Published in final edited form as:

Contemp Clin Trials. 2012 May ; 33(3): 550–556.

Sample Size Calculation for Time-Averaged Differences in the Presence of Missing Data

Song Zhang^a and Chul Ahn^{a,*}

^aDepartment of Clinical Sciences, UT Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX, 75390-9066

Abstract

Sample size calculations based on two-sample comparisons of slopes in repeated measurements have been reported by many investigators. In contrast, the literature has paid relatively little attention to the sample size calculations for time-averaged differences in the presence of missing data in repeated measurements studies. Diggle *et al.* (2002) provided a sample size formula comparing time-averaged differences for continuous outcomes in repeated measurement studies assuming no missing data and the compound symmetry (CS) correlation structure among outcomes from the same subject. In this paper we extend Diggle *et al.*'s time-averaged difference sample size formula by allowing missing data and various correlations structures. We propose to use the generalized estimating equation (GEE) method to compare the time-averaged differences in repeated measurement studies and introduce a closed form formula for sample size and power. Simulation studies were conducted to investigate the performance of GEE sample size formula with small sample sizes, damped exponential family of correlation structure and missing data. The proposed sample size formula is illustrated using a clinical trial example.

Keywords

damped exponential correlation; missing data

1. Introduction

In controlled clinical trials, subjects are often evaluated at baseline and intervals over a treatment period. For example, Aronow and Ahn (1994) investigated how blood pressure levels vary after taking meals in 499 elderly residents in a long-term health care facility. Blood pressure was measured at baseline and then every 15 minutes over a two-hour period after taking a meal. The researchers showed that the mean maximal reduction in post-prandial systolic blood pressure was significantly greater in those treated with angiotensin-converting-enzyme inhibitors, calcium channel blockers, diuretics, nitrates, and psychotropic drugs than that in elderly residents not treated with these drugs. By measuring the response (e.g., blood pressure reduction) at multiple time points, researchers hope that the time-averaged response (e.g., averaged reduction in blood pressure over time points)

© 2012 Elsevier Inc. All rights reserved.

Correspondence to: Song Zhang.

*Corresponding author Chul.Ahn@UTSouthwestern.edu (Chul Ahn). Song.Zhang@utsouthwestern.edu (Song Zhang).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

within each group can provide a more precise assessment of the treatment effect. Comparing two treatments based on the time-averaged difference, defined as the difference between the time-averaged responses under two treatments, might consequently offer a greater testing power. Care must be taken in the analysis because of the correlation introduced when several measurements are taken from the same individual. The analysis might be further complicated by the occurrence of missing data.

In this paper we investigate sample size determination based on the test of time-averaged difference between treatment groups over a period of a fixed duration. Diggle et al. (2002) provided closed-form sample size formulas to compare the time-averaged responses and the rates of change in studies with continuous outcomes, assuming no missing data, an equal number of subjects between two groups, and the compound symmetry (CS) correlation among observations from the same subject. Liu and Wu (2005) extended the sample size formula for time-averaged differences to unbalanced clinical trials. Zhang and Ahn (2011) investigated how the number of repeated measurements affects the sample size requirement in repeated measurement studies, where statistical inference is obtained based on time-averaged differences. Here we further extend the sample size calculation for time-averaged difference to allow for missing data, general correlation structures, and unequal sample sizes between study groups. Liang and Zeger (1986) developed the generalized estimating equation (GEE) method which has been widely used to analyze repeated measurements data due to its ability to accommodate missing data and robustness against mis-specification of the true correlation structure. We will employ the GEE method to derive a closed-form sample size formula for repeated measurement studies.

We briefly review the GEE method for the analysis of repeated measurements data in Section 2. A closed-form sample size formula for comparing the time-averaged differences between treatment groups will be derived in Section 3. The proposed sample size formula is general enough to accommodate various missing data patterns, such as random missing or monotone missing, and various correlation structures, represented by a damped exponential family that includes autoregressive correlation with order 1 (AR(1)) and compound symmetry (CS) correlation as special cases. In Section 4, we compare sample size adjustment for missing data by the proposed approach (through theoretical derivation which appropriately accounts for the impact of missing pattern, observation probabilities over time, and correlation structure) with that by the traditional approach (a conservative strategy such that even after excluding patients with partial measurements, the subset of patients with complete measurements are sufficient to achieve the nominal power and type I error). We perform simulations to assess the performance of the sample sizes in Section 5. We applied the sample size formula to a clinical trial example in Section 6. The final section is devoted to discussion.

2. Generalized Estimating Equation Estimator

Let Y_{ij} denote the continuous response measurement obtained at time t_j ($j = 1, \dots, m$) from subject i ($i = 1, \dots, n$), where m is the number of repeated measurements per subject, and n is the total number of subjects enrolled in the study. Without loss of generality, we set the length of the follow-up period $T = t_m - t_1 = 1$. The subjects are randomly assigned to a control or a treatment group, indicated by $r_i = 0$ or 1, respectively. We use $\bar{r} = \sum_{i=1}^n r_i / n$ to denote the proportion of subjects in the treatment group. To make inference based on the difference in time-averaged responses between two groups, we assume the following statistical model (Diggle et al., 2002),

$$Y_{ij} = \beta_1 + \beta_2 r_i + \epsilon_{ij}, \quad (1)$$

where parameter β_1 models the intercept effect, parameter β_2 models the difference in time-averaged responses between two groups, and ε_{ij} denotes random error. The primary interest is to test hypothesis $H_0 : \beta_2 = 0$.

It is usually assumed that $E(\varepsilon_{ij})=0$ and $\text{Var}(\varepsilon_{ij}) = \sigma^2$. Different within-subject correlation structures can be assumed for the error terms. For example, the AR(1) model assumes an exponentially decaying pattern of correlation according to the temporal distance between the repeated measurements while the compound symmetry (CS) model assumes a constant correlation between two distinct measurements regardless of distance. It is also assumed that the error terms are independent between subjects.

