Power for Balanced Linear Mixed Models with Complex Missing Data Processes

Kevin P. Josey[1], Brandy M. Ringham[2], Anna E. Barón[1], Margaret Schenkman[3], Katherine A. Sauder[4], Keith E. Muller[5], Dana Dabelea[6], and Deborah H. Glueck[4,*]

[1]Department of Biostatistics and Informatics, Colorado School of Public Health, University of Colorado Denver
[2]Lifecourse Epidemiology of Adiposity and Disease (LEAD) Center, Colorado School of Public Health, University of Colorado Denver
[3]Physical Therapy Program, University of Colorado School of Medicine, University of Colorado Denver
[4]Department of Pediatrics, University of Colorado School of Medicine, University of Colorado Denver
[5]Department of Health Outcomes and Biomedical Informatics, College of Medicine, University of Florida
[6]Department of Epidemiology and the Lifecourse Epidemiology of Adiposity and Diabetes (LEAD) Center, Colorado School of Public Health, University of Colorado Denver
[*]Corresponding Author: Deborah H. Glueck, Deborah.Glueck@ucdenver.edu

Abstract

When designing repeated measures studies, both the amount and the pattern of missing outcome data can affect power. The chance that an observation is missing may vary across measurements, and missingness may be correlated across measurements. Such complex missing data patterns affect both observational and interventional studies. For example, in a randomized trial of physiotherapy in patients with Parkinson's disease, increasing intermittent dropout over time yielded missing measurements of physical function. In this example, we assume data are missing completely at random, since the chance that a data point was missing appears to be unrelated to either outcomes or covariates. For data missing completely at random, we propose noncentral F power approximations for the Wald test for balanced linear mixed models with Gaussian responses. The noncentral F power approximations are based on moments of missing data summary statistics. The moments were derived assuming a conditional linear process for missingness. The approach provides approximate power for both complete-case and observed-case analyses. A complete-case analysis uses only the independent sampling units with all measurements present. An observed-case analysis uses all independent sampling units with at least one observation. Monte Carlo simulations demonstrate the accuracy of the power methods in small samples. We illustrate the utility of the method by computing power for proposed replications of the Parkinson's disease study.

Keywords: power, missing data, mixed model, multilevel, longitudinal

1 Introduction

In repeated measures studies, researchers observe multiple responses on a set of independent sampling units. The repeated measurements may be longitudinal (across time), spatial (across location), or multivariate (on different scales). Outcome data are missing when a measurement is not

observed at a particular planned occasion. The chance that an observation is missing may vary across measurements, and missingness may be correlated across measurements. Both the amount and the pattern of missing outcome data can affect power in repeated measures studies. Such complex missing data patterns occur frequently in biomedical studies. For example, in a randomized controlled clinical trial of physiotherapy in patients with Parkinson's disease (ClinicalTrials.gov Identifier: NCT01257945) [1], increasing dropout over a 16 month follow-up period produced missing measurements of physical function. In this study, the chance that a data point was missing appeared to be unrelated to either outcomes or covariates.

Published power approximations use assumptions that do not match the complex missing data patterns like the one in the Parkinson's disease study. Muller *et al.* [2, 3] assumed no missing data. Li and McKeague [4] use local alternative distributions which rely on asymptotic assumptions for accuracy. Further compounding the issues with local alternatives, the use of GEE estimates has been shown to suffer from inflated Type I error rates in small samples, weakening its otherwise general appeal [5]. Tu *et al.* [6] accounts for missing data in their power approximation, but they too assume local alternative distributions. Ringham *et al.* [7] accounts for missing data and uses a noncentral F-distribution proposed by Muller *et al.* [2, 3], but they assume that every response was independently and equally likely to be missing.

To counteract these challenges, we provide power approximations that accommodate complex missing data processes characterized by a conditional linear model [8]. We modify a noncentral F power approximation for a class of mixed models with no missing data described by Muller *et al.* [2, 3] and Edwards *et al.* [9]. The modifications incorporate the expected value of one of several missing data summary statistics. [7, 10, 11] The choice of which missing data summary statistic to use depends on whether scientists plan *complete-* or *observed-*case data analytic approaches. The complete-case approach uses only independent sampling units with all measurements present. The observed-case approach uses any independent sampling unit with at least one measurement. There are two missing data summary statistics in particular that we favor for these objective analyses.

The work applies to a broad and useful subset of missing processes, models and hypothesis tests. The data are assumed to be missing completely at random (MCAR) [12]. Under the MCAR assumption, the distribution of the missing responses does not depend on any of the observed outcomes, unobserved outcomes or explanatory variables. The models are assumed to be a class of general linear mixed models [13] referred to as balanced linear mixed models[3, 7]. A general linear mixed model is a balanced linear mixed model if it can be expressed as an equivalent general linear multivariate model when there is no missing data. We restrict attention to the Wald test and the Kenward-Roger reference distribution [10, 11, 7] due to its equivalence with a Hotelling-Lawley trace test [14]. We use a covariance estimator that assumes an unstructured correlation between responses which has been shown to be quite favorable in clinical trial settings with random dropout. [15] Balanced linear mixed models fulfill the following criteria. For a particular independent sampling unit, the predictors have the same value for all outcomes. In addition, each independent sampling unit has the same number of planned predictors and the same error covariance matrix. In balanced linear mixed models with no missing data, each independent sampling unit has the same number and type of planned outcome measurements. In this manuscript, we relax the last assumption, and allow for missing data.

The paper has six sections. Section 2 contains general notation, definitions, and model assumptions, while Section 3 contains the power approximations. In Section 4, Monte Carlo simulations are used to demonstrate the accuracy of the method in small samples. Section 5 contains example power analyses for a proposed clinical trial. The implications and limitations of the results are discussed in Section 6.

2 Notation, Definitions, and Assumptions

2.1 General Notation

For an $(m \times o)$ matrix $\boldsymbol{A}$ and an $(o \times n)$ matrix $\boldsymbol{B}$, $\boldsymbol{AB}$ is the matrix product. Throughout $\text{vec}(\boldsymbol{A})$ is the matrix function that stacks the columns of an $(m \times n)$ matrix $\boldsymbol{A}$ into an $(mn \times 1)$ vector. For an $(m \times n)$ matrix $\boldsymbol{A}$ and a $(p \times q)$ matrix $\boldsymbol{B}$, the Kronecker product $\boldsymbol{A} \otimes \boldsymbol{B}$ generates the $(mp \times qn)$ matrix $\boldsymbol{C} = \{a_{ij}\boldsymbol{B}\}$. The trace of a square matrix $\boldsymbol{A}$ is $\text{tr}(\boldsymbol{A}) = \sum_{i=1}^{N} a_{ii}$. A square and full rank matrix $\boldsymbol{A}$ has inverse $\boldsymbol{A}^{-1}$. Also, for any matrix $\boldsymbol{A}$, the Moore-Penrose inverse is $\boldsymbol{A}^{+}$ and $\boldsymbol{A}'$ is the transpose. An $(a \times a)$ identity matrix is $\boldsymbol{I}_a$, an $(a \times 1)$ vector with every value set to 1 is $\boldsymbol{1}_a$ and an $(a \times 1)$ vector with every value set to 0 is $\boldsymbol{0}_a$. The minimum and maximum values of a set $\mathcal{S}$ are denoted by $\min\{\mathcal{S}\}$ and $\max\{\mathcal{S}\}$, respectively.

