicating that the estimated sample size reduction was accurate. Similarly, Table 4 shows very similar results between unadjusted MMRM and PROCOVA-MMRM using 13% fewer subjects. We did not use a deflation factor for this exploratory analysis, but it can be useful in cases where the original estimate may be slightly optimistic. Regardless, this minor difference in variance would not have affected inference. Additionally, while the point estimates for the treatment effects were modified to some extent when prognostic score adjustment was applied, the changes were minimal relative to the size of the estimated standard errors. Overall, we demonstrate

that the PROCOVA-MMRM model can be applied prospectively using an accurate estimate of the sample size reduction to the Target Trial.

C Case Study 2 - Amyotrophic Lateral Sclerosis

Data from the ALS Target Trial (Cudkowicz et al., 2014) were obtained from the National Institute of Neurological Disorders and Stroke (NINDS). The study consisted of Stage 1 (pharmacokinetics of ceftriaxone 2g and 4g over 7 days), Stage 2 (safety of the same doses over 20 weeks), and Stage 3 (efficacy and safety), where all patients on active treatment received 4g ceftriaxone regardless of their dose in the earlier stages. The final Stage 3 dataset comprised 513 participants; 340 in the ceftriaxone group and 173 in the placebo group. Participants were followed for 52 weeks after the randomization of the last participant.

The co-primary outcome measures used in the ALS Target Trial were the rate of functional decline as quantified by change in the ALS Functional Rating Scale-Revised (ALSFRS-R) over time, and the time to death, tracheostomy, or the start of permanent assisted ventilation, of which the former served as the primary outcome of interest. During Stages 1 and 2, a statistically significant difference was observed in the rate of functional decline (as measured via ALSFRS-R) for 4g ceftriaxone vs placebo. However, during Stage 3, no significant differences between the active and placebo groups were detected on either of the primary endpoints. For our re-analyses, we focused on the final phase 3 dataset of 513 participants; 340 in the ceftriaxone group and 173 in the placebo group (2:1 randomization ratio).

Historical de-identified ALS study data were obtained from four different sources and curated for model training. The Pooled Resource Open-Access ALS Clinical Trials (PRO-ACT) dataset contained data from 26 phase 2 and 3 RCTs encompassing 11675 trial participants. This data has been volunteered by members of the PRO-ACT Consortium, which was formed in 2011, Prize4Life, in collaboration with the Northeast ALS Consortium (NEALS), and with funding from the ALS Therapy Alliance. The NEALS dataset contained 5 clinical RCTs with 814 participants. The Pooled Resource Open-Access Clinical Research ALS (PRO-ACE) dataset included observational data for 4197 participants. The Witzel et al. (2021) dataset contained one observational and two interventional trials with 128 participants.

The data were processed to promote consistency and quality. Since the PRO-ACT dataset does not identify trials, it is possible that data from other sources is replicated in PRO-ACT. To correct for this issue, we applied a deduplication algorithm: participants were removed from the PRO-ACT dataset if a pre-defined subset of baseline and longitudinal variables (shared with other sources) matched or closely reproduced values observed in the NEALS dataset or in the ALS Target Trial dataset. In particular, we matched and removed all 513 subjects of the ALS Target Trial from PRO-ACT. Additional steps were taken to harmonize the data. Observations were set to standard units and data falling outside a predefined range were removed to promote quality and remove extreme outliers. For participants who dropped out from the study, the time to death was censored to the time of the last visit for which study data was reported. Participants were excluded if they had no available data for model training or if they died before study start (day 0). All 70 of the variables used to train the model were available for at least a subset of the clinical trials in the training dataset. Finally, longitudinal data was binned in time using a 28-days window centered around nominal visits with a 28 days cadence. The final curated dataset used for training and cross-validation consisted of 10,887 subjects.

C.1 ALS Experiment 1

After fitting the prognostic models and obtaining prognostic scores for each study visit and participant in the ALS Target Trial, we analyzed the results for the treatment effect of ceftriaxone on the change in ALSFRS-R at 48 weeks using four approaches: MMRM; MMRM with covariate adjustment for BL ALSFRS-R only; PROCOVA-MMRM; and PROCOVA-MMRM with BL ALSFRS-R as an additional covariate. This experiment used the same number of participants and randomization ratio as the original study reported by Cudkowicz et al. (2014). We compared the resulting point estimates and variance obtained with these four conditions for the effect of treatment on the changes in ALSFRS-R at 48 weeks (Table 5). All data presented are heteroskedasticity consistent variance estimated with the Huber-White estimator.

Table 5 Estimated treatment effect and variance for ALSFRS-R at week 48 in reanalysis of the ALS Target Trial.

Analysis Method	Treatment Effect Estimate	Estimated Variance
Unadjusted MMRM	1.519	0.797
MMRM + BL ALSFRS-R	1.520	0.795
PROCOVA-MMRM	2.007	0.637
PROCOVA-MMRM + BL ALSFRS-R	2.069	0.627

We also evaluated how well the outcomes predicted by our prognostic model correlated with the subsequent study observations. The correlation between time-matched prognostic scores generated from baseline data and the study visit observations in the ALS Target Trial was 0.27, 0.37, 0.34, 0.38, 0.38 and 0.41 at weeks 8, 16, 24, 32, 40 and 48, respectively.

C.2 ALS Experiment 2

In this second experiment, we analyze the performance of PROCOVA-MMRM vs MMRM when following the prospective implementation procedure as described in Section III. To use the PROCOVA formula as a conservative estimate for PROCOVA-MMRM, we first needed to understand the Pearson correlation between the ALS prognostic model predictions and actual outcomes in historical data that has similar characteristics to the ALS Target Trial. Thus, a subset of the modeling data with diagnosis within three years of baseline and 18 was used to obtain a Pearson correlation coefficient of 0.391 for ALSFRS-R. This translates to an estimated sample size reduction of 15.3% using the PROCOVA formula with no deflation factor applied and assuming a gamma (inflation factor for the standard deviation) value of 1. This sample size reduction was then applied by randomly removing participants from both treatment and control arm of the ALS Target Trial for PROCOVA-MMRM analysis. This process was repeated 1000 times and the average of the treatment effect variances with PROCOVA-MMRM was calculated. The estimated potential sample size reduction was accurate if the calculated treatment effect variance with PROCOVA-MMRM was equal to or less than the unadjusted MMRM estimate.

Table 6 Comparison of sample size, estimated treatment effect, and variance for ALSFRS-R at week 48 in reanalysis of the ALS Target Trial.

Analysis Method	Sample Size	Treatment Effect Estimate	Estimated Variance
Unadjusted MMRM	513	1.519	0.797
PROCOVA-MMRM	427	1.997	0.753

Results in Table 6 show that the endpoint treatment effect variance for PROCOVA-MMRM is lower than the unadjusted MMRM, indicating the same capacity for inference with a 15.3% reduction in sample size. While the point estimates for the treatment effects were modified to some extent when prognostic score adjustment was applied, the changes were minimal relative to the size of the estimated standard errors. Similar to the AD case study, changes in treatment effect estimates were insignificant relative to the variance. This confirms that the PROCOVA-MMRM model can be applied prospectively using the PROCOVA formula to estimate an appropriate sample size reduction for the Target Trial.