To simplify derivation, we rewrite the model as

$$Y_{ij} = b_1 + b_2 (r_i - \bar{r}) + \varepsilon_{ij},$$

where $b_1 = \beta_1 + \beta_2 \bar{r}$, and $b_2 = \beta_2$. Define the vector of covariates $Z_{ij} = (1, r_i - \bar{r})'$. The GEE estimator $\hat{\mathbf{b}} = (\hat{b}_1, \hat{b}_2)$ solves equation $S_n(\mathbf{b}) = 0$, where

$$S_n(\mathbf{b}) = n^{-1/2} \sum_{i=1}^n \sum_{j=1}^m (Y_{ij} - \mathbf{b}' Z_{ij}) Z_{ij}.$$

That is,

$$\hat{\mathbf{b}} = \left(\sum_i \sum_j Z_{ij} Z_{ij}' \right)^{-1} \sum_i \sum_j Z_{ij} Y_{ij}. \quad (2)$$

We use the independent working correlation structure to derive the GEE estimator.

Following Liang and Zeger (1986), $\sqrt{n}(\hat{\mathbf{b}} - \mathbf{b})$ is approximately normal with mean 0 and variance $\sum_n = A_n^{-1} V_n A_n^{-1}$, with

$$\begin{aligned} A_n &= n^{-1} \sum_{i=1}^n \sum_{j=1}^m Z_{ij} Z_{ij}' = n^{-1} \sum_{i=1}^n \sum_{j=1}^m \begin{pmatrix} 1 & r_i - \bar{r} \\ r_i - \bar{r} & (r_i - \bar{r})^2 \end{pmatrix}, \\ V_n &= n^{-1} \sum_{i=1}^n \left(\sum_j Z_{ij} \hat{\varepsilon}_{ij} \right) \left(\sum_j Z_{ij} \hat{\varepsilon}_{ij} \right)' \\ &= n^{-1} \sum_{i=1}^n \sum_{j=1}^m \sum_{j'=1}^m \hat{\varepsilon}_{ij} \hat{\varepsilon}_{ij}' \begin{pmatrix} 1 & r_i - \bar{r} \\ r_i - \bar{r} & (r_i - \bar{r})^2 \end{pmatrix}, \end{aligned} \quad (3)$$

and $\hat{\varepsilon}_{ij} = Y_{ij} - \hat{\mathbf{b}}' Z_{ij}$. Hence we reject hypothesis $H_0 : b_2 = 0$, which is equivalent to reject hypothesis $H_0 : \beta_2 = 0$, if the absolute value of $\sqrt{n} \hat{b}_2 / \hat{\sigma}_2$ is greater than $z_{1-\alpha/2}$. Here $\hat{\sigma}_2^2$ is the (2, 2)th element in $\sum_n$ and $z_{1-\alpha/2}$ is the 100(1 - $\alpha/2$)th percentile of a standard normal distribution.

Let σ_2^2 be the variance of the GEE estimator for β_2 . Given type I error α , power 1 - γ , and the true value of time-averaged difference β_2 , the required sample size is

$$n = \frac{\sigma_2^2 (z_{1-\alpha/2} + z_{1-\gamma})^2}{\beta_2^2}.$$

3. A Closed Form Sample Size Formula

To account for missing data, we assume that measurements are made at scheduled times unless missing, and the missing probability only depends on time. We introduce an indicator δ_{ij} which takes value 1 if subject i has an observation at t_j and 0 otherwise. Under the assumption of missing completely at random (MCAR), $(\delta_{i1}, \dots, \delta_{im})$ is independent from $(Y_{i1}, \dots, Y_{im})$. We define $p_j = E(\delta_{ij})$ to be the probability of a subject contributing a measurement at time t_j . We further define $p_{jj'} = E(\delta_{ij}\delta_{ij'})$ to be the probability of a subject having observations at both time t_j and $t_{j'}$. Then a general expression of A_n and V_n in the presence of missing data is

$$A_n = n^{-1} \sum_{i=1}^n \sum_{j=1}^m \delta_{ij} \begin{pmatrix} 1 & r_i - \bar{r} \\ r_i - \bar{r} & (r_i - \bar{r})^2 \end{pmatrix}$$

and

$$V_n = n^{-1} \sum_{i=1}^n \sum_{j=1}^m \sum_{j'=1}^m \delta_{ij} \delta_{ij'} \widehat{\epsilon}_{ij} \widehat{\epsilon}_{ij'} \begin{pmatrix} 1 & r_i - \bar{r} \\ r_i - \bar{r} & (r_i - \bar{r})^2 \end{pmatrix}.$$

Let $\rho_{jj'} = \text{Corr}(\epsilon_{ij}, \epsilon_{ij'})$ for $j \neq j'$ and $\rho_{jj} = 1$. It can be shown that, as $n \rightarrow \infty$, $A_n \rightarrow A$ and $V_n \rightarrow V$, with

$$A = \lambda \begin{pmatrix} 1 & 0 \\ 0 & \sigma_r^2 \end{pmatrix} \quad \text{and} \quad V = \sigma^2 \begin{pmatrix} \eta & 0 \\ 0 & \eta \sigma_r^2 \end{pmatrix},$$

where $\sigma_r^2 = \bar{r}(1 - \bar{r})$, $\lambda = \sum_{j=1}^m p_j$, and $\eta = \sum_{j=1}^m \sum_{j'=1}^m p_{jj'} \rho_{jj'}$. It is easy to show that the (2, 2)th element of $\Sigma = A^{-1} V A^{-1}$ is

$$\sigma_2^2 = \frac{\sigma^2 \eta}{\lambda^2 \sigma_r^2}.$$

Then the required sample size is

$$n = \frac{\sigma^2 \eta (z_{1-\alpha/2} + z_{1-\gamma})^2}{\beta_2^2 \lambda^2 \sigma_r^2}. \quad (4)$$

This closed-form sample size formula provides a flexible approach to sample size estimation because it can accommodate a broad spectrum of experimental designs, missing patterns, and correlation structures through the specifications of $(\bar{r}, p_j, p_{jj'}, \rho_{jj'})$.

