Received: 15 June 2020 | Revised: 2 November 2020 | Accepted: 7 December 2020

DOI: 10.1002/bimj.202000184

Biometrical Journal

RESEARCH PAPER

# Power and sample size for random coefficient regression models in randomized experiments with monotone missing data

**Nan Hu**[1] | **Howard Mackey**[1] | **Ronald Thomas**[2]

[1] Department of Biostatistics, Genentech Inc., San Francisco, CA, USA

[2] Department of Neurosciences, University of California San Diego, San Diego, CA, USA

**Correspondence**
Nan Hu, Department of Biostatistics, Genentech Inc., 1 DNA Way, South San Francisco, CA 94080, USA.
Email: hu.nan@gene.com

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

**Funding information**
National Institutes of Health, Grant/Award Number: U01 AG024904; Department of Defense, Grant/Award Number: W81XWH-12-2-0012

This article has earned an open data badge "**Reproducible Research**" for making publicly available the code necessary to reproduce the reported results. The results reported in this article could fully be reproduced.

**Abstract**

Random coefficient regression (also known as random effects, mixed effects, growth curve, variance component, multilevel, or hierarchical linear modeling) can be a natural and useful approach for characterizing and testing hypotheses in data that are correlated within experimental units. Existing power and sample size software for such data are based on two variance component models or those using a two-stage formulation. These approaches may be markedly inaccurate in settings where more variance components (i.e., intercept, rate of change, and residual error) are warranted. We present variance, power, sample size formulae, and software (R Shiny app) for use with random coefficient regression models with possible missing data and variable follow-up. We illustrate sample size and study design planning using data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. We additionally examine the drivers of variability to better inform study design.

**KEYWORDS**
growth curve, mixed effects model, mixed model with repeated measures, random coefficient regression model, statistical power

## 1 | INTRODUCTION

Laird and Ware (1982) introduced the two-stage random effects model, which has led to a rich body of research into random effects and mixed effects modeling (Diggle et al., 2002; Fitzmaurice et al., 2012). The random coefficient regression model (RCRM), a special case of the random effects model with time as a continuous covariate, has been a useful method for

  

modeling longitudinal data with correlations within experimental units of interest. It has been employed in an array of research areas (Curran et al., 2010; Diggle et al., 2002; Fitzmaurice et al., 2012; Gelman & Hill, 2006; Harrison et al., 2018) and in regulatory settings such as the U.S. Food and Drug Administration (FDA) approval on nintedanib in idiopathic pulmonary fibrosis and systemic sclerosis-associated interstitial lung disease (Flaherty et al., 2019; Richeldi et al., 2014).

Power and sample size formulae for RCRMs have historically been approximated by variance component models containing two components or motivated by a two-stage formulation (`longpower`, Donohue et al., 2013; `EaST`, East 6, 2020). While these approaches may be useful in some settings, they can markedly over- or underestimate power and sample size in more general settings where more variance components (i.e., intercept, rate of change, and residual error) are warranted (Fitzmaurice et al., 2012). Extensive work has been done based on longitudinal modeling frameworks other than the RCRM. For example, Roy et al. (2007) use the general matrix notation to provide the sample size formulae under the three-level (center, experimental group, and subject) mixed effects modeling framework; Heo and Leon (2009) also present the sample size formulae assuming no dropout in the experimental design under the three-level modeling framework; Hedeker et al. (1999) derive the sample size needed when the model treats assessment time points as categorical.

We assume a parallel-arm repeated measure experimental design with two experimental levels that can be adequately described with constant rates of change over time or as a first-order approximation of a potentially nonlinear model. The remainder of this article is organized as follows. In Section 2, we describe the RCRM and its underlying assumptions. In Section 3, we provide the covariance matrix for the fixed effects in the RCRM with and without missingness (at random) and in the special case of equally spaced design points; we provide insights on how model parameters might inform the experimental design; we also compare the RCRM variance with other commonly used longitudinal model variances. In Section 4, we present a power expression based on the covariance matrix presented in Section 3 and in Section 5, the resulting sample size formula. Section 6 contains simulations to assess the sensitivity of our power formula to various factors including sample size, the correlation between intercept and rate of change random effects, missing data, nonnormality, and nonlinearity. In Section 7, we illustrate in detail how the sample size formula can be applied and how more efficient experiments might be designed. A discussion follows in Section 8.

## 2 | RCRM MODEL CONSTRUCTION

Throughout this article, for illustration purposes, we assume two experimental levels: a control group and an experimental group from a randomized experiment such as a clinical trial. We construct the RCRM assuming constant rates of change in both the control and experimental groups as follows:

$$Y_{ij} = \alpha + u_{0i} + \left(\beta + \beta_x I_{x,i} + u_{1i}\right) t_j + \varepsilon_{ij}, \tag{1}$$

where $Y_{ij}$ is the continuous outcome measure for subject $i$ at the $j^{\text{th}}$ visit, defined for subjects $i = 1, 2, \ldots, N$, $j = 0, 1, 2, \ldots, k_i \leqslant J$, and $t_0 \equiv 0$ & $t_J \equiv T$ (time at the last visit among all subjects). Note that $j = 0$ represents the baseline visit. In randomized clinical trials, the schedule of postbaseline visits is usually preplanned and is the same for all experimental units (or subjects as we will refer to them going forth). Therefore, the assessment times are not subject dependent. In an ideal setting where every subject finishes all assessments, $t_{k_i} = t_J = T$, for $i = 1, 2, \ldots, N$.

The intercept $\alpha$ is the overall baseline population mean, which is assumed to be the same between experimental and control groups. This assumption is reasonable for randomized experiments where overall balance at baseline is achieved with randomization (Rosenberger & Lachin, 2015). The same assumption is made in the constrained longitudinal data analysis (cLDA) model (Liang & Zeger, 2000) where timepoints are modeled as categorical. The $I_{x,i}$ is an indicator variable (where "x" is short for "treatment," or more generally "experimental") that is defined as

$$I_{x,i} = \begin{cases} 1, & \text{if subject } i \text{ is in treatment (experimental) group} \\ 0, & \text{otherwise.} \end{cases}$$

The random effect vectors $\{\begin{bmatrix} u_{0i}, & u_{1i} \end{bmatrix}^T\}$ are $2 \times 1$, and are assumed to be independent and identically distributed (i.i.d.) zero-mean normal vectors with $2 \times 2$ covariance matrix $R = \begin{bmatrix} \sigma_\alpha^2 & \rho\sigma_\alpha\sigma_\beta \\ \rho\sigma_\alpha\sigma_\beta & \sigma_\beta^2 \end{bmatrix}$. The parameter $\rho$ is the correlation between

$u_{0i}$ and $u_{1i}$. The parameter $\beta$ represents the constant rate of change for the control group population, whereas $\beta_x$ denotes the treatment benefit in slowing the rate of change. The rate of change for the experimental group population is then $\beta + \beta_x$. The $\{\varepsilon_{ij}\}$ are i.i.d. pure errors with mean zero and variance $\sigma^2$, representing the variation within the experiment unit. We further assume that the random effect vectors and the pure errors are independent.

For simplicity of the methodological development, we assume the only two variables in the model are time and group information. Other covariates could be easily incorporated into the results throughout the article, however, in practice, it will be sufficient to use parameter and variance component estimates (using available data) in the power and sample size formulae that have been modeled using the same covariates expected in the planned experiment. When adjusting for the desired covariates is not possible, these results will be slightly conservative and, in many cases, negligibly so. The model presented in Equation (1) treats the response variable change over time linearly in both the experimental and control groups. This can often be justified, especially in cases where the outcome variable changes slowly over time. The linear function in time $t$ can also be viewed as a first order Taylor expansion of any true trajectory. We note that the RCRM can be generalized by adding higher order terms in time, and the results in this article can be generalized accordingly.

We further assume monotone missingness, defined as missing subsequent assessments after the first missingness occurs; this is also the assumption made in Lu et al. (2008). Granted, this assumption is not absolutely necessary for the methodological development in this article; however, the assumption is usually an appropriate approximation in clinical settings. The assumption also helps produce a both succinct and practical power and sample size formulae. Lastly, we assume the missing data are missing at random, so that the likelihood-based statistical inference we draw from the RCRM is valid (Little & Rubin, 2019).

Since the parameters $u_{0i}$ and $u_{1i}$ are both random variables, the model described in Equation (1) is classified as a random effects model (Diggle et al., 2002). Further, $u_{0i}$ enters in as an intercept coefficient (or baseline; we will treat them interchangeably) and $u_{1i}$ enters in as a rate of change coefficient (or slope; we will treat them interchangeably), hence the name "random coefficient" in RCRM. The pure error $\varepsilon_{ij}$ contributes to the within-subject variability; the random variables $u_{0i}$ and $u_{1i}$ contribute to the between-subject variability.

## 3 | COVARIANCE MATRIX OF RCRM ESTIMATORS

We start this section by presenting the results where we assume all subjects have all assessments up until $t_k$ $(k \geq 1)$ with no missing values. The results show the explicit form of the covariance matrix of the RCRM maximum likelihood estimators (MLEs), with the focus on the treatment effect estimator. We examine how the timing of the assessments affects the variance of the treatment effect estimator. We also provide comparisons between the RCRM variance and other popular longitudinal model variances. We close this section by extending the covariance matrix of the RCRM estimators to the more general context of monotone missing data.

Similar to what has been done in Lu et al. (2008), we also assume that all the variance components $\sigma_\alpha^2$, $\sigma_\beta^2$, $\rho$, and $\sigma^2$ are known in the subsequent derivation. We start from the simple case where all subjects have exactly the same assessment times without missing values; the covariance matrix of the RCRM MLE is stated in Lemma A1; see Appendix A.1.