5 Simulation Studies of PROCOVA-MMRM

We utilize several realistic simulated scenarios to demonstrate that PROCOVA-MMRM reduces variance and provides more precise estimates of treatment effects relative to MMRM (without adjustment for baseline covariates), while controlling type I error rate. Since the treatment effect is known, the difference between the true effect and the estimated result can be measured directly in various scenarios to characterize how treatment effect estimation differs between MMRM and PROCOVA-MMRM. Additional simulations and details on the simulation studies can be found in Appendix C.

A Simulation Study Methods

Assuming a treatment effect of -1.2, we simulated four different scenarios that model different situations encountered in clinical trials, and that enable us to probe the sensitivity of PROCOVA-MMRM to particular assumptions.

- The Linear simulation describes a scenario in which the treatment effect is constant and the relationship between the time-matched prognostic scores and the expected outcomes is linear in both the treatment and control arms.
- The Additional Covariates simulation describes a scenario in which the relationships are the same as the Linear scenario, but three additional baseline covariates (ranging from 0.15-0.25 correlation with the outcome) have been added to the MMRM and PROCOVA-MMRM analyses.
- The Shifted simulation is similar to the Linear scenario, except that the time-matched prognostic scores have been systematically shifted. This simulates an event in which the prognostic scores may not fully apply to the Target Trial because of a population shift.

- The Heterogeneous simulation describes a scenario in which the conditional average effect $E[Y_{i,t}(1) - Y_{i,t}(0)|\mathbf{X}_i] = \mu_{1,t}(\mathbf{X}_i) - \mu_{0,t}(\mathbf{X}_i)$ is not constant (i.e., $E[Y_{i,t}(1) - Y_{i,t}(0)|\mathbf{X}_i] \neq \mu_{1,t} - \mu_{0,t}$). This represents an event where there is a nonconstant treatment effect.

The first two simulation scenarios fall under the assumptions in our theoretical results. Therefore, we expect PROCOVA-MMRM to perform well, so long as we use a prognostic model capable of capturing linear relationships. In contrast, the Heterogeneous scenario violates the constant additive treatment effect assumption of PROCOVA-MMRM, probing the sensitivity of PROCOVA-MMRM to such violations. Although the Shifted scenario does not violate any of the assumptions of PROCOVA-MMRM, a population shift in the trial participants' characteristics may affect how strongly the time-matched prognostic scores are correlated with the expected outcomes in the Target Trial. Therefore, this scenario probes the sensitivity of PROCOVA-MMRM to the predictive performance of the trained prognostic model.

To compare treatment effect estimation precision, bias, confidence interval coverage and power between MMRM and PROCOVA-MMRM, we generated 2500 longitudinal datasets for each scenario with a total of 5 post-baseline measurements simulated for each participant. Observed outcomes were generated conditional on the prognostic score at a correlation of 0.5 with the final visit outcome. Missingness was assumed to follow the missing at random (MAR) assumption where the missingness is fully explained by baseline covariates only. Data were missing from approximately 30% of participants, and the missingness was monotone. The alternative hypothesis was constructed such that there was 80% power under the standard t-test. Additional details on the data generating process for each of the simulation scenarios are provided in Appendix C.

B Simulation Study Results and Discussion

Table 7 presents the results obtained with MMRM and PROCOVA-MMRM in each of the scenarios described above, assuming a treatment effect of -1.2. Results include the point estimates of treatment effect with deviation from the true value, average variance of the estimate, the 95% interval coverage probability, and the power.

Table 7 Simulation results with treatment effect = -1.2

Scenario	Treatment Effect Estimate	Bias	Average Variance	95% CI Coverage	Power
Linear					
MMRM	-1.188	0.012	0.165	0.954	0.827
PROCOVA-MMRM	-1.191	0.009	0.122	0.944	0.920
Additional Covariates					
MMRM	-1.191	0.009	0.141	0.945	0.886
PROCOVA-MMRM	-1.195	0.005	0.098	0.946	0.967
Shifted					
MMRM	-1.209	-0.009	0.165	0.947	0.841
PROCOVA-MMRM	-1.211	-0.011	0.145	0.954	0.887
Heterogenous					
MMRM	-1.173	0.027	0.166	0.951	0.814
PROCOVA-MMRM	-1.176	0.024	0.123	0.949	0.910

In agreement with our case study results, PROCOVA-MMRM provides unbiased estimators of treatment effects and smaller average variance estimates that increase power while maintaining type

I error control in all scenarios. The variances of the point estimators of the treatment effect under PROCOVA-MMRM is smaller than or equal to those obtained with MMRM. Our results show that PROCOVA-MMRM does not perform any worse than traditional approaches (nor does it lose its key statistical properties) even in circumstances when linear model assumptions (that are not specific to PROCOVA-MMRM) are challenged. Thus, PROCOVA-MMRM is a robust technique for estimating treatment effects from RCTs.

6 Discussion

MMRM has been growing in popularity for longitudinal data analysis as it includes the intermediate observation data that is collected across study visits, simultaneously enhancing treatment estimation precision and handling the common RCT issue of missing data without explicit imputation (Garcia and Marder, 2017; EMA, 2010). While covariate adjustment is a standard procedure for RCT analysis, PROCOVA enhances its benefit through incorporating AI-generated optimal adjustment covariates (Schuler et al., 2021). These predicted control outcomes, from individual baseline data, can be seamlessly integrated into a regulatory aligned framework for any clinical program where sufficient historical data is available to train and test a prognostic model (EMA, 2022). Here, we introduce PROCOVA-MMRM as a synergistic union of these two methods that is positioned to improve treatment effect estimation precision in longitudinal RCTs.

We provide mathematical, case study, and simulation evidence to support the advantageous properties of PROCOVA-MMRM. Namely, PROCOVA-MMRM yields unbiased estimators of treatment effects, smaller average variance estimates, increased power, and controlled type I error rates, highlighting its superiority over the unadjusted MMRM analysis. We also described an implementation of PROCOVA-MMRM that allows for prospective sample size reduction. This procedure allowed us to remove at random up to 15.3% of clinical trial participants in a case study and still obtain the same results.

As described above, to enable the application of PROCOVA-MMRM, prognostic scores must be generated from the baseline characteristics of participants in the Target Trial using a prognostic model. Our procedure theoretically attains the largest decrease in variance when the prognostic model accurately predicts the expected control outcomes over time in the Target Trial population. However, some methods for fitting predictive models may overfit to the population in the training data, leading to a scenario in which the prognostic score has a much larger correlation with observed outcomes in the training dataset than in the Target Trial population. Therefore, to maximize the efficiencies gained through application of PROCOVA-MMRM, the prognostic model should be trained in a wide variety of data, but also perform well in a population similar to that of the Target Trial. While a lower performing model may result in a smaller efficiency gain, control of the type I error rate and unbiased treatment effect estimation will still be preserved with PROCOVA-MMRM in the large-sample setting.

Here we have outlined a streamlined implementation of PROCOVA-MMRM that utilizes the PROCOVA sample size formula as a conservative estimate for sample size reduction. Although this does not capture the maximum conversion from increased precision to prospective sample size reduction, it provides an additional buffer against underpowering the study and a logical path to validation of the prognostic model. Simulation studies may also provide insight into prospective estimation of sample size reduction/efficiency gain achievable through the use of any specific prog-

nostic model with PROCOVA-MMRM, which may enable a greater reduction in sample size than the streamlined approach. Regardless of the method chosen for prospective implementation, the statistical properties of PROCOVA-MMRM are maintained.