For example, $\bar{r} = 0.5$ indicates a balanced design while $\bar{r} = 0.33$ implies a 1 : 2 randomization ratio between the treatment and control group. The temporal trend of missingness is described by $\mathbf{P} = (p_1, \dots, p_m)$. We can use $p_j = 1 - \theta t_j$ with $\theta \geq 0$ to model the trend that the proportion of missing data increases steadily over time. Other types of temporal trends (e.g., missing probability accelerates or plateaus over time) can also be specified. The specification of $p_{jj'}$ implies different missing patterns. If a subject misses his/her scheduled measurements randomly, that is, δ_{ij} is independent from $\delta_{ij'}$, we have $p_{jj'} = E(\delta_{ij}\delta_{ij'}) = p_j p_{j'}$ for $j \neq j'$. Note that $p_{jj} = p_j$. We call it a random missing (RM) pattern. In some studies, it is likely that if a subject misses a measurement at time t_j , he/she will miss all subsequently scheduled measurements. In such cases, we have $p_{jj'} = p_{j'}$ for $j < j'$ ($p_1 \dots p_m$), and we call it a monotone missing (MM) pattern. Finally, different correlation structures are modeled by $\rho_{jj'}$. We explore a damped exponential family of correlation structures introduced by Munoz et al. (1992), where the $m \times m$ correlation matrix is generally expressed as

$$\rho = (\rho_{jj'})_{m \times m} = \begin{pmatrix} 1 & \rho^{|t_1 - t_2|^\phi} & \dots & \rho^{|t_1 - t_m|^\phi} \\ \rho^{|t_2 - t_1|^\phi} & 1 & \dots & \rho^{|t_2 - t_m|^\phi} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{|t_m - t_1|^\phi} & \rho^{|t_m - t_2|^\phi} & \dots & 1 \end{pmatrix}. \quad (5)$$

Note that under $\phi = 0$, it describes a CS correlation structure with $\rho_{jj'} = \rho$ for $j \neq j'$. That is, the strength of within-subject correlation is independent of the temporal distance between measurements. Under $\phi = 1$, however, it describes an AR(1) correlation structure with $\rho_{jj'} = \rho^{|t_j' - t_j|}$, where the within-subject correlation decays exponentially over time. Thus both the CS and the AR(1) correlation structures are special cases of the damped exponential family. In practice, we can set ϕ to any value between 0 and 1 to flexibly model various types of correlation structures.

4. Traditional Adjustment for Missingness

In the design of clinical trials with repeated measurements, a common practice to account for missing data is to estimate sample size by $n_0/(1 - q)$, where n_0 is the sample size estimated assuming no missing measurement and q is the expected dropout rate (Patel and Rowe, 1999). If we specify $\mathbf{P} = (p_1, \dots, p_m)$ with $p_1 \dots p_m$ the expected dropout rate at the end of study will be $q = 1 - p_m$. Thus the traditional approach would produce a sample size estimate of n_0/p_m . Such a crude adjustment might be unsatisfactory. Specifically, Equation (4) suggests that missing data (characterized by probabilities p_j and $p_{jj'}$) affects sample size requirement through η and λ , which also involves within-subject correlation ($\rho_{jj'}$). For example, under the same dropout rate $1 - p_m$, the actual proportion of missing data can vary depending on the values of $p_1, \dots, p_{m-1}$. Furthermore, missing data might result in less information loss if there are high correlation among repeated measurements. These important factors are ignored by the traditional approach. The following theorem indicates that, under realistic scenarios, the tradition adjustment for missing data always leads to an overestimated sample size to maintain the nominal power and type I error.

Theorem 1. In designing a repeated measurement study to compare the time-averaged difference between the control and treatment groups, the sample size estimated by (4) is smaller than that obtained by traditional adjustment for missing data,

$$n \leq n_0/p_m, \quad (6)$$

as long as the following two conditions hold:

1. The probability of observation is non-increasing over time, $p_1 \geq p_2 \geq \dots \geq p_m$.
2. The within-subject correlation is positive, $\rho_{jj'} > 0$. Furthermore, the equality sign in (6) only holds for complete data ($p_1 = \dots = p_m = 1$) with independent measurements within subjects ($\rho_{jj'} = 0$ for $j \neq j'$).

Proof. See Appendix A.1.

Condition 1 assumes a non-decreasing trend in the probability of missing data, which is a realistic assumption in most clinical trials. The positive correlation assumption stated by Condition 2 agrees with the general perception that the within-subject correlation between measurements is generally non-negative. These two conditions do not put restriction on missing patterns (complete data, RM, MM) or correlation structures (CS, AR(1)). The proposed sample size formula (4) appropriately accounts for factors affecting statistical inference in repeated measurement studies. As a result, Theorem 1 shows that it preserves the power and type I error better than the traditional approach in the presence of missing data.

5. Simulation Study

5.1. Effects of various design configurations

We conduct Simulation 1 to demonstrate the effect of various design configurations on sample size in repeated measurement studies. The nominal levels of power and type I error are set at $1 - \gamma = 0.8$ and $\alpha = 0.05$, respectively. We consider various correlation structures from the damped exponential family with $\phi = 0, 1/4, 1/2, 3/4$ and 1, representing a gradual transition from the CS to the AR(1) structure. Different values of ρ are explored, ranging from 0.1, 0.25, to 0.5. We also assess the effect of missing patterns, RM and MM, with various trends in the observation probability. Assuming a total number of $m = 6$ scheduled measurements from each subject, we consider

$$\begin{aligned}
 \mathbf{P}_1 &= (1, 0.82, 0.79, 0.76, 0.73, 0.7), \\
 \mathbf{P}_2 &= (1, 0.94, 0.88, 0.82, 0.76, 0.7), \\
 \mathbf{P}_3 &= (1, 1, 1, 0.9, 0.8, 0.7), \\
 \mathbf{P}_4 &= (1, 1, 1, 1, 1, 1).
 \end{aligned} \tag{7}$$

