Journal of Educational and Behavioral Statistics
Spring 1999, Vol. 24, No. 1, pp. 70–93

Sample Size Estimation for Longitudinal Designs with Attrition: Comparing Time-Related Contrasts Between Two Groups

Donald Hedeker
University of Illinois at Chicago

Robert D. Gibbons
University of Illinois at Chicago

Christine Waternaux
Columbia University

Keywords: *clustering, hierarchical linear models, missing data, multilevel models, mixed-effects models, repeated measures, statistical power*

Formulas for estimating sample sizes are presented to provide specified levels of power for tests of significance from a longitudinal design allowing for subject attrition. These formulas are derived for a comparison of two groups in terms of single degree-of-freedom contrasts of population means across the study timepoints. Contrasts of this type can often capture the main and interaction effects in a two-group repeated measures design. For example, a two-group comparison of either an average across time or a specific trend across time (e.g., linear or quadratic) can be considered. Since longitudinal data with attrition are often analyzed using an unbalanced repeated measures model (with a structured variance-covariance matrix for the repeated measures) or a random-effects model for incomplete longitudinal data, the variance-covariance matrix of the repeated measures is allowed to assume a variety of forms. Tables are presented listing sample size determinations assuming compound symmetry, a first-order autoregressive structure, and a non-stationary random-effects structure. Examples are provided to illustrate use of the formulas, and a computer program implementing the procedure is available from the first author.

Longitudinal studies occupy an important role in applied research in many fields. In planning and designing longitudinal studies, researchers must determine the number of subjects necessary for the study by calculating statistical power associated with a proposed sample size and analysis plan. As Muller, LaVange, Ramey, and Ramey (1992) note, determining statistical power is important to avoid two serious errors. Insufficient sample size can lead to inadequate sensitivity, whereas an excessive sample size can be a waste of the researchers' (and study participants') time and money.

As longitudinal designs in applied research have proliferated, there has been more focus on power calculations and sample size determination. Muller and Barton (1989) and Muller and Peterson (1984) discuss the univariate approach to repeated measures and power calculations for the general linear multivariate model. O'Brien and Muller (1992) provide a tutorial on power for linear models and for tests ranging from t-tests to multivariate tests. Much of this work was reviewed and expanded upon in Muller et al. (1992), which presents a comprehensive approach to the calculation of statistical power for longitudinal studies. Their treatment includes power formulas applicable to both multivariate and univariate repeated measures analysis, and for between-group comparisons involving multiple groups. Overall and Doyle (1994) and Kirby, Galai, and Muñoz (1994) also discuss sample size estimation for repeated measures models involving two (between-subjects) groups.

For random-effects models, Snijders and Bosker (1993) present approximate formulas for the standard errors of estimated regression coefficients. These standard errors can be used for power calculations for explanatory variables (e.g., group effects, or group by time interactions) in random-effects models involving repeated observations (level-1) nested within subjects (level-2). Their approach assumes that the sample size at either level (i.e., subjects and timepoints) is sufficiently large and does not change over time, and that there are no autocorrelated errors. These assumptions are generally reasonable for clustered data (e.g., students within schools, or patients within clinics), however, in the longitudinal context they may be less so.

In general, these power formulas assume that the sample size is constant across time. Since this assumption is rarely met in practice, it is often recommended to conservatively use the minimum expected sample size at any study timepoint in applying these formulas. In this article, we will develop formulas for sample size estimation and assessment of statistical power for longitudinal designs allowing for subject attrition. These formulas will be developed for comparing two groups in terms of single degree-of-freedom contrasts across time. Though simple, single degree-of-freedom contrasts can often be used to capture the main and interaction effects in a two-group repeated measures design, particularly in the design phase of a study. For example, a one degree-of-freedom contrast might represent a hypothesized overall group difference across time or a group by time interaction of a specific type (e.g., a linear or quadratic trend). With regards to missing data, missingness is assumed not to moderate or mediate the effect of model terms for which power is being determined.

In addition to allowing for attrition, we will consider the variance-covariance matrix of the repeated measures to assume a general form, to be restricted to a specific form (e.g., compound symmetry, autoregressive, or Toeplitz structures), or to follow a structure resulting from a posited random-effects model of the repeated measures. Thus, these power formulas are aimed at two-group repeated measures studies where specific time-related contrasts are of primary interest and the number of observations vary across time. Since only single degree-of-freedom contrasts across time are considered, the assumption of circularity or sphericity regarding the variance-covariance matrix of repeated measurements is not an issue. As is shown, these formulae can be applied under a variety of variance-covariance structures for the repeated measures.[1]

Models for Unbalanced Longitudinal Data

Assume that there are n timepoints ($j = 1, 2, \ldots, n$), corresponding to a fixed number of occasions or experimental conditions, however not all subjects are observed at all timepoints. Let n_i equal the number of observations for subject i. Jennrich and Schluchter (1986) describe models for unbalanced longitudinal data under a variety of variance-covariance structures. For this, consider the following linear model in which the $n_i \times 1$ response vector $\mathbf{y}_i$ for subject i is modeled in terms of p covariates (including the intercept):

$$\mathbf{y}_i = X_i\beta + \mathbf{e}_i \tag{1}$$

where X_i is a known $n_i \times p$ covariate matrix, β is a $p \times 1$ vector of unknown regression parameters, and $\mathbf{e}_i$ is a $n_i \times 1$ vector of residuals distributed independently as $N(0, \Sigma_i)$. The matrix Σ_i depends on i only through its dimension; it depends on q unknown covariance parameters. Besides the unstructured (i.e., $q = n(n + 1)/2$) common choices for Σ_i include a compound-symmetry structure ($q = 2$), a first-order autoregressive structure ($q = 2$), and a Toeplitz structure ($q = n$).

Alternatively, the variance-covariance matrix can be assumed to follow a structure based on a particular random-effects model. Under these models, also known as multilevel or hierarchical linear models (Laird & Ware, 1982; Goldstein, 1995; Bock, 1989; Bryk & Raudenbush, 1992; Longford, 1993), the variance-covariance matrix of the repeated measures is expressed as

$$\Sigma_i = Z_i\Sigma_\theta Z_i' + \sigma_\epsilon^2\Omega_i, \tag{2}$$

where Σ_θ represents the variance covariance matrix of the person-varying random effects θ_i, Z_i represents the design matrix corresponding to the random effects, σ_ϵ^2 represents the error variance, and Ω_i represents a possible autocorrelated error structure (e.g., if autocorrelated errors are not considered, Ω_i is assumed to be the identity matrix I_i). This structure arises by assuming

$$\mathbf{e}_i = Z_i\theta_i + \epsilon_i \tag{3}$$

where the random effects θ_i and errors ϵ_i are assumed to be distributed independently as $N(0, \Sigma_\theta)$ and $N(0, \sigma_\epsilon^2\Omega_i)$.

For the two-group design, suppose that X_i, has dimension $n_i \times 2n$ and is partitioned into two $n_i \times n$ component matrices. The first component matrix (designated W_i) includes a constant term as the first column followed by $n - 1$ contrasts for the within-subjects design. The second component matrix (designated B_i) represents the between-subjects design and is obtained by multiplying W_i by -1 (if subject i belongs to the first group) or 1 (if subject i belongs to the second group). Similarly, partition the $2n \times 1$ parameter vector β into β_W and β_B. The first element of β_W is the grand mean of the model and the remaining $n - 1$ elements represents the parameters corresponding to the main effect of time. Similarly, the first element of β_B is the parameter representing the main effect of group and the remaining $n - 1$ elements represent the parameters of the group by time interaction.