We indicate the expected value and variance of $\boldsymbol{y}$ as $\boldsymbol{\mathcal{E}}(\boldsymbol{y})$ and $\boldsymbol{\mathcal{V}}(\boldsymbol{y})$, respectively. Similarly, for scalar y, we indicate the expected value and variance as $\mathcal{E}(y)$ and $\mathcal{V}(y)$. Scalar random variables y_i and y_j have covariance $\mathcal{C}(y_i, y_j)$. For positive integer n, we write $\boldsymbol{y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Xi})$ to indicate the vector normal probability distribution [16] with $\boldsymbol{\mu} \in \Re^n$ an $(n \times 1)$ vector of means and $\boldsymbol{\Xi} \in (\Re^n \times \Re^n)$ a positive definite symmetric $(n \times n)$ matrix of covariances. Similarly, with N and p positive integers, $N > p$, we write $\boldsymbol{Y} \sim N_{N,p}(\boldsymbol{M}, \boldsymbol{\Omega}, \boldsymbol{\Xi})$ to indicate the matrix normal [16], with $\boldsymbol{M} \in (\Re^N \times \Re^p)$ an $(N \times p)$ matrix of means, $\boldsymbol{\Omega} \in (\Re^p \times \Re^p)$ a positive definite symmetric $(p \times p)$ matrix of column covariances and $\boldsymbol{\Xi} \in (\Re^N \times \Re^N)$ a positive definite symmetric $(N \times N)$ matrix of row covariances. Consequently, $\boldsymbol{Y} \sim N_{N,p}(\boldsymbol{M}, \boldsymbol{I}_N, \boldsymbol{\Xi})$ implies that $\text{vec}(\boldsymbol{Y}) \sim N_{Np}[\text{vec}(\boldsymbol{M}), \boldsymbol{\Xi} \otimes \boldsymbol{I}_N]$.

2.2 Balanced General Linear Mixed Model and General Linear Hypothesis

Under the assumptions in Section 1, we define a balanced linear mixed model [3, 7]. We use the subscript m to indicate mixed model components.

Independent sampling unit $i \in \{1, 2, \ldots, N\}$ at measurement $j \in \{1, 2, \ldots, p\}$ has response $y_{i,j}$, with $\boldsymbol{y}_i = \begin{bmatrix} y_{i,1} & y_{i,2} & \ldots & y_{i,p} \end{bmatrix}'$ the $(p \times 1)$ vector of responses for independent sampling unit i, and $\boldsymbol{y} = \begin{bmatrix} \boldsymbol{y}_1' & \boldsymbol{y}_2' & \ldots \boldsymbol{y}_N' \end{bmatrix}'$ the $(Np \times 1)$ vector of responses for all independent sampling units. The fixed and known $(N \times q)$ design matrix for a single outcome is $\boldsymbol{X}$, and $\boldsymbol{X}_m = \boldsymbol{X} \otimes \boldsymbol{I}_p$ is the corresponding $(Np \times qp)$ design matrix for all outcomes. The $(qp \times 1)$ vector $\boldsymbol{\beta}$ contains the unknown and constant model coefficients. In turn, $\boldsymbol{\Sigma}$ is a positive definite symmetric $(p \times p)$ covariance matrix and is the same for all independent sampling units. We define $\boldsymbol{\Psi} = (\boldsymbol{I}_N \otimes \boldsymbol{\Sigma})$. Throughout, $\boldsymbol{e} \sim N_{Np}(\boldsymbol{0}, \boldsymbol{\Psi})$ is the $(Np \times 1)$ vector of errors, which implies statistical independence of $\boldsymbol{e}_i$ and $\boldsymbol{e}_{i'}$ for $i, i' \in \{1, 2, \ldots, N\}$, $i \neq i'$. The balanced general linear mixed model is

$$\boldsymbol{y} = \boldsymbol{X}_m \boldsymbol{\beta} + \boldsymbol{e}, \tag{1}$$

with $\boldsymbol{y} \sim N_{Np}(\boldsymbol{X}_m \boldsymbol{\beta}, \boldsymbol{\Psi})$ and $\boldsymbol{y}_i$ independent of $\boldsymbol{y}_{i'}$ for $i, i' \in \{1, 2, \ldots, N\}$, $i \neq i'$.

The $(a \times q)$ matrix $\boldsymbol{C}$ contains the between-independent-sampling-unit contrasts and the $(p \times b)$ matrix $\boldsymbol{U}$ contains the within-independent-sampling-unit contrasts. The $(ab \times pq)$ contrast $\boldsymbol{L} = \boldsymbol{C} \otimes \boldsymbol{U}'$ defines the secondary parameter $\boldsymbol{\theta} = \boldsymbol{L}\boldsymbol{\beta}$ and the corresponding null hypothesis

$$H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0. \tag{2}$$

The Wald parameter is defined as

$$\zeta_W(\boldsymbol{\theta}, \boldsymbol{\Psi}) = (ab)^{-1} (\boldsymbol{\theta} - \boldsymbol{\theta}_0)' \left[\boldsymbol{L} \left(\boldsymbol{X}_m' \boldsymbol{\Psi}^{-1} \boldsymbol{X}_m \right)^{-1} \boldsymbol{L}' \right]^{-1} (\boldsymbol{\theta} - \boldsymbol{\theta}_0), \tag{3}$$

with $\boldsymbol{X}_m$ the design matrix speculated to exist were the study to be observed.

2.3 Vector of Non-missing Data Indicators

The balanced linear mixed model described in Equation 1 has no missing response data. Often, in biomedical studies, responses at planned measurements are not observed, resulting in missing response data, and a sample size that may be smaller than the planned sample size, N. We define notation for the probability that $y_{i,j}$ is missing or non-missing.

The indicator variable $d_{i,j}$ has realization $d_{i,j}^*$. If $y_{i,j}$ is missing then $d_{i,j}^* = 0$. Otherwise, $d_{i,j}^* = 1$. For $d_{i,j} \sim \text{Bernoulli}(\pi_j)$ it follows that

$$\Pr\left\{d_{i,j} = d_{i,j}^*\right\} = \begin{cases} \pi_j & d_{i,j}^* = 1 \\ 1 - \pi_j & d_{i,j}^* = 0. \end{cases} \tag{4}$$

With $j \neq j'$ and $j, j' \in \{1, 2, \ldots, p\}$, the joint probability mass function is defined by

$$\Pr\left\{d_{i,j} = d_{i,j}^*, d_{i,j'} = d_{i,j'}^*\right\} = \begin{cases} \pi_{jj'} & d_{i,j}^* = 1 \text{ and } d_{i,j'}^* = 1 \\ \pi_j - \pi_{jj'} & d_{i,j}^* = 1 \text{ and } d_{i,j'}^* = 0 \\ \pi_{j'} - \pi_{jj'} & d_{i,j}^* = 0 \text{ and } d_{i,j'}^* = 1 \\ 1 - \pi_j - \pi_{j'} + \pi_{jj'} & d_{i,j}^* = 0 \text{ and } d_{i,j'}^* = 0. \end{cases} \tag{5}$$

An implicit requirement is that $\pi_{jj'} \leq \min\left(\pi_j, \pi_{j'}\right)$. It is important to recognize that implicit in Equations 4 and 5 is a homogeneity assumption: the probability that a response is non-missing is the same for all i.

The $(p \times 1)$ vector of non-missing data indicators for independent-sampling-unit i is $\boldsymbol{d}_i = \begin{bmatrix} d_{i,1} & d_{i,2} & \ldots & d_{i,p} \end{bmatrix}'$, the $(1 \times N)$ matrix of non-missing data indicators for measurement j is $\boldsymbol{D}_j = \begin{bmatrix} d_{1,j} & d_{2,j} & \ldots & d_{N,j} \end{bmatrix}$ and the $(Np \times 1)$ vector of non-missing data indicators for all independent sampling units and measurements is $\boldsymbol{d} = \begin{bmatrix} \boldsymbol{d}_1' & \boldsymbol{d}_2' & \ldots & \boldsymbol{d}_N' \end{bmatrix}'$.

2.4 Conditional Linear Missing Data Process

In the previous section, we defined notation to describe whether a response was present or missing. We turn now to examining the underlying probability processes which dictate the presence or absence of the responses. For the sake of clarity, we use the word "process" to describe missingness. This choice is in contrast to the word "model" used in Section 2 to describe the association between outcomes and predictors.