We chose to assess PROCOVA-MMRM in ALS and AD here for several reasons, which included the availability of historical data, the unmet medical need, the widespread use of MMRM analysis in AD and ALS RCTs (Sims et al., 2023; van Dyck et al., 2022; Haeberlein et al., 2022; Mintun et al., 2021; Ostrowitzki et al., 2017; Shefner et al., 2021), and known challenges in recruitment and enrollment. However, the utility of PROCOVA-MMRM is not limited to incurable neurodegenerative diseases. Any clinical indication where a model with high predictive accuracy is available (Unlearn.AI, 2024) can benefit from the integration of PROCOVA-MMRM into the design and analysis.

Financial Disclosure

JLR, AS, RZ, and DB are equity-holding employees of Unlearn.AI, Inc., a company that creates digital twin generators to forecast patient outcomes.

Acknowledgements

The authors would like to recognize Anna Vidovszky, Alyssa Vanderbeek, Stefanie Millar, Coco Kusiak, Ryan Douglas, and Eric Tramel for their ongoing support of this project, and David Walsh and David Miller for their previous work on this project.

Certain data used in the preparation of this submission were obtained from the University of California, San Diego Alzheimer's Disease Cooperative Study (ADCS) Legacy database (National Institute on Aging Grant Number U19AG010483).

Certain data used in the preparation of this submission were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for NeuroImaging at the University of Southern California.

Certain data used in the preparation of this submission were obtained from the Critical Path for Alzheimer's Disease (CPAD) database. In 2008, Critical Path Institute, in collaboration with the Engelberg Center for Health Care Reform at the Brookings Institution, formed the Coalition Against Major Diseases (CAMD), which was then renamed to CPAD in 2018. The Coalition brings together patient groups, biopharmaceutical companies, and scientists from academia, the U.S. Food and Drug Administration (FDA), the European Medicines Agency (EMA), the National Institute of Neurological Disorders and Stroke (NINDS), and the National Institute on Aging (NIA). CPAD currently includes over 200 scientists, drug development and regulatory agency professionals, from member and non-member organizations. The data available in the CPAD database has been volunteered by CPAD member companies and non-member organizations.

Certain data used in this submission were provided by the Knight Alzheimer Disease Research Center and OASIS-4: Clinical Cohort (Principal Investigators: T. Benzinger, L. Koenig, P. LaMontagne).

Certain data used in preparation of this article were obtained from the EPAD LCS data set V.IMI, doi:10.34688/epadlcs_v.imi_20.10.30. This work used data from the EPAD project which received support from the EU/EFPIA Innovative Medicines Initiative Joint Undertaking EPAD grant agreement n° 115736 and an Alzheimer's Association Grant (SG-21-818099-EPAD). EPAD LCS is registered at www.clinicaltrials.gov Identifier: NCT02804789.

Certain data used in this submission were obtained from the NACC database. The NACC database is funded by NIA/NIH Grant U24 AG072122. NACC data are contributed by the NIA-funded ADCs: P50 AG005131 (PI James Brewer, MD, PhD); P50 AG005133 (PI Oscar Lopez, MD); P50 AG005134 (PI Bradley Hyman, MD, PhD); P50 AG005136 (PI Thomas Grabowski, MD); P50 AG005138 (PI Mary Sano, PhD); P50 AG005142 (PI Helena Chui, MD); P50 AG005146 (PI Marilyn Albert, PhD); P50 AG005681 (PI John Morris, MD); P30 AG008017 (PI Jeffrey Kaye, MD); P30 AG008051 (PI Thomas Wisniewski, MD); P50 AG008702 (PI Scott Small, MD); P30 AG010124 (PI John Trojanowski, MD, PhD); P30 AG010129 (PI Charles DeCarli, MD); P30 AG010133 (PI Andrew Saykin, PsyD); P30 AG010161 (PI David Bennett, MD); P30 AG012300 (PI Roger Rosenberg, MD); P30 AG013846 (PI Neil Kowall, MD); P30 AG013854 (PI Robert Vassar, PhD); P50 AG016573 (PI Frank LaFerla, PhD); P50 AG016574 (PI Ronald Petersen, MD, PhD); P30 AG019610 (PI Eric Reiman, MD); P50 AG023501 (PI Bruce Miller, MD); P50 AG025688 (PI Allan Levey, MD, PhD); P30 AG028383 (PI Linda Van Eldik, PhD); P50 AG033514 (PI Sanjay Asthana, MD, FRCP); P30 AG035982 (PI Russell Swerdlow, MD); P50 AG047266 (PI Todd Golde, MD, PhD); P50 AG047270 (PI Stephen Strittmatter, MD, PhD); P50 AG047366 (PI Victor Henderson, MD, MS); P30 AG049638 (PI Suzanne Craft, PhD); P30 AG053760 (PI Henry Paulson, MD, PhD); P30 AG066546 (PI Sudha Seshadri, MD); P20 AG068024 (PI Erik Roberson, MD, PhD); P20 AG068053 (PI Marwan Sabbagh, MD); P20 AG068077 (PI Gary Rosenberg, MD); P20 AG068082 (PI Angela Jefferson, PhD); P30 AG072958 (PI Heather Whitson, MD); and P30 AG072959 (PI James Leverenz, MD).

Certain data used in this submission were obtained from the National Institute of Neurological Diseases and Stroke's Archived Clinical Research data (Title: Clinical Trial Ceftriaxone in Subjects with Amyotrophic Lateral Sclerosis, PI: E. Cudkowicz, MD, Funding: NINDS 5 U01-NS-049640) received from the Archived Clinical Research Dataset web site (ninds.nih.gov/currentresearch/research-funded-ninds/clinical-research/archived-clinical-research-datasets).

Certain data used in this submission were obtained from the PRO-ACT Database, where they

have been volunteered by members of the PRO-ACT Consortium: ALS Therapy Alliance, Cytokinetics, Inc., Amylyx Pharmaceuticals, Inc., Knopp Biosciences, Neuraltus Pharmaceuticals, Inc., Neurological Clinical Research Institute, MGH, Northeast ALS Consortium, Novartis, Prize4Life Israel, Regeneron Pharmaceuticals, Inc., Sanofi, Teva Pharmaceutical Industries, Ltd., and The ALS Association.

A subset of the data used in preparation of this submission were obtained from NEALS. We acknowledge the NEALS Biorepository for providing all or part of the biofluids from the ALS, healthy controls, and non-ALS neurological controls used in this study.

A subset of the data used in preparation of this document were obtained from the authors of 'Neurofilament light and heterogeneity of disease progression in amyotrophic lateral sclerosis'. We acknowledge the authors Simon Witzel, Felix Frauhammer, and Petra Steinacker for their work (Witzel et al 2021). The data used in this project is attributed under the Creative Commons license: https://creativecommons.org/licenses/by/4.0/

Certain data used in the preparation of this submission were obtained from the PRO-ACE Database. The data available in the PRO-ACE Database have been volunteered by PRO-ACE Consortium members.

References

Bertolini, D., Loukianov, A. D., Smith, A. M., Li-Bland, D., Pouliot, Y., Walsh, J. R., and Fisher, C. K. (2020). Modeling Disease Progression in Mild Cognitive Impairment and Alzheimer's Disease with Digital Twins. *arXiv*.