Contrasts for W_i (and thus for B_i as well) based on an orthonormal $n \times n$ contrast matrix W have many advantages. In this case, assuming a compound symmetry structure and no missing data across time (i.e., $n_i = n$ for all i), maximum likelihood estimates of β_W and β_B are invariant with respect to the addition or deletion of W and B terms, respectively. Additionally, for the full model (i.e., with $2n$ regression parameters), the estimates of β_W and β_B are invariant with respect to the variance-covariance structure of Σ (see Appendix) if $n_i = n$ for all i. Since, in general, $n_i \neq n$ for all subjects, neither W_i or B_i are mathematically orthogonal (or orthonormal) for subject i when $n_i \neq n$. However, if it is assumed that there is no mediating or moderating effect of missingness on the terms in W or B in the population, then approximately the same estimates of β_W and β_B are obtained. For longitudinal studies, a common choice of W is to use orthogonal polynomial contrasts. To obtain orthonormal contrast coefficients for specific polynomial trends across time, a table listing orthogonal polynomials (Bock, 1975) or programmed subroutines (Cooper, 1971) can be used.

In planning a two-group study, interest usually centers around the power associated with hypothesis testing of the parameters β_B. If the number of observations n per subject is equal, the methods described by Muller et al., (1992) can be used to determine power for both the overall group effect and the group by time interaction, assuming a variety of forms for the variance-covariance structure of the repeated measures. In what follows, we provide formulas for power calculations for testing specific elements of β_B, allowing the number of observations per subject to vary and for a variety of variance-covariance structures. These methods can be used when interest focuses on testing the main effect of group (i.e., the first element of β_B) or a specific form for the group by time interaction (i.e., a single element of the remaining $n - 1$ parameters in β_B).

Power under a General Variance-Covariance Structure

Consider the case where the variance-covariance structure of repeated measures is of a general form, however, the matrix is assumed to be homogeneous across the two groups. Further, it is assumed that the variances and covariances of the matrix are known or can be specified. To begin, we will consider the formula for a two-group comparison at a single timepoint and extend this formula for a two-group comparison based on a specified contrast across time.

Comparison of Two Groups at a Single Timepoint

For the case of a comparison of two groups at a single timepoint, the following formula can be used to approximate the required number of subjects (N) in each of the two groups (Fleiss, 1986, page 369):

$$N = \frac{2(z_\alpha + z_\beta)^2\sigma^2}{(\mu_1 - \mu_2)^2} \tag{4}$$

where z_α is the value of the standardized score cutting off $\alpha/2$ proportion of each tail of a standard normal distribution (for a two-tailed hypothesis test), z_β is the value of the standardized score cutting off the upper β proportion, σ^2 is the assumed common variance in the two groups, and $\mu_1 - \mu_2$ is the difference in means of the two groups. This normal approximation of the noncentral t distribution is reasonable if the degrees of freedom exceed 30. Thus, this approximation can be used in sample size determination for most educational and behavioral studies.

Comparison of Two Groups Across Timepoints—Balanced Case

As noted by Overall and Doyle (1994), for longitudinal designs where the number of timepoints is equal to n, the above formula can be modified for determining the sample size corresponding to a contrast (denoted Ψ_c) of the group population means across the n timepoints as:

$$N = \frac{2(z_\alpha + z_\beta)^2\sigma_c^2}{\Psi_c^2} \tag{5}$$

with

$$\Psi_c = \sum_{j=1}^{n} c_j(\mu_{1j} - \mu_{2j}) \tag{6}$$

and

$$\sigma_c^2 = \sum_{j=1}^{n} c_j^2\sigma_j^2 + 2\sum_{j<j'}^{n} c_j c_{j'} \sigma_{j,j'} \tag{7}$$

Here, σ_j^2 refers to the assumed common variance in the two groups at timepoint j, $\sigma_{j,j'}$ refers to the assumed common covariance in the two groups between timepoints j and j', and c_j refers to the within-subjects contrast applied at timepoint j. Since equation (6) includes the coding of -1 and 1 for the comparison of the two groups, it incorporates the multiplication of the group indicator (i.e., -1 or 1) by the within-subjects design matrix $\boldsymbol{W}$ that was used to yield the between-subjects design matrix $\boldsymbol{B}$ described earlier. As such, the contrast coefficients c_j are the elements of a particular column of the design matrix $\boldsymbol{W}$. Note that in assuming an equal number of subjects in each of the two groups (i.e., N) across all timepoints, the variance of the sample contrast $V(\hat{\Psi}_c)$ equals $2\sigma_c^2/N$.

Formula (5) can be used to calculate the number of subjects necessary in each of the two groups to achieve the desired level of power for a between-group difference in terms of a specific contrast across time. If the sample size is known and the degree of power is to be determined, the formula can be re-expressed as:

$$z_\beta = \sqrt{\frac{N\Psi_c^2}{2\sigma_c^2}} - z_\alpha = \sqrt{\frac{\Psi_c^2}{V(\hat{\Psi}_c)}} - z_\alpha . \tag{8}$$

Single degree-of-freedom contrasts across time (i.e., Ψ_c) can often be used to represent the main and interaction effects in a two-group repeated measures design. The choice of these contrasts clearly depends upon the analysis that is planned for the study. For example, if $c_j = 1/\sqrt{n}$ ($j = 1, \ldots, n$), testing $\Psi_c = 0$ provides a test of an overall group difference across time (i.e., the first element of $\beta_B = 0$). Here, the contrast is standardized so that $\Sigma_{j=1}^n c_j^2 = 1$. Notice that if $c_j = 1/\sqrt{n}$, then

$$\sigma_c^2 = \frac{1}{n}\left(\sum_{j=1}^{n} \sigma_j^2 + 2\sum_{j<j'}^{n} \sigma_{j,j'}\right), \tag{9}$$

which when substituted in (5) shows that, all other things being equal, more subjects are necessary to detect an overall group effect as the correlation of the repeated measures increases.

To calculate power for a group by time interaction, a simple but often reasonable option is to consider a specific form for the interaction (i.e., a particular column of β_B). For example, a group by linear time interaction is a common choice. While the degrees of freedom for the group by time interaction equals $n - 1$, if a particular type of group by time interaction is hypothesized in the study design, the above formula for a single degree of freedom contrast can be used to assess power for this particular component of the group by time interaction. For example, with two timepoints, the orthonormal coefficients for the linear component equal $c_1 = 1/\sqrt{2}$ and $c_2 = -1/\sqrt{2}$, and so

$$\sigma_c^2 = \frac{1}{2}(\sigma_1^2 + \sigma_2^2) - \sigma_{12}, \tag{10}$$

illustrating that, all other things being equal, less subjects are necessary to detect a group by linear time effect as the correlation of the repeated measures increases. If the timepoints are not equally spaced, the coefficients for the (orthonormal) linear contrast can be modified accordingly (Emerson, 1968). Higher-order polynomials can also be used to provide power for a test of, say, a group by quadratic time interaction. For two-period crossover designs, the coefficients c_j can be set equal to $-1/\sqrt{n}$ and $1/\sqrt{n}$ for the first and second periods, respectively. Then the contrast represents the group by period interaction. Since the number of timepoints within each period may not be equal, standardization within each period can be performed (e.g., $c_j = -1/\sqrt{2n_1}$ for the first period with n_1 timepoints, and $c_j = 1/\sqrt{2n_2}$ for the second period with n_2 timepoints) to achieve equal weighting of the two periods. Other types of contrasts for two-period crossover studies are described in Bock (1983).

Once a set of time-related contrast coefficients (c_j, $j = 1, \ldots, n$) has been determined, usage of formula (5) or (8) requires the assumption of a common sample size (N) for each group across the n timepoints. Additionally, expected group mean differences at each of the n timepoints and the variance-covariance matrix of the repeated measures must be specified. For specific contrasts, using formula (8) gives the same results as those obtained using the methods described in Muller et al., (1992) for moderate to large sample size (i.e., $N > 30$ or so).