The probabilities $\mathbf{P}_1$, $\mathbf{P}_2$ and $\mathbf{P}_3$ describe different scenarios where an increasing number of subjects miss visits over the follow-up period, with the same dropout rate at the end of study $q = 1 - p_m = 0.3$. Specifically, $\mathbf{P}_1$ corresponds to an accelerating trend in missing data, $\mathbf{P}_2$ corresponds to a linear trend (a constant drop over time), while $\mathbf{P}_3$ corresponds to a high observation rate initially but sharp drops near the end of study. We use $\mathbf{P}_4$ to indicate the scenario of complete data. Thus the sample size under $\mathbf{P}_4$ is denoted by n_0 in Theorem 1. Finally, we set the true values of regression coefficients $\boldsymbol{\beta} = (\beta_1, \beta_2)' = (0.3, 0.2)'$ and variance $\sigma^2 = 1$. Note that the sample size estimation does not depend on the value of β_1 and $\beta_2 = 0.2$ indicates an effect size of 0.2 comparing treatment versus control.

To assess the performance of the proposed sample size approach, for every combination of the aforementioned factors (σ^2 , ρ , ϕ , observation probability, missing pattern), the simulation study proceeds as follows: a) Estimate sample size (n) based on Equation (4); b) Generate 5000 null (under $\beta_2 = 0$) and 5000 alternative (under $\beta_2 = 0.2$) data sets each containing n subjects. Every subject has a vector of measurements, $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{im})'$, generated from Model (1), $Y_{ij} = \beta_1 + \beta_2 r_i + e_{ij}$. The random error vectors $(e_{i1}, e_{i2}, \dots, e_{im})$ are generated from a multivariate normal distribution with mean $\mathbf{0}$, variance σ^2 , and correlation matrix $\boldsymbol{\rho}$ as defined in (5); c) The binary missing indicators $(\delta_{ij}, \delta_{i2}, \dots, \delta_{im})$ are generated differently according to missing patterns. Under IM, we generate δ_{ij}

independently from a Bernoulli distribution with probability p_j for $j = 1, \dots, m$. Under MM, δ_{ij} are generated similarly except that when $\delta_{ij} = 0$ we set $\delta_{ij'} = 0$ for $j < j' \leq m$. Different trends of observation probabilities p_j are presented in (7). d) For each data set, obtain $\hat{b}_2$ and $\hat{\sigma}_2$ from Equations (2) and (3), respectively. e) Calculate empirical type I error and empirical power as the proportion of $\sqrt{nb_2}/\hat{\sigma}_2 > z_{1-0.05/2}$ under the null and alternative hypotheses.

The empirical type I error and empirical power being close to the nominal levels ($\alpha = 0.05$, $1 - \gamma = 0.8$) indicate good performance of the proposed method. Table 1 presents sample size estimates and empirical power from Simulation 1. The empirical type I errors are all close to 0.05 and omitted due to space limit. The rows with observation probability P_4 , are identical under missing patterns RM and MM since there is no missing data. We have several observations. First, sample sizes increase in the order of P_4, P_3, P_2 and P_1 , under all correlation structures and missing patterns, which is understandable because the severity of missing increases in the same order. It should be pointed out that the traditional adjustment for missing data greatly overestimates sample size requirement. For example, under the CS correlation with $\rho = 0.1$, the sample size for complete data is $n_0 = 197$. The traditional adjustment for a dropout rate at 0.3 would require a sample size of 282. The proposed approach, however, requires at most 229 and 240 subjects, under the RM and MM missing patterns, respectively. The saving in sample size not only reduces waste in time and resources, more importantly, it exposes less subjects to the potential risk of experimental treatments. Second, along each row, the sample sizes increase with ϕ within the damped exponential family, regardless of missing pattern. The magnitude of relative increase, however, is smaller under higher correlation. For example, under the RM pattern and observation probability P_1 , with $\rho = 0.1$, the sample size increases from 229 to 419 as ϕ increases from 0 to 1, an 83% increase. With a higher correlation ($\rho = 0.5$), the sample size increases by 31%, from 490 to 641. Third, given any missing pattern and correlation structure, a stronger correlation always leads to a larger sample size requirement. For example, under the MM pattern with observation probability P_2 and the AR(1) structure, the required sample sizes are 439, 551, and 677, for correlation $\rho = 0.1, 0.25$, and 0.5 , respectively. Finally, Table 1 demonstrates the importance of correctly specifying the correlation structure and missing pattern. For example, in the first row of Table 1, all other factors being the same, a CS correlation structure ($\phi = 0$) leads to a sample size of 229 while an AR(1) structure ($\phi = 1$) leads to a sample size of 419. Thus if the true correlation structure is AR(1) but we mistakenly assume a CS structure in experimental design, the study will be severely under powered. Similarly, by comparing the corresponding sample sizes under the RM and MM patterns in the table, we can assess the impact of missing pattern. In Table 1, a sample size under MM is always larger than that under RM. Thus if the true missing pattern is random but we mistakenly assume an MM pattern in sample size calculation, the study will be over powered, resulting waste in resource and excess risk for patients. We conducted another simulation with power 0.9 (results presented in Table B.2). Except for larger sample sizes, we have similar observations as from Table B.1. The following theorem formally summarizes the empirical observations from the two tables.

Theorem 2. For a repeated measurement study to compare the time-averaged differences between the control and treatment groups, when Condition 1 and 2 listed in Theorem 1 hold, the required sample size

1. increases with the power parameter ϕ in the damped exponential family;
2. increases with correlation ρ ;
3. is larger under MM than that under RM, when all the other factors hold equal.

Proof. See Appendix A.2.