Besser, L., Kukull, W., Knopman, D. S., Chui, H., Galasko, D., Weintraub, S., Jicha, G., Carlsson, C., Burns, J., Quinn, J., Sweet, R. A., Rascovsky, K., Teylan, M., Beekly, D., Thomas, G., Bollenbeck, M., Monsell, S., Mock, C., Zhou, X. H., Thomas, N., Robichaud, E., Dean, M., Hubbard, J., Jacka, M., Schwabe-Fry, K., Wu, J., Phelps, C., Morris, J. C., and Centers, Neuropsychology Work Group, Directors, and Clinical Core leaders of the National Institute on Aging-funded US Alzheimer's Disease (2018). Version 3 of the National Alzheimer's Coordinating Center's Uniform Data Set. *Alzheimer Disease & Associated Disorders*, 32(4):351–358.

Brøgger-Mikkelsen, M., Ali, Z., Zibert, J. R., Andersen, A. D., and Thomsen, S. F. (2020). Online Patient Recruitment in Clinical Trials: Systematic Review and Meta-Analysis. *Journal of Medical Internet Research*, 22(11):e22179.

Coon, K. D., Myers, A. J., Craig, D. W., Webster, J. A., Pearson, J. V., Lince, D. H., Zismann, V. L., Beach, T. G., Leung, D., Bryden, L., Halperin, R. F., Marlowe, L., Kaleem, M., Walker, D. G., Ravid, R., Heward, C. B., Rogers, J., Papassotiropoulos, A., Reiman, E. M., Hardy, J., and Stephan, D. A. (2007). A high-density whole-genome association study reveals that APOE is the major susceptibility gene for sporadic late-onset Alzheimer's disease. *The Journal of Clinical Psychiatry*, 68(4):613–618. PMID: 17474819 PMID: 17474819.

Cudkowicz, M. E., Titus, S., Kearney, M., Yu, H., Sherman, A., Schoenfeld, D., Hayden, D., Shui, A., Brooks, B., Conwit, R., Felsenstein, D., Greenblatt, D. J., Keroack, M., Kissel, J. T., Miller, R., Rosenfeld, J., Rothstein, J. D., Simpson, E., Tolkoff-Rubin, N., Zinman, L., Shefner, J. M., and Investigators, f. t. C. S. (2014). Safety and efficacy of ceftriaxone for amyotrophic lateral

sclerosis: a multi-stage, randomised, double-blind, placebo-controlled trial. *The Lancet Neurology*, 13(11):1083–1091.

Diggle, P., Heagerty, P., Liang, K.-Y., and Zeger, S. (2013). *Analysis of Longitudinal Data - Second Edition*. Oxford University Press.

EMA (2010). Missing data in confirmatory clinical trials. https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-missing-data-confirmatory-clinical-trials_en.pdf.

EMA (2015). Guideline on adjustment for baseline covariates in clinical trials. https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-adjustment-baseline-covariates-clinical-trials_en.pdf.

EMA (2020). ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials - Step 5. https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e9-r1-addendum-estimands-and-sensitivity-analysis-clinical-trials-guideline-statistical-principles-clinical-trials-step-5_en.pdf.

EMA (2022). Qualification opinion for Prognostic Covariate Adjustment (PROCOVA). https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/qualification-opinion-prognostic-covariate-adjustment-procovatm_en.pdf.

FDA (2023). Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biological Products. https://www.fda.gov/media/148910/download.

Fisher, C. K., Smith, A. M., Walsh, J. R., Simon, A. J., Edgar, C., Jack, C. R., Holtzman, D., Russell, D., Hill, D., Grosset, D., Wood, F., Vanderstichele, H., Morris, J., Blennow, K., Marek, K., Shaw, L. M., Albert, M., Weiner, M., Fox, N., Aisen, P., Cole, P. E., Petersen, R., Sherer, T., and Kubick, W. (2019). Machine learning for comprehensive forecasting of Alzheimer's Disease progression. *Scientific Reports*, 9(1):13622. Number: 1 Publisher: Nature Publishing Group Number: 1 Publisher: Nature Publishing Group.

Fogel, D. B. (2018). Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: A review. *Contemporary Clinical Trials Communications*, 11:156–164.

Food and Drug Administration, US Department of Health and Human Services, Center for Devices and Radiological Health (CDRH), and Center for Biologics Evaluation and Research (CBER) (2010). Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/guidance-use-bayesian-statistics-medical-device-clinical-trials.

Garcia, T. P. and Marder, K. (2017). Statistical Approaches to Longitudinal Data Analysis in Neurodegenerative Diseases: Huntington's Disease as a Model. *Current Neurology and Neuroscience Reports*, 17(2):14.

Haeberlein, S. B., Aisen, P., Barkhof, F., Chalkias, S., Chen, T., Cohen, S., Dent, G., Hansson, O., Harrison, K., Hehn, C. v., Iwatsubo, T., Mallinckrodt, C., Mummery, C., Muralidharan, K., Nestorov, I., Nisenbaum, L., Rajagovindan, R., Skordos, L., Tian, Y., Dyck, C. v., Vellas, B., Wu,

S., Zhu, Y., and Sandrock, A. (2022). Two Randomized Phase 3 Studies of Aducanumab in Early Alzheimer's Disease. *The Journal of Prevention of Alzheimer's Disease*, 9(2):197–210.

Hansen, B. B. (2008). The prognostic analogue of the propensity score. *Biometrika*, 95(2):481–488.

Huber, P. J. (1967). The Behavior of Maximum Likelihood Estimates under Nonstandard Conditions. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1:221–233.

Jastreboff, A. M., Aronne, L. J., Ahmad, N. N., Wharton, S., Connery, L., Alves, B., Kiyosue, A., Zhang, S., Liu, B., Bunck, M. C., Stefanski, A., and Investigators, S.-. (2022). Tirzepatide Once Weekly for the Treatment of Obesity. *New England Journal of Medicine*, 387(3):205–216.

Kabashima, K., Matsumura, T., Komazaki, H., Kawashima, M., and Group, N.-J. S. (2020). Trial of Nemolizumab and Topical Agents for Atopic Dermatitis with Pruritus. *New England Journal of Medicine*, 383(2):141–150.

Koenig, L. N., Day, G. S., Salter, A., Keefe, S., Marple, L. M., Long, J., LaMontagne, P., Massoumzadeh, P., Snider, B. J., Kanthamneni, M., Raji, C. A., Ghoshal, N., Gordon, B. A., Miller-Thomas, M., Morris, J. C., Shimony, J. S., Benzinger, T. L., and (DIAN), for the Alzheimer's Disease Neuroimaging Initiative (ADNI) and the Dominantly Inherited Alzheimer Network (2020). Select Atrophied Regions in Alzheimer disease (SARA): An improved volumetric model for identifying Alzheimer disease dementia. *NeuroImage: Clinical*, 26:102248.

Lang, A. H., Loukianov, A. D., and Fisher, C. K. (2023). Neural Boltzmann Machines. *arXiv*.

Liublinska, V. and Rubin, D. B. (2014). Sensitivity analysis for a partially missing binary outcome in a two-arm randomized clinical trial. *Statistics in Medicine*, 33(24):4170–4185.

Mallinckrodt, C. H., Clark, W. S., and David, S. R. (2001). Type I Error Rates from Mixed Effects Model Repeated Measures Versus Fixed Effects Anova with Missing Values Imputed Via Last Observation Carried Forward. *Drug information journal : DIJ / Drug Information Association*, 35(4):1215–1225.