Comparison of Two Groups Across Timepoints—Unbalanced Case

In many studies sample size does not remain constant over time, it generally decreases due to subject attrition or non-response. One option is to apply the above formula using the minimum expected sample size at any timepoint of the study. However, this approach will generally underestimate power. Alternatively, the above formula can be modified to appropriately account for the sample size at each of n timepoints. We also distinguish between the sample size in the first group (N_{1j}) and the second group (N_{2j}) at timepoint $j (j = 1, \ldots, n)$. Allowing for varying sample sizes between groups and across timepoints, variance of the sample contrast $\hat{\Psi}_c$ is now equal to

$$V(\hat{\Psi}_c) = \sum_{j=1}^{n} c_j^2 \sigma_j^2 \left(\frac{1}{N_{1j}} + \frac{1}{N_{2j}} \right) + 2 \sum_{j<j'}^{n} c_j c_{j'} \sigma_{j,j'} \left(\frac{1}{\sqrt{N_{1j} N_{1j'}}} + \frac{1}{\sqrt{N_{2j} N_{2j'}}} \right). \quad (11)$$

Using the sample size in the first group at the first timepoint (n_{11}) as a reference, let us define retention rates for this group as r_{1j} for timepoints $j = 1, \ldots, n$, which indicate the proportion of N_1 subjects observed at timepoint i (note that $r_{11} = 1$ and $N_{1j} = r_{1j} N_{11}$). Similarly we define N_{21} and r_{2j} for group two. Then the above formula is rewritten as:

$$V(\hat{\Psi}_c) = \frac{1}{N_{11}} \left[\sum_{j=1}^{n} c_j^2 \sigma_j^2 \left(\frac{1}{r_{1j}} + \frac{1}{r_{2j}} \frac{N_{11}}{N_{21}} \right) + 2 \sum_{j<j'}^{n} c_j c_{j'} \sigma_{j,j'} \left(\frac{1}{\sqrt{r_{1j} r_{1j'}}} + \frac{N_{11}}{N_{21}} \frac{1}{\sqrt{r_{2j} r_{2j'}}} \right) \right], \quad (12)$$

and if we denote the ratio of sample sizes at the first timepoint (N_{11}/N_{21}) as N_{-1}, then

$$V(\hat{\Psi}_c) = \frac{1}{N_{11}}\left[\sum_{j=1}^{n} c_j^2\sigma_j^2\left(\frac{1}{r_{1j}} + \frac{N_{-1}}{r_{2j}}\right) + \right.$$
$$\left. 2\sum_{j<j'}^{n} c_j c_{j'} \sigma_{j,j'}\left(\frac{1}{\sqrt{r_{1j}r_{1j'}}} + N_{-1}\frac{1}{\sqrt{r_{2j}r_{2j'}}}\right)\right]. \tag{13}$$

Notice this formula simplifies if the retention rates are equal for the two groups (i.e., $r_{1j} = r_{2j} = r_j$),

$$V(\hat{\Psi}_c) = \frac{N_{-1}+1}{N_{11}}\left[\sum_{j=1}^{n}\frac{c_j^2\sigma_j^2}{r_j} + 2\sum_{j<j'}^{n}\frac{c_j c_{j'}\sigma_{j,j'}}{\sqrt{r_j r_{j,}}}\right]$$
$$= \frac{N_{-1}+1}{N_{11}}\sigma_{rc}^2. \tag{14}$$

where σ_{rc}^2 can be seen as an extension of σ^2 given in (7) allowing for varying sample size across timepoints (although group retention rates are assumed equal).

To calculate power for any of the above variance formulations of the sample contrast, we can use the previously noted relationship

$$z_\beta = \sqrt{\frac{\Psi_c^2}{V(\hat{\Psi}_c)}} - z_\alpha. \tag{15}$$

In particular, for the case of equal group retention rates,

$$z_\beta = \sqrt{\left(\frac{N_{11}}{N_{-1}+1}\right)\frac{\Psi_c^2}{\sigma_{rc}^2}} - z_\alpha, \tag{16}$$

which reduces to the formula given in (8) when the sample sizes are the same for the two groups ($N_{-1} = 1$) at all timepoints ($r_j = 1$ for all i). Formula (16) can be re-expressed as number of subjects needed in the first group at the first time-point:

$$N_{11} = \frac{(N_{-1}+1)(z_\alpha + z_\beta)^2\sigma_{rc}^2}{\Psi_c^2}. \tag{17}$$

Based on the sample size ratio between groups N_{-1} and retention rates r_j, required sample size at each timepoint for both groups can be calculated.

Specific Variance-Covariance Structures

The presentation thus far has assumed that the variance-covariance matrix of the repeated measures is of a general form, and so requires specification of the unique elements of this $n \times n$ matrix. It is often more parsimonious to assume a restricted form for the variance-covariance matrix. As described above, analysis of unbalanced repeated-measures data with different types of covariance structures is discussed by Jennrich and Schluchter (1986). In terms of the previous formulas, the variance of the sample contrast needs to be respecified in accordance with the specific variance-covariance structure assumed. For example, for a compound symmetry structure, (11) can be rewritten as:

$$V(\hat{\Psi}_c) = \sigma^2 \left[\sum_{j=1}^{n} c_j^2 \left(\frac{1}{N_{1j}} + \frac{1}{N_{2j}} \right) + 2\rho \sum_{j<j'}^{n} c_j c_{j'} \left(\frac{1}{\sqrt{N_{1j} N_{1j'}}} + \frac{1}{\sqrt{N_{2j} N_{2j'}}} \right) \right], \qquad (18)$$

requiring specification of only the variance and correlation parameters σ^2 and ρ that are assumed homogeneous across time. Assuming that the retention rates are equal for the two groups, and denoting the effect size $d_j = (\mu_{1j} - \mu_{2j})/\sigma$, formula (16) becomes:

$$z_\beta = \sqrt{\left(\frac{N_{11}}{N_{-1} + 1} \right) \frac{(\Sigma_{j=1}^{n} c_j d_j)^2}{\Sigma_{j=1}^{n} c_j^2 / r_j + 2\rho \Sigma_{j<j'}^{n} c_j c_{j'} / \sqrt{r_j r_{j'}}}} - z_\alpha, \qquad (19)$$

and so, the number of subjects necessary in the first group at the first timepoint (N_{11}) is equal to

$$N_{11} = \frac{(N_{-1} + 1)(z_\alpha + z_\beta)^2 \left(\Sigma_{j=1}^{n} c_j^2 / r_j + 2\rho \Sigma_{j<j'}^{n} c_j c_{j'} / \sqrt{r_j r_{j'}} \right)}{(\Sigma_{j=1}^{n} c_j d_j)^2}. \qquad (20)$$

Further simplification is possible for an overall comparison of the groups across time (i.e., a group main effect). In this case, specifying $c_j = 1/\sqrt{n}$, yields

$$z_\beta = \sqrt{\left(\frac{N_{11}}{N_{-1} + 1} \right) \frac{(\Sigma_{j=1}^{n} d_j)^2}{\Sigma_{j=1}^{n} 1 / r_j + 2\rho \Sigma_{j<j'}^{n} 1 / \sqrt{r_j r_{j'}}}} - z_\alpha, \qquad (21)$$

and

$$N_{11} = \frac{(N_{-1} + 1)(z_\alpha + z_\beta)^2 \left(\Sigma_{j=1}^{n} 1 / r_j + 2\rho \Sigma_{j<j'}^{n} 1 / \sqrt{r_j r_{j'}} \right)}{(\Sigma_{j=1}^{n} d_j)^2}. \qquad (22)$$

Similarly, assuming equal retention rates in the two groups, for a stationary first-order autoregressive structure, denoted AR(1), we get:

$$N_{11} = \frac{(N_{-1} + 1)(z_\alpha + z_\beta)^2 \left(\Sigma_{j=1}^{n} c_j^2 / r_j + 2\Sigma_{j<j'}^{n} \rho^{(j-j')} c_j c_{j'} / \sqrt{r_j r_{j'}} \right)}{(\Sigma_{j=1}^{n} c_j d_j)^2}, \qquad (23)$$

which simplifies for a main group effect to

$$N_{11} = \frac{(N_{-1} + 1)(z_\alpha + z_\beta)^2 \left(\Sigma_{j=1}^{n} 1 / r_j + 2\Sigma_{j<j'}^{n} \rho^{(j-j')} / \sqrt{r_j r_{j'}} \right)}{(\Sigma_{j=1}^{n} d_j)^2}. \qquad (24)$$

For the AR(1) structure, one must specify the variance σ^2 (to obtain d_j) that is homogeneous across time, and first-order autocorrelation parameter ρ (where $\rho^{(j-j')}$ indicates the correlation between timepoints j and j').