5.2. Performance under random measurement times

In clinical trials patients might have difficulty following the exact schedule of clinical visits. In other words, the measurement times are more likely to be random instead of being fixed. We conduct simulation 2 to examine the performance of the proposed sample size in a more realistic scenario, where the sample size is calculated assuming a fixed visit schedule based on (4), but the data is simulated in such a fashion that the first and last measurements are made at time 0 and T , and the other measurements are made at random time points in between. The simulation results are presented in Table 3. Because the CS structure ($\phi = 0$) assumes a constant within-subject correlation, the measurement time being random or fixed does not affect the statistical inference of time-averaged differences. Thus the first columns of Table B.1 and Table B.3 are identical. When $\phi > 0$, the temporal difference between measurements affects the strength of within-subject correlation, and in turn affects the time-averaged difference. The effect of randomness in measurement times, however, seems to be small. In Table B.3, the empirical power remains close to the nominal level ($1 - \gamma = 0.8$), indicating that the proposed sample size performs well when clinical visits deviate from the planned schedule.

5.3. Performance under a small sample size

The statistical inference under the GEE method is based on a large sample approximation. It is thus worthwhile to examine the performance of the proposed approach in a small-sample-size scenario, which might be encountered by practitioners due to a large treatment effect or a limited disease population. We conduct Simulation 3 where the sample size is fixed at 30 subjects per group with $n = 60$. Here is the basic idea. From Equation (4) we can derive a relationship between sample size n and treatment effect β_2 :

$$\beta_2 = \sqrt{\frac{\sigma^2 \eta (z_{1-\alpha/2} + z_{1-\gamma})^2}{n \lambda^2 \sigma_r^2}}. \quad (8)$$

Thus by setting $n = 60$, we can compute from (8) the corresponding values of β_2 . In other words, instead of fixing the value of the treatment effect β_2 , in Simulation 3 we allow β_2 to vary across different trial configurations so that the sample size estimated by the proposed approach is always small ($n = 60$).

The results of Simulation 3 are presented in Table B.4, which is conducted in the similar fashion as Simulation 1 except for letting β_2 vary to fix sample size at $n = 60$. The empirical powers are close to the nominal level (0.8) under all trial configurations. It provides assurance to researchers that the proposed approach is widely applicable to clinical trial with repeated measurements, even when the sample size is relatively small (30 subjects per group with a 30% dropout rate).

The simulation studies are conducted using statistical software R 2.13.1 (R Foundation for Statistical Computing, Vienna, Austria). The R code is available upon request from the first author.

6. Example

We apply the sample size calculation method to a labor pain study (Davis, 1991), where women in labour were randomly assigned to the pain medication or the placebo group. At 30 minutes intervals, the self-reported amount of pain was marked on a 100mm line with 0mm=no pain and 100mm=extreme pain. The maximum number of measurements for each woman was $m = 6$, but there were numerous missing values at later measurement times with the monotone missing pattern. The observation probabilities are $\mathbf{P} = (1, 0.90, 0.77, 0.67,$

0.54, 0.41). Exploratory analysis indicates that the measurements has a AR(1) correlation structure with $\rho = 0.38$ and standard deviation $\sigma = 21.3\text{mm}$. In order to detect a time-averaged difference of 15mm with a two-sided type I error of 5% and power 80%, the proposed sample size formula suggests that we need to enroll 55 patients in each group. Under the traditional adjustment for missing data, the estimated sample size will be 64 per group, 16.4% greater than the proposed approach.

7. Discussion

In this study we have derived a closed-form sample size formula for clinical trials with repeated measurements, where the primary goal is to compare the time-averaged difference between the control and treatment groups. This formula is flexible enough to accommodate different missing pattern, severity of missing data, and correlation structures. We further show that the proposed sample size formula, by taking into account of the various designing factors, performs better in adjusting the sample size for missing data than the traditional approach. It should be pointed out that sometimes it might be difficult to know the dropout rates at different time points or to accurately specify the within-subject correlation structure. In such cases the traditional adjustment for missing data remains a practical solution. The closed-form formula also provides insight into the relationship between sample size requirement and the various factors, as summarized in Theorem 2.

The proposed approach is readily applicable to cluster randomization trials with missing data. We have assumed that the missing data arises from a missing completely at random (MCAR) mechanism. In some clinical trials this assumption may not hold. To appropriately account for a non-MCAR mechanism, however, an additional model is required which characterizes the true missing mechanism adequately well. In such cases a general sample size formula is usually unavailable and a specially designed numerical study is required to estimate the sample size for a particular missing mechanism.

Acknowledgments

This work was supported in part by NIH grants UL1 RR024982, P30CA142543, P50CA70907, and DK081872.

Appendix A. Proof of Theorem 1

First we derive the expression of n_0 , which can be obtained from Equation (4) by setting all $p_j = 1$ and all $p_{jj'} = 1$,

$$n_0 = \frac{\sigma^2 \left(\sum_{j=1}^m \sum_{j'=1}^m \rho_{jj'} \right) (z_{1-\alpha/2} + z_{1-\gamma})^2}{\beta_2 m^2 \sigma_r^2}.$$

Thus

$$\frac{n}{n_0/p_m} = \frac{m^2 p_m}{\left(\sum_{j=1}^m p_j \right)^2} \cdot \frac{\sum_{j=1}^m \sum_{j'=1}^m p_{jj'} \rho_{jj'}}{\sum_{j=1}^m \sum_{j'=1}^m \rho_{jj'}}.$$

Defining $\bar{p} = \sum_{j=1}^m p_j / m$ and using the fact that $\rho_{jj} = 1$, we have

$$\frac{n}{n_0/p_m} = \frac{m^2 p_m}{m^2 \bar{p}} \cdot \frac{m \bar{p} + 2 \sum_{j=1}^{m-1} \sum_{j'=j+1}^m p_{jj'} \rho_{jj'}}{m + 2 \sum_{j=1}^{m-1} \sum_{j'=j+1}^m \rho_{jj'}}.$$