Mallinckrodt, C. H., Lane, P. W., Schnell, D., Peng, Y., and Mancuso, J. P. (2008). Recommendations for the Primary Analysis of Continuous Endpoints in Longitudinal Clinical Trials. *Drug information journal : DIJ / Drug Information Association*, 42(4):303–319.

Mintun, M. A., Lo, A. C., Evans, C. D., Wessels, A. M., Ardayfio, P. A., Andersen, S. W., Shcherbinin, S., Sparks, J., Sims, J. R., Brys, M., Apostolova, L. G., Salloway, S. P., and Skovronsky, D. M. (2021). Donanemab in Early Alzheimer's Disease. *New England Journal of Medicine*, 384(18):1691–1704.

Morris, J. C. (1993). The Clinical Dementia Rating (CDR): Current version and scoring rules. *Neurology*, 43(11):2412.2–2412–a. PMID: 8232972.

Neville, J., Kopko, S., Broadbent, S., Avilés, E., Stafford, R., Solinsky, C. M., Bain, L. J., Cisneroz, M., Romero, K., Stephenson, D., and Diseases, C. A. M. (2015). Development of a unified clinical trial database for Alzheimer's disease. *Alzheimer's & Dementia*, 11(10):1212–1221. PMID: 25676387.

Ostrowitzki, S., Lasser, R. A., Dorflinger, E., Scheltens, P., Barkhof, F., Nikolcheva, T., Ashford, E., Retout, S., Hofmann, C., Delmar, P., Klein, G., Andjelkovic, M., Dubois, B., Boada, M., Blennow, K., Santarelli, L., Fontoura, P., and Investigators, S. R. (2017). A phase III randomized trial of gantenerumab in prodromal Alzheimer's disease. *Alzheimer's Research & Therapy*, 9(1):95.

Petersen, R. C., Aisen, P. S., Beckett, L. A., Donohue, M. C., Gamst, A. C., Harvey, D. J., Jack, C. R., Jagust, W. J., Shaw, L. M., Toga, A. W., Trojanowski, J. Q., and Weiner, M. W. (2010). Alzheimer's Disease Neuroimaging Initiative (ADNI) Clinical characterization. *Neurology*, 74(3):201–209. Publisher: Wolters Kluwer Health, Inc. on behalf of the American Academy of Neurology Section: Articles PMID: 20042704 http://www.ncbi.nlm.nih.gov/pubmed/20042704.

Quinn, J. F., Raman, R., Thomas, R. G., Yurko-Mauro, K., Nelson, E. B., Dyck, C. V., Galvin, J. E., Emond, J., Jack, C. R., Weiner, M., Shinto, L., and Aisen, P. S. (2010). Docosahexaenoic Acid Supplementation and Cognitive Decline in Alzheimer Disease: A Randomized Trial. *JAMA*, 304(17):1903–1911.

Richeldi, L., Azuma, A., Cottin, V., Hesslinger, C., Stowasser, S., Valenzuela, C., Wijsenbeek, M. S., Zoz, D. F., Voss, F., Maher, T. M., and Investigators, .-. T. (2022). Trial of a Preferential Phosphodiesterase 4B Inhibitor for Idiopathic Pulmonary Fibrosis. *New England Journal of Medicine*, 386(23):2178–2187.

Romero, K., Mars, M., Frank, D., Anthony, M., Neville, J., Kirby, L., Smith, K., and Woosley, R. L. (2009). The Coalition Against Major Diseases: Developing Tools for an Integrated Drug Development Process for Alzheimer's and Parkinson's Diseases. *Clinical Pharmacology & Therapeutics*, 86(4):365–367. PMID: 19763117.

Rosen, W. G., Mohs, R. C., and Davis, K. L. (1984). A new rating scale for Alzheimer's disease. *The American Journal of Psychiatry*, 141(11):1356–1364. PMID: 6496779 PMID: 6496779.

Safieh, M., Korczyn, A. D., and Michaelson, D. M. (2019). ApoE4: an emerging therapeutic target for Alzheimer's disease. *BMC Medicine*, 17(1):64.

Schuler, A. (2022). Mixed Models for Repeated Measures Should Include Time-by-Covariate Interactions to Assure Power Gains and Robustness Against Dropout Bias Relative to Complete-Case ANCOVA. *Therapeutic Innovation & Regulatory Science*, 56(1):145–154.

Schuler, A., Walsh, D., Hall, D., Walsh, J., Fisher, C., Disease, T. C. P. f. A., Initiative, T. A. D. N., and Study, T. A. D. C. (2021). Increasing the efficiency of randomized trial estimates via linear adjustment for a prognostic score. *The International Journal of Biostatistics*, 0(0). https://www.degruyter.com/document/doi/10.1515/ijb-2021-0072/html?lang=en https://www.degruyter.com/document/doi/10.1515/ijb-2021-0072/html?lang=en.

Shefner, J. M., Andrews, J. A., Genge, A., Jackson, C., Lechtzin, N., Miller, T. M., Cockroft, B. M., Meng, L., Wei, J., Wolff, A. A., Malik, F. I., Bodkin, C., Brooks, B. R., Caress, J., Dionne, A., Fee, D., Goutman, S. A., Goyal, N. A., Hardiman, O., Hayat, G., Heiman-Patterson, T., Heitzman, D., Henderson, R. D., Johnston, W., Karam, C., Kiernan, M. C., Kolb, S. J., Korngut, L., Ladha, S., Matte, G., Mora, J. S., Needham, M., Oskarsson, B., Pattee, G. L., Pioro, E. P., Pulley, M., Quan, D., Rezania, K., Schellenberg, K. L., Schultz, D., Shoesmith, C., Simmons, Z., Statland,

J., Sultan, S., Swenson, A., Berg, L. H. V. D., Vu, T., Vucic, S., Weiss, M., Whyte-Rayson, A., Wymer, J., Zinman, L., and Rudnicki, S. A. (2021). A Phase 2, Double-Blind, Randomized, Dose-Ranging Trial Of Reldesemtiv In Patients With ALS. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, 22(3-4):287–299.