For data analysis, Qaqish [8] described binary responses using a class of conditional linear models with constrained covariance parameters. We can adapt his model to characterize a missing data process, referred to hereafter as the conditional linear process (CLP). The conditional linear process has two parameters. For $j \in \{1, 2 \ldots, p\}$, we define the $(p \times 1)$ vector of the marginal probabilities that a response at a given repeated measurement j is non-missing as $\boldsymbol{\pi} = \begin{bmatrix} \pi_1 & \pi_2 & \ldots & \pi_p \end{bmatrix}'$. Further, we write $\boldsymbol{\pi}_{j-1}$ to indicate the $(j-1)$ vector $\begin{bmatrix} \pi_1 & \pi_2 & \ldots & \pi_{j-1} \end{bmatrix}'$. We use $\gamma_{jj'}$ to denote the correlation between $d_{i,j}$ and $d_{i,j'}$ for $j \neq j'$. With $\psi_j = [\pi_j / (1 - \pi_j)]^{1/2}$, Qaqish [8] showed that bounding

$$\max\left\{-\psi_j\psi_{j'}, \frac{-1}{\psi_j\psi_{j'}}\right\} \leq \gamma_{jj'} \leq \min\left\{\frac{\psi_j}{\psi_{j'}}, \frac{\psi_{j'}}{\psi_j}\right\} \tag{6}$$

yields a positive semidefinite covariance matrix while strict inequalities produce positive definite covariance matrices.

For the conditional linear process, we write the $(p \times 1)$ vector of expected values $\boldsymbol{\mathcal{E}}(\boldsymbol{d}_i) = \boldsymbol{\pi}$ and the $(p \times p)$ covariance $\boldsymbol{\mathcal{V}}(\boldsymbol{d}_i) = \boldsymbol{\Phi}$. Both $\boldsymbol{\pi}$ and $\boldsymbol{\Phi}$ are defined by a conditional model, which gives the result for measurement j, conditioned on the previous $(j-1)$ measurements.

To describe the missing states for the first j measurements on independent sampling unit i, we define the $(j \times 1)$ vector $\boldsymbol{d}_{i,j} = \begin{bmatrix} d_{i,1} & d_{i,2} \ldots & d_{i,j} \end{bmatrix}'$. As described above, we use $\boldsymbol{d}_{i,j}^* = \begin{bmatrix} d_{i,1}^* & d_{i,2}^* & \ldots & d_{i,j}^* \end{bmatrix}'$ to denote the vector of realized values for $\boldsymbol{d}_{i,j}$. Similarly, we define the $(j \times j)$ matrix $\boldsymbol{\mathcal{V}}(\boldsymbol{d}_{i,j}) = \boldsymbol{\Phi}_j$ to contain the variance and covariance parameters for the first j measurements on independent sampling unit i. With $\mathcal{C}(d_{i,k}, d_{i,j}) = \tau_{k,j}$, $k \in \{1, 2, \ldots, j-1\}$, and the $[(j-1) \times 1]$ vector of covariances $\boldsymbol{\tau}_{j-1} = \begin{bmatrix} \tau_{1,j} & \tau_{2,j} & \ldots & \tau_{j-1,j} \end{bmatrix}'$, we obtain

$$\boldsymbol{\Phi}_j = \begin{bmatrix} \underset{(j-1)\times(j-1)}{\boldsymbol{\Phi}_{j-1}} & \underset{(j-1)\times 1}{\boldsymbol{\tau}_{j-1}} \\ \underset{1\times(j-1)}{\boldsymbol{\tau}'_{j-1}} & \underset{1\times 1}{\pi_j (1 - \pi_j)} \end{bmatrix}. \tag{7}$$

Variance elements in $\boldsymbol{\Phi}$ fall in $[0, 0.25]$. Covariance elements are a function of the chance that both elements are non-missing. Using the result in Equation 7, we write

$$\mathcal{E}\left(d_{i,j} | \boldsymbol{d}_{i,j-1} = \boldsymbol{d}_{i,j-1}^*\right) = \pi_j + \boldsymbol{\tau}'_{j-1}\boldsymbol{\Phi}_{j-1}^+ \left(\boldsymbol{d}_{i,j-1}^* - \boldsymbol{\pi}_{j-1}\right). \tag{8}$$

2.5 Missing Data Summary Statistics

Barton and Cramer [10] and Catellier and Muller [11] suggested using missing data summary statistics to summarize an observed missing data pattern. However, when designing a study and calculating power, the pattern of missingness has not yet been observed. Instead, one must consider the expected value of the missing data summary statistic with respect to the missing data process. The expected value is a weighted average over all possible realizations of the missing data process.

For independent sampling unit i, we define the number of missing responses $r_i \in \{0, 1, \ldots, p\}$ as

$$r_i = p - \mathbf{1}'_p \boldsymbol{d}_i. \tag{9}$$

A complete case is indicated by

$$I(r_i = 0) = \begin{cases} 1 & \text{if } r_i = 0 \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

For $j \in \{1, 2, \ldots, p\}$, the number of independent sampling units who have measurement j present is given by

$$N_{jj} = \boldsymbol{D}_j \boldsymbol{D}'_j, \tag{11}$$

for $N_{jj} \in \{0, 1, \ldots, N\}$.

The missing data summary statistics N_k, $k \in \{c, m, p\}$ are defined in Table 1. For the formulae presented in Table 1, we define

$$\phi = \pi_1 \prod_{j=2}^{p} \left[\pi_j + \boldsymbol{\tau}'_{j-1}\boldsymbol{\Phi}_{j-1}^+ \left(\mathbf{1}_{j-1} - \boldsymbol{\pi}_{j-1}\right)\right] \tag{12}$$

The moments of the missing data summary statistics appear in Table 1.

3   Power for Balanced Linear Mixed Models with Missing Data

We provide power approximations for balanced linear mixed models with no missing data, and for the complete- and observed-case analyses. We restrict attention to the Wald test with the Kenward-Roger [17, 18] reference distribution.

The new power approximation is based on a sequence of previous published works. Muller *et al.* [3] and Edwards *et al.* [9] demonstrated that a multivariate model can be converted to a balanced linear mixed model, and that the Hotelling-Lawley trace statistic can be converted to the Wald test statistic. Under the alternative hypothesis, $H_A : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, for balanced linear mixed models with no missing data, a function of the Wald statistic has an approximate noncentral F distribution [9]. Ringham *et al.* [7] presented a power approximation for balanced linear mixed models using a multivariate noncentrality parameter. In this work, we demonstrate that the multivariate non-centrality parameter [7] can be transformed into the noncentrality parameter for balanced linear mixed models, given below in Equation 13. The transformation is detailed in Section A.1 of the appendix.

In the remainder of Section 3 we outline the approximation method. The error degrees of freedom are $\nu_e = N - \text{rank}(\boldsymbol{X})$. With missing data, with $\mathcal{E}(N_k)$ as given in Table 1, we write the adjusted error degrees of freedom as $\nu_k = \mathcal{E}(N_k) - \text{rank}(\boldsymbol{X})$. The noncentrality parameter is given by

$$\omega(\nu_k) = \left(\frac{ab \cdot \nu_k}{\nu_e}\right) \zeta_W(\boldsymbol{\theta}, \boldsymbol{\Psi}). \tag{13}$$

As defined previously, a and b are the ranks of the between- and within-independent-sampling-unit contrast matrices and that $\zeta_W(\boldsymbol{\theta}, \boldsymbol{\Psi})$ is the Wald parameter (Equation 3).

Let F_f^{-1} indicate the quantile function for a noncentral F random variable. Following the approach in Muller *et al.* [2, 3] and Ringham *et al.* [7], we use the McKeon [14] approximation for the Hotelling-Lawley trace to define the degrees of freedom. With ν_k replacing ν_e in the forms for no missing data, the numerator and denominator degrees of freedom are $\nu_1 = ab$ and

$$\nu_2(\nu_k) = \frac{(ab+2)\left[\nu_k^2 - \nu_k(2b+3) + b(b+3)\right]}{\nu_k(a+b+1) - (a+2b+b^2-1)} + 4, \tag{14}$$

respectively. For $f_0(\nu_k) \approx F_f^{-1}[1-\alpha; ab, \nu_2(\nu_k)]$, we approximate power by

$$\text{Power}(\nu_k) = 1 - F_f[f_0(\nu_k); ab, \nu_2(\nu_k), \omega(\nu_k)]. \tag{15}$$

We use the notation $\nu_2(\nu_k)$ and $\omega(\nu_k)$ to emphasize the dependence of the parameters on ν_k.

Equation 15 gives approximate power for two missing data analysis approaches. To calculate the approximate power for the complete-case analysis, one uses Equation 15 evaluated with $\nu_k = \nu_c = \mathcal{E}(N_c) - \text{rank}(\boldsymbol{X})$. In this setting $\mathcal{E}(N_c)$ represents the expected number of complete cases. To calculate the approximate power for the observed-case analysis, one uses Equation 15 evaluated with $\nu_k = \nu_m = \mathcal{E}(N_m) - \text{rank}(\boldsymbol{X})$. We substitute $\mathcal{E}(N_m)$ for N in the observed-case as it characterizes the expected number of observations at each repeated measurement.