Using Lemma A1, the variance of the RCRM treatment effect estimate is

$$\operatorname{var}\left(\hat{\beta}_x\right) = \left(\frac{m+n}{mn}\right) \times \frac{\sigma^4 + (k+1)\sigma^2\left(\sigma_\beta^2\overline{\boldsymbol{t}_{\{k\}}^2} + \sigma_\alpha^2 + 2\rho\sigma_\alpha\sigma_\beta\overline{\boldsymbol{t}_{\{k\}}}\right) + (k+1)^2\sigma_\alpha^2\sigma_\beta^2(1-\rho^2)\left(\overline{\boldsymbol{t}_{\{k\}}^2} - \overline{\boldsymbol{t}_{\{k\}}}\,\overline{\boldsymbol{t}_{\{k\}}}\right)}{(k+1)\left[\sigma^2\overline{\boldsymbol{t}_{\{k\}}^2} + (k+1)\sigma_\alpha^2\left(\overline{\boldsymbol{t}_{\{k\}}^2} - \overline{\boldsymbol{t}_{\{k\}}}\,\overline{\boldsymbol{t}_{\{k\}}}\right)\right]}, \tag{2}$$

with

$$\overline{\boldsymbol{t}_{\{k\}}^2} = \frac{1}{k+1}\sum_{i=0}^{k} t_i^2 \text{ and } \overline{\boldsymbol{t}_{\{k\}}} = \frac{1}{k+1}\sum_{i=0}^{k} t_i,$$

where $m$ and $n$ are the sample sizes of the experimental group and control group, respectively. When $k = J$, Equation (A.1) in Appendix A.1 provides the covariance matrix as a special case where all subjects finish all assessments in the experiment without any missing values. From Equation (A.1), both the sum of assessment times and the squared sum of

**TABLE 1** Parameters, variance of treatment effect estimator, and experimental design

| Parameter | Impact on var($\hat{\beta}_x$) | Implications for Experimental Design |
|---|---|---|
| $\sigma^2$ | • Variance is bounded below by $\left(\frac{m+n}{mn}\right)\sigma_\beta^2(1-\rho^2)$ when $\sigma^2 = 0$<br>• Variance monotonically increases as $\sigma^2$ increases | Reducing the pure error or choosing endpoints with smaller pure error increases power. |
| $\sigma_\alpha^2$ | • When $\sigma_\alpha^2 = 0$, variance reduces to $\left(\frac{m+n}{mn}\right)\left(\frac{\sigma^2}{(k+1)\overline{t_{\{k\}}^2}} + \sigma_\beta^2\right)$ | If population is very homogeneous (in terms of the outcome measure) at baseline, then placing more intermediate points closer to the end increases power. |
| | • When $\sigma_\alpha^2 \to \infty$, variance converges to $\left(\frac{m+n}{mn}\right)\left(\frac{\sigma^2}{(k+1)(\overline{t_{\{k\}}^2} - \overline{t_{\{k\}}}\,\overline{t_{\{k\}}})} + \sigma_\beta^2(1-\rho^2)\right)$ | If population is very heterogeneous at baseline, then maximizing the spread of design points increases power where maximum power is achieved with half of the design points at baseline and half at the end. |
| $\sigma_\beta^2$ | • If $\rho \geq 0$, variance is minimized when $\sigma_\beta = 0$ | When baseline and slope are not negatively correlated, smaller slope variability always increases power. However, this is not true in cases of negatively correlated baseline & slope. |
| | • If $\rho < 0$, variance is minimized when $\sigma_\beta = \frac{-\sigma^2\rho\sigma_\alpha\overline{t_{\{k\}}}}{\sigma^2\overline{t_{\{k\}}^2} + (k+1)\sigma_\alpha^2(1-\rho^2)(\overline{t_{\{k\}}^2} - \overline{t_{\{k\}}}\,\overline{t_{\{k\}}})}$ | |
| $\rho$ | • Variance is a quadratic function of $\rho$ | All else being equal, endpoints with a more negatively correlated baseline & slope will be more powerful than those with a positively correlated baseline & slope. |
| | • Variance is minimized when $\rho = -1$ | |
| | • Variance is maximized when $\rho = min\{\frac{\sigma^2\overline{t_{\{k\}}}}{(k+1)\sigma_\alpha\sigma_\beta(\overline{t_{\{k\}}^2} - \overline{t_{\{k\}}}\,\overline{t_{\{k\}}})}, 1\}$ | |
| Duration of follow-up | • Variance monotonically decreases as duration increases | Longer follow-up has bigger impact on power in homogeneous populations (i.e. $\sigma_\alpha$ & $\sigma_\beta$ small) vs. heterogeneous populations. |
| | • When $\sigma_\alpha^2 = 0$ and $\sigma_\beta^2 = 0$, variance reduces to $\left(\frac{m+n}{mn}\right)\frac{\sigma^2}{(k+1)\overline{t_{\{k\}}^2}}$ | |
| Placement of intermediate points | • Impact is non-linear without a closed-form solution. However, the variance is the ratio of polynomial functions of $\overline{t_{\{k\}}}$ and var($\{t_0, t_1, \ldots, t_k\}$). In a closed region of variables $\overline{t_{\{k\}}}$ and var($\{t_0, t_1, \ldots, t_k\}$), the global minimum value can be achieved at a certain combination of $\overline{t_{\{k\}}}$ and var($\{t_0, t_1, \ldots, t_k\}$). A numerical grid search on placement of intermediate points can be used to satisfy the combination of these 2 variables. | Equispaced design appears to have good properties at first glance, however, optimal placement of design points requires further research. |

them play a role in the covariance matrix. When designing an experiment, researchers are interested in the properties of the treatment effect estimator variance var($\hat{\beta}_x$); a design objective is to minimize the treatment estimator variance. Formula (2) provides a quantitative objective function to explore where to place postbaseline assessment timepoints. We mention without proving that var($\hat{\beta}_x$) decreases as $k$ increases, since more information is collected with a longer duration experiment. Also note that assigning equal sample sizes of experimental and control groups is optimal in terms of minimizing Formula (2). We summarize in Table 1 how different parameters in the model affect var($\hat{\beta}_x$) and possible implications for experimental design.

In addition to the RCRM when dealing with correlation within experimental units, another popular model is the mixed model with repeated measures (MMRM) (Gadbury et al., 2003; Little & Rubin, 2019). The MMRM treats the assessment time as categorical, and the response variable is usually "change-from-baseline" with baseline value being added as a covariate. Under the RCRM generating model, Mackey et al. (2017) compares the RCRM and MMRM in power and sample size. If $\hat{\beta}_{x,\text{MMRM}}$ is the treatment effect estimator (annualized for fair comparison) at time $t_k$ under MMRM fitting, it can be shown (Mackey et al., 2017) that

$$\text{var}(\hat{\beta}_{x,\text{MMRM}}) = \left(\frac{m+n}{mn}\right)\frac{\sigma^4 + \sigma_\beta^2\left(\sigma_\alpha^2(1-\rho^2)+\sigma^2\right)t_k^2 + 2\sigma^2\left(\sigma_\alpha^2 + \rho\sigma_\alpha\sigma_\beta t_k\right)}{\left(\sigma_\alpha^2+\sigma^2\right)t_k^2},$$

under the assumption that all subjects finish assessments at $t_0, t_1, \ldots, t_k$. It can be shown that $\text{var}(\hat{\beta}_{x,\text{MMRM}})$ from the MMRM is identical to $\text{var}(\hat{\beta}_x)$ from the RCRM, when the generating model is the RCRM and that there is only one postassessment ($k = 1$). In other cases where there are at least two postbaseline assessments, it can be shown that the RCRM is more efficient than the MMRM (Mackey et al., 2017). Granted, one can construct a contrast across different timepoints based on the MMRM, however the focus is usually on the last timepoint in primary analyses of clinical trials (Biogen, 2015; Hoffmann-La Roche, 2016; Honig et al., 2018). Chen et al. (2018) also compares the RCRM and MMRM through simulations, concluding that in general both are type I error controlled under model misspecification and that the RCRM (without random effects added) has a moderate power advantage over the MMRM.

We also compare the RCRM treatment effect estimate variance with the two-stage formulation variance (Fitzmaurice et al., 2012). Essentially, Stage 1 fits separate linear regression models for each subject, and Stage 2 further assumes that the fitted individual-specific effects are random. Yang et al. (2001) shows that the two-stage estimators for the model parameters are consistent estimators under the assumption that the fourth moments of $\beta$, $\beta_x$, and $\varepsilon_{ij}$ exist. The two-stage formulation variance on the treatment effect $\text{var}(\widehat{\beta}_{x,\text{2-stage}})$ is expressed as

$$\text{var}(\widehat{\beta}_{x,\text{2-stage}}) = \left(\frac{m+n}{mn}\right)\left(\frac{\sigma^2}{(k+1)\left(\overline{\mathbf{t}_{\{k\}}^2} - \overline{\mathbf{t}_{\{k\}}}\,\overline{\mathbf{t}_{\{k\}}}\right)} + \sigma_\beta^2\right).$$

Clearly, $\text{var}(\widehat{\beta}_{x,\text{2-stage}})$ does not take into account the impact of the variability $\sigma_\alpha^2$ at baseline or its correlation with the slope, whereas $\text{var}(\hat{\beta}_x)$ from the RCRM provides a full description of how different variability components and placement of points affect the treatment effect variance.

Next, we extend the RCRM covariance matrix to the case of monotone missingness.

**Theorem 1.** *Under the RCRM generating model described in (1), assuming monotone missingness for assessments $t_0, t_1, t_2, \ldots, t_J$ and assuming same dropout rate between experimental and control groups, the covariance matrix of the MLE vector $\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \\ \hat{\beta}_x \end{bmatrix}$ is*

$$\begin{aligned} var(\hat{\boldsymbol{\beta}}) &= \begin{bmatrix} (m+n)A^* & (m+n)B^* & mB^* \\ (m+n)B^* & (m+n)C^* & mC^* \\ mB^* & mC^* & mC^* \end{bmatrix}^{-1} \\ &= \begin{bmatrix} -\dfrac{C^*}{(B^{*2}-A^*C^*)(m+n)} & \dfrac{B^*}{(B^{*2}-A^*C^*)(m+n)} & 0 \\ \dfrac{B^*}{(B^{*2}-A^*C^*)(m+n)} & -\dfrac{A^*C^*(m+n)-B^{*2}m}{C^*(B^{*2}-A^*C^*)n(m+n)} & -\dfrac{1}{C^*n} \\ 0 & -\dfrac{1}{C^*n} & \dfrac{m+n}{C^*mn} \end{bmatrix}, \end{aligned} \tag{3}$$