For the random-effects structure, as noted above, the variants-covariance matrix of the repeated measures is expressed as $\Sigma = Z\Sigma_\theta Z' + \sigma_\epsilon^2\Omega$, and so the design matrix Z and assumed values for the parameter matrices Σ_θ and $\sigma_\epsilon^2\Omega$ need to be specified. Either formula (13) or (14) for the variance of a specific contrast can then be used to calculate the power given sample size (16), or required sample size for a given level of power (17). For example, for a random-intercepts model (without autocorrelated errors) $Z = 1_n$, yielding the compound symmetry structure for Σ with $\sigma_\theta^2 + \sigma_\epsilon^2$ on the diagonal and σ_θ^2 for all off-diagonal elements. In this case, assuming equal retention rates in the two groups, the same results are obtained as given above in formulas (19) and (20) with $\sigma^2 = \sigma_\theta^2 + \sigma_\epsilon^2$ and $\rho = \sigma_\theta^2/(\sigma_\theta^2 + \sigma_\epsilon^2)$.

Examples of Sample Size Determination

To illustrate use of the formulas: suppose that a two-group study with equal sample sizes in the two groups ($N_{\_1} = 1$) is being proposed with three timepoints ($n = 3$), and that the group difference is expected to be .5 standard deviation units at each timepoint ($d_j = .5$ for all j). Cohen (1988) denotes a difference of this size a "medium" effect size. Interest is in determining the number of subjects necessary for power equal to .8 ($z_\beta = .842$) for a two-tailed .05 hypothesis test ($z_\alpha = 1.96$) of an overall group difference ($c_j = 1/\sqrt{3}$ for all three timepoints). Further assume the attrition rate is expected to be 10% at each timepoint after the first, and is assumed to be the same for both groups. ($r_1 = 1$, $r_2 = .9$, $r_3 = .81$).

If all pairwise correlations of the three repeated measures are assumed to be .5 ($\rho = .5$), then formula (22) can be used. Given the assumptions above, note that $(N_{\_1} + 1)(z_\alpha + z_\beta)^2/(\Sigma_{j=1}^n d_j)^2$ equals $2(1.96 + 0.842)^2/1.5^2 = 6.98$, and so

$$N_{11} = 6.98\left[1/1 + 1/.9 + 1/.81 + 2(.5)\left(1/\sqrt{.9} + 1/\sqrt{.81} + 1/\sqrt{.729}\right)\right] = 46.6.$$

For sample size determination, it is reasonable to round up to the nearest integer, thus N_{11} can be set to 47. Alternatively, if the association of the repeated measures across time is represented by an AR(1) process with $\rho = .5$, then applying (24) yields

$$N_{11} = 6.98\left[1/1 + 1/.9 + 1/.81 + 2\left(.5/\sqrt{.9} + .25/\sqrt{.81} + .5/\sqrt{.729}\right)\right] = 42.8 \approx 43.$$

This lower relative sample size for the AR(1) structure reflects the lower correlation between timepoints 1 and 3, relative to the compound symmetry structure (i.e., $r_{13} = .25$ versus .50).

Finally, suppose a previous study analyzed a similar dataset using a random-effects model with subject-varying intercepts and trends across time. In terms of Z, this previous study coded the intercept as 1 and the time variable as −1, 0, and 1 for the three timepoints, respectively. Also, from this previous study, the estimated intercept variance was .4, the slope variance .1, the intercept slope covariance .1, and the error variance .5. With these results, the variance-covariance matrix is estimated as

$$\hat{\Sigma} = \begin{bmatrix} .8 & .3 & .3 \\ .3 & .9 & .5 \\ .3 & .5 & 1.2 \end{bmatrix}, \tag{25}$$

indicating a pattern of increasing variances and covariances across time. Assuming that the variance-covariance structure for the proposed study is the same as this estimated matrix, applying (17) yields $N_{11} = 41.8 \approx 42$. Thus, between 42 to 47 subjects would need to be randomized to each group depending on the variance-covariance structure that is assumed.

Alternatively, suppose that the group means are expected to diverge linearly across time. For example, the expected effect sizes are $d_1 = 0$, $d_2 = 1/3$, and $d_3 = 2/3$. Here, interest focuses on determining the number of subjects necessary for a group by linear time interaction, so the orthonormal linear contrast $\sqrt{2} \times c_j = -1$, 0, and 1 can be applied for the three timepoints, respectively. Assuming all else is as specified above, applying (17) yields N_{11} = 40, 60, and 48 for compound symmetry with $\rho = .5$, AR(1) with $\rho = .5$, and the above random-effects structure, respectively.

General Sample Size Determination

More general sample size determinations are now given for three variance-covariance structures: compound symmetry, a first-order autoregressive structure, and a random-effects structure. For the first two, the degree of correlation among the repeated measures is varied to provide a reasonable range of potential patterns of correlations. For the random-effects structure, the parameter estimates (of Σ_θ and Ω) that are necessary to determine Σ are based on published results from a large psychiatric clinical trial. While the compound symmetry and AR(1) structures are stationary structures in that the correlation of the repeated measurements depends on, at most, the time lag, a non-stationary structure is considered in illustrating application of the random-effects model. For all structures, sample sizes are determined while varying the number of timepoints, rate of attrition, and desired effect size. These sample sizes are given for tests of a constant group effect over time and a group by (linear) time interaction.

Compound Symmetry Structure

Table 1 presents necessary group sample sizes based on formula (20) to achieve .80 power (on a two-tailed .05 test) for a test of an overall group effect across time for various levels of attrition, effect sizes, number of timepoints, and correlation of the repeated measures. Attrition rates between each pair of timepoints are assumed equal between groups and are specified as 0, 0.05, and 0.10. Constant effect sizes (d_i) across time of 0.2, 0.5, and 0.8 standard deviation units

TABLE 1
Required Group Sample Size at the First Timepoint
Power = .80 for a Two-tailed .05 Test
Test of a Constant Group Effect across Time

		CS ρ = .1			CS ρ = .3			CS ρ = .5		
Number of Timepoints	Attrition Rate	small effect	medium effect	large effect	small effect	medium effect	large effect	small effect	medium effect	large effect
4	0.00	128	21	8	187	30	12	246	40	16
4	0.05	138	23	9	202	33	13	266	43	17
4	0.10	151	24	10	220	36	14	289	46	18
6	0.00	99	16	7	164	27	11	229	37	15
6	0.05	112	18	7	187	30	12	261	42	17
6	0.10	130	21	9	216	35	14	301	49	19
8	0.00	84	14	6	153	25	10	221	36	14
8	0.05	101	17	7	183	30	12	266	43	17
8	0.10	124	20	8	224	36	14	325	52	21

Attrition rate refers to rate of attrition between each pair of timepoints
CS level is for the correlational structure of the repeated measure
small effect = a between-groups difference of .2 *SD* units at each timepoint
medium effect = a between-groups difference of .5 *SD* units at each timepoint
large effect = a between-groups difference of .8 *SD* units at each timepoint

were chosen based on the classification of "small," "medium," and "large" effect sizes described by Cohen (1988). Four, six, and eight timepoints are considered and the (homogeneous) correlation among the repeated measures ρ equals .1, .3, and .5. While correlation as low as .1 is unlikely in repeated measures designs, it can be considered as an upper bound for other types of clustered data (e.g., nesting of students within classrooms) that are often analyzed using the same general statistical methods as repeated measures problems.