Lemma 1. Under Conditions 1 and 2,

$$\sum_{j=1}^{m-1} \sum_{j'=j+1}^m p_{jj'} \rho_{jj'} \leq \bar{p} \sum_{j=1}^{m-1} \sum_{j'=j+1}^m \rho_{jj'}. \quad (\text{A.1})$$

Proof. Inequality (A.1) is equivalent to

$$\frac{\sum_{j=1}^{m-1} \sum_{j'=j+1}^m p_{jj'} \rho_{jj'}}{\sum_{j=1}^{m-1} \sum_{j'=j+1}^m \rho_{jj'}} \leq \bar{p}.$$

Because $\rho_{jj'} = 0$ and $p_{jj'} = p_{j'}$, we have

$$\begin{aligned} \frac{\sum_{j=1}^{m-1} \sum_{j'=j+1}^m p_{jj'} \rho_{jj'}}{\sum_{j=1}^{m-1} \sum_{j'=j+1}^m \rho_{jj'}} &\leq \frac{\sum_{j=1}^{m-1} \sum_{j'=j+1}^m p_{j'} \rho_{jj'}}{\sum_{j=1}^{m-1} \sum_{j'=j+1}^m \rho_{jj'}} \\ &= \frac{\sum_{j'=2}^m \left(p_{j'} \sum_{j=1}^{j'-1} \rho_{jj'} \right)}{\sum_{j'=2}^m \sum_{j=1}^{j'-1} \rho_{jj'}} \\ &= \sum_{j'=2}^m p_{j'} w_{j'}. \end{aligned}$$

That is, the right hand side of inequality is a weighted average of $\{p_{j'}: j' = 2, \dots, m\}$, with weights

$$w_{j'} = \frac{\sum_{j=1}^{j'-1} \rho_{jj'}}{\sum_{l=2}^m \sum_{j=1}^{l-1} \rho_{jl}}$$

With $\rho_{jj'} = 0$ from Condition (2), we have $w_2 = w_3 = \dots = w_m$. Further more, Condition (1) indicates that $p_1 = p_2 = \dots = p_m$. In the weighted average, the weights decrease with the values of the elements. Thus

$$\sum_{j'=2}^m p_{j'} w_{j'} \leq \bar{p}_{(-1)} \leq \bar{p}.$$

Here $\bar{p}_{(-1)} = \sum_{j'=2}^m p_{j'} / (m-2)$ is the unweighted average of $\{p_{j'}: j' = 2, \dots, m\}$. We have the last “ $\leq$ ” sign because $\bar{p}$ includes an additional element (p_1) which is no less than any of the elements in $\bar{p}_{(-1)}$. Thus we complete the proof.

Using Lemma 1, we have

$$\frac{n}{n_0/p_m} \leq \frac{p_m}{\bar{p}^2} \cdot \frac{m\bar{p} + 2\bar{p} \sum_{j=1}^{m-1} \sum_{j'=j+1}^m \rho_{jj'}}{m + 2 \sum_{j=1}^{m-1} \sum_{j'=j+1}^m \rho_{jj'}} = \frac{p_m \bar{p}}{\bar{p}^2} = \frac{p_m}{\bar{p}} \leq 1.$$

The last “ $\leq$ ” sign comes from the fact that p_m is the smallest element in $\{p_j: j = 1, \dots, m\}$, so $p_m \leq \bar{p}$.

It is obvious from the above derivation that the equality sign in (6) only holds when $p_1 = \dots = p_m = 1$ and $\rho_{jj'} = 0$ ($j' > j$).

Appendix B. Proof of Theorem 2

To prove Point 1, it is obvious from Equation (4) that parameter ϕ affects sample size

through $\eta = \sum_{j=1}^m \sum_{j'=1}^m p_{jj'} \rho_{jj'}$. Because $\rho_{jj'} = \rho^{c_{jj'}}$ with $c_{jj'} = |t_j - t_{j'}|^\phi$, for a given ρ (Condition 2 implies that $\rho > 0$), the within-subject correlation coefficients $\rho_{jj'}$ are increasing functions of ϕ . Thus η increases with ϕ , or equivalent, the sample size n increases with ϕ in the damped exponential family.

To prove Point 2, it is easy to show that, given power ϕ , the correlation coefficients $\rho_{jj'}$ are increasing functions of ρ . Thus η increases with ρ , or the sample size n increases with ρ .

To prove Point 3, first note that the missing patterns (RM and MM) affect sample size through $p_{jj'}$ in η . Under missing pattern RM, we have $p_{jj'}^{RM} = p_j p_{j'}$, while under MM, we have $p_{jj'}^{MM} = p_{j'}$ for $j' > j$. With $0 < p_j < 1$, it is always true that $p_{jj'}^{RM} \leq p_{jj'}^{MM}$. With $\rho_{jj'} > 0$ implied by Condition 2, we have $\eta^{RM} \leq \eta^{MM}$. In other words, the MM missing pattern leads to a larger sample size requirement than the RM missing pattern.

Table B.1

Sample Size (Empirical Power) from Simulation 1, Power = 0.8

ρ	P	$\phi = 0$ CS	$\phi = 0.25$	$\phi = 0.5$	$\phi = 0.75$	$\phi = 1$ AR(I)
(a) Random missing						
0.1	P_1	229(0.808)	269(0.809)	318(0.807)	370(0.805)	419(0.804)
	P_2	220(0.808)	261(0.805)	311(0.795)	363(0.812)	413(0.798)
	P_3	211(0.801)	253(0.798)	304(0.798)	357(0.795)	408(0.811)
	P_4	197(0.801)	237(0.807)	287(0.791)	340(0.802)	389(0.797)
0.25	P_1	327(0.785)	380(0.802)	433(0.805)	481(0.800)	522(0.807)
	P_2	317(0.797)	371(0.802)	425(0.809)	474(0.794)	517(0.802)
	P_3	309(0.810)	364(0.812)	418(0.792)	468(0.799)	512(0.799)
	P_4	295(0.795)	349(0.797)	402(0.800)	451(0.804)	493(0.809)
0.5	P_1	490(0.807)	538(0.807)	579(0.802)	613(0.801)	641(0.807)
	P_2	480(0.797)	529(0.798)	571(0.806)	606(0.807)	634(0.793)
	P_3	472(0.791)	522(0.802)	564(0.806)	600(0.791)	628(0.808)
	P_4	458(0.793)	507(0.804)	549(0.802)	583(0.801)	611(0.804)
(b) Monotone missing						
0.1	P_1	240(0.802)	288(0.795)	346(0.803)	408(0.797)	466(0.807)