Sims, J. R., Zimmer, J. A., Evans, C. D., Lu, M., Ardayfio, P., Sparks, J., Wessels, A. M., Shcherbinin, S., Wang, H., Nery, E. S. M., Collins, E. C., Solomon, P., Salloway, S., Apostolova, L. G., Hansson, O., Ritchie, C., Brooks, D. A., Mintun, M., Skovronsky, D. M., Investigators, T.-A. ., Abreu, R., Agarwal, P., Aggarwal, P., Agronin, M., Allen, A., Altamirano, D., Alva, G., Andersen, J., Anderson, A., Anderson, D., Arnold, J., Asada, T., Aso, Y., Atit, V., Ayala, R., Badruddoja, M., Badzio-jagiello, H., Bajacek, M., Barton, D., Bear, D., Benjamin, S., Bergeron, R., Bhatia, P., Black, S., Block, A., Bolouri, M., Bond, W., Bouthillier, J., Brangman, S., Brew, B., Brisbin, S., Brisken, T., Brodtmann, A., Brody, M., Brosch, J., Brown, C., Brownstone, P., Bukowczan, S., Burns, J., Cabrera, A., Capote, H., Carrasco, A., Yepez, J. C., Chavez, E., Chertkow, H., Chyrchel-paszkiewicz, U., Ciabarra, A., Clemmons, E., Cohen, D., Cohen, R., Cohen, I., Concha, M., Costell, B., Crimmins, D., Cruz-pagan, Y., Cueli, A., Cupelo, R., Czarnecki, M., Darby, D., Dautzenberg, P. l. j., Deyn, P. D., Gandara, J. D. L., Deck, K., Dibenedetto, D., Dibuono, M., Dinnerstein, E., Dirican, A., Dixit, S., Dobryniewski, J., Drake, R., Drysdale, P., Duara, R., Duffy, J., Ellenbogen, A., Faradji, V., Feinberg, M., Feldman, R., Fishman, S., Flitman, S., Forchetti, C., Fraga, I., Frank, A., Frishberg, B., Fujigasaki, H., Fukase, H., Fumero, I., Furihata, K., Galloway, C., Gandhi, R., George, K., Germain, M., Gitelman, D., Goetsch, N., Goldfarb, D., Goldstein, M., Goldstick, L., Rojas, Y. G., Goodman, I., Greeley, D., Griffin, C., Grigsby, E., Grosz, D., Hafner, K., Hart, D., Henein, S., Herskowitz, B., Higashi, S., Higashi, Y., Ho, G., Hodgson, J., Hohenberg, M., Hollenbeck, L., Holub, R., Hori, T., Hort, J., Ilkowski, J., Ingram, K. J., Isaac, M., Ishikawa, M., Janu, L., Johnston, M., Julio, W., Justiz, W., Kaga, T., Kakigi, T., Kalafer, M., Kamijo, M., Kaplan, J., Karathanos, M., Katayama, S., Kaul, S., Keegan, A., Kerwin, D., Khan, U., Khan, A., Kimura, N., Kirk, G., Klodowska, G., Kowa, H., Kutz, C., Kwentus, J., Lai, R., Lall, A., Lawrence, M., Lee, E., Leon, R., Linker, G., Lisewski, P., Liss, J., Liu, C., Losk, S., Lukaszyk, E., Lynch, J., Macfarlane, S., Macsweeney, J., Mannering, N., Markovic, O., Marks, D., Masdeu, J., Matsui, Y., Matsuishi, K., Mcallister, P., Mcconnehey, B., Mcelveen, A., Mcgill, L., Mecca, A., Mega, M., Mensah, J., Mickielewicz, A., Minaeian, A., Mocherla, B., Murphy, C., Murphy, P., Nagashima, H., Nair, A., Nair, M., Nardandrea, J., Nash, M., Nasreddine, Z., Nishida, Y., Norton, J., Nunez, L., Ochiai, J., Ohkubo, T., Okamura, Y., Okorie, E., Olivera, E., O'mahony, J., Omidvar, O., Ortiz-Cruz, D., Osowa, A., Papka, M., Parker, A., Patel, P., Patel, A., Patel, M., Patry, C., Peckham, E., Pfeffer, M., Pietras, A., Plopper, M., Porsteinsson, A., Robitaille, R. P., Prins, N., Puente, O., Ratajczak, M., Rhee, M., Ritter, A., Rodriguez, R., Ables, L. R., Rojas, J., Ross, J., Royer, P., Rubin, J., Russell, D., Rutgers, S. M., Rutrick, S., Sadowski, M., Safirstein, B., Sagisaka, T., Scharre, D., Schneider, L., Schreiber, C., Schrift, M., Schulz, P., Schwartz, H., Schwartzbard, J., Scott, J., Selem, L., Sethi, P., Sha, S., Sharlin, K., Sharma, S., Shiovitz, T., Shiwach, R., Sladek, M., Sloan, B., Smith, A., Solomon, P., Sorial, E., Sosa, E., Stedman, M., Steen, S., Stein, L., Stolyar, A., Stoukides, J., Sudoh, S., Sutton, J., Syed, J., Szigeti, K., Tachibana, H., Takahashi, Y., Tateno, A., Taylor, J. D., Taylor, K., Tcheremissine, O., Thebaud, A., Thein, S., Thurman, L., Toenjes, S., Toji, H., Toma, M., Tran, D., Trueba, P., Tsujimoto, M., Turner, R., Uchiyama, A., Ussorowska, D., Vaishnavi, S., Valor, E., Vandersluis, J., Vasquez, A., Velez, J., Verghese, C., Vodickova-borzova, K., Watson, D., Wei-

dman, D., Weisman, D., White, A., Willingham, K., Winkel, I., Winner, P., Winston, J., Wolff, A., Yagi, H., Yamamoto, H., Yathiraj, S., Yoshiyama, Y., and Zboch, M. (2023). Donanemab in Early Symptomatic Alzheimer Disease. *JAMA*, 330(6).

Solomon, A., Kivipelto, M., Molinuevo, J. L., Tom, B., Ritchie, C. W., and Consortium, E. (2018). European Prevention of Alzheimer's Dementia Longitudinal Cohort Study (EPAD LCS): study protocol. *BMJ Open*, 8(12):e021017.

Unlearn.AI (2024). Digital Twins for Disease Modeling. *arXiv*.

van Dyck, C. H., Swanson, C. J., Aisen, P., Bateman, R. J., Chen, C., Gee, M., Kanekiyo, M., Li, D., Reyderman, L., Cohen, S., Froelich, L., Katayama, S., Sabbagh, M., Vellas, B., Watson, D., Dhadda, S., Irizarry, M., Kramer, L. D., and Iwatsubo, T. (2022). Lecanemab in Early Alzheimer's Disease. *New England Journal of Medicine*, 388(1):9–21.

Wang, B. and Du, Y. (2023). Improving the mixed model for repeated measures to robustly increase precision in randomized trials. *Int J Biostat*.

White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica*, 48(4):817–838.

White, I. R., Royston, P., and Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine*, 30(4):377–399.

Witzel, S., Frauhammer, F., Steinacker, P., Devos, D., Pradat, P.-F., Meininger, V., Halbgebauer, S., Oeckl, P., Schuster, J., Anders, S., Dorst, J., Otto, M., and Ludolph, A. C. (2021). Neurofilament light and heterogeneity of disease progression in amyotrophic lateral sclerosis: development and validation of a prediction model to improve interventional trials. *Translational Neurodegeneration*, 10(1):31.

7 Appendix

A Missing Data Assumptions

In order to compare the precisions of different methods for repeated measures analyses in the case of missing data, we generally assume that the data are missing completely at random (MCAR) so as to have a fair basis for comparison. For example, if the data were missing at random (MAR), then the traditional MMRM analysis that does not involve any adjustment would correspond to an analysis under a missing not at random (MNAR) mechanism, and would likely yield biased treatment effect inferences. As such, it would be difficult to justify a fair comparison of PROCOVA-MMRM (or any MMRM method that adjusts for baseline covariates) to the traditional MMRM analysis, because the latter would be biased whereas the former could be unbiased. We shall also assume that the missing data pattern is monotone.

B Relating Sample Size Reduction to the Effective Sample Size

We provide a simple proof, based on the concept of the effective sample size (ESS), to demonstrate how the sample size for a repeated measures trial calculated according to the PROCOVA formula would be a conservative estimate of the sample size that would be necessary for a PROCOVA-MMRM analysis.