*where $J$ is the maximum number of postbaseline assessments, and*

$$A^* = \sum_{k=0}^{J} p_k \left\{ \frac{(k+1)\left[\sigma^2 + (k+1)\sigma_\beta^2\left(\overline{\boldsymbol{t}_{\{k\}}^2} - \overline{\boldsymbol{t}_{\{k\}}}\,\overline{\boldsymbol{t}_{\{k\}}}\right)\right]}{\sigma^4 + (k+1)\sigma^2\left(\sigma_\beta^2\overline{\boldsymbol{t}_{\{k\}}^2} + \sigma_\alpha^2 + 2\rho\sigma_\alpha\sigma_\beta\overline{\boldsymbol{t}_{\{k\}}}\right) + (k+1)^2\sigma_\alpha^2\sigma_\beta^2(1-\rho^2)\left(\overline{\boldsymbol{t}_{\{k\}}^2} - \overline{\boldsymbol{t}_{\{k\}}}\,\overline{\boldsymbol{t}_{\{k\}}}\right)} \right\},$$

$$B^* = \sum_{k=0}^{J} p_k \left\{ \frac{(k+1)\left[\sigma^2\overline{\boldsymbol{t}_{\{k\}}} - (k+1)\rho\sigma_\alpha\sigma_\beta\left(\overline{\boldsymbol{t}_{\{k\}}^2} - \overline{\boldsymbol{t}_{\{k\}}}\,\overline{\boldsymbol{t}_{\{k\}}}\right)\right]}{\sigma^4 + (k+1)\sigma^2\left(\sigma_\beta^2\overline{\boldsymbol{t}_{\{k\}}^2} + \sigma_\alpha^2 + 2\rho\sigma_\alpha\sigma_\beta\overline{\boldsymbol{t}_{\{k\}}}\right) + (k+1)^2\sigma_\alpha^2\sigma_\beta^2(1-\rho^2)\left(\overline{\boldsymbol{t}_{\{k\}}^2} - \overline{\boldsymbol{t}_{\{k\}}}\,\overline{\boldsymbol{t}_{\{k\}}}\right)} \right\},$$

$$C^* = \sum_{k=0}^{J} p_k \left\{ \frac{(k+1)\left[\sigma^2\overline{\boldsymbol{t}_{\{k\}}^2} + (k+1)\sigma_\alpha^2\left(\overline{\boldsymbol{t}_{\{k\}}^2} - \overline{\boldsymbol{t}_{\{k\}}}\,\overline{\boldsymbol{t}_{\{k\}}}\right)\right]}{\sigma^4 + (k+1)\sigma^2\left(\sigma_\beta^2\overline{\boldsymbol{t}_{\{k\}}^2} + \sigma_\alpha^2 + 2\rho\sigma_\alpha\sigma_\beta\overline{\boldsymbol{t}_{\{k\}}}\right) + (k+1)^2\sigma_\alpha^2\sigma_\beta^2(1-\rho^2)\left(\overline{\boldsymbol{t}_{\{k\}}^2} - \overline{\boldsymbol{t}_{\{k\}}}\,\overline{\boldsymbol{t}_{\{k\}}}\right)} \right\},$$

*where*

$$\overline{\boldsymbol{t}_{\{k\}}^2} = \frac{1}{k+1}\sum_{j=0}^{k} t_j^2$$

$$\overline{\boldsymbol{t}_{\{k\}}} = \frac{1}{k+1}\sum_{j=0}^{k} t_j$$

*with $t_0 \equiv 0$; $m$ and $n$ are baseline sample sizes of the experimental group and control group, respectively; $p_k$ $(k = 0, 1, 2, \ldots, J)$ is the percentage of subjects who only have assessments at $t_0, t_1, \ldots, t_k$ (i.e., no additional assessments), with $\sum_{k=0}^{J} p_k = 1$.*

*Proof.* See Appendix A.2. □

Note that $p_k$ $(k = 0, 1, \ldots, J)$ is the same between the experimental and control groups, under the assumption of same dropout rate between groups. In cases where the dropout rates between groups are not the same, the result in Theorem 1 can be generalized accordingly; we present this generalized result in Theorem A1 in Appendix A.3. Unless specifically mentioned, we assume the same dropout rate between groups in this article. Since many experiments are designed with equally spaced assessment gaps, we also derive the variance results in this specific setting, presented in Corollary A1 in Appendix A.4.

In clinical trials, it is often reasonable to assume certain dropout patterns based on either the knowledge of historical trials or the nature of the therapeutic area. Next in this section, we will give an example of how a dropout assumption can be incorporated in the variance calculation in the setting with equally spaced assessment gaps. We further extend the formula to a special setting called common-close in an experimental design.

## 3.1 | Dropout and common-close

A commonly used dropout model in clinical trials is the exponential model. Under this assumption, the number of subjects in the trial at time $t$ is expressed as $N(t) = N(0)e^{-\lambda t}$, where $N(0)$ is the number of subjects at baseline. In practice, the dropout rate $\lambda$ can often be estimated from historical data. Again assuming monotone missingness and equal assessment gaps denoted as $h$, the probabilities $p_k$'s are calculated as

$$p_k = \begin{cases} e^{-kh\lambda}\left(1 - e^{-h\lambda}\right), & k = 0, 1, 2, \ldots, J-1 \\ e^{-Jh\lambda}, & k = J. \end{cases} \tag{4}$$

For other dropout assumptions, $p_k$'s can also be calculated accordingly; we will not further enumerate other cases.

**TABLE 2** Parameters, variance of treatment effect estimator, and experimental design (Continued)

| Parameter | Impact on var($\hat{\beta}_x$) | Implications for Experimental Design |
|---|---|---|
| $p_k$ | • Variance decreases as larger $p_k$'s are distributed to larger $k$'s (later time points) | Smaller drop-out rate (i.e. assigning more $p_k$'s to later time points) will lead to power gain. |
| Equal follow-up vs. common-close | • Common-close always provides smaller variance (some subjects gain extra assessments and thus $p_k$'s will be distributed to later time points) | Common-close designs and analyses may provide substantial power increase over equal follow-up. |
| | • Longer enrollment with more subjects enrolled early in the common-close design always provides smaller variance (more assessment points; $p_k$'s being assigned more to later time points) | |

As the next part of this section, we point out that the variance formula described in either Theorem 1 or Corollary A1 can accommodate the common-close experimental design. A common-close specifies that all subjects remain in their randomized group and are assessed until the last enrolled subject in the experiment reaches a certain landmark in time (as opposed to a common duration from baseline) (Tariot et al., 2018). In the common-close setting, the probabilities $p_k$ are affected by not only the dropout rate, but also enrollment (e.g., duration and distribution). This design feature adds additional computation to the power and sample size estimation and can qualitatively affect power and sample size. This makes intuitive sense since the RCRM is provided with extra information coming from additional assessments that are further from baseline than all other assessments not part of the common-close period. We next present how to calculate $p_k$'s in a common-close design, under the assumption of exponential dropout and equal assessment gap $h$.

We assume the enrollment duration is $E$ with $E = (e + \delta)h$ where $e$ is an integer such that $0 \leq \delta < 1$. The experiment is designed such that every subject remains in the experiment until the last enrolled subject reaches time $Jh$. Without loss of generality, we further assume the enrollment duration $E$ is shorter than $Jh$. As a result, subjects who enrolled early could be assessed up to $t = (J + e)h$. For example, if the enrollment duration is 1.7 years and every subject is assessed with 0.5-year equal gaps until the last enrolled subject reaches 2 years, then $h = 0.5$, $J = 4$, $E = 1.7$ with $e = 3$ and $\delta = 0.4$; early-enrollers could be assessed up to 3.5 years.

Further, we assume $100 * q_0\%$ of the subjects are enrolled in the enrollment interval $[0, \delta h]$ and $100 * q_j\%$ are enrolled in the enrollment interval $((\delta + j - 1)h, (\delta + j)h]$ $(j = 1, 2, \ldots, e)$. Note that $\sum_{j=0}^{e} q_j = 1$. To allow for flexibility, we will not assume any specific enrollment distribution. To present the variance, power, and sample size formula in the common-close setting, all the upper bound of summation indices in the previous Theorem 1 and Corollary A1 will be extended from $J$ to $J + e$. We first categorize subjects into groups according to the maximum number of possible assessments and then apply the exponential dropout assumption, we present the $p_k$'s in the common-close setting.

$$p_k = \begin{cases} e^{-kh\lambda}\left(1 - e^{-h\lambda}\right), & k = 0, 1, 2, \ldots, J-1 \\ \left(\sum_{j=0}^{J+e-1-k} q_j\right) e^{-kh\lambda}\left(1 - e^{-h\lambda}\right) + q_{J+e-k}e^{-kh\lambda}, & k = J, J+1, \ldots, J+e-1 \\ q_0 e^{-(J+e)h\lambda}, & k = J+e. \end{cases} \tag{5}$$

Since common-close always provides smaller variance than follow-up $= Jh$ for all subjects (by examining Equations (3), (4), and (5); details not shown in this article), common-close designs have the potential to provide a substantial power increase over equal follow-up designs. More specifically, longer enrollment with more subjects enrolled toward the beginning of a common-close design, resulting in larger $e$ (more assessment points) and more $p_k$'s being assigned to later time points, will also increase power. We summarize the implications of $p_k$'s and common-close design in Table 2, as the continuation of Table 1.

## 4 | POWER FORMULA

We start this section by stating the hypothesis test in the following where we assume the true $\beta_x > 0$.

$$H_0 : \beta_x = 0,$$
$$H_1 : \beta_x > 0.$$

We assume the one-sided confidence level is $a/2$. Under the RCRM, the statistical power is expressed as:

$$\text{Power} = \text{P}(\text{Reject } H_0 | \beta_x > 0) \tag{6}$$

$$= \Phi\left( \frac{\beta_x}{\sqrt{\text{var}\left(\hat{\beta}_x\right)}} - z_{1-\frac{a}{2}} \right), \tag{7}$$

where

$$\text{var}\left(\hat{\beta}_x\right) = \frac{m+n}{C^{**}mn}$$

with notations from Theorem 1 (in which case $C^{**} = C^*$) or Corollary A1 (in which case $C^{**} = C^\dagger$), and $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. In the case where the true $\beta_x < 0$, then $\beta_x$ in Formula (7) can be replaced with $-\beta_x$.

While our assumption of mutivariate normality with known variance components ($\sigma_\alpha^2$, $\sigma_\beta^2$, $\sigma^2$, and $\rho$) implies exact normality of the estimated parameters, the result will be asymptotically normal in cases where the variance components are consistently estimated. In small samples, the power formula can be modified to incorporate the noncentral $t$-distribution with the appropriate degrees of freedom, for example, by Kenward-Roger (Kenward & Roger, 1997). In Section 6, we will illustrate the accuracy of asymptotic estimates of power when variance components are not assumed to be known.

## 5 | SAMPLE SIZE FORMULA

We present the following theorem on sample size planning.