ρ	P	$\phi = 0$ CS	$\phi = 0.25$	$\phi = 0.5$	$\phi = 0.75$	$\phi = 1$ AR(I)
0.25	P_2	225(0.805)	270(0.805)	326(0.804)	384(0.804)	439(0.796)
	P_3	213(0.811)	256(0.803)	309(0.802)	364(0.799)	417(0.798)
	P_4	197(0.814)	237(0.811)	287(0.808)	340(0.801)	389(0.808)
	P_1	353(0.794)	416(0.798)	479(0.811)	536(0.791)	586(0.799)
0.5	P_2	331(0.806)	391(0.803)	450(0.803)	504(0.805)	551(0.807)
	P_3	313(0.800)	370(0.813)	426(0.799)	478(0.803)	523(0.806)
	P_4	295(0.801)	349(0.813)	402(0.804)	451(0.806)	493(0.803)
	P_1	542(0.803)	599(0.797)	648(0.798)	689(0.808)	722(0.806)
	P_2	507(0.790)	562(0.799)	608(0.796)	646(0.806)	677(0.798)
	P_3	480(0.800)	531(0.806)	575(0.807)	612(0.805)	641(0.797)
	P_4	458(0.809)	507(0.799)	549(0.808)	583(0.793)	611(0.798)

Table B.2

Sample Size (Empirical Power) from Simulation 1, Power = 0.9

ρ	P	$\phi = 0$ CS	$\phi = 0.25$	$\phi = 0.5$	$\phi = 0.75$	$\phi = 1$ AR(1)
(a) Random missing						
0.1	P_1	307(0.893)	360(0.903)	426(0.899)	495(0.901)	560(0.895)
	P_2	294(0.902)	349(0.903)	416(0.906)	486(0.900)	553(0.902)
	P_3	282(0.905)	338(0.909)	406(0.899)	478(0.899)	546(0.900)
	P_4	263(0.907)	318(0.904)	384(0.898)	454(0.902)	521(0.902)
0.25	P_1	438(0.898)	508(0.895)	579(0.900)	643(0.897)	699(0.899)
	P_2	425(0.897)	497(0.902)	569(0.898)	635(0.904)	691(0.894)
	P_3	413(0.900)	487(0.903)	560(0.904)	627(0.901)	685(0.902)
	P_4	395(0.899)	466(0.903)	538(0.900)	603(0.899)	660(0.895)
0.5	P_1	656(0.905)	720(0.904)	775(0.896)	821(0.904)	858(0.898)
	P_2	643(0.904)	709(0.899)	765(0.901)	811(0.902)	849(0.888)
	P_3	631(0.906)	698(0.896)	775(0.902)	802(0.905)	841(0.905)
	P_4	613(0.903)	679(0.901)	735(0.905)	781(0.897)	818(0.907)
(b) Monotone missing						
0.1	P_1	321(0.899)	385(0.905)	463(0.901)	545(0.904)	624(0.899)
	P_2	301(0.902)	362(0.904)	436(0.903)	514(0.905)	588(0.910)
	P_3	284(0.900)	342(0.900)	413(0.896)	488(0.899)	558(0.910)
	P_4	263(0.905)	318(0.906)	384(0.907)	454(0.901)	521(0.904)
0.25	P_1	473(0.902)	557(0.898)	641(0.911)	718(0.904)	784(0.900)
	P_2	443(0.893)	523(0.898)	602(0.898)	675(0.902)	737(0.896)
	P_3	418(0.901)	495(0.905)	571(0.909)	640(0.898)	700(0.906)
	P_4	395(0.898)	466(0.898)	538(0.899)	603(0.911)	660(0.903)
0.5	P_1	726(0.898)	802(0.907)	868(0.900)	922(0.892)	966(0.897)
	P_2	679(0.897)	752(0.896)	814(0.898)	865(0.903)	906(0.905)
	P_3	642(0.906)	711(0.903)	770(0.895)	819(0.898)	858(0.894)
	P_4	613(0.904)	679(0.904)	735(0.900)	781(0.908)	818(0.905)