For no missing data, the results of Equation 15 reduces to the power approximation suggested by Edwards *et al.* [9]. The results are equivalent because $\nu_k = \nu_p = N_p - \text{rank}(\boldsymbol{X}) = N - \text{rank}(\boldsymbol{X}) = \nu_e$, $\omega(\nu_k) = ab \cdot \zeta_W$ and thus, $\text{Power}(\nu_k) = \text{Power}(\nu_e)$. In an even more restrictive case, with no missing data and $s = \min\{a, b\} = 1$, the noncentral distribution of the test statistic is known and the power results are exact, rather than approximate.

4 Numerical Evaluation of the Accuracy of the Power Computations

4.1 Simulation Methods

We evaluated the accuracy of the power approximations via simulation. To compute empirical power, we defined α, the Type I error rate, $\boldsymbol{X}$, $\boldsymbol{C}$, $\boldsymbol{U}$, $\boldsymbol{\beta}$, $\boldsymbol{\theta}_0$ and $\boldsymbol{\Psi}$. We computed $\boldsymbol{X}_m = \boldsymbol{X} \otimes \boldsymbol{I}_p$, $\boldsymbol{L} = \boldsymbol{C} \otimes \boldsymbol{U}'$ and $\boldsymbol{\theta} = \boldsymbol{L}\boldsymbol{\beta}$. We generated realizations of $\boldsymbol{e}$ and computed realizations of $\boldsymbol{y}$ as in Equation 1.

We randomly generated realizations of $\boldsymbol{d}_i$, the vector of non-missing data indicators, using different missing data processes. We then created $\boldsymbol{y}^*$ by setting each corresponding value in $\boldsymbol{y}$ to present or missing, as indicated by $\boldsymbol{d}_i$. The process was repeated for 10,000 replications.

For each replication, we attempted to fit a general linear mixed model, with $\boldsymbol{y}^*$ as the response vector. Mixed models were fit in SAS 9.4 [19] using the PROC GLIMMIX procedure with double dogleg optimization. The modeling approach used an unstructured covariance, the REPEATED statement and the KR2 (Kenward Roger) option for the denominator degrees of freedom.

We evaluated the accuracy of the approximations separately for the complete- and observed-case approaches. For the complete-case approach, we analyzed only the independent sampling units with all measurements present. For the observed-case approach, we analyzed all independent sampling units with at least one measurement present.

For some replicates, we were unable to fit the mixed model. The parameters $\boldsymbol{\beta}$ and hence $\boldsymbol{\theta}$ are estimable only if there is at least one data point for each within-by-between cross-classification. In some cases, there were sufficient data so that $\boldsymbol{\beta}$ was technically estimable, but the model failed to converge, usually because the Hessian matrix was not positive definite. We recorded the number of replicates for which $\boldsymbol{\beta}$ was estimable, and the number of replicates for which the model converged. Convergence rate was calculated as the number of replicates for which the model converged divided by the number of replicates for which $\boldsymbol{\beta}$ was estimable.

For each model that converged, we calculated the Wald statistic and computed a p-value using the Kenward-Roger [18] degrees of freedom. Empirical power was computed as the number of p-values less than or equal to α, divided by the number of replicates for which the model converged. The simulation study used $10,000$ replications so that the error in the estimation of the empirical power would occur in the second decimal place.

The analytic power approximations for the complete-case and observed-case analyses were computed as in Equation 15. The absolute deviation was calculated as the absolute value of the difference between the approximate power and the empirical power. The maximum absolute deviation was computed as the maximum of the absolute deviations across a specified range of experimental conditions.

We examined the accuracy of the power approximations for five of the nine scenarios previously considered by Johnson *et al.* [20] and then by Kreidler *et al.* [21]. The subset was chosen to ensure that the Wald statistic and the mixed model were appropriate. Below, we provide results for two of the five scenarios: Scenario 4 and Scenario 5. Scenario 4 involves a test of a time-by-treatment interaction in a design with four exposure groups and three repeated measurements over time. Scenario 5 also involves a test of a time-by-treatment interaction, but for a design with two exposure groups, and five repeated measurements over time. Detailed descriptions of the study designs for Scenarios 4 and 5 appear in Section A.5 in the appendix. The online supplemental material contains study designs for the remaining scenarios.

Scenarios 4 and 5 were chosen to allow evaluation of the accuracy of the power approximations in two important cases. In Section 3, we defined $s = \min\{a, b\}$ to allow us to distinguish between

single and multiple degree of freedom hypothesis tests. For the design in Scenario 4, $s > 1$, and the power results would be approximate even with complete data. For the design in Scenario 5, $s = 1$. This means that the analytic power results would be exact with no missing data, but are approximate when some measurements are missing.

For each experimental scenario, we varied the analysis approach (complete- or observed-case), the planned sample size, the missing data process, the parameter inputs for the missing data process (π and Φ) and the scaling parameters for the error variance and mean difference (δ_σ and δ_β, respectively).

4.2   Simulation Results

Over all experimental conditions, the minimum number of replications for which β was estimable was $9,431$ out of the planned $10,000$ replicates. The convergence rates across most experimental conditions were greater than 0.99. The exception were conditions for Scenario 5 with a planned sample size of 20, which gave the minimum convergence rate of 0.78.

A subset of the results for Scenario 4 appears in Table 2. A subset of the results for Scenario 5 appears in Table 3. The online supplemental material contains results for the remaining scenarios and experimental conditions. We report $\mathcal{E}(N_c)$ for the complete-case approach, $\mathcal{E}(N_m)$ for the observed-case approach, the approximate analytic power, the empirical power and the absolute deviation.

Across both experimental scenarios, and all experimental conditions, the maximum absolute deviation between the approximate and analytic power was less than 0.056 (data not shown). Most scientists design studies so that they have power between 0.8 and 0.95. For all scenarios where the analytic power was greater than or equal to 0.8, the maximum absolute deviation was less than 0.052. The accuracy of the approximations was similar across different sample sizes, values of variance, mean differences, types of missing data processes, inputs for the missing data processes and choice of analysis approach. Results for the other five scenarios shown in the online supplemental material were similar.

5   Randomized Controlled Clinical Trial in Parkinson's Disease

Schenkman *et al.* [1] conducted a randomized controlled clinical trial of three different forms of exercise therapy: flexibility/balance/function, standard aerobic exercises and a home-based exercise regimen. The outcome was the Continuous Scale Physical Function test at baseline, 4, 10, and 16 months after randomization.

Investigators propose a similar follow-up study with a different sample from the same population. For the follow-up study, investigators will fit a general linear mixed model with indicator variables for the three treatments as predictors and the four repeated measurements of the physical function test as outcomes. The goal will be to evaluate the time-by-treatment interaction, using the mixed model Wald test with Kenward-Roger degrees of freedom and an α level of 0.05.

We provide a power analysis for the follow-up study. Model and hypothesis parameters are given in Section A.6 in the appendix. We assumed a conditional linear missing data process. The missing data process parameter estimates are taken from the preliminary data analysis, with the

covariance matrix of the missingness given by

$$\mathbf{\Phi} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0.116 & 0.101 & 0.081 \\ 0 & 0.101 & 0.150 & 0.129 \\ 0 & 0.081 & 0.129 & 0.165 \end{bmatrix}. \tag{16}$$

We consider three patterns for $\boldsymbol{\pi}$. Pattern A corresponds to complete data, with $\boldsymbol{\pi} = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}'$. Pattern B corresponds to the missing data pattern observed in the preliminary data analysis, with $\boldsymbol{\pi} = \begin{bmatrix} 1 & 0.868 & 0.818 & 0.793 \end{bmatrix}'$. Pattern C corresponds to half the missing data observed in the preliminary analysis, with $\boldsymbol{\pi} = \begin{bmatrix} 1 & 0.736 & 0.636 & 0.587 \end{bmatrix}'$.

Power curves for the three patterns are shown in Figure 1. Power is shown on the y axis. The x axis is in units of the largest element in the interaction coefficient matrix observed in the previous study, 4.153. No units are shown because the outcome is unitless.