**Theorem 2.** *Let $\gamma$ denote the sample size allocation ratio between the experimental group and control group (experimental over control) at baseline, and let $N$ denote the total sample size needed with $1-\eta$ power at $a/2$ one-sided confidence level, we have*

$$N = \lceil \frac{(1+\gamma)^2}{\gamma C^{**}} \left( \frac{z_{1-\frac{a}{2}} + z_{1-\eta}}{\beta_x} \right)^2 \rceil, \tag{8}$$

*where $C^{**}$ is from either Theorem 1 or Corollary A1, and $\lceil \cdot \rceil$ is the ceiling function.*

*Proof.* The proof is straightforward based on Equation (7) and therefore omitted. □

## 6 | SIMULATIONS

In this section, we perform simulation studies to evaluate the performance of the power formula from Section 4. We assume $\alpha = 0$, $\beta = -1$, $\sigma_\alpha^2 = 2$, $\sigma_\beta^2 = 0.5$, and $\sigma^2 = 1$. We choose $\rho = -0.6$ and 0.3 so that the overall pairwise correlation between visits ($Y_{ij}$'s) is about 0.5 and 0.7, respectively (Lu et al., 2009); we also choose $\rho = 0$ to evaluate the performance

**TABLE 3** Data distribution in the simulation study

| Total Dropout (%) | $\lambda$ | Baseline $p_0$ (%) | $t = 0.5$ $p_1$ (%) | $t = 1$ $p_2$ (%) | $t = 1.5$ $p_3$ (%) | $t = 2$ $p_4$ (%) |
|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 | 100 |
| 15 | 0.081 | 4.0 | 3.8 | 3.7 | 3.5 | 85 |
| 30 | 0.178 | 8.5 | 7.8 | 7.1 | 6.5 | 70 |

of the power formula under the assumption of independence between intercept and slope random effects. To examine the sample size sensitivity of the RCRM power formula in Section 4, we consider two sample size scenarios: 100 and 500. The sample size of 100 aims to reflect a potential Phase II Alzheimer's disease (AD) trial (Salloway et al., 2018; Cummings et al., 2018), and the size of 500 reflects a potential Phase III AD trial (Honig et al., 2018; Ostrowitzki et al., 2017). In both scenarios, we assume the experimental and control allocation is 1:1 and the total trial duration is two years with a 0.5-year assessment gap. We calculate $\beta_x$ so that the theoretical power to detect the treatment effect from Equation (7) is 90% and 80%, respectively, when the total sample size is 500 and 100, assuming no dropouts. The parameters are summarized in the following.

1. Total size 500 with 1:1 allocation ratio: theoretical power is 90% with one-sided 0.025 level hypothesis test, assuming no dropouts. $\beta_x = 0.208$ in the case of $\rho = -0.6$; $\beta_x = 0.274$ in the case of $\rho = 0.3$; $\beta_x = 0.265$ in the case of $\rho = 0$.
2. Total size 100 with 1:1 allocation ratio: theoretical power is 80% with one-sided 0.1 level hypothesis test, assuming no dropouts. $\beta_x = 0.305$ when $\rho = -0.6$; $\beta_x = 0.402$ when $\rho = 0.3$; $\beta_x = 0.389$ when $\rho = 0$.

We further assume three exponential dropout scenarios with 0%, 15%, and 30% end-of-study dropout percentages. For simplicity, the exponential dropout assumption is used. Table 3 summarizes the exponential yearly dropout rate $\lambda$ and the data distribution at each assessment. For example, in the case of 15% total dropout at the end of the trial, the exponential yearly dropout rate $\lambda$ is 0.081; about 3.7% of the subjects will only have data up until $t = 1$ (corresponds to $p_2$) and 85% of the subjects will complete the study (corresponds to $p_4$). In our simulation, we generate missing values in each replicate based on the multinomial distribution with probabilities specified in Table 3.

We generate data from both the correctly specified model and contaminated models to, respectively, evaluate the validity and robustness of the RCRM power formula. We summarize the five model generating mechanisms below:

1. Generate data from normal distributions ($u_{0i}$, $u_{1i}$, and $\varepsilon_{ij}$); mean trajectory is generated from the specified RCRM model.
2. Generate baseline data from a truncated normal distribution, forcing the baseline score $\geq 0.8$; mean trajectory is generated from the specified RCRM model. This is to mimic trial inclusion criterion of the baseline score needing to meet a certain threshold; this leads to the probability of enrollment being around 68%.
3. Generate $u_{0i}$ and $u_{1i}$ from a multivariate $t$ distribution with degrees of freedom 3; mean trajectory is generated from the specified RCRM model. This is to mimic a heavy tailed distribution.
4. Generate $u_{0i}$ and $u_{1i}$ from a truncated multivariate $t$ distribution with degrees of freedom 3, forcing the baseline score $\geq 0.8$; mean trajectory is generated from the specified RCRM model. This is to mimic both trial inclusion criterion and heavy tails.
5. Generate data from normal distributions ($u_{0i}$, $u_{1i}$, and $\varepsilon_{ij}$). However, the mean trajectory is generated from a quadratic model where a $t^2$ term with coefficient of $0.05\beta$ is added to both experimental and control groups. As a result, the mean change from baseline at $t = 2$ in both groups gets inflated by 10% compared with that from the specified RCRM model. This scenario is to mimic the case in which the time trajectory is not linear.

We use the `lme` R function from the `nlme` package (Pinheiro et al., 2017) to fit the generated data and the number of iterations in our simulation studies is 10,000. We also calculate the two-stage formulation power described in Fitzmaurice, Laird, and Ware (Fitzmaurice et al., 2012). Although acknowledging the full model should come from Equation (1) where four variance components are warranted, the calculation of two-stage formulation power only takes into account the slope random effect $u_{1i}$ and the pure error $\varepsilon_{ij}$ (Fitzmaurice et al., 2012). We extend the two-stage formulation power formula to

**TABLE 4** Large sample size simulation setup and results (total sample size = 500 with 1:1 allocation; 10,000 replications). $\alpha = 0, \beta = -1, \sigma_\alpha^2 = 2, \sigma_\beta^2 = 0.5, \sigma^2 = 1$

| | | | | Simulated Power | | | | |
|---|---|---|---|---|---|---|---|---|
| Parameters | Total Dropout (%) | RCRM Theoretical Power | Two-Stage Formulation Power | Normal | Truncated Normal | $t_3$ | Truncated $t_3$ | Quad |
| $\rho = -0.6$ $\beta_x = 0.208$ | 0 | 0.900 | 0.689 | 0.899 | 0.900 | 0.907 | 0.862 | 0.898 |
| | 15 | 0.863 | 0.639 | 0.862 | 0.865 | 0.870 | 0.823 | 0.865 |
| | 30 | 0.813 | 0.581 | 0.803 | 0.812 | 0.830 | 0.781 | 0.808 |
| $\rho = 0.3$ $\beta_x = 0.274$ | 0 | 0.900 | 0.899 | 0.892 | 0.893 | 0.911 | 0.865 | 0.899 |
| | 15 | 0.865 | 0.863 | 0.868 | 0.860 | 0.880 | 0.829 | 0.864 |
| | 30 | 0.818 | 0.815 | 0.818 | 0.818 | 0.826 | 0.773 | 0.817 |
| $\rho = 0$ $\beta_x = 0.265$ | 0 | 0.900 | 0.879 | 0.900 | 0.903 | 0.902 | 0.860 | 0.898 |
| | 15 | 0.865 | 0.840 | 0.863 | 0.866 | 0.874 | 0.825 | 0.870 |
| | 30 | 0.819 | 0.789 | 0.822 | 0.814 | 0.840 | 0.770 | 0.817 |

**TABLE 5** Small sample size simulation setup and results (total sample size = 100 with 1:1 allocation; 10,000 replications). $\alpha = 0, \beta = -1, \sigma_\alpha^2 = 2, \sigma_\beta^2 = 0.5, \sigma^2 = 1$

| | | | | Simulated Power | | | | |
|---|---|---|---|---|---|---|---|---|
| Parameters | Total Dropout (%) | RCRM Theoretical Power | Two-Stage Formulation Power | Normal | Truncated Normal | $t_3$ | Truncated $t_3$ | Quad |
| $\rho = -0.6$ $\beta_x = 0.305$ | 0 | 0.800 | 0.627 | 0.800 | 0.805 | 0.822 | 0.786 | 0.802 |
| | 15 | 0.764 | 0.593 | 0.762 | 0.760 | 0.799 | 0.754 | 0.768 |
| | 30 | 0.721 | 0.554 | 0.713 | 0.721 | 0.753 | 0.705 | 0.721 |
| $\rho = 0.3$ $\beta_x = 0.402$ | 0 | 0.800 | 0.799 | 0.792 | 0.798 | 0.821 | 0.781 | 0.797 |
| | 15 | 0.766 | 0.764 | 0.766 | 0.758 | 0.795 | 0.745 | 0.768 |
| | 30 | 0.725 | 0.722 | 0.728 | 0.719 | 0.745 | 0.715 | 0.720 |
| $\rho = 0$ $\beta_x = 0.389$ | 0 | 0.800 | 0.779 | 0.796 | 0.793 | 0.822 | 0.780 | 0.799 |
| | 15 | 0.766 | 0.743 | 0.770 | 0.761 | 0.791 | 0.755 | 0.770 |
| | 30 | 0.725 | 0.701 | 0.722 | 0.723 | 0.750 | 0.716 | 0.719 |

also accommodate the dropout information:

$$\text{Power}_{2\text{-stage}} = \Phi\left(\frac{\beta_x}{\sqrt{\text{var}\left(\widehat{\beta}_{x,2\text{-stage}}\right)}} - z_{1-\frac{a}{2}}\right),$$

where

$$\text{var}\left(\widehat{\beta}_{x,2\text{-stage}}\right) = \left(\frac{m+n}{mn}\right) \Big/ \sum_{k=0}^{J} p_k \left(\frac{(k+1)\left(\overline{\mathbf{t}_{\{k\}}^2} - \overline{\mathbf{t}_{\{k\}}}\,\overline{\mathbf{t}_{\{k\}}}\right)}{\sigma^2 + (k+1)\left(\overline{\mathbf{t}_{\{k\}}^2} - \overline{\mathbf{t}_{\{k\}}}\,\overline{\mathbf{t}_{\{k\}}}\right)\sigma_\beta^2}\right).$$

The simulation setup and results are summarized in Tables 4 and 5. In general, the theoretical RCRM power is very close to the simulated power regardless of the sample size.

**TABLE 6** ADNI RCRM fitting results

| $\alpha$ | $\beta$ | $\sigma_\alpha^2$ | $\sigma_\beta^2$ | $\rho$ | $\sigma^2$ |
|---|---|---|---|---|---|
| 1.53 | 1.10 | 0.54 | 1.01 | 0.07 | 0.81 |

We first examine Table 4 where the total size is 500. When data are simulated from normal distribution, truncated normal distribution, $t_3$ distribution, and the quadratic trajectory, the theoretical RCRM power is bounded within 0.01 of the simulated power, except for some cases of $t_3$ distribution (off up to 0.021 in difference). Of note, when data are generated from the $t_3$ distribution, the theoretical RCRM power consistently underestimates the simulated power. In our simulation setup, the case of truncated $t_3$ distribution yields the farthest departure from the model assumption in terms of variance structure, in which case the theoretical RCRM power is the most inaccurate; the theoretical RCRM power overestimates the simulated power by about 0.03 to 0.04 in difference. When the data are generated from the quadratic trajectory, the theoretical RCRM power is still very close to the simulated power.