The ESS is a metric that can capture the sample size efficiency of one method of estimation versus a benchmark method. An explanation of ESS is provided in the Food and Drug Administration et al. (2010, p. 39–40) *Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials*. It can be interpreted for a Bayesian method as a measure of the number of patients that are "borrowed" from a historical trial in the specification of the prior distribution for the Bayesian analysis. To formally define this quantity, let $V_{\text{benchmark}}$ denote the variance of the treatment effect estimator under a benchmark method, and V_{new} denote the variance of the treatment effect estimator under a new method. Then the ESS is defined as

$$\text{ESS} = \frac{N V_{\text{benchmark}}}{V_{\text{new}}}, \tag{2}$$

where N denotes the trial sample size. The core of this definition is the ratio of the variances, which is our focus as well.

The ESS can also be related to the sample size reduction of a new method compared to a benchmark method. This broader scope of the ESS is demonstrated by means of a Taylor series expansion of the sample size formula for the new method. Let N_0 denote the number of control participants and N_1 denote the number of treated participants in a trial. We consider a sample size formula for a new method that is just a function of N_0, i.e., the number of treated participants N_1 is held fixed throughout. Also, let $f : \mathbb{R}^+ \to \mathbb{R}^+$ denote the function whose input is the control group sample size N_0, and whose output is the *precision* of the treatment effect estimator under the new method. We consider precision (the inverse of variance) because of the comparative ease in considering the derivatives f', f'', etc., in this case.

Suppose n_0 is the control arm sample size that was identified for the trial under the analysis by the benchmark method. For example, when the benchmark analysis is ANOVA, n_0 would denote the control arm sample size under which ANOVA would have power 80%. We assume that an expression exists for the variance of the treatment effect estimator under the benchmark method at the sample

size n_0. Let $V_{\text{benchmark}}$ denote the variance of the treatment effect estimator under the benchmark method with a control arm sample size of n_0.

The Taylor series expansion for f at n_0 is

$$f(N_0) = f(n_0) + (N_0 - n_0)f'(n_0) + \frac{1}{2}(N_0 - n_0)^2 f''(n_0) + \cdots . \tag{3}$$

Suppose $f''(n_0), f^{(3)}(n_0)$, and the other higher-order derivatives in this expansion are negligible. Then

$$f(N_0) \approx f(n_0) + (N_0 - n_0)f'(n_0). \tag{4}$$

We use equation (4) to find the approximate value of N_0 such that $f(N_0) = V_{\text{benchmark}}^{-1}$:

$$N_0 \approx n_0 - \left\{ \frac{V_{\text{benchmark}} f(n_0) - 1}{V_{\text{benchmark}} f'(n_0)} \right\}. \tag{5}$$

Let $f(n_0) = V_{\text{new}}^{-1}$ denote the precision of the new method under the control arm sample size of n_0. Then

$$N_0 \approx n_0 - \left\{ \frac{V_{\text{benchmark}}/V_{\text{new}} - 1}{V_{\text{benchmark}} f'(n_0)} \right\} \tag{6}$$

$$\approx n_0 - \left\{ \frac{(\text{ESS} - N)}{N V_{\text{benchmark}} f'(n_0)} \right\}, \tag{7}$$

where $N = n_0 + N_1$. Thus, N_0 can be approximated using the ESS. It is important to note the following points that are involved in this derivation.

- We are considering the case in which $V_{\text{benchmark}}/V_{\text{new}} > 1$ under the control arm sample size n_0 and treatment arm sample size N_1.
- We assume $f'(n_0) > 0$ for all n_0 and $f'(n_0) \to 0$ as $n_0 \to \infty$.
  - For a well-behaved treatment effect estimator, we can expect its precision to increase monotonically as a function of the sample size.
  - Furthermore, we can expect the rate of increase for the precision to diminish, and ultimately converge to zero, as the sample size increases.
- As ESS increases, N_0 decreases (i.e., there is greater sample size reduction).

B.1 Simple Proof for the Effective Sample Size for PROCOVA-MMRM Versus PROCOVA

Let $\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_0 \\ \boldsymbol{\beta}_w \\ \boldsymbol{\beta}_x \end{pmatrix}$ denote the vector of regression coefficients for PROCOVA-MMRM, and $\boldsymbol{\phi}$ be the vector containing all the parameters in the participants' covariance matrices $\boldsymbol{\Psi}$. Both the maximum likelihood estimator (MLE) and restricted maximum likelihood estimator (REML) for $\begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\phi} \end{pmatrix}$ are

obtained by first identifying the vector $\hat{\boldsymbol{\phi}}$ that maximizes the respective profile likelihood, and then identifying $\boldsymbol{\beta}$ via

$$\hat{\boldsymbol{\beta}}\left(\hat{\boldsymbol{\phi}}\right) = \left(\mathbf{X}^{\mathsf{T}}\hat{\boldsymbol{\Omega}}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathsf{T}}\hat{\boldsymbol{\Omega}}^{-1}\mathbf{y}, \tag{8}$$

where $\mathbf{X}$ is the row-wise concatenation of the predictor matrices for the participants, and $\boldsymbol{\Omega}$ is the $\left(\sum_{i=1}^{N} T_i\right) \times \left(\sum_{i=1}^{N} T_i\right)$ block-diagonal matrix with diagonal components $\mathbf{R}_1, \ldots, \mathbf{R}_N$. For the purposes of simplifying the proof, for now we consider the case that $\boldsymbol{\Omega}$ is known and focus on $\hat{\boldsymbol{\beta}}(\boldsymbol{\phi})$. The covariance matrix of this estimator is

$$\left(\mathbf{X}^{\mathsf{T}}\boldsymbol{\Omega}^{-1}\mathbf{X}\right)^{-1} = \left\{\sum_{i=1}^{N}\left(\mathbf{X}_i^{\mathsf{T}}\mathbf{R}_i^{-1}\mathbf{X}_i\right)\right\}^{-1}, \tag{9}$$

where $\mathbf{X}_i$ is the $T_i \times 3T$ predictor matrix for participant i. The precision matrix is thus

$$\sum_{i=1}^{N}\left(\mathbf{X}_i^{\mathsf{T}}\mathbf{R}_i^{-1}\mathbf{X}_i\right). \tag{10}$$

We note that

$$\mathbf{X}_i^{\mathsf{T}}\mathbf{R}_i^{-1}\mathbf{X}_i = \begin{pmatrix} \mathbf{R}_i^{-1} & \mathbf{0} & w_i\mathbf{R}_i^{-1} & \mathbf{0} & \mathbf{R}_i^{-1}\mathbf{D}_i & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ w_i\mathbf{R}_i^{-1} & \mathbf{0} & w_i\mathbf{R}_i^{-1} & \mathbf{0} & w_i\mathbf{R}_i^{-1}\mathbf{D}_i & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{D}_i^{\mathsf{T}}\mathbf{R}_i^{-1} & \mathbf{0} & w_i\mathbf{D}_i^{\mathsf{T}}\mathbf{R}_i^{-1} & \mathbf{0} & \mathbf{D}_i^{\mathsf{T}}\mathbf{R}_i^{-1}\mathbf{D}_i & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}$$

where $\mathbf{D}_i = \mathrm{diag}\left(x_{i,1}, \ldots, x_{i,T_i}\right)$ is the $T_i \times T_i$ diagonal matrix consisting of the T_i prognostic scores for participant i, and the $\mathbf{0}$ matrices correspond to the time points that are missing for participant i.