6 Discussion

Missing data occurs everywhere in biomedical research, from longitudinal epidemiologic observational studies to randomized controlled clinical trials. Missing data can strongly affect the power of biomedical research studies and thus the choice of sample size. If the sample size is larger than needed, too many participants are exposed to the potential risk of research. If the sample size is smaller than needed, resources, investigator and participant time are wasted on a study that has insufficient statistical power to answer the question of interest. Accurate power and sample size calculations must take into account the chance of missing data.

The methods proposed in the current work allow power calculations for some important classes of models and missing data processes. We consider balanced linear mixed models and the Wald test with Kenward-Rogers degrees of freedom. We also used a flexible model of the missing data process to allow us to examine how different missing data patterns affect the results of the power analysis.

The work extends previous power approximations for data without missing values [3, 9] and for data with values missing completely at random with equal probability [7]. The new methods accommodate complex missing data processes [8] to better mirror processes likely to occur in real repeated measures studies. The missing data processes we consider allow a different chance of missingness for each measurement in a repeated measures study. In addition, the missing data processes allow for the chance of missingness for one measurement to be correlated with the chance of missingness of another measurement. We provide different power approximations for the complete- and observed-case data analytic approaches, both of which are commonly used in biomedical research. The approximation has adequate accuracy for the scenarios scientists care about the most, with true power between 0.8 and 0.95. The error in the power approximation is in the second decimal place.

A version of the approximations described here will be included in Version 3 of GLIMMPSE [21], available at SampleSizeShop.org. The GLIMMPSE power and sample size software for multilevel and longitudinal studies is user-friendly, point-and-click and available to run without cost from a web browser. Users who prefer R code to implement the methods may request a copy from the authors.

To approximate power for a study with no missing data, a scientist needs to specify the Type I error rate, predictor matrix, error variance, between- and within-independent sampling unit contrasts and sample size. To conduct an accurate power calculation with outcome data missing

completely at random, a scientist must, in addition, choose an appropriate missing data process. The choice should depend on the scientist's assumptions about the missing data process. The conditional linear missing data process can accommodate several assumptions, including autoregressive and unstructured correlation structures. Estimates for the chance that the response is missing at each measurement, and the correlation between missingness at each measurement can be obtained from literature reports, from previous unpublished data, or from an understanding of the rationale for missing data in a certain study design, disease state, or treatment regimen.

While the power approximations cover a useful set of models and missing data processes, there are limitations to the work. First, the assumption that the data are missing completely at random (MCAR) may not hold. The MCAR assumption means that the chance that a response is missing does not depend on observed data, including covariates. Secondly, the manuscript only proposes power analyses for balanced linear mixed models with Gaussian outcomes. This means that the results may not hold for studies with mistimed data, time-dependent covariates or non-Gaussian outcomes.

In future work, we hope to develop a maximum likelihood approach to estimate marginal missingness, and correlation between missingness from completed studies. It would be useful to develop a corresponding likelihood ratio test to aid scientists in characterizing a missing data process. We also plan to relax the assumption that the outcome data are missing completely at random. We hope to consider a larger class of models and outcomes, including non-linear mixed models, and binary and Poisson distributed outcomes. It is important to consider how power calculations are affected by the use of observed (and hence random) percentages of missingness at each measurement drawn from previous experiments. An additional open question is how to compute confidence intervals for power for studies with missing data. For ethical reasons, some investigators may prefer to consider quantiles of power, rather than power calculated for an expected amount of missingness.

Acknowledgement

Disclaimer: This manuscript was submitted to the Department of Biostatistics and Informatics in the Colorado School of Public Health, University of Colorado Denver, in partial fulfillment of the requirements for the degree of Master of Science in Biostatistics for KPJ. The content of this paper is solely the responsibility of the authors, and does not necessarily represent the official views of the Eunice Kennedy Shriver National Institute of Child Health and Human Development, the National Center for Advancing Translational Sciences, the National Institute of Diabetes and Digestive and Kidney Diseases, the National Institute of General Medical Sciences, the National Institute of Neurological Disorders and Stroke, the National Library of Medicine, the Office of the Director, the National Institutes of Health, the American Heart Association nor the Parkinson's Disease Foundation. The authors have no conflicts of interest to disclose.

References

[1] Schenkman M, Hall DA, Barón AE et al. Exercise for people in early- or mid-stage Parkinson disease: a 16-month randomized controlled trial. *Physical Therapy* 2012; 92(11): 1395–1410.

[2] Muller KE, Lavange LM, Ramey SL et al. Power calculations for general linear multivariate models including repeated measures applications. *Journal of the American Statistical Association* 1992; 87(420): 1209–1226.

[3] Muller KE, Edwards LJ, Simpson SL et al. Statistical tests with accurate size and power for balanced linear mixed models. *Statistics in Medicine* 2007; 26(19): 3639–3660.

[4] Li Z and McKeague IW. Power and Sample Size Calculations for Generalized Estimating Equations via Local Asymptotics. *Statistica Sinica* 2013; DOI:10.5705/ss.2011.081. URL http://www3.stat.sinica.edu.tw/statistica/j23n1/J23N111/J23N111.html.

[5] Stiger TR, Kosinski AS, Barnhart HX et al. Anova for repeated ordinal data with small sample size? a comparison of anova, manova, wls and gee methods by simulation. *Communications in Statistics - Simulation and Computation* 1998; 27(2): 357–375. DOI: 10.1080/03610919808813485. URL https://doi.org/10.1080/03610919808813485.

[6] Tu XM, Zhang J, Kowalski J et al. Power analyses for longitudinal study designs with missing data. *Statistics in Medicine* 2007; 26(15): 2958–2981.

[7] Ringham BM, Kreidler SM, Muller KE et al. Multivariate test power approximations for balanced linear mixed models in studies with missing data. *Statistics in Medicine* 2016; 35(17): 2921–2937.

[8] Qaqish BF. A family of multivariate binary distributions for simulating correlated binary variables with specified marginal means and correlations. *Biometrika* 2003; 90(2): 455–463.

[9] Edwards LJ, Muller KE, Wolfinger RD et al. An R2 statistic for fixed effects in the linear mixed model. *Statistics in Medicine* 2008; 27(29): 6137–6157.

[10] Barton CN and Cramer EC. Hypothesis Testing in Multivariate Linear Models with Randomly Missing Data. *Communications in Statistics - Simulation and Computation* 1989; 18(3): 875–895.

[11] Catellier DJ and Muller KE. Tests for Gaussian repeated measures with missing data in small samples. *Statistics in Medicine* 2000; 19(8): 1101–1114.

[12] Little RJA and Rubin DB. *Statistical Analysis with Missing Data*. 2nd edition ed. Hoboken, N.J: Wiley-Interscience, 2002.

[13] Laird NM and Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982; 38(4): 963–974.

[14] McKeon JJ. F Approximations to the Distribution of Hotelling's T2 0. *Biometrika* 1974; 61(2): 381–383.

[15] Gosho M, Hirakawa A, Noma H et al. Comparison of bias-corrected covariance estimators for MMRM analysis in longitudinal data with dropouts. *Statistical Methods in Medical Research* 2017; 26(5): 2389–2406. DOI:10.1177/0962280215597938.

[16] Arnold SF. *The Theory of Linear Models and Multivariate Analysis*. 1st edition ed. New York: John Wiley & Sons Inc, 1981.

[17] Kenward MG and Roger JH. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* 1997; 53(3): 983–997.

[18] Kenward MG and Roger JH. An improved approximation to the precision of fixed effects from restricted maximum likelihood. *Computational Statistics & Data Analysis* 2009; 53(7): 2583–2595.

[19] SAS Institute Inc. SAS 9.4, 2017.

[20] Johnson JL, Muller KE, Slaughter JC et al. POWERLIB: SAS/IML Software for Computing Power in Multivariate Linear Models. *Journal of Statistical Software* 2009; 30(5).

[21] Kreidler SM, Muller KE, Grunwald GK et al. GLIMMPSE: Online Power Computation for Linear Models with and without a Baseline Covariate. *Journal of Statistical Software* 2013; 54(10).