Table B.3

Sample Size (Empirical Power) from Simulation 2, Random Measurement Times

ρ	P	$\phi = 0$ CS	$\phi = 0.25$	$\phi = 0.5$	$\phi = 0.75$	$\phi = 1$ AR(1)
(a) Random missing						
0.1	P_1	229(0.808)	269(0.774)	318(0.775)	370(0.775)	419(0.780)
	P_2	220(0.808)	261(0.783)	311(0.772)	363(0.770)	413(0.784)
	P_3	211(0.801)	253(0.776)	304(0.764)	357(0.766)	408(0.780)
	P_4	197(0.801)	237(0.778)	287(0.769)	340(0.767)	389(0.766)
0.25	P_1	327(0.785)	380(0.795)	433(0.782)	481(0.790)	522(0.792)
	P_2	317(0.797)	371(0.784)	425(0.779)	474(0.782)	517(0.790)
	P_3	309(0.810)	364(0.790)	418(0.787)	468(0.776)	512(0.792)
	P_4	295(0.795)	349(0.792)	402(0.787)	451(0.787)	493(0.791)
0.5	P_1	490(0.807)	538(0.798)	579(0.794)	613(0.795)	641(0.795)
	P_2	480(0.797)	529(0.796)	571(0.795)	606(0.791)	634(0.799)
	P_3	472(0.791)	522(0.789)	564(0.789)	600(0.786)	628(0.806)
	P_4	458(0.793)	507(0.798)	549(0.786)	583(0.802)	611(0.801)
(b) Monotone missing						
0.1	P_1	240(0.802)	288(0.772)	346(0.764)	408(0.764)	466(0.784)
	P_2	225(0.805)	270(0.776)	326(0.774)	384(0.768)	439(0.782)
	P_3	213(0.811)	256(0.765)	309(0.773)	364(0.773)	417(0.779)
	P_4	197(0.814)	237(0.780)	287(0.775)	340(0.780)	389(0.787)
0.25	P_1	353(0.794)	416(0.789)	479(0.781)	536(0.793)	586(0.795)
	P_2	331(0.806)	391(0.791)	450(0.793)	504(0.784)	551(0.785)
	P_3	313(0.800)	370(0.792)	426(0.793)	478(0.791)	523(0.806)
	P_4	295(0.801)	349(0.788)	402(0.788)	451(0.784)	493(0.791)
0.5	P_1	542(0.803)	599(0.785)	648(0.801)	689(0.793)	722(0.799)
	P_2	507(0.790)	562(0.789)	608(0.794)	646(0.795)	677(0.798)
	P_3	480(0.800)	531(0.790)	575(0.788)	612(0.807)	641(0.796)
	P_4	458(0.809)	507(0.792)	549(0.794)	583(0.802)	611(0.786)

Table B.4

Sample Size (Empirical Power) From Simulation 3, Small Sample Sizes

ρ	P	$\phi = 0$ CS	$\phi = 0.25$	$\phi = 0.5$	$\phi = 0.75$	$\phi = 1$ AR(1)
(a) Random missing						
0.1	P_1	60(0.809)	60(0.807)	60(0.798)	60(0.817)	60(0.809)
	P_2	60(0.809)	60(0.814)	60(0.816)	60(0.803)	60(0.816)
	P_3	60(0.808)	60(0.818)	60(0.811)	60(0.810)	60(0.802)
	P_4	60(0.815)	60(0.798)	60(0.809)	60(0.814)	60(0.819)
0.25	P_1	60(0.808)	60(0.808)	60(0.814)	60(0.824)	60(0.809)
	P_2	60(0.809)	60(0.812)	60(0.804)	60(0.812)	60(0.804)
	P_3	60(0.814)	60(0.812)	60(0.805)	60(0.813)	60(0.809)

ρ	P	$\phi = 0$ CS	$\phi = 0.25$	$\phi = 0.5$	$\phi = 0.75$	$\phi = 1$ AR(1)
0.5	P_4	60(0.806)	60(0.813)	60(0.810)	60(0.820)	60(0.807)
	P_1	60(0.799)	60(0.808)	60(0.824)	60(0.810)	60(0.814)
	P_2	60(0.824)	60(0.804)	60(0.810)	60(0.803)	60(0.805)
	P_3	60(0.808)	60(0.817)	60(0.810)	60(0.806)	60(0.815)
	P_4	60(0.808)	60(0.809)	60(0.820)	60(0.817)	60(0.806)
(b) Monotone missing						
0.1	P_1	60(0.808)	60(0.803)	60(0.815)	60(0.826)	60(0.812)
	P_2	60(0.812)	60(0.808)	60(0.810)	60(0.825)	60(0.818)
	P_3	60(0.807)	60(0.810)	60(0.804)	60(0.800)	60(0.801)
	P_4	60(0.797)	60(0.797)	60(0.815)	60(0.807)	60(0.817)
0.25	P_1	60(0.806)	60(0.814)	60(0.806)	60(0.811)	60(0.812)
	P_2	60(0.812)	60(0.810)	60(0.812)	60(0.801)	60(0.814)
	P_3	60(0.802)	60(0.804)	60(0.810)	60(0.813)	60(0.814)
	P_4	60(0.800)	60(0.798)	60(0.817)	60(0.807)	60(0.812)
0.5	P_1	60(0.815)	60(0.817)	60(0.812)	60(0.816)	60(0.816)
	P_2	60(0.815)	60(0.815)	60(0.824)	60(0.818)	60(0.813)
	P_3	60(0.807)	60(0.803)	60(0.799)	60(0.804)	60(0.798)
	P_4	60(0.808)	60(0.806)	60(0.809)	60(0.813)	60(0.798)

References

- Aronow WS, Ahn C. Postprandial hypotension in 499 elderly persons in a long-term health care facility. *Journal of the American Geriatrics Society*. 1994; 42(9):930–932. [PubMed: 8064099]
- Davis CS. Semi-parametric and non-parametric methods for the analysis of repeated measurements with applications to clinical trials. *Statistics in Medicine*. 1991; 10(12):1959–1980. [PubMed: 1805321]
- Diggle, P.; Heagerty, P.; Liang, K.; Zeger, S. *Analysis of longitudinal data*. 2nd ed.. Oxford University Press; 2002.
- Liang K, Zeger SL. Longitudinal data analysis for discrete and continuous outcomes using generalized linear models. *Biometrika*. 1986; 84:3–32.
- Liu H, Wu T. Sample size calculation and power analysis of time-averaged difference. *Journal of Modern Applied Statistical Methods*. 2005; 4(2):434–445.
- Munoz A, Carey V, Schouten JP, Segal M, Rosner B. A parametric family of correlation structures for the analysis of longitudinal data. *Biometrics*. 1992; 48(3):733–742. [PubMed: 1420837]
- Patel H, Rowe E. Sample size for comparing linear growth curves. *Journal of Biopharmaceutical Statistics*. 1999; 9(2):339–50. [PubMed: 10379697]
- Zhang S, Ahn C. How many measurements for time-averaged differences in repeated measurement studies? *Contemporary Clinical Trials*. 2011; 32(3):412–417. [PubMed: 21241827]