We next examine Table 5 where the total size is relatively small: 100. When data are simulated from normal distribution, truncated normal distribution, and the quadratic trajectory, the theoretical RCRM power is bounded within 0.01 of the simulated power. Again, as with what is observed in the case of sample size = 500, when data are generated from the $t_3$ distribution, the theoretical RCRM power consistently underestimates the simulated power. However, when data are generated from the truncated $t_3$ distribution, the theoretical RCRM power slightly overestimates the simulated power. With the smaller sample size of 100, the theoretical RCRM power still performs reasonably well when the mean trajectory is generated by a quadratic curve.

In both sample size cases where $\rho = -0.6$, the two-stage formulation power dramatically underestimates the simulated power by about 0.2 to 0.3 in difference. When $\rho = 0.3$ and $\rho = 0$, however, the two-stage formulation power is very close to the theoretical RCRM and also the simulated power. We conclude that the theoretical RCRM power formula will be useful in general practice. The two-stage formulation power, however, can be dramatically inaccurate and thus should be used with caution.

## 7 | AN APPLICATION

To illustrate how to apply the RCRM sample size formula for trial planning, we use the "ADNIMERGE" data set from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu) for parameter estimation. 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. We take a subset of the data using the following simple inclusion criteria, aiming to mimic some of the most common assumptions in recent AD Phase III trials (Biogen, 2015; Hoffmann-La Roche, 2016; Honig et al., 2018).

1. Amyloid PET positive at baseline (variable "PET.bl.ind" in the data set)
2. AD status is late mild cognitive impairment (LMCI) at baseline (variable "DX.bl" in the data set)

The outcome measure chosen is the Clinical Dementia Rating sum of boxes (CDR-SB) (Hughes et al., 1982). To maintain the population of interest, we only include subject records within five years after baseline. After applying the inclusion criteria, there are 255 subjects with 1305 observations remaining. We then fit the RCRM model described in Equation (1), except that the $\beta_x$ term is removed because all subjects in this database did not receive any disease-modifying treatment. Table 6 summarizes the fitting results.

Now we design a trial to detect the effect of a disease-modifying treatment. From Table 6, $\beta = 1.10$, which will be assumed as the placebo population yearly slope in the trial design. If we assume the disease-modifying treatment has a true 30% relative reduction effect on slowing the disease progression, $\beta_x$ can be calculated as $-30\%\beta = -0.33$.

Instead of using a traditional trial design where all subjects are scheduled to be assessed for a fixed duration in time, we illustrate the use of a common-close design. The aim is to showcase how the common-close design could reduce the sample size to maintain a power target. In particular, we design a trial where every subject remains in the 1:1 randomized group (experimental or control) until the last enrolled subject reaches two years, at which point the trial will end. Further,

**TABLE 7** Data distribution in the application

| | Baseline | $t = 0.5$ | $t = 1$ | $t = 1.5$ | $t = 2$ | $t = 2.5$ | $t = 3$ | $t = 3.5$ |
|---|---|---|---|---|---|---|---|---|
| $\lambda$ | $p_0$ (%) | $p_1$ (%) | $p_2$ (%) | $p_3$ (%) | $p_4$ (%) | $p_5$ (%) | $p_6$ (%) | $p_7$ (%) |
| 0.081 | 3.97 | 3.81 | 3.66 | 3.51 | 27.40 | 25.35 | 23.43 | 8.86 |

**TABLE 8** Schedule of assessments and sample size (exponential dropout rate $\lambda = 0.081$)

| Schedule of Assessments | | | | | | | | Total Sample Size |
|---|---|---|---|---|---|---|---|---|
| Baseline | 0.5 | 1 | 1.5 | 2 | 2.5 | 3 | 3.5 | 368 |
| Baseline | 0.25 | 0.5 | 1.5 | 2 | 3 | 3.25 | 3.5 | 378 |
| Baseline | 0.25 | 0.5 | 0.75 | 1 | 2 | 3.25 | 3.5 | 388 |
| Baseline | 0.5 | 1 | 2 | 2.1 | 2.2 | 2.3 | 3.5 | 366 |

we assume that the total enrollment duration is 1.7 years and that subjects are entering the trial based on a Poisson process. Thus, the earliest-enrolled subjects can be assessed in the trial for up to 3.7 years. We further assume that the exponential dropout rate is $\lambda = 0.081$, which corresponds to 7.78% dropout in the first year of enrollment. We point out that many assumptions in the proceeding are mainly for illustration and can be easily adapted to other settings.

We first calculate the sample size required for a trial design where assuming equal 0.5-year assessment gaps, to achieve 80% power at one-sided $a/2 = 0.025$ level detecting 30% treatment relative reduction (i.e., $\beta_x = -0.33$). Without showing details in calculation, we present in Table 7 the data distribution based on our previously stated assumptions. Approximately 57.6% of the subjects will reach at least year 2.5 in the study.

Using the sample size formula in Equation (8) along with the variance formula in Theorem 1 and the data distribution from Table 7, the total sample size needed is 368 with 184 subjects in each group.

We then explore the sample size needed in the scenario of unequal assessment gaps; we keep the number of assessments the same and we fix the last assessment at $t = 3.5$. The total sample size needed under several assumptions of schedule of assessments is summarized in Table 8. The first row serves as the basis where 0.5-year equal gap is assumed. The schedule of assessments in the second row replaces $t = 1$ with 0.25 and $t = 2.5$ with 3.25, which leads to 10 extra subjects needed in total. In the next row, more frequent assessments are assumed in the beginning and the end, which leads to 20 extra subjects needed in total; the last row reduces two subjects needed in total, putting more assessments around year 2. We point out that the sample size depends on the schedule of assessments, dropout pattern, and enrollment pattern. In this example, the equally gapped schedule of assessments appears to be performing well in terms of the sample size. However, further research on the optimal placement of assessments is warranted.

## 8 | DISCUSSION

In this article, we have provided the theoretical variance, power, and sample size formulae in the context of the RCRM. Simulations have illustrated the accuracy and robustness of the theoretical results in both small and large samples. We have also provided an application to demonstrate sample size and assessment planning, based on comprehensive assumptions including enrollment pattern and dropout rate.

The use of the RCRM assumes linearity, or more generally, further assumes the validity of partitioning the total variability into four components: baseline, slope, correlation between baseline and slope, and pure error. We believe that in many nonlinear cases, the linear framework provides a useful first-order approximation, which is especially true when the response variable changes slowly within the experiment's timescale. In addition, the four-part variability partition provides flexibility to accommodate many underlying variance–covariance structures. In cases where the linearity assumption is in question and/or the control of type I error is critical, the robust sandwich variance estimator proposed by White (1982) can be used in addition to, or in place of, the likelihood-based variance during the analysis stage.

We have provided insights on how to design more efficient experiments. Common-close design, perhaps, is the most obvious choice to help boost power. The problem of optimizing placement of intermediate point(s), however, is nonlinear without a closed-form solution; further research is needed.

## 9 | R SHINY APPLICATION

To help researchers directly use the theoretical results from this article, we have built an R Shiny application, hosted on the cloud through shinyapps.io. To access the web version of the Shiny application, visit https://rcrm-power-size.shinyapps.io/rcrm_power_size/. For R code details, visit the GitHub site https://github.com/nan-hu-personal/RCRM_power_size.

## ACKNOWLEDGMENTS

We thank Paul Delmar, Paul Manser, and Bali Toth for providing insightful feedback to this article. We also thank Zhe Qu for helping build the R Shiny app based on the theoretical results in this article.

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

## CONFLICT OF INTEREST

The authors have declared no conflict of interest.

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in ADNI at https://ida.loni.usc.edu/login.jsp?project=ADNI.

## OPEN RESEARCH BADGES

This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available in the Supporting Information section.

This article has earned an open data badge "**Reproducible Research**" for making publicly available the code necessary to reproduce the reported results. The results reported in this article could fully be reproduced.

## ORCID

*Nan Hu* https://orcid.org/0000-0003-4439-0680

*Ronald Thomas* https://orcid.org/0000-0003-1686-4965

## REFERENCES

Biogen. (2015). 221ad301 phase 3 study of aducanumab (biib037) in early Alzheimer's disease (engage). Identification No. NCT02477800. https://clinicaltrials.gov/ct2/show/study/NCT02477800.

Chen, Y.-F., Ni, X., Fleisher, A. S., Zhou, W., Aisen, P., & Mohs, R. (2018). A simulation study comparing slope model with mixed-model repeated measure to assess cognitive data in clinical trials of Alzheimer's disease. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, *4*, 46–53.

Cummings, J. L., Cohen, S., van Dyck, C. H., Brody, M., Curtis, C., Cho, W., Ward, M., Friesenhahn, M., Rabe, C., Brunstein, F., Quartino, A., Honigberg, L. A., Fuji, R. N., Clayton, D., Mortensen, D., Ho, C., & Paul, R. (2018). Abby: A phase 2 randomized trial of crenezumab in mild to moderate Alzheimer disease. *Neurology*, *90*, e1889–e1897.

Curran, P. J., Obeidat, K., & Losardo, D. (2010). Twelve frequently asked questions about growth curve modeling. *Journal of Cognition and Development*, *11*, 121–136.

Diggle, P., Heagerty, P., Liang, K.-Y., & Zeger, S. (2002). *Analysis of longitudinal data* (2nd ed.). Oxford University Press.

Donohue, M., Gamst, A. C., & Edland, S. (2013). Longpower: Sample size calculations for longitudinal data. *R package version 1.0-19*.

East 6. (2020). Statistical software for the design, simulation and monitoring clinical trials. Cambridge MA: Cytel Inc.

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2012). *Applied longitudinal analysis* (2nd ed., Vol. 998). John Wiley & Sons.

Flaherty, K. R., Wells, A. U., Cottin, V., Devaraj, A., Walsh, S. L., Inoue, Y., Richeldi, L., Kolb, M., Tetzlaff, K., Stowasser, S., Coeck, C., Clerisme-Beaty, E., Rosenstock, B., Quaresma, M., Haeufel, T., Goeldner, R. G., Schlenker-Herceg, R., Brown, K. K., INBUILD Trial Investigators,. (2019). Nintedanib in progressive fibrosing interstitial lung diseases. *New England Journal of Medicine*, *381*, 1718–1727.

Gadbury, G., Coffey, C., & Allison, D. (2003). Modern statistical methods for handling missing repeated measurements in obesity trial data: Beyond locf. *Obesity Reviews*, *4*, 175–184.

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models* (1st ed.). Cambridge University Press.

Harrison, X. A., Donaldson, L., Correa-Cano, M. E., Evans, J., Fisher, D. N., Goodwin, C. E., Robinson, B. S., Hodgson, D. J., & Inger, R. (2018). A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ*, *6*, e4794.

Hedeker, D., Gibbons, R. D., & Waternaux, C. (1999). Sample size estimation for longitudinal designs with attrition: Comparing time-related contrasts between two groups. *Journal of Educational and Behavioral Statistics*, *24*, 70–93.