In Table 1, without attrition, the necessary sample size decreases as the effect size is increased, the number of timepoints is increased, and the (assumed homogeneous) correlation among the repeated measures is decreased. When attrition is present, however, increasing the number of timepoints does not necessarily decrease the required sample size, especially as the correlation among repeated measures increases (e.g., when the correlation equals 0.5 and the attrition rate equals 0.10). In general, for a given correlation level, as attrition is increased a more positive relationship develops between number of timepoints and required sample size. That is, if the relationship is negative without attrition, so that adding timepoints reduces the necessary sample size (e.g., ρ = .1 and ρ .3), increasing attrition level makes this negative relationship less pronounced or disappear. Alternatively, if the relationship is less pronounced when there is no attrition, so that adding timepoints has little effect on the required sample size in the absence of attrition (e.g., ρ = .5), then increasing attrition level produces a positive relationship so that more subjects are required as timepoints are increased.

For the case of a linear group difference across time (i.e., a group by linear time interaction), Table 2 presents group sample sizes based on formula (20) to achieve .80 power (on a two-tailed .05 test) for similar levels of attrition, number of timepoints, and correlation of the repeated measures, as was presented in Table 1. Effect sizes were assumed equal to 0 at the first timepoint and to increase linearly to levels of 0.2, 0.5, and 0.8 standard deviation (*SD*) units at the last timepoint. For example, in order to achieve a 0.5 *SD* unit difference over six timepoints the group differences would equal, 0, .1, .2, .3, .4, and .5 *SD* units. To achieve this same difference over eight timepoints, group differences would equal 0, .0714, .1429, .2143, .2857, .3571, .4286, and .5 sd units. Fixing the effect size to 0 at the first timepoint and 0.5 *SD* units at the last timepoint allows comparison of sample size requirements when additional measurements are made between these two timepoints. That is, with four, six, and eight timepoints, the results in Table 2 contrast the change in sample size requirement by adding two to four to six intermittent measurements between the first and last timepoints.

TABLE 2
Required Group Sample Size at the First Timepoint
Power = .80 for a Two-tailed .05 Test
Test of a Between Groups Linear Trend Effect

		CS $\rho = .1$			CS $\rho = .3$			CS $\rho = .5$		
Number of Timepoints	Attrition Rate	small effect	medium effect	large effect	small effect	medium effect	large effect	small effect	medium effect	large effect
4	0.00	636	102	40	495	80	31	354	57	23
4	0.05	689	111	44	537	86	34	384	62	24
4	0.10	755	121	48	590	95	37	425	68	27
6	0.00	505	81	32	393	63	25	281	45	18
6	0.05	579	93	37	452	73	29	325	52	21
6	0.10	679	109	43	536	86	34	393	63	25
8	0.00	413	66	26	321	52	20	229	37	14
8	0.05	501	81	32	393	63	25	285	46	18
8	0.10	635	102	40	511	82	32	388	62	25

Attrition rate refers to rate of attrition between each pair of timepoints
CS level is for the correlational structure of the repeated measure
small effect = a between-groups difference increasing linearly from 0 (first time) to .2 *SD* units (last time)
medium effect = a between-groups difference increasing linearly from 0 (first time) to .5 *SD* units (last time)
large effect = a between-groups difference increasing linearly from 0 (first time) to .8 *SD* units (last time)

To test a between-group linear trend effect (a group by linear time interaction) in the absence of attrition, as Table 2 illustrates, the necessary sample size decreases as the effect size increases, the number of intermittent timepoints increases, and the correlation among the repeated measures increases. Thus, in contrast to the test of a constant group effect, for the group by time interaction the greater the correlation among the measures, the better. With attrition the same general conclusions hold, however the effect of increasing intermittent timepoints on decreasing the necessary sample size is not as pronounced. For example, when the correlation between all repeated measures is assumed to be 0.5 and the attrition rate between timepoints is 0.10, increasing the number of timepoints from six to eight leads to a minor decrease in the number of necessary subjects per group.

First-Order Autoregressive Structure

The conditions for Tables 3 and 4 are identical to Tables 1 and 2 with the exception of the assumed form of the variance-covariance structure of the repeated measures. In Tables 3 and 4, the assumed form of the variance covariance for the repeated measures was a first-order autoregressive (AR1) structure with ρ set equal to .3, .5, and .7. Group sample sizes were based on formula (23) to achieve .80 power for a two-tailed .05 test.

TABLE 3
Required Group Sample Size at the First Timepoint
Power = .80 for a Two-Tailed .05 Test
Test of a Constant Group Effect across Time

		AR1 $\rho = .3$			AR1 $\rho = .5$			AR1 $\rho = .7$		
Number of Timepoints	Attrition Rate	small effect	medium effect	large effect	small effect	medium effect	large effect	small effect	medium effect	large effect
4	0.00	153	25	10	203	33	13	267	43	17
4	0.05	165	27	11	219	35	14	288	46	18
4	0.10	180	29	12	239	39	15	313	51	20
6	0.00	109	18	7	154	25	10	222	36	14
6	0.05	124	20	8	175	28	11	252	41	16
6	0.10	143	23	9	202	33	13	291	47	19
8	0.00	84	14	6	123	20	8	189	31	12
8	0.05	101	17	7	148	24	10	227	37	15
8	0.10	124	20	8	182	30	12	278	45	18

Attrition rate refers to rate of attrition between each pair of timepoints
AR1 level is for the correlational structure of the repeated measure
small effect = a between-groups difference of .2 *SD* units at each timepoint
medium effect = a between-groups difference of .5 *SD* units at each timepoint
large effect = a between-groups difference of .8 *SD* units at each timepoint

TABLE 4
Required Group Sample Size at the First Timepoint
Power = .80 for a Two-tailed .05 Test
Test of a Between Groups Linear Trend Effect

		AR1 ρ = .3			AR1 ρ = .5			AR1 ρ = .7		
Number of Timepoints	Attrition Rate	small effect	medium effect	large effect	small effect	medium effect	large effect	small effect	medium effect	large effect
4	0.00	758	122	48	698	112	44	528	85	33
4	0.05	821	132	52	756	121	48	573	92	36
4	0.10	898	144	57	828	133	52	630	101	40
6	0.00	722	116	46	777	125	49	698	112	44
6	0.05	826	133	52	889	143	56	800	128	50
6	0.10	965	155	61	1039	167	65	940	151	59
8	0.00	649	104	41	769	124	49	787	126	50
8	0.05	785	126	49	931	149	59	954	153	60
8	0.10	983	158	62	1167	187	73	1201	193	75

Attrition rate refers to rate of attrition between each pair of timepoints
AR1 level is for the correlational structure of the repeated measure
small effect = a between groups difference increasing linearly from 0 (first time) to .2 *SD* units (last time)
medium effect = a between groups difference increasing linearly from 0 (first time) to .5 *SD* units (last time)
large effect = a between groups difference increasing linearly from 0 (first time) to .8 *SD* units (last time)

When there is no attrition and the test is for an overall group difference (Table 3), the necessary sample size decreases as the effect size is increased, the number of timepoints is increased, and the correlation among the repeated measures is decreased. As attrition levels increase, there is a less pronounced decrease in the required sample size with increasing number of timepoints, especially as the correlation of repeated measurements is increased. This agrees with the general pattern of results obtained from the compound symmetry structure given in Table 1.

For the test of a group by linear time interaction, results for the CS and AR(1) structures are not the same, even in the absence of attrition. With an AR(1) structure, while effect size increases always lead to fewer required subjects, results pertaining to the level of the AR(1) term and the number of intermittent timepoints are not so clear-cut. Focusing (in Table 4) on the required sample sizes when attrition is not present, increasing the number of intermittent timepoints reduces the required number of subjects when the AR(1) term is small (.3), but increases the required number of subjects when the AR(1) term is large (.7). With increasing attrition level, again, the relationship between sample size and number of timepoints either becomes positive (ρ = .3) or becomes more positive (ρ = .5 and ρ =.7).