Let $\mathcal{S}$ denote those participants who have complete outcomes. Then the component of the precision matrix above corresponding to these subjects is

$$\sum_{i\in\mathcal{S}}\left(\mathbf{X}_i^{\mathsf{T}}\mathbf{R}_i^{-1}\mathbf{X}_i\right) = \begin{pmatrix} N\boldsymbol{\Psi}^{-1} & N_1\boldsymbol{\Psi}^{-1} & \boldsymbol{\Psi}^{-1}\left(\sum_{i\in\mathcal{S}}\mathbf{D}_i\right) \\ N_1\boldsymbol{\Psi}^{-1} & N_1\boldsymbol{\Psi}^{-1} & \boldsymbol{\Psi}^{-1}\left(\sum_{i\in\mathcal{S}}w_i\mathbf{D}_i\right) \\ \left(\sum_{i\in\mathcal{S}}\mathbf{D}_i^{\mathsf{T}}\right)\boldsymbol{\Psi}^{-1} & \left(\sum_{i\in\mathcal{S}}w_i\mathbf{D}_i^{\mathsf{T}}\right)\boldsymbol{\Psi}^{-1} & \left(\sum_{i\in\mathcal{S}}\mathbf{D}_i^{\mathsf{T}}\boldsymbol{\Psi}^{-1}\mathbf{D}_i\right) \end{pmatrix}$$

Finally, as $\sum_{i=1}^{N}\left(\mathbf{X}_i^{\mathsf{T}}\mathbf{R}_i^{-1}\mathbf{X}_i\right) \geq \sum_{i\in\mathcal{S}}\left(\mathbf{X}_i^{\mathsf{T}}\mathbf{R}_i^{-1}\mathbf{X}_i\right)$ in terms of positive semi-definiteness, we then have that $\left\{\sum_{i=1}^{N}\left(\mathbf{X}_i^{\mathsf{T}}\mathbf{R}_i^{-1}\mathbf{X}_i\right)\right\}^{-1} \leq \left\{\sum_{i\in\mathcal{S}}\left(\mathbf{X}_i^{\mathsf{T}}\mathbf{R}_i^{-1}\mathbf{X}_i\right)\right\}^{-1}$ in terms of positive semidefiniteness. The latter inequality enables us to establish the inequality between the variance of the treatment effect estimator under PROCOVA-MMRM to the treatment effect estimator under PROCOVA.

C Details of Simulation Studies

In our simulation studies, we considered a RCT where the endpoint is continuous and each patient has measurements of the endpoint at 5 total endpoints. The data were analyzed using a MMRM

model using all available measurements. Data generation was simulated using the block matrices of a 10x10 covariance matrix where the upper left (Quadrant II) block matrix was the 5x5 covariance matrix of observed outcomes, the bottom right (Quadrant IV) block matrix was the 5x5 covariance matrix of the time-matched prognostic scores, and the off diagonal (Quadrants I, III) block matrices was the 5x5 covariance matrix of the observed outcome and time-matched prognostic scores.

The time-matched prognostic scores were first generated from a multivariate normal distribution with mean 0 and covariance defined by the 5x5 block matrix in Quadrant IV. The observed outcomes were generated conditional on the time-matched prognostic scores using a multivariate normal distribution with conditional mean and covariance derived from the block covariance matrices and mean prognostic score. In all simulation scenarios except the shifted scenario, the mean deviation of the time-matched prognostic scores from the observed control outcome was 0. Missingness was simulated as a function of 3 baseline covariates, resulting in approximately 30% missingness at the final timepoint. Missingness was also assumed to be monotone. A total of 1,000 participants before dropout was simulated in each arm, and each scenario was repeated 2,500 times.

For the analysis, the prognostic score and time interaction was included to account for the time-matched prognostic scores. An unstructured covariance matrix was used to model the within-participant correlations. The estimated Least Square Means (LSM) and variance estimator of the LSM were used to assess the statistical properties. The Kenward-Roger approximation was used to adjust the degrees of freedom for hypothesis testing. The longitudinal datasets and analysis were completed in R using the "mmrm" package.

Further details regarding the scenarios:

- Linear: Follows exactly the text above.
- Shifted: A random shift was added to the time-matched prognostic scores, mimicking the scenario where there are population differences between the historical training set and Target Trial. The time-matched prognostic scores at each timepoint were shifted by a normal distribution with mean vector -.3, -.5, -1, -2, -2.5 and standard deviation 3, 4 ,6, 7, 8. Adding this shift changes the covariance matrix of the prognostic score and observed outcomes. While the original simulation targeted a final correlation of 0.5, the resulting mean correlation was only 0.37.
- Additional Covariates: Three continuous baseline covariates were generated with moderate prognostic ability. Each baseline covariate had an approximate correlation of 0.25 with the outcome. Therefore, the covariance matrix to generate all observed in this scenario was a 13x13 matrix. Baseline covariate and time interactions were included in these analyses.
- Heterogeneous: A non-constant treatment effect was simulated by generating outcomes that were conditional on the baseline covariate by treatment interaction. Three baseline covariates were used, and the true treatment effect was estimated by generating the complete set of potential outcomes under the hypothetical scenario for all participants. The baseline by treatment interaction terms were not included in the analysis model.

Here we provide additional simulation results for the aforementioned scenarios in the condition that the null hypothesis is correct.

Table 8 Simulation results with treatment effect = 0

Scenario	Treatment Effect Estimate	Bias	Average Variance	95% CI Coverage	Type I Error
Linear					
MMRM	0.014	0.014	0.165	0.954	0.046
PROCOVA-MMRM	0.009	0.009	0.122	0.944	0.054
Additional Covariates					
MMRM	0.009	0.009	0.141	0.945	0.055
PROCOVA-MMRM	0.005	0.005	0.098	0.946	0.054
Shifted					
MMRM	-0.009	-0.009	0.165	0.947	0.053
PROCOVA-MMRM	-0.011	-0.011	0.146	0.954	0.046
Heterogenous					
MMRM	0.007	0.007	0.165	0.953	0.048
PROCOVA-MMRM	0.005	0.005	0.122	0.946	0.053

In this simulation setting, the correlation between the time-matched prognostic scores and observed outcomes was easily controlled using the block covariance matrices. For all prior simulations (Tables 7 and 8), we assumed a correlation of 0.5 between the prognostic model predictions and observed outcomes. Here we provide additional linear simulation results using a correlation value of 0.1 to demonstrate how a lower performing prognostic model affects treatment effect estimation in PROCOVA-MMRM.

Table 9 Additional Linear Simulation Results for a Less Correlated Prognostic Model

Condition	Treatment Effect Estimate	Bias	Average Variance	95% CI Coverage	Null Rejection Probability
H1 Linear					
MMRM	-1.191	0.009	0.164	0.944	0.833
PROCOVA-MMRM	-1.191	0.009	0.163	0.944	0.836
H0 Linear					
MMRM	0.009	0.009	0.164	0.944	0.056
PROCOVA-MMRM	0.009	0.009	0.163	0.944	0.054

The H1 linear condition in Table 9 corresponds with the linear scenario results in Table 7 of the main text (i.e., treatment effect), and the H0 Linear condition corresponds with the linear scenario results in Table 8 of the appendix (i.e., no treatment effect). Note that the unadjusted MMRM results have also changed due to the change in data generating mechanism for the expected outcome. Regardless if a correlation of 0.1 or 0.5 was used, PROCOVA-MMRM returned similar results for point estimate of treatment effect, 95% confidence interval coverage, and type I error rate. As expected, the lower performing model showed much less variance reduction than the higher performing model.