A   Appendix

A.1   Approximate Distribution of the Wald Statistic

Some multivariate model notation is needed. Here $y_{i,j}$ is the response for independent sampling unit $i \in \{1, 2, \ldots, N\}$ at measurement $j \in \{1, 2, \ldots, p\}$, and $\boldsymbol{Y} = \{y_{i,j}\}$ is the $(N \times p)$ matrix of responses for all independent sampling units. The corresponding $(N \times q)$ design matrix is $\boldsymbol{X}$, while $\boldsymbol{B}$ is the $(q \times p)$ primary parameter matrix, and $\boldsymbol{E}$ is the $(N \times p)$ error matrix. For $\boldsymbol{\Sigma}$ a positive definite symmetric $(p \times p)$ covariance matrix, $\boldsymbol{Y} \sim N_{N,p}\left(\boldsymbol{XB}, \boldsymbol{I}_N, \boldsymbol{\Sigma}\right)$. The general linear multivariate model is written

$$\boldsymbol{Y} = \boldsymbol{XB} + \boldsymbol{E}. \tag{17}$$

The $(a \times q)$ between-independent-sampling-unit contrast matrix $\boldsymbol{C}$, the $(p \times b)$ within-independent-sampling-unit contrast matrix $\boldsymbol{U}$, and the $(a \times b)$ $\boldsymbol{\Theta}_0$ are all matrices of known constants. The multivariate null hypothesis is

$$H_0 : \boldsymbol{CBU} = \boldsymbol{\Theta} = \boldsymbol{\Theta}_0. \tag{18}$$

Also

$$\boldsymbol{\Sigma}_* = \boldsymbol{U}'\boldsymbol{\Sigma}\boldsymbol{U}. \tag{19}$$

The constant matrix

$$\boldsymbol{M} = \boldsymbol{C}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{C}' \tag{20}$$

helps define

$$\boldsymbol{\Delta} = \left(\boldsymbol{\Theta} - \boldsymbol{\Theta}_0\right)' \boldsymbol{M}^{-1} \left(\boldsymbol{\Theta} - \boldsymbol{\Theta}_0\right). \tag{21}$$

The Hotelling-Lawley trace parameter then follows

$$\zeta_H\left(\boldsymbol{\Delta}, \boldsymbol{\Sigma}_*\right) = \mathrm{tr}\left[\boldsymbol{\Delta}\left(\nu_e \boldsymbol{\Sigma}_*\right)^{-1}\right]. \tag{22}$$

With the Hotelling-Lawley parameter, we derive the Wald parameter as follows:

$$\zeta_H(\boldsymbol{\Delta}, \boldsymbol{\Sigma}_*) = \text{tr}\left[(\boldsymbol{\Theta} - \boldsymbol{\Theta}_0)' \boldsymbol{M}^{-1} (\boldsymbol{\Theta} - \boldsymbol{\Theta}_0)(\nu_e \boldsymbol{\Sigma}_*)^{-1}\right] \tag{23}$$

$$= \nu_e^{-1}\text{tr}\left[(\boldsymbol{\Theta} - \boldsymbol{\Theta}_0)' \boldsymbol{M}^{-1} (\boldsymbol{\Theta} - \boldsymbol{\Theta}_0)\boldsymbol{\Sigma}_*^{-1}\right]$$

$$= \nu_e^{-1}\text{vec}(\boldsymbol{\Theta} - \boldsymbol{\Theta}_0)' \text{vec}\left[\boldsymbol{\Sigma}_*^{-1} (\boldsymbol{\Theta} - \boldsymbol{\Theta}_0)' \boldsymbol{M}^{-1}\right]$$

$$= \nu_e^{-1}\text{vec}(\boldsymbol{\Theta} - \boldsymbol{\Theta}_0)' \left(\boldsymbol{M}^{-1} \otimes \boldsymbol{\Sigma}_*^{-1}\right) \text{vec}(\boldsymbol{\Theta} - \boldsymbol{\Theta}_0)$$

$$= \nu_e^{-1}\text{vec}(\boldsymbol{\Theta} - \boldsymbol{\Theta}_0)' \left(\boldsymbol{M} \otimes \boldsymbol{\Sigma}_*\right)^{-1} \text{vec}(\boldsymbol{\Theta} - \boldsymbol{\Theta}_0)$$

$$= \nu_e^{-1}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)' \left(\boldsymbol{M} \otimes \boldsymbol{\Sigma}_*\right)^{-1} (\boldsymbol{\theta} - \boldsymbol{\theta}_0)$$

$$= \nu_e^{-1}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)' \left[\boldsymbol{L}\left(\boldsymbol{X}_m' \boldsymbol{\Psi}^{-1} \boldsymbol{X}_m\right)^{-1} \boldsymbol{L}'\right]^{-1} (\boldsymbol{\theta} - \boldsymbol{\theta}_0)$$

$$= \nu_e^{-1}(ab)\,\zeta_W(\boldsymbol{\theta}, \boldsymbol{\Psi}).$$

Under an alternative hypothesis, Ringham *et al.* [7] proposed using $\omega_*(\nu_k) = \nu_k\text{tr}\left[\boldsymbol{\Delta}\left(\nu_e\boldsymbol{\Sigma}_*\right)^{-1}\right]$ as the noncentrality parameter in their F power approximation for a multivariate model with missing data. Following Equation 23, $\omega_*(\nu_k)$ is equivalent to $\omega(\nu_k)$ defined in Equation 13.

To obtain the Hotelling-Lawley trace statistic, use the ordinary least squares estimator to obtain $\hat{\boldsymbol{B}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$ with which we get $\hat{\boldsymbol{\Sigma}} = \left(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{B}}\right)'\left(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{B}}\right)/\nu_e$ and $\hat{\boldsymbol{\Theta}} = \boldsymbol{C}\hat{\boldsymbol{B}}\boldsymbol{U}$. Then $\hat{\boldsymbol{\Delta}} = \left(\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}_0\right)'\boldsymbol{M}^{-1}\left(\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}_0\right)$ and $\hat{\boldsymbol{\Sigma}}_* = \boldsymbol{U}'\hat{\boldsymbol{\Sigma}}\boldsymbol{U}$. The Hotelling-Lawley statistic is then defined $\zeta_H(\hat{\boldsymbol{\Delta}}, \hat{\boldsymbol{\Sigma}}_*) = \text{tr}\left[\hat{\boldsymbol{\Delta}}\left(\nu_e\hat{\boldsymbol{\Sigma}}_*\right)^{-1}\right]$. The analogous Wald statistic is defined $\zeta_W\left(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\Psi}}\right)$ where $\hat{\boldsymbol{\beta}} = \text{vec}\left(\hat{\boldsymbol{B}}'\right)$, $\hat{\boldsymbol{\theta}} = \boldsymbol{L}\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\Psi}} = \boldsymbol{I}_N \otimes \hat{\boldsymbol{\Sigma}}$. Using Equation 23, it can be shown that $\zeta_H\left(\hat{\boldsymbol{\Delta}}, \hat{\boldsymbol{\Sigma}}_*\right) = \nu_e^{-1}(ab)\,\zeta_W(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\Psi}})$.

Assuming the null hypothesis is true, Mckeon [14] showed that

$$f_*(\nu_e) = \frac{(\nu_e - b - 1)\,\hat{\zeta}_H/\nu_1}{[\nu_2(\nu_e) - 2]/\nu_2(\nu_e)} \tag{24}$$

is approximately distributed as $F[ab, \nu_2(\nu_e)]$ using method of moments to approximate $\nu_2(\nu_e)$ and a constant scaling component with ν_1 given. Thus,

$$f(\nu_e) = \frac{(\nu_e - b - 1)\,\hat{\zeta}_W/\nu_e}{[\nu_2(\nu_e) - 2]/\nu_2(\nu_e)} \tag{25}$$

is also approximately distributed as $F[ab, \nu_2(\nu_e)]$. When missing data is present, we plug-in ν_k for ν_e.