Random-Effects (RE) Structure

Gibbons et al., (1993) used a random-effects analysis to analyze data from the National Institute of Mental Health (NIMH) Treatment of Depression Collaborative Research Project (TDCRP), a longitudinal study which examined the relative effectiveness of cognitive behavior therapy, interpersonal psychotherapy, imipramine with clinical maintenance, and placebo with clinical maintenance in the treatment of outpatient depression. While 250 subjects were randomized to one of four treatment groups, 239 entered treatment and were measured at baseline (week 0) and monthly thereafter for four months (weeks 4, 8, 12, and 16), though not all subjects were measured at all timepoints. Outcome was measured on the commonly-used Hamilton Rating Scale for Depression (HRSD).

In their analysis of these data, two random-effects structures were considered: the first (RE structure I) included a random (e.g., a person-varying) linear trend across time, while the second (RE structure II) augmented the random linear trend with a term for autocorrelated residuals, specifically allowing residuals to follow a nonstationary first-order autoregressive [$AR(1)$] process as described by Mansour, Nordheim, and Rutledge (1985). Since analysis of these data did not reveal person-specific deviations in severity at baseline, random intercepts were not considered in either model. To achieve approximate linearity in trend across time, a log transformation on time (ln [weeks + 1]) was used. Thus, the design matrix (a vector since there is only one random effect) for the random effects was $\mathbf{Z}' = \ln[1\ \ 5\ \ 9\ \ 13\ \ 17] = [0\ \ 1.609\ \ 2.197\ \ 2.565\ \ 2.833]$. Estimates from the first analysis of these data (RE structure I) yielded $\hat{\sigma}^2_\theta = 4.69138$ for random linear slope and $\hat{\sigma}^2 = 18.39606$ for residual variance. For the model adding autocorrelated residuals (RE structure II), estimates were $\hat{\sigma}^2{}_\theta = 3.14408$ for random linear slope, $\hat{\sigma}^2 = 20.82148$ for residual variance, and $\hat{\rho} = 0.3637$ for the nonstationary AR(1) parameter; RE structure II significantly improved the fit as compared to structure I by the likelihood-ratio χ^2 test.

Table 5 lists the necessary group sample sizes based on formula (17) to achieve .80 power (on a two-tailed .05 test) for a test of an overall group effect across time for various levels of attrition (0, 5%, and 10% between all timepoints), effect sizes (small, medium, and large), number of timepoints (four, six, and eight), and for variance-covariance form given by (2) using estimates from RE structures I and II. To use (2) in determining Σ, values of the design matrix Z must be specified. To achieve compatibility with the scale of the parameter estimates, the eight potential timepoints were specified as weeks 0 through 28 in increments of four weeks, and then transformed using logs (ln[weeks + 1]). Since in the previous study the maximum value of time equaled 16 weeks (the fifth timepoint), sample size calculations for six and eight timepoints represent an extrapolation based on parameter estimates obtained in this previous study. Also, given the estimates reported for σ^2, σ^2_θ and ρ, correlations among the repeated measures increase with time. This can be seen clearly by using (2) to obtain the correlations of the repeated measures based on RE structure I:

week 0	1.000							
week 4	0.000	1.000						
week 8	0.000	0.468	1.000					
week 12	0.000	0.499	0.588	1.000				
week 16	0.000	0.517	0.609	0.649	1.000			
week 20	0.000	0.529	0.623	0.664	0.687	1.000		
week 24	0.000	0.537	0.633	0.674	0.698	0.714	1.000	
week 28	0.000	0.544	0.640	0.682	0.706	0.723	0.734	1.000

and based on RE structure II:

week 0	1.000							
week 4	0.305	1.000						
week 8	0.101	0.523	1.000					
week 12	0.035	0.376	0.589	1.000				
week 16	0.012	0.334	0.459	0.629	1.000			
week 20	0.004	0.325	0.421	0.508	0.656	1.000		
week 24	0.002	0.327	0.414	0.472	0.542	0.676	1.000	
week 28	0.001	0.331	0.417	0.465	0.508	0.568	0.692	1.000

This increasing correlation structure across time is an example of a non-stationary structure, and is in contrast to the stationary compound symmetry and AR(1) structures that assume equal correlations across time and time differences, respectively. Thus, Tables 5 and 6 illustrate the change in necessary sample size when increasingly correlated repeated measures are added.

As Table 5 reveals, while increasing the effect size reduces the necessary number of subjects, the result of increasing the number of timepoints on the required sample size is not always consistent between RE structures I and II. For structure I, increasing the number of timepoints leads to an increase in the required number of subjects, and this effect is more pronounced as attrition increases. Alternatively, for structure II, only when attrition is present does increasing the number of timepoints lead to an increase in the required number of subjects. When there is no attrition and structure II is assumed, increasing the number of timepoints from four to six reduces the number of required subjects, while minimal change is observed between six and eight timepoints. For both structures though, as attrition is increased, adding increasingly correlated measurements across time leads to greater sample size requirements for the overall group effect.

For the case of a linear group difference across time (i.e., a group by linear time interaction), Table 6 presents group sample sizes based on formula (17) to

TABLE 5
Required Group Sample Size at the First Timepoint
Power = .80 for a Two-tailed .05 Test
Test of a Constant Group Effect across Time

		RE structure I			RE structure II		
Number of Timepoints	Attrition Rate	small effect	medium effect	large effect	small effect	medium effect	large effect
4	0.00	192	31	12	200	32	13
4	0.05	213	34	14	219	35	14
4	0.10	237	38	15	242	39	16
6	0.00	215	35	14	190	31	12
6	0.05	253	41	16	222	36	14
6	0.10	303	49	19	263	42	17
8	0.00	233	38	15	191	31	12
8	0.05	292	47	19	238	38	15
8	0.10	373	60	24	301	48	19

Attrition rate refers to rate of attrition between each pair of timepoints
RE structure I refers to a random-effects structure with random slope and residual term (see text)
RE structure II refers to a random-effects structure with random slope, residual term, and autocorrelated residuals (see text)
small effect = a between-groups difference of .2 *SD* units at each timepoint
medium effect = a between-groups difference of .5 *SD* units at each timepoint
large effect = a between-groups difference of .8 *SD* units at each timepoint

achieve .80 power (on a two-tailed .05 test) for similar levels of attrition, number of timepoints, and variance-covariance structures of the repeated measures, as was presented in Table 5. Effect sizes were assumed to be equal to 0 at the first timepoint and to increase linearly to levels of 0.2, 0.5, and 0.8 standard deviation (*SD*) units at the last timepoint.

From Table 6, it is clear that when attrition is not present, increasing timepoints and effect size lowers the required number of subjects to detect the group by linear time interaction. However, as attrition level is increased, the negative relationship between number of timepoints and required sample size generally either becomes less pronounced or disappears (RE structure I) or becomes positive (RE structure II). Interestingly, the required number of subjects is quite a bit larger for RE structure II compared to I, indicating the effect on sample size determination of misspecification of the variance-covariance structure.

Discussion

Formulas for estimating sample size and power levels were presented for significance tests based on longitudinal designs with the possibility of attrition. We focused on the case of a two-group contrast of group population means across study timepoints. Single degree-of-freedom contrasts of this type can

TABLE 6
Required Group Sample Size at the First Timepoint
Power = .80 for a Two-tailed .05 Test
Test of a Between Groups Linear Trend Effect

		RE structure I			RE structure II		
Number of Timepoints	Attrition Rate	small effect	medium effect	large effect	small effect	medium effect	large effect
4	0.00	491	79	31	570	92	36
4	0.05	553	89	35	632	102	40
4	0.10	629	101	40	710	114	45
6	0.00	385	62	24	517	83	33
6	0.05	481	77	31	622	100	39
6	0.10	614	99	39	766	123	48
8	0.00	321	52	20	454	73	29
8	0.05	453	73	29	599	96	38
8	0.10	657	106	41	817	131	52

Attrition rate refers to rate of attrition between each pair of timepoints
RE structure I refers to a random-effects structure with random slope and residual term (see text)
RE structure II refers to a random-effects structure with random slope, residual term, and autocorrelated residuals (see text)
small effect = a between groups difference increasing linearly from 0 (first time) to .2 *SD* units (last time)
medium effect = a between groups difference increasing linearly from 0 (first time) to .5 *SD* units (last time)
large effect = a between groups difference increasing linearly from 0 (first time) to .8 *SD* units (last time)

often capture the main and interaction effects in a two-group repeated measures design. Assuming that missingness does not alter the model terms considered in the power calculations, equations (16) and (17) can be used to perform study-specific power computations for specific comparisons. Simplified formulas for testing an overall group effect were also provided and their usage illustrated.