Heo, M., & Leon, A. C. (2009). Sample size requirements to detect an intervention by time interaction in longitudinal cluster randomized clinical trials. *Statistics in Medicine*, *28*, 1017–1027.

Higham, N. J. (2002). *Accuracy and stability of numerical algorithms* (2nd ed.). SIAM: Society for Industrial and Applied Mathematics.

Hoffmann-La Roche. (2016). Cread study: A study of crenezumab versus placebo to evaluate the efficacy and safety in participants with prodromal to mild Alzheimer's disease (AD). Identification No. NCT02670083. https://clinicaltrials.gov/ct2/show/NCT02670083.

Honig, L. S., Vellas, B., Woodward, M., Boada, M., Bullock, R., Borrie, M., Hager, K., Andreasen, N., Scarpini, E., Liu-Seifert, H., Case, M., Dean, R. A., Hake, A., Sundell, K., Poole Hoffmann, V., Carlson, C., Khanna, R., Mintun, M., DeMattos, R., ...Siemers, E. (2018). Trial of solanezumab for mild dementia due to Alzheimer's disease. *New England Journal of Medicine*, *378*, 321–330.

Hughes, C. P., Berg, L., Danziger, W., Coben, L. A., & Martin, R. L. (1982). A new clinical scale for the staging of dementia. *The British Journal of Psychiatry*, *140*, 566–572.

Kenward, M. G., & Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, *53*(3), 983–997.

Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, *38*, 963–974.

Liang, K.-Y., & Zeger, S. L. (2000). Longitudinal data analysis of continuous and discrete responses for pre-post designs. *Sankhyā: The Indian Journal of Statistics, Series B*, *62*, 134–148.

Little, R. J., & Rubin, D. B. (2019). *Statistical analysis with missing data* (2nd ed., Vol. 793). Wiley.

Lu, K., Luo, X., & Chen, P.-Y. (2008). Sample size estimation for repeated measures analysis in randomized clinical trials with missing data. *The International Journal of Biostatistics*, *4*(1), Art. no. 9.

Lu, K., Mehrotra, D. V., & Liu, G. (2009). Sample size determination for constrained longitudinal data analysis. *Statistics in Medicine*, *28*, 679–699.

Mackey, H., Hu, N., Malek-Ahmadi, M., Chen, Y., Tariot, P., Reiman, E., Lopera, F., Chen, K., & Thomas, R. (2017). An examination of rate of decline as an alternative to change from baseline. Talk presented at: The 10th Clinical Trials on Alzheimer's Disease. November 02, 2017. Boston, MA.

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., & SCarlet RoAD Investigators,. (2017). A phase III randomized trial of gantenerumab in prodromal Alzheimer's disease. *Alzheimer's Research & Therapy*, *9*, 95.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., Heisterkamp, S., Van Willigen, B., & Maintainer, R. (2017). Package "nlme."*Linear and Nonlinear Mixed Effects Models, version*, 3.

Richeldi, L., Du Bois, R. M., Raghu, G., Azuma, A., Brown, K. K., Costabel, U., Cottin, V., Flaherty, K. R., Hansell, D. M., Inoue, Y., Kim, D. S., Kolb, M., Nicholson, A. G., Noble, P. W., Selman, M., Taniguchi, H., Brun, M., Le Maulf, F., Girard, M., ... INPULSIS Trial Investigators,. (2014). Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. *New England Journal of Medicine*, *370*, 2071–2082.

Rosenberger, W. F., & Lachin, J. M. (2015). *Randomization in clinical trials: Theory and practice* (2nd ed.). John Wiley & Sons.

Roy, A., Bhaumik, D. K., Aryal, S., & Gibbons, R. D. (2007). Sample size determination for hierarchical longitudinal designs with differential attrition rates. *Biometrics*, *63*, 699–707.

Salloway, S., Honigberg, L. A., Cho, W., Ward, M., Friesenhahn, M., Brunstein, F., Quartino, A., Clayton, D., Mortensen, D., Bittner, T., Ho, C., Rabe, C., Schauer, S. P., Wildsmith, K. R., Fuji, R. N., Suliman, S., Reiman, E. M., Chen, K., & Paul, R. (2018). Amyloid positron emission tomography and cerebrospinal fluid results from a crenezumab anti-amyloid-beta antibody double-blind, placebo-controlled, randomized phase II study in mild-to-moderate Alzheimer's disease (BLAZE). *Alzheimer's Research & Therapy*, *10*, 1–13.

Tariot, P. N., Lopera, F., Langbaum, J. B., Thomas, R. G., Hendrix, S., Schneider, L. S., Rios-Romenets, S., Giraldo, M., Acosta, N., Tobon, C., Ramos, C., Espinosa, A., Cho, W., Ward, M., Clayton, D., Friesenhahn, M., Mackey, H., Honigberg, L., Sanabria Bohorquez, S., ...Reiman, E. M. (2018). The Alzheimer's prevention initiative autosomal-dominant Alzheimer's disease trial: A study of crenezumab versus placebo in preclinical PSEN1 E280A mutation carriers to evaluate efficacy and safety in the treatment of autosomal-dominant Alzheimer's disease, including a placebo-treated noncarrier cohort. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, *4*, 150–160.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica: Journal of the Econometric Society*, *50*, 1–25.

Yang, S.-S., Yu, Q., & Al-Zaid, M. A. (2001). A two-stage estimation procedure for linear mixed-effects models. *Communications in Statistics - Theory and Methods*, *30*, 2637–2653.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**How to cite this article:** Hu N, Mackey H, Thomas R. Power and sample size for random coefficient regression models in randomized experiments with monotone missing data. *Biometrical Journal*. 2021;1–19. https://doi.org/10.1002/bimj.202000184

## APPENDIX

### A.1 | Lemma 1 and proof

**Lemma A1.** *Under the RCRM generating model described in (1), assuming all subjects only complete assessments at* $t_0, t_1, t_2, \dots, t_k$, *the covariance matrix of the RCRM MLE vector* $\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \\ \hat{\beta}_x \end{bmatrix}$ *is*

$$
\begin{aligned}
var(\hat{\boldsymbol{\beta}}) &= \begin{bmatrix} (m+n)A & (m+n)B & mB \\ (m+n)B & (m+n)C & mC \\ mB & mC & mC \end{bmatrix}^{-1} \\
&= \begin{bmatrix} -\dfrac{C}{(B^2-AC)(m+n)} & \dfrac{B}{(B^2-AC)(m+n)} & 0 \\ \dfrac{B}{(B^2-AC)(m+n)} & -\dfrac{AC(m+n)-B^2m}{C(B^2-AC)n(m+n)} & -\dfrac{1}{Cn} \\ 0 & -\dfrac{1}{Cn} & \dfrac{m+n}{Cmn} \end{bmatrix},
\end{aligned} \tag{A.1}
$$

*where* $k$ *represents the number of postbaseline assessments, and*

$$
A = \frac{(k+1)\left[\sigma^2 + (k+1)\sigma_\beta^2\left(\overline{\boldsymbol{t}_{\{k\}}^2} - \overline{\boldsymbol{t}_{\{k\}}}\,\overline{\boldsymbol{t}_{\{k\}}}\right)\right]}{\sigma^4 + (k+1)\sigma^2\left(\sigma_\beta^2\overline{\boldsymbol{t}_{\{k\}}^2} + \sigma_\alpha^2 + 2\rho\sigma_\alpha\sigma_\beta\overline{\boldsymbol{t}_{\{k\}}}\right) + (k+1)^2\sigma_\alpha^2\sigma_\beta^2(1-\rho^2)\left(\overline{\boldsymbol{t}_{\{k\}}^2} - \overline{\boldsymbol{t}_{\{k\}}}\,\overline{\boldsymbol{t}_{\{k\}}}\right)},
$$

$$
B = \frac{(k+1)\left[\sigma^2\overline{\boldsymbol{t}_{\{k\}}} - (k+1)\rho\sigma_\alpha\sigma_\beta\left(\overline{\boldsymbol{t}_{\{k\}}^2} - \overline{\boldsymbol{t}_{\{k\}}}\,\overline{\boldsymbol{t}_{\{k\}}}\right)\right]}{\sigma^4 + (k+1)\sigma^2\left(\sigma_\beta^2\overline{\boldsymbol{t}_{\{k\}}^2} + \sigma_\alpha^2 + 2\rho\sigma_\alpha\sigma_\beta\overline{\boldsymbol{t}_{\{k\}}}\right) + (k+1)^2\sigma_\alpha^2\sigma_\beta^2(1-\rho^2)\left(\overline{\boldsymbol{t}_{\{k\}}^2} - \overline{\boldsymbol{t}_{\{k\}}}\,\overline{\boldsymbol{t}_{\{k\}}}\right)},
$$

$$
C = \frac{(k+1)\left[\sigma^2\overline{\boldsymbol{t}_{\{k\}}^2} + (k+1)\sigma_\alpha^2\left(\overline{\boldsymbol{t}_{\{k\}}^2} - \overline{\boldsymbol{t}_{\{k\}}}\,\overline{\boldsymbol{t}_{\{k\}}}\right)\right]}{\sigma^4 + (k+1)\sigma^2\left(\sigma_\beta^2\overline{\boldsymbol{t}_{\{k\}}^2} + \sigma_\alpha^2 + 2\rho\sigma_\alpha\sigma_\beta\overline{\boldsymbol{t}_{\{k\}}}\right) + (k+1)^2\sigma_\alpha^2\sigma_\beta^2(1-\rho^2)\left(\overline{\boldsymbol{t}_{\{k\}}^2} - \overline{\boldsymbol{t}_{\{k\}}}\,\overline{\boldsymbol{t}_{\{k\}}}\right)},
$$

*where*

$$
\overline{\boldsymbol{t}_{\{k\}}^2} = \frac{1}{k+1}\sum_{i=0}^{k} t_i^2,
$$

$$
\overline{\boldsymbol{t}_{\{k\}}} = \frac{1}{k+1}\sum_{i=0}^{k} t_i,
$$

*with* $t_0 \equiv 0$*;* $m$ *and* $n$ *are the sample sizes of the experimental group and control group, respectively.*

*Proof.* We start by writing the RCRM in the matrix form. Specifically, Equation (1) can be rewritten as

$$
\boldsymbol{Y}_i = X^{(l)}\boldsymbol{\beta} + U\boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}_i, \tag{A.2}
$$

where $\boldsymbol{Y}_i = [Y_{i0}, Y_{i1}, \dots, Y_{ik}]^T$, $\boldsymbol{\beta} = [\alpha, \beta, \beta_x]^T$, $\boldsymbol{\gamma}_i = [u_{0i}, u_{1i}]^T$, $\boldsymbol{\varepsilon}_i = [\varepsilon_{i0}, \varepsilon_{i1}, \dots, \varepsilon_{ik}]^T$, and

$$X^{(l)} = \begin{cases} \begin{bmatrix} 1 & t_0 & t_0 \\ 1 & t_1 & t_1 \\ \vdots & \vdots & \vdots \\ 1 & t_k & t_k \end{bmatrix}, & \text{if } 1 \le i \le m \, (l = 1) \\ \begin{bmatrix} 1 & t_0 & 0 \\ 1 & t_1 & 0 \\ \vdots & \vdots & \vdots \\ 1 & t_k & 0 \end{bmatrix}, & \text{else if } m+1 \le i \le m+n \, (l = 2) \end{cases}$$

with $U = \begin{bmatrix} 1 & t_0 \\ 1 & t_1 \\ \vdots & \vdots \\ 1 & t_k \end{bmatrix}$. Here, without loss of generality, we assume the first $m$ subjects ($l = 1$) are in the experimental group and the last $n$ subjects ($l = 2$) are in the control group. We further denote the variance matrix within each subject as

$$V = UgU^T + R, \tag{A.3}$$

where $g = \begin{bmatrix} \sigma_\alpha^2 & \rho\sigma_\alpha\sigma_\beta \\ \rho\sigma_\alpha\sigma_\beta & \sigma_\beta^2 \end{bmatrix}$ and $R = \sigma^2 I_{k+1}$. Notice that $V$ is the same across all subjects.