A.2   Moments for the Univariate Indicator of the Non-missing Responses

The expected value of the missing state for a single repeated measurement is given by

$$\mathcal{E}(d_{i,j}) = \sum_{d_{i,j}^*=0}^{1} d_{i,j}^* (\pi_j)^{d_{i,j}^*}(1 - \pi_j)^{1-d_{i,j}^*} = \pi_j. \tag{26}$$

The second moment is

$$\mathcal{E}(d_{i,j}^2) = \sum_{d_{i,j}^*=0}^{1} \left(d_{i,j}^*\right)^2 (\pi_j)^{d_{i,j}^*}(1 - \pi_j)^{1-d_{i,j}^*} = \pi_j, \tag{27}$$

yielding $\mathcal{V}(d_{i,j}) = \pi_j(1 - \pi_j)$. For $d_{i,j}d_{i,j'} \in \{0, 1\}$, $\mathcal{E}(d_{i,j}d_{i,j'}) = \pi_{jj'}$ since $d_{i,j}d_{i,j'}$ has a Bernoulli distribution. This gives $\mathcal{C}(d_{i,j}, d_{i,j'}) = \pi_{jj'} - \pi_j\pi_{j'}$.

A.3   Moments for N_c

The moments can be derived by a process of induction. To begin, we show that $\mathcal{E}\left(\prod_{j=1}^{p} d_{i,j}\right) = \pi_1 \prod_{j=2}^{p} \left[\pi_j + \boldsymbol{\tau}'_{j-1}\boldsymbol{\Phi}^{+}_{j-1}(\mathbf{1}_{j-1} - \boldsymbol{\pi}_{j-1})\right]$. The initial step gives

$$\mathcal{E}(d_{i,1}) = \pi_1. \tag{28}$$

For the base case, allow $p = 2$. Then

$$\mathcal{E}(d_{i,1}d_{i,2}) = \mathcal{E}(d_{i,1})\mathcal{E}(d_{i,2}) + \mathcal{C}(d_{i,1}, d_{i,2}) \tag{29}$$

$$= \pi_1\pi_2 + \pi_{12} - \pi_1\pi_2$$

$$= \pi_1\left[\pi_2 + \frac{\pi_{12} - \pi_1\pi_2}{\pi_1(1 - \pi_1)}(1 - \pi_1)\right]$$

$$= \pi_1\left[\pi_2 + \boldsymbol{\tau}'_1\boldsymbol{\Phi}^{+}_1(1 - \pi_1)\right].$$

For the induction step, suppose that

$$\mathcal{E}\left(\prod_{j=1}^{p} d_{i,j}\right) = \pi_1 \prod_{j=2}^{p} \left[\pi_j + \boldsymbol{\tau}'_{j-1}\boldsymbol{\Phi}^{+}_{j-1}(\mathbf{1}_{j-1} - \boldsymbol{\pi}_{j-1})\right]. \tag{30}$$

Then

$$\mathcal{E}\left[\left(\prod_{j=1}^{p} d_{i,j}\right) \cdot d_{i,p+1}\right] = \Pr\left\{d_{i,p+1} = 1 \middle| \prod_{j=1}^{p} d_{i,j} = 1\right\} \Pr\left\{\prod_{j=1}^{p} d_{i,j} = 1\right\} \tag{31}$$

$$= \left[\pi_{p+1} + \boldsymbol{\tau}'_p\boldsymbol{\Phi}^{+}_p(\mathbf{1}_p - \boldsymbol{\pi}_p)\right]$$

$$\times \left\{\pi_1 \prod_{j=2}^{p} \left[\pi_j + \boldsymbol{\tau}'_{j-1}\boldsymbol{\Phi}^{+}_{j-1}(\mathbf{1}_{j-1} - \boldsymbol{\pi}_{j-1})\right]\right\}$$

$$= \pi_1 \prod_{j=2}^{p+1} \left[\pi_j + \boldsymbol{\tau}'_{j-1}\boldsymbol{\Phi}^{+}_{j-1}(\mathbf{1}_{j-1} - \boldsymbol{\pi}_{j-1})\right]$$

$$= \mathcal{E}\left(\prod_{j=1}^{p+1} d_{i,j}\right)$$

The complete-case indicator variable has a Bernoulli distribution such that

$$I(r_i = 0) = \begin{cases} 1 & \pi_1 \prod_{j=2}^{p} \left[\pi_j + \boldsymbol{\tau}'_{j-1}\boldsymbol{\Phi}^{+}_{j-1}(\mathbf{1}_{j-1} - \boldsymbol{\pi}_{j-1})\right] \\ 0 & 1 - \left\{\pi_1 \prod_{j=2}^{p} \left[\pi_j + \boldsymbol{\tau}'_{j-1}\boldsymbol{\Phi}^{+}_{j-1}(\mathbf{1}_{j-1} - \boldsymbol{\pi}_{j-1})\right]\right\}. \end{cases} \tag{32}$$

Since we assume that for all $i, i' \in \{1, 2, \ldots, N\}, i \neq i'$ and $j, j' \in \{1, 2, \ldots, p\}$, $\mathcal{C}(d_{i,j}, d_{i',j'}) = 0$, the sum of complete-case indicator variables has a binomial distribution with

$$\mathcal{E}\left[\sum_{i=1}^{N} I(r_i = 0)\right] = N\pi_1 \prod_{j=2}^{p} \left[\pi_j + \boldsymbol{\tau}'_{j-1}\boldsymbol{\Phi}^{+}_{j-1}(\mathbf{1}_{j-1} - \boldsymbol{\pi}_{j-1})\right] \tag{33}$$

14

and

$$\mathcal{V}\left[\sum_{i=1}^{N} I\left(r_i = 0\right)\right] = N\pi_1 \prod_{j=2}^{p} \left[\pi_j + \boldsymbol{\tau}'_{j-1}\boldsymbol{\Phi}^+_{j-1}\left(\mathbf{1}_{j-1} - \boldsymbol{\pi}_{j-1}\right)\right] \tag{34}$$

$$\times \left\{1 - \pi_1 \prod_{j=2}^{p} \left[\pi_j + \boldsymbol{\tau}'_{j-1}\boldsymbol{\Phi}^+_{j-1}\left(\mathbf{1}_{j-1} - \boldsymbol{\pi}_{j-1}\right)\right]\right\}.$$

A.4 Moments for N_m

We derive the moments for N_m as follows:

$$\mathcal{E}(\bar{N}_{jj}) = \frac{1}{p}\mathcal{E}\left(\sum_{j=1}^{p} N_{jj}\right) \tag{35}$$

$$= \frac{N}{p}\sum_{j=1}^{p} \pi_j,$$

and

$$\mathcal{V}(\bar{N}_{jj}) = \frac{1}{p^2}\mathcal{V}\left(\sum_{j=1}^{p} N_{jj}\right) \tag{36}$$

$$= \frac{1}{p^2}\left[\sum_{j=1}^{p} \mathcal{V}\left(N_{jj}\right) + \sum_{\{(j,j'):j<j'\}} 2\mathcal{C}\left(N_{jj}, N_{j'j'}\right)\right]$$

$$= N\left(\mathbf{1}'_p \boldsymbol{\Phi} \mathbf{1}_p\right)/p^2.$$

A.5 Specification of Scenarios for Numerical Evaluation of Accuracy

A.5.1 Scenario 4

Here, we specify parameters for the complete data. Let $\alpha = 0.01$, $N \in \{20, 40\}$, $\delta_\beta \in \{0.0, 0.5, 1.0, 1.5, 2\}$, $\delta_\sigma \in \{1.0, 2.0\}$ and $\boldsymbol{\theta}_0 = \mathbf{0}_6$. Define the following matrices:

$$\boldsymbol{X}_m = \boldsymbol{I}_4 \otimes \mathbf{1}_N \otimes \boldsymbol{I}_3, \tag{37}$$

$$\boldsymbol{\beta} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}', \tag{38}$$

$$\boldsymbol{C} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{bmatrix}, \tag{39}$$

$$\boldsymbol{U} = \begin{bmatrix} 1 & 1 \\ -1 & 0 \\ 0 & -1 \end{bmatrix}, \tag{40}$$

$$\boldsymbol{\Psi} = \boldsymbol{I}_N \otimes \begin{bmatrix} 1.0 & 0.4 & 0.4 \\ 0.4 & 1.0 & 0.4 \\ 0.4 & 0.4 & 1.0 \end{bmatrix}. \tag{41}$$

We now specify the missing processes that act on the complete data. For the conditional linear missing data process [8], we specified $\boldsymbol{\pi} = \begin{bmatrix} 0.8 & 0.8 & 0.8 \end{bmatrix}$ and $\boldsymbol{\Phi}$ as compound symmetric covariance pattern with $\gamma_{jj'} = 0.25$.