More generally, the formulas were used to yield sample size requirements for an overall group effect over time and a group by (linear) time interaction under variance-covariance structures corresponding to a compound symmetry structure, a first-order autoregressive structure, and a non-stationary structure based on random-effects modeling of the repeated measures. As opposed to the stationary compound symmetry and first-order autoregressive structures, the assumed variance-covariance structure based on the random-effects modeling allowed for increasing correlations across time. Tables 5 and 6 present these results and might be particularly useful to behavioral researchers since they are based on a typical psychiatric longitudinal clinical trial using a common outcome measure of clinical efficacy. Tables 1-4 provide more generic results assuming compound symmetry and autocorrelation and should be of use in situations where the assumption of stationarity is reasonable. As was seen, the required sample size varied considerably depending on test (overall group effect or group by time interaction), variance-covariance structure of the repeated measures, number of timepoints, and effect size.

The effect of attrition was seen most clearly in terms of its influence on the relationship between number of timepoints and required sample size. In all cases, increasing the level of attrition made this relationship between number of timepoints and required sample size more positive. In general, for a given test, if increasing the number of timepoints lowered the sample size requirement, then increasing attrition level produced a less pronounced or marginal negative relationship between these two factors. Alternatively, for a given test, if adding timepoints had little effect or increased the required sample size, then increasing attrition level resulted in a more positive relationship so that even more subjects were required with increasing number of timepoints. Thus, ignoring attrition level in sample size determination for a longitudinal study is clearly risky. Hopefully, use of the formulas and results in this article will help researchers avoid that risk.

The formulas presented here are model specific. However, we present results for a wide variety of models (i.e., variance-covariance structures). Thus, they can be used to assess sample size requirements for many different models. Two recommendations of Muller et al., (1992) are relevant in this context. The first is to align the models for power calculation and planned data analysis as closely as possible. This avoids making what Kimball (1957) describes as a Type III error, that is, getting the right answer to the wrong problem. Secondarily, a sensitivity analysis should be conducted to assess the degree to which sample size requirements vary with key model assumptions. For this, the formulas presented here can be used to assess sample size requirements under a variety of variance-covariance structures and attrition patterns. The scope of the sensitivity analysis may depend on the degree of uncertainty regarding the anticipated variance-covariance structure and/or attrition.

Though the formulas allow for attrition, it is assumed that missingness does not alter the model terms considered in the power calculations. Specifically, the proposed treatment differences across time ($\mu_{1j} - \mu_{2j}$) and variance-covariance structure (Σ) are assumed to be the same for subjects who complete the study as well as those who dropout. Note that it is the mean differences across time and not the actual means that are assumed independent of missingness. In longitudinal studies, this type of ignorable nonresponse falls under Rubin's (1976) "missing at random" (MAR) assumption. As pointed out by Shih (1992), an important ancillary condition to MAR is the distinct parameters condition, namely, that the parameters of the missingness are distinct from the parameters of interest. This distinct parameters assumption is critical in application of the formulas presented in this article, since it is assumed that the missingness is independent of the proposed treatment differences and variance-covariance structure (i.e., the parameters of interest).

Notes

The authors thank Siu Chi Wong for computer programming assistance, and the associate editor and an anonymous reviewer for several helpful and constructive comments. Preparation of this article was supported by National Institutes of Mental Health grants R01-MH44826 and R01-MH56146.

[1]A FORTRAN program that performs the sample size and/or power calculations corresponding to the methods presented in this paper can be obtained at internet location http://www.uic.edu/~hedeker/mix.html.

Appendix

Following Jennrich and Schluchter (1986) and others, maximum likelihood estimates of β for the model specified in (1) are obtained by

$$\hat{\beta} = \left(\sum_{i=1}^{N} X_i' \Sigma_i^{-1} X_i \right)^{-1} \left(\sum_{i=1}^{N} X_i' \Sigma_i^{-1} y_i \right).$$

For the two-group problem discussed in this article, $\boldsymbol{X}_i$ is partitioned into two $n_i \times n$ component matrices. The first component matrix $\boldsymbol{W}_i$ includes a constant term as the first column followed by $n - 1$ contrasts for the within-subjects design. The second component matrix $\boldsymbol{B}_i$ represents the between-subjects design and is obtained by multiplying $\boldsymbol{W}_i$ by -1 (if subject i belongs to the first group)_ or 1 (if subject i belongs to the second group). The parameter vector β is similarly partitioned into β_W and β_B. Thus, $\boldsymbol{X}_i = (\boldsymbol{W}_i | \boldsymbol{W}_i)$ for subjects belonging to the first group, and $\boldsymbol{X}_i = (\boldsymbol{W}_i | -\boldsymbol{W}_i)$ for subjects of the second group. If it is assumed that $n_i = n$ for all subjects, and that N_1 and N_2 are the numbers of subjects in the two groups, then

$$\hat{\beta} = \left(N_1 \begin{bmatrix} \boldsymbol{W}' \\ \boldsymbol{W}' \end{bmatrix} \Sigma^{-1} [\boldsymbol{W} \mid \boldsymbol{W}] + N_2 \begin{bmatrix} \boldsymbol{W}' \\ -\boldsymbol{W}' \end{bmatrix} \Sigma^{-1} [\boldsymbol{W} \mid -\boldsymbol{W}] \right)^{-1}$$

$$\left(N_1 \begin{bmatrix} \boldsymbol{W}' \\ \boldsymbol{W}' \end{bmatrix} \Sigma^{-1} \bar{y}_1 + N_2 \begin{bmatrix} \boldsymbol{W}' \\ -\boldsymbol{W}' \end{bmatrix} \Sigma^{-1} \bar{y}_2 \right)$$

$$= \left(N_1 \begin{bmatrix} \boldsymbol{W}'\Sigma^{-1}\boldsymbol{W} & \boldsymbol{W}'\Sigma^{-1}\boldsymbol{W} \\ \boldsymbol{W}'\Sigma^{-1}\boldsymbol{W} & \boldsymbol{W}'\Sigma^{-1}\boldsymbol{W} \end{bmatrix} + N_2 \begin{bmatrix} \boldsymbol{W}'\Sigma^{-1}\boldsymbol{W} & -\boldsymbol{W}'\Sigma^{-1}\boldsymbol{W} \\ -\boldsymbol{W}'\Sigma^{-1}\boldsymbol{W} & \boldsymbol{W}'\Sigma^{-1}\boldsymbol{W} \end{bmatrix} \right)^{-1}$$

$$\left(N_1 \begin{bmatrix} \boldsymbol{W}'\Sigma^{-1}\bar{y}_1 \\ \boldsymbol{W}'\Sigma^{-1}\bar{y}_1 \end{bmatrix} + N_2 \begin{bmatrix} \boldsymbol{W}'\Sigma^{-1} y_2 \\ -\boldsymbol{W}'\Sigma^{-1}\bar{y}_2 \end{bmatrix} \right)$$

$$= \begin{pmatrix} (N_1 + N_2)\, \boldsymbol{W}'\Sigma^{-1}\boldsymbol{W} & (N_1 - N_2)\boldsymbol{W}'\Sigma^{-1}\boldsymbol{W} \\ (N_1 - N_2)\, \boldsymbol{W}'\Sigma^{-1}\boldsymbol{W} & (N_1 + N_2)\boldsymbol{W}'\Sigma^{-1}\boldsymbol{W} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{W}'\Sigma^{-1}(N_1\bar{y}_1 + N_2\bar{y}_2) \\ \boldsymbol{W}'\Sigma^{-1}(N_1\bar{y}_1 - N_2\bar{y}_2) \end{pmatrix}$$