The MLE of $\boldsymbol{\beta}$ is then expressed as

$$\hat{\boldsymbol{\beta}} = \left(X^T\Sigma^{-1}X\right)^{-1}X^T\Sigma^{-1}\boldsymbol{Y}$$

with the variance matrix given by

$$\mathrm{var}\left(\hat{\boldsymbol{\beta}}\right) = \left(X^T\Sigma^{-1}X\right)^{-1},$$

where $X = \mathrm{col}\{X^{(1)T}, X^{(1)T}, \dots, X^{(2)T}\}^T$, $\Sigma = \mathrm{diag}\{V, V, \dots, V\}$, and $\boldsymbol{Y} = [\boldsymbol{Y}_1^T, \boldsymbol{Y}_2^T, \dots, \boldsymbol{Y}_{m+n}^T]^T$. Based on the assumption that subjects are mutually independent, we have

$$\begin{aligned} X^T\Sigma^{-1}X &= \sum_{i=1}^{m} X^{(1)T}V^{-1}X^{(1)} + \sum_{i=m+1}^{m+n} X^{(2)T}V^{-1}X^{(2)} \\ &= mX^{(1)T}V^{-1}X^{(1)} + nX^{(2)T}V^{-1}X^{(2)}. \end{aligned}$$

Next, we will derive the explicit form of $V^{-1}$. Applying the Woodbury matrix identity (Higham, 2002) to Equation (A.3) yields

$$V^{-1} = R^{-1} - R^{-1}U\left(g^{-1} + U^TR^{-1}U\right)^{-1}U^TR^{-1}.$$

Notice that $(g^{-1} + U^TR^{-1}U)$ is a $2 \times 2$ matrix and thus it is easy to calculate the inverse. After some algebraic calculation and simplification, we have

$$\begin{aligned} V^{-1} =& \frac{1}{\sigma^2}I_{k+1} - \frac{1}{\sigma^4 M} \times \\ & \left[ \left( \frac{\sigma_\alpha^2 + \rho\sigma_\alpha\sigma_\beta(t_{i-1} + t_{j-1}) + \sigma_\beta^2 t_{i-1}t_{j-1}}{\sigma_\alpha^2\sigma_\beta^2(1-\rho^2)} + \frac{(k+1)\left(\overline{\boldsymbol{t}_{\{k\}}^2} - \overline{\boldsymbol{t}_{\{k\}}}(t_{i-1} + t_{j-1}) + t_{i-1}t_{j-1}\right)}{\sigma^2} \right)_{ij} \right], \end{aligned}$$

with

$$M = \frac{\sigma^4 + (k+1)\sigma^2\left(\sigma_\beta^2\overline{\boldsymbol{t}_{\{k\}}^2} + \sigma_\alpha^2 + 2\rho\sigma_\alpha\sigma_\beta\overline{\boldsymbol{t}_{\{k\}}}\right) + (k+1)^2\sigma_\alpha^2\sigma_\beta^2(1-\rho^2)\left(\overline{\boldsymbol{t}_{\{k\}}^2} - \overline{\boldsymbol{t}_{\{k\}}}\,\overline{\boldsymbol{t}_{\{k\}}}\right)}{\sigma^4\sigma_\alpha^2\sigma_\beta^2(1-\rho^2)}.$$

For $1 \le i \le m$ (subjects from the experimental group), $X^{(1)T}V^{-1}X^{(1)}$ can be further calculated as

$$X^{(1)T}V^{-1}X^{(1)} = \begin{bmatrix} A & B & B \\ B & C & C \\ B & C & C \end{bmatrix},$$

and for $(m+1) \le i \le (m+n)$ (subjects from the control group),

$$X^{(2)T}V^{-1}X^{(2)} = \begin{bmatrix} A & B & 0 \\ B & C & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

where

$$A = \frac{(k+1)\left[\sigma^2 + (k+1)\sigma_\beta^2\left(\overline{\boldsymbol{t}_{\{k\}}^2} - \overline{\boldsymbol{t}_{\{k\}}}\,\overline{\boldsymbol{t}_{\{k\}}}\right)\right]}{\sigma^4 + (k+1)\sigma^2\left(\sigma_\beta^2\overline{\boldsymbol{t}_{\{k\}}^2} + \sigma_\alpha^2 + 2\rho\sigma_\alpha\sigma_\beta\overline{\boldsymbol{t}_{\{k\}}}\right) + (k+1)^2\sigma_\alpha^2\sigma_\beta^2(1-\rho^2)\left(\overline{\boldsymbol{t}_{\{k\}}^2} - \overline{\boldsymbol{t}_{\{k\}}}\,\overline{\boldsymbol{t}_{\{k\}}}\right)},$$

$$B = \frac{(k+1)\left[\sigma^2\overline{\boldsymbol{t}_{\{k\}}} - (k+1)\rho\sigma_\alpha\sigma_\beta\left(\overline{\boldsymbol{t}_{\{k\}}^2} - \overline{\boldsymbol{t}_{\{k\}}}\,\overline{\boldsymbol{t}_{\{k\}}}\right)\right]}{\sigma^4 + (k+1)\sigma^2\left(\sigma_\beta^2\overline{\boldsymbol{t}_{\{k\}}^2} + \sigma_\alpha^2 + 2\rho\sigma_\alpha\sigma_\beta\overline{\boldsymbol{t}_{\{k\}}}\right) + (k+1)^2\sigma_\alpha^2\sigma_\beta^2(1-\rho^2)\left(\overline{\boldsymbol{t}_{\{k\}}^2} - \overline{\boldsymbol{t}_{\{k\}}}\,\overline{\boldsymbol{t}_{\{k\}}}\right)},$$

$$C = \frac{(k+1)\left[\sigma^2\overline{\boldsymbol{t}_{\{k\}}^2} + (k+1)\sigma_\alpha^2\left(\overline{\boldsymbol{t}_{\{k\}}^2} - \overline{\boldsymbol{t}_{\{k\}}}\,\overline{\boldsymbol{t}_{\{k\}}}\right)\right]}{\sigma^4 + (k+1)\sigma^2\left(\sigma_\beta^2\overline{\boldsymbol{t}_{\{k\}}^2} + \sigma_\alpha^2 + 2\rho\sigma_\alpha\sigma_\beta\overline{\boldsymbol{t}_{\{k\}}}\right) + (k+1)^2\sigma_\alpha^2\sigma_\beta^2(1-\rho^2)\left(\overline{\boldsymbol{t}_{\{k\}}^2} - \overline{\boldsymbol{t}_{\{k\}}}\,\overline{\boldsymbol{t}_{\{k\}}}\right)},$$

with

$$\overline{\boldsymbol{t}_{\{k\}}^2} = \frac{1}{k+1}\sum_{i=0}^{k} t_i^2,$$

$$\overline{\boldsymbol{t}_{\{k\}}} = \frac{1}{k+1}\sum_{i=0}^{k} t_i.$$

The result in Lemma A1 then follows. □

### A.2 | Proof for Theorem 1

*Proof.* In this more general case, subjects will be grouped into different categories based on their last assessment time. We now generalize the matrix form in Equation (A.2) as

$$\boldsymbol{Y}_i = X_{(k)}^{(l)}\boldsymbol{\beta} + U_{(k)}\boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}_i.$$

The superscript $l$ is the index of the randomized group as in Equation (A.2); the subscript $k$ denotes the number of post-baseline assessments for subject $i$. Again, we assume $l = 1$ corresponds to the experimental group and $l = 2$ corresponds to the control group.

We now categorize subjects into $J+1$ groups, based on their last assessment time. Specifically, we assume there are $p_k * 100\%$ of subjects that have and only have assessments $t_0, t_1, \dots, t_k$ $(k = 0, 1, \dots, J)$ with $\sum_{k=0}^{J} p_k = 1$. The assumption of monotone missingness guarantees there are only these $J+1$ groups in the experiment. In each category (without loss of generality, we assume $t_k$ being the last assessment time), the corresponding design matrix $X_{(k)}^{(l)}$ and $U_{(k)}$ are defined in Appendix A.1. Vectors $\boldsymbol{Y}_i$ and $\boldsymbol{\varepsilon}_i$ also have the corresponding length $k+1$. The corresponding variance matrix for each subject within this category is

$$V_{(k)} = U_{(k)} g U_{(k)}^T + R_{(k)},$$

where $g$ and $R_{(k)}$ are defined in Appendix A.1. Similarly, the MLE of $\boldsymbol{\beta}$ is expressed as

$$\hat{\boldsymbol{\beta}} = \left(X^T \Sigma^{-1} X\right)^{-1} X^T \Sigma^{-1} \boldsymbol{Y}$$

with the variance matrix given by

$$\mathrm{var}\left(\hat{\boldsymbol{\beta}}\right) = \left(X^T \Sigma^{-1} X\right)^{-1},$$

where

$$X = \mathrm{col}\left\{X_{(0)}^{(1)T}, X_{(0)}^{(1)T}, \dots, X_{(J)}^{(1)T}, X_{(J)}^{(1)T}, \dots X_{(0)}^{(2)T}, X_{(0)}^{(2)T}, \dots, X_{(J)}^{(2)T}, X_{(J)}^{(2)T}\right\}^T,$$

$$\Sigma = \mathrm{diag}\left\{V_{(0)}, V_{(0)}, \dots, V_{(J)}, V_{(J)}, V_{(0)}, V_{(0)}, \dots, V_{(J)}, V_{(J)}\right\},$$

and $\boldsymbol{Y} = [\boldsymbol{Y}_1^T, \boldsymbol{Y}_2^T, \dots, \boldsymbol{Y}_{m+n}^T]^T$. Grouping subjects first into randomized groups and then into categories based on the last assessment time, we rewrite the Information matrix $X^T \Sigma^{-1} X$ as

$$\begin{aligned} X^T \Sigma^{-1} X &= \sum_{k=0}^{J} m p_k X_{(k)}^{(1)T} V_{(k)}^{-1} X_{(k)}^{(1)} + \sum_{k=0}^{J} n p_k X_{(k)}^{(2)T} V_{(k)}^{-1} X_{(k)}^{(2)} \\ &= \sum_{k=0}^{J} p_k \left(m X_{(k)}^{(1)T} V_{(k)}^{-1} X_{(k)}^{(1)} + n X_{(k)}^{(2)T} V_{(k)}^{-1} X_{(k)}^{(2)}\right). \end{aligned}$$