A.5.2 Scenario 5

Here, we specify parameters for the complete data. Let $\alpha = 0.05$, $N \in \{20, 40, 80\}$, $\delta_\beta \in \{1\}$, $\delta_\sigma \in \{1\}$ and $\boldsymbol{\theta}_0 = \boldsymbol{0}_4$. Define the following matrices:

$$\boldsymbol{X}_m = \boldsymbol{I}_2 \otimes \boldsymbol{1}_N \otimes \boldsymbol{I}_5, \tag{42}$$

$$\boldsymbol{\beta} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}', \tag{43}$$

$$\boldsymbol{C} = \begin{bmatrix} 1 & -1 \end{bmatrix}, \tag{44}$$

$$\boldsymbol{U} = \begin{bmatrix} -0.6325 & 0.5345 & -0.3162 & 0.1195 \\ -0.3162 & -0.2673 & 0.6325 & -0.4781 \\ 0.0000 & -0.5345 & 0.0000 & 0.7171 \\ 0.3162 & -0.2673 & -0.6325 & -0.4781 \\ 0.6325 & 0.5345 & 0.3162 & 0.1195 \end{bmatrix}, \tag{45}$$

$$\boldsymbol{\Psi} = \boldsymbol{I}_N \otimes \begin{bmatrix} 1.500 & 0.375 & 0.375 & 0.375 & 0.375 \\ 0.375 & 1.500 & 0.375 & 0.375 & 0.375 \\ 0.375 & 0.375 & 1.500 & 0.375 & 0.375 \\ 0.375 & 0.375 & 0.375 & 1.500 & 0.375 \\ 0.375 & 0.375 & 0.375 & 0.375 & 1.500 \end{bmatrix}. \tag{46}$$

The missing data processes that act on the complete data are parallel to those described in Scenario 4 with one exception. For Scenario 5, $\boldsymbol{\pi} = \begin{bmatrix} 0.8 & 0.8 & 0.8 & 0.8 & 0.8 \end{bmatrix}$ and $\boldsymbol{\Phi}$ was an AR(1) covariance pattern with $\gamma_{12} = 0.25$.

A.6 Specification of Scenarios for Parkinson's Randomized Controlled Trial Illustrative Example

We give the designs for complete data. Let $\alpha = 0.01$, $N = 121$, δ_β range from 0 to 10 by 0.01, $\delta_\sigma = 1$ and $\boldsymbol{\theta}_0 = \boldsymbol{0}_6$. Define the following matrices:

$$\boldsymbol{X}_m = \begin{bmatrix} \boldsymbol{1}_{41} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1}_{39} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{1}_{41} \end{bmatrix} \otimes \boldsymbol{I}_3, \tag{47}$$

$$\boldsymbol{C} = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \end{bmatrix}, \tag{48}$$

$$\boldsymbol{\beta} = \mathrm{vec}\left(\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}' \right), \tag{49}$$

$$\boldsymbol{U} = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \tag{50}$$

$$\boldsymbol{\theta} = \mathrm{vec}\left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}' \right), \tag{51}$$

$$\boldsymbol{\Psi} = \boldsymbol{I}_N \otimes \begin{bmatrix} 261.86 & 253.47 & 264.80 & 283.52 \\ 253.47 & 286.83 & 284.42 & 302.19 \\ 264.80 & 284.42 & 334.80 & 340.80 \\ 283.52 & 302.19 & 340.80 & 377.03 \end{bmatrix}. \tag{52}$$

Tables & Figures

Summary Statistic	Function	Description	$\mathcal{E}\left(N_{k}\right)$	$\mathcal{V}\left(N_{k}\right)$
N_c	$\sum_{i=1}^{N} I\left(r_i = 0\right)$	number of complete cases	$N\phi$	$N\phi\left(1 - \phi\right)$
N_m	$\bar{N}_{jj}$	mean number of non-missing measurements	$N\bar{\pi}$	$N\left(\mathbf{1}_p'\boldsymbol{\Phi}\mathbf{1}_p\right)/p^2$
N_p	N	planned number of measurements	N	0

Table 1: Missing Data Summary Statistics

Complete/Observed Case Analysis	δ_σ	δ_β	N	$\mathcal{E}(N_c)$	Analytic Power	Empirical Power	Absolute Deviation
Complete	1	1	20	11.5600	0.0213	0.0246	0.0033
Complete	1	2	20	11.5600	0.0771	0.1098	0.0326
Complete	2	1	20	11.5600	0.0152	0.0147	0.0004
Complete	1	1	40	23.1200	0.0777	0.0811	0.0033
Complete	1	2	40	23.1200	0.5329	0.5455	0.0126
Complete	2	1	40	23.1200	0.0356	0.0433	0.0078
Observed	1	1	20	16.0000	0.0407	0.0561	0.0155
Observed	1	2	20	16.0000	0.2440	0.2359	0.0082
Observed	2	1	20	16.0000	0.0227	0.0336	0.0108
Observed	1	1	40	32.0000	0.1542	0.1423	0.0119
Observed	1	2	40	32.0000	0.8408	0.7852	0.0556
Observed	2	1	40	32.0000	0.0604	0.0615	0.0011

Table 2: Selections from Scenario 4

| Complete/Observed Case Analysis	δ_σ	δ_β	N	$\mathcal{E}(N_c)$	Analytic Power	Empirical Power	Absolute Deviation
Complete	1	1	20	8.3521	0.1109	0.1564	0.0454
Complete	1	1	40	16.704	0.3737	0.3855	0.0117
Complete	1	1	80	33.4084	0.8072	0.8128	0.0056
Observed	1	1	20	16.0000	0.3675	0.3672	0.0002
Observed	1	1	40	32.0000	0.7945	0.7618	0.0327
Observed	1	1	80	64.0000	0.9919	0.9883	0.0036

Table 3: Selections from Scenario 5
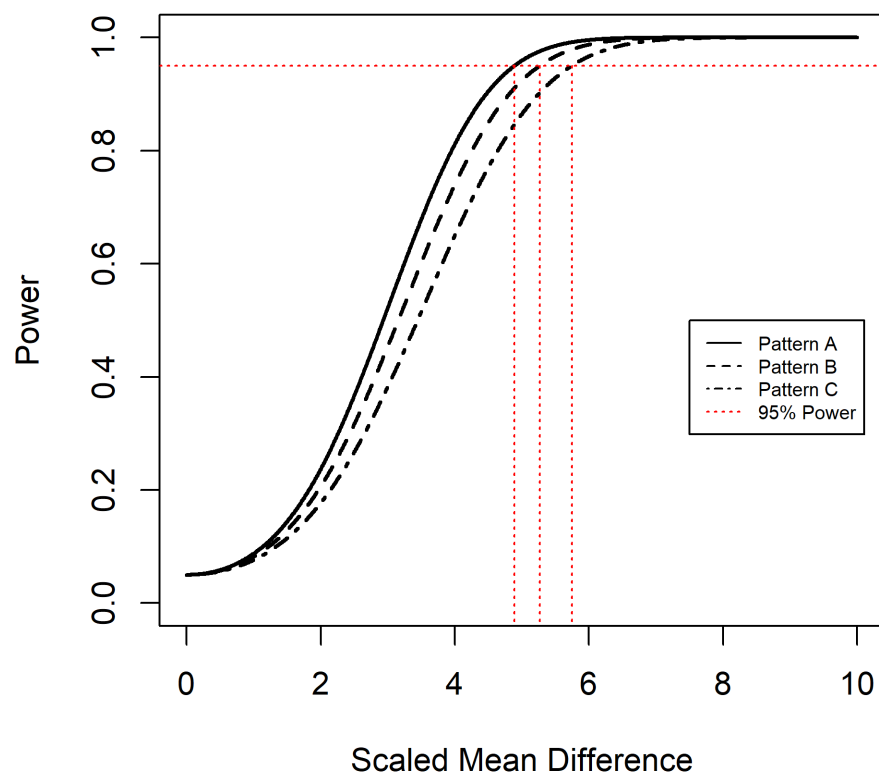
Figure 1: Power for a randomized controlled trial