where $\bar{y}_1$ is the $n \times 1$ vector of means for group 1, and $\bar{y}_2$ is the $n \times 1$ vector of means for group 2. Denoting $k = (N_2 - N_1)/(N_1 + N_2)$ and using the formula for the inverse of a partitioned matrix (see Bock, 1975, page 38), we get for β

$$\begin{bmatrix} \hat{\beta}_W \\ \hat{\beta}_B \end{bmatrix} = \frac{N_1 + N_2}{4N_1N_2} \begin{bmatrix} (W'\Sigma^{-1}W)^{-1} & k(W'\Sigma^{-1}W)^{-1} \\ k(W'\Sigma^{-1}W)^{-1} & (W'\Sigma^{-1}W)^{-1} \end{bmatrix} \begin{bmatrix} W'\Sigma^{-1}(N_1\bar{y}_1 + N_2\bar{y}_2) \\ W'\Sigma^{-1}(N_1\bar{y}_1 - N_2\bar{y}_2) \end{bmatrix}$$

Notice that if the design is completely balanced so that $N_1 = N_2$, then $k = 0$ and the first matrix on the right side of the equality is quasidiagonal, that is, diagonal by blocks. If the compound symmetry structure is assumed for Σ and if W is an orthonormal matrix, then the diagonal blocks are diagonal matrices (see Bock, 1975). In this case, the maximum likelihood estimates of β are invariant with respect to the addition or deletion of model terms in X. In the slightly more general case of $N_1 \neq N_2$, if compound symmetry is assumed, then the first matrix on the right side of the equality is diagonal within blocks. The maximum likelihood estimates within β_W and β_B are then invariant with respect to addition or deletion of terms within W and B, respectively. Relaxing the compound symmetry assumption, under the full model, the maximum likelihood estimates are invariant to the structure of Σ. Since, in this case, W is a $n \times n$ square matrix, and so

$$\begin{bmatrix} \hat{\beta}_W \\ \hat{\beta}_B \end{bmatrix} = \frac{N_1 + N_2}{4N_1N_2} \begin{bmatrix} W^{-1}(W'\Sigma^{-1})^{-1} & kW^{-1}(W'\Sigma^{-1})^{-1} \\ kW^{-1}(W'\Sigma^{-1})^{-1} & W^{-1}(W'\Sigma^{-1})^{-1} \end{bmatrix} \begin{bmatrix} W'\Sigma^{-1}(N_1\bar{y}_1 + N_2\bar{y}_2) \\ W'\Sigma^{-1}(N_1\bar{y}_1 - N_2\bar{y}_2) \end{bmatrix}$$

$$= \frac{N_1 + N_2}{4N_1N_2} \begin{bmatrix} W^{-1}(N_1\bar{y}_1 + N_2\bar{y}_2) + kW^{-1}(N_1\bar{y}_1 - N_2\bar{y}_2) \\ kW^{-1}(N_1\bar{y}_1 + N_2\bar{y}_2) + W^{-1}(N_1\bar{y}_1 - N_2\bar{y}_2) \end{bmatrix}$$

which does not depend on Σ.

References

Bock, R. D. (1975). *Multivariate statistical methods in behavioral research.* New York: McGraw-Hill.

Bock, R. D. (1983). Within-subject experimentation in psychiatric research. In R. D. Gibbons & M. W. Dysken (Eds.), *Statistical and methodological advances in psychiatric research* (pp. 59–90). New York: Spectrum.

Bock, R. D. (1989). Measurement of human variation: A two stage model. In R. D. Bock (Ed.), *Multilevel analysis of educational data* (pp. 319–342). San Diego: Academic Press.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical Linear Models: Applications and Data Analysis Methods.* Newbury Park, CA: Sage Publications.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences, 2nd edition.* Hillsdale, NJ: Erlbaum.

Cooper, B. E. (1971). The use of orthogonal polynomials with equal *x*-values. *Applied Statistics, 20,* 209–213.

Emerson, P. L. (1968). Numerical construction of orthogonal polynomials from a general recurrence formula. *Biometrics, 24,* 695–701.

Fleiss, J. L. (1986). *The design and analysis of clinical experiments.* New York: Wiley.

Gibbons, R. D., Hedeker, D., Elkin, I., Watemaux, C., Kraemer, H. C., Greenhouse, J. B., Shea, M. T., Imber, S. D., Sotsky, S. M., & Watkins, J. T. (1993). Some conceptual and statistical issues in analysis of longitudinal psychiatric data. *Archives of General Psychiatry, 50,* 739–750.

Goldstein, H. (1995). *Multilevel statistical models* (2nd edition). New York: Halstead Press.

Jennrich, R. I., & Schluchter, M. D. (1986). Unbalanced repeated-measures models with structured covariance matrices. *Biometrics, 42,* 805–820.

Kimball, A. W. (1957). Errors of the third kind in statistical consulting. *Journal of the American Statistical Association, 52,* 133–142.

Kirby, A. J., Galai, N., & Muñoz, A. (1994). Sample size estimation using repeated measurements on biomarkers as outcomes. *Controlled Clinical Trials, 15,* 165–172.

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

Longford, N.T. (1993). *Random coefficient models.* New York: Oxford University Press.

Mansour, H., Nordheim, E. V., & Rutledge, J. J. (1985). Maximum likelihood estimation of variance components in repeated measures designs assuming autoregressive errors. *Biometrics, 41,* 287–294.

Muller, K. E., LaVange, L. M., Ramey, S. L., & Ramey, C. T. (1992). Power calculations for general linear multivariate models including repeated measures applications. *Journal of the American Statistical Association, 87,* 1209–1226.

Muller, K. E., & Barton, C. N. (1989). Approximate power for repeated measures ANOVA lacking sphericity. *Journal of the American Statistical Association, 84,* 549–555.

Muller, K. E., & Peterson, B. L. (1984). Practical methods for computing power in testing the multivariate general linear hypothesis. *Computational Statistics and Data Analysis, 2,* 143–158.

O'Brien, R. G., & Muller, K. E. (1992). Unified power analysis for *t*-tests through multivariate hypotheses. In L. K. Edwards (Ed.), *Applied Analysis of Variance in Behavioral Sciences,* pp. 297–344. New York: Marcel Dekker.

Overall, J. E., & Doyle, S. R. (1994). Estimating sample sizes for repeated measurement designs. *Controlled Clinical Trials, 15,* 100–123.

Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63,* 581–592.

Shih, W. J. (1992). On informative and random dropouts in longitudinal studies. *Biometrics, 48,* 970–972.

Snijders, T. A. B., & Bosker, R. J. (1993). Standard errors and sample sizes for two-level research. *Journal of Educational Statistics, 18,* 237–259.

Authors

DONALD HEDEKER is Associate Professor, Division of Epidemiology & Biostatistics and Health Research and Policy Centers, School of Public Health, University of Illinois at Chicago, 2121 West Taylor Street, Room 510, Chicago, IL, 60612-7260; hedeker@uic.edu. He specializes in longitudinal data analysis and mixed-effects models.

ROBERT D. GIBBONS is Professor, Department of Psychiatry and Division of Epidemiology and Biostatistics, University of Illinois at Chicago, 912 South Wood Street, Chicago, IL, 60612; rdgib@uic.edu. He specializes in analysis of longitudinal data with particular emphasis on problems in mental health research.

CHRISTINE WATERNAUX is Associate Professor of Clinical Biostatistics and Psychiatry, Division of Biostatistics, Columbia University, Box 48, New York State Psychiatric Institute, 722 West 168th Street, New York, NY, 10032; cwatern@neuron.cpmc.columbia.edu. She specializes in variance component models for longitudinal data and meta analysis.

Received June 19, 1995
Revision received January 15, 1997
Accepted November 25, 1997