Note that we assume the probabilities $p_0, p_1, \dots, p_J$ and the parameters in $\Sigma$ (i.e. $\sigma^2, \sigma_\alpha^2, \sigma_\beta^2, \rho$) are fixed and known in the derivation above. Since we assume the data are missing at random, parameters can be consistently estimated by the MLE (Little & Rubin, 2019). Using the results from Appendix A.1 then completes the proof. □

### A.3 | Theorem 3

**Theorem 1.** *We now assume different dropout probabilities between the experimental group and control group. Denote $p_{exp,k}$ and $p_{ctl,k}$ $(k = 0, 1, 2, \dots, J)$ as the percentages of subjects who only have assessments at $t_0, t_1, \dots, t_k$ (i.e., no additional assessments) from the experimental group and control group, respectively. Other assumptions and notations are the same as those in Theorem 1. The covariance matrix of the MLE vector $\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \\ \hat{\beta}_x \end{bmatrix}$ is*

$$var(\hat{\boldsymbol{\beta}}) = \begin{bmatrix} mA_{exp} + nA_{ctl} & mB_{exp} + nB_{ctl} & mB_{exp} \\ mB_{exp} + nB_{ctl} & mC_{exp} + nC_{ctl} & mC_{exp} \\ mB_{exp} & mC_{exp} & mC_{exp} \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} \frac{C_{exp}C_{ctl}}{C_{ctl}\left(C_{exp}\left(A_{exp}m+A_{ctl}n\right)-B_{exp}^2m\right)-B_{ctl}^2C_{exp}n} & \frac{B_{ctl}C_{exp}}{C_{ctl}\left(B_{exp}^2m-C_{exp}\left(A_{exp}m+A_{ctl}n\right)\right)+B_{ctl}^2C_{exp}n} & 0 \\ \frac{B_{ctl}C_{exp}}{C_{ctl}\left(B_{exp}^2m-C_{exp}\left(A_{exp}m+A_{ctl}n\right)\right)+B_{ctl}^2C_{exp}n} & \frac{B_{exp}^2m-C_{exp}\left(A_{exp}m+A_{ctl}n\right)}{n\left(C_{ctl}\left(B_{exp}^2m-C_{exp}\left(A_{exp}m+A_{ctl}n\right)\right)+B_{ctl}^2C_{exp}n\right)} & \frac{C_{exp}\left(A_{exp}m+A_{ctl}n\right)-B_{exp}\left(B_{exp}m+B_{ctl}n\right)}{n\left(C_{ctl}\left(B_{exp}^2m-C_{exp}\left(A_{exp}m+A_{ctl}n\right)\right)+B_{ctl}^2C_{exp}n\right)} \\ 0 & \frac{C_{exp}\left(A_{exp}m+A_{ctl}n\right)-B_{exp}\left(B_{exp}m+B_{ctl}n\right)}{n\left(C_{ctl}\left(B_{exp}^2m-C_{exp}\left(A_{exp}m+A_{ctl}n\right)\right)+B_{ctl}^2C_{exp}n\right)} & \frac{\left(A_{exp}m+A_{ctl}n\right)\left(C_{exp}m+C_{ctl}n\right)-\left(B_{exp}m+B_{ctl}n\right)^2}{mn\left(C_{ctl}\left(C_{exp}\left(A_{exp}m+A_{ctl}n\right)-B_{exp}^2m\right)-B_{ctl}^2C_{exp}n\right)} \end{bmatrix},$$

*where*

$$A_{\{\cdot\}} = \sum_{k=0}^{J} p_{\{\cdot\},k} \left\{ \frac{(k+1)\left[\sigma^2 + (k+1)\sigma_\beta^2\left(\overline{\boldsymbol{t}_{\{k\}}^2} - \overline{\boldsymbol{t}_{\{k\}}}\,\overline{\boldsymbol{t}_{\{k\}}}\right)\right]}{\sigma^4 + (k+1)\sigma^2\left(\sigma_\beta^2\overline{\boldsymbol{t}_{\{k\}}^2} + \sigma_\alpha^2 + 2\rho\sigma_\alpha\sigma_\beta\overline{\boldsymbol{t}_{\{k\}}}\right) + (k+1)^2\sigma_\alpha^2\sigma_\beta^2(1-\rho^2)\left(\overline{\boldsymbol{t}_{\{k\}}^2} - \overline{\boldsymbol{t}_{\{k\}}}\,\overline{\boldsymbol{t}_{\{k\}}}\right)} \right\},$$

$$B_{\{\cdot\}} = \sum_{k=0}^{J} p_{\{\cdot\},k} \left\{ \frac{(k+1)\left[\sigma^2\overline{\boldsymbol{t}_{\{k\}}} - (k+1)\rho\sigma_\alpha\sigma_\beta\left(\overline{\boldsymbol{t}_{\{k\}}^2} - \overline{\boldsymbol{t}_{\{k\}}}\,\overline{\boldsymbol{t}_{\{k\}}}\right)\right]}{\sigma^4 + (k+1)\sigma^2\left(\sigma_\beta^2\overline{\boldsymbol{t}_{\{k\}}^2} + \sigma_\alpha^2 + 2\rho\sigma_\alpha\sigma_\beta\overline{\boldsymbol{t}_{\{k\}}}\right) + (k+1)^2\sigma_\alpha^2\sigma_\beta^2(1-\rho^2)\left(\overline{\boldsymbol{t}_{\{k\}}^2} - \overline{\boldsymbol{t}_{\{k\}}}\,\overline{\boldsymbol{t}_{\{k\}}}\right)} \right\},$$

$$C_{\{\cdot\}} = \sum_{k=0}^{J} p_{\{\cdot\},k} \left\{ \frac{(k+1)\left[\sigma^2\overline{\boldsymbol{t}_{\{k\}}^2} + (k+1)\sigma_\alpha^2\left(\overline{\boldsymbol{t}_{\{k\}}^2} - \overline{\boldsymbol{t}_{\{k\}}}\,\overline{\boldsymbol{t}_{\{k\}}}\right)\right]}{\sigma^4 + (k+1)\sigma^2\left(\sigma_\beta^2\overline{\boldsymbol{t}_{\{k\}}^2} + \sigma_\alpha^2 + 2\rho\sigma_\alpha\sigma_\beta\overline{\boldsymbol{t}_{\{k\}}}\right) + (k+1)^2\sigma_\alpha^2\sigma_\beta^2(1-\rho^2)\left(\overline{\boldsymbol{t}_{\{k\}}^2} - \overline{\boldsymbol{t}_{\{k\}}}\,\overline{\boldsymbol{t}_{\{k\}}}\right)} \right\},$$

*with $\{\cdot\}$ takes the values of "exp" or "ctl," indicating subjects' group information with "exp" representing the experimental/treatment group and "ctl" representing the control group.*

*Proof.* The proof is similar to the matrix calculation presented in A.2, and is therefore omitted. □

### A.4 | Corollary 1

**Corollary 1.** *Assuming the time gap between any two consecutive assessments is h, then the variance matrix (3) reduces to*

$$var\left(\hat{\boldsymbol{\beta}}\right) = \begin{bmatrix} (m+n)A^\dagger & (m+n)B^\dagger & mB^\dagger \\ (m+n)B^\dagger & (m+n)C^\dagger & mC^\dagger \\ mB^\dagger & mC^\dagger & mC^\dagger \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} -\frac{C^\dagger}{(B^{\dagger 2}-A^\dagger C^\dagger)(m+n)} & \frac{B^\dagger}{(B^{\dagger 2}-A^\dagger C^\dagger)(m+n)} & 0 \\ \frac{B^\dagger}{(B^{\dagger 2}-A^\dagger C^\dagger)(m+n)} & -\frac{A^\dagger C^\dagger(m+n)-B^{\dagger 2}m}{C^\dagger(B^{\dagger 2}-A^\dagger C^\dagger)n(m+n)} & -\frac{1}{C^\dagger n} \\ 0 & -\frac{1}{C^\dagger n} & \frac{m+n}{C^\dagger mn} \end{bmatrix},$$

*where J is the maximum number of postbaseline assessments, and*

$$A^\dagger = \sum_{k=0}^{J} p_k \left\{ \frac{(k+1)\left[12\sigma^2 + h^2k(k+1)(k+2)\sigma_\beta^2\right]}{12\sigma^4 + (k+1)\sigma^2\left[2h^2k(2k+1)\sigma_\beta^2 + 12\sigma_\alpha^2 + 12hk\rho\sigma_\alpha\sigma_\beta\right] + h^2k(k+1)^2(k+2)\sigma_\alpha^2\sigma_\beta^2(1-\rho^2)} \right\},$$

$$B^\dagger = \sum_{k=0}^{J} p_k \left\{ \frac{(k+1)\left[6hk\sigma^2 - h^2k(k+1)(k+2)\rho\sigma_\alpha\sigma_\beta\right]}{12\sigma^4 + (k+1)\sigma^2\left[2h^2k(2k+1)\sigma_\beta^2 + 12\sigma_\alpha^2 + 12hk\rho\sigma_\alpha\sigma_\beta\right] + h^2k(k+1)^2(k+2)\sigma_\alpha^2\sigma_\beta^2(1-\rho^2)} \right\},$$

$$C^\dagger = \sum_{k=0}^{J} p_k \left\{ \frac{(k+1)\left[2h^2k(2k+1)\sigma^2 + h^2k(k+1)(k+2)\sigma_\alpha^2\right]}{12\sigma^4 + (k+1)\sigma^2\left[2h^2k(2k+1)\sigma_\beta^2 + 12\sigma_\alpha^2 + 12hk\rho\sigma_\alpha\sigma_\beta\right] + h^2k(k+1)^2(k+2)\sigma_\alpha^2\sigma_\beta^2(1-\rho^2)} \right\}.$$

*Proof.* Using Theorem 1, the proof is fairly straightforward and therefore omitted. □