Sample sizes required to detect interactions between two binary fixed-effects in a mixed-effects linear regression model

Andrew C. Leon [a,b,*], Moonseong Heo [a]

[a] Department of Psychiatry, Weill Medical College of Cornell University, United States
[b] Department of Public Health, Weill Medical College of Cornell University, United States

ARTICLE INFO

ABSTRACT

Mixed-effects linear regression models have become more widely used for analysis of repeatedly measured outcomes in clinical trials over the past decade. There are formulae and tables for estimating sample sizes required to detect the main effects of treatment and the treatment by time interactions for those models. A formula is proposed to estimate the sample size required to detect an interaction between two binary variables in a factorial design with repeated measures of a continuous outcome. The formula is based, in part, on the fact that the variance of an interaction is fourfold that of the main effect. A simulation study examines the statistical power associated with the resulting sample sizes in a mixed-effects linear regression model with a random intercept. The simulation varies the magnitude (Δ) of the standardized main effects and interactions, the intraclass correlation coefficient (ρ), and the number (k) of repeated measures within-subject. The results of the simulation study verify that the sample size required to detect a 2×2 interaction in a mixed-effects linear regression model is fourfold that to detect a main effect of the same magnitude.

1. Introduction

The mixed-effects linear regression model (Harville, 1977; Laird and Ware, 1982) is widely used in observational studies and randomized controlled clinical trials (RCT) in which there are repeated measures over time. In designing a study, the Ethical Guidelines of the American Statistical Association (1999) advise statisticians to provide informed recommendations for sample size such that a research protocol will neither propose an inadequate nor an excessive number of subjects to detect a scientifically noteworthy result with acceptable statistical power. Several authors have examined the sample sizes required to detect the main effects and interaction of treatment and time in longitudinal studies with repeated measures (e.g., Hsieh (1988), Rochon (1991), Overall and Doyle (1994), Hedeker et al. (1999), Raudenbush and Liu (2001) and Diggle et al. (2002)). Yet a study that is designed to detect the main effect of treatment will not have sufficient power to detect the interaction between two binary fixed effects. In a 2×2 factorial fixed-effects ANOVA with equal cell sizes and an assumption of independence among observations, for instance, the sample size required to detect an interaction is four times that for a main effect of the same magnitude (Fleiss, 1986). However, we are not aware of formulae to estimate the sample size needed to detect an interaction between two binary fixed effects in a mixed-effects linear regression model for analysis of repeatedly measured correlated data.

---

* Corresponding address: Department of Psychiatry, Weill Medical College of Cornell University, Box 140, 525 East 68th Street, New York, NY 10065, United States. Tel.: +1 212 746 3872; fax: +1 212 746 8754.
   E-mail address: acleon@med.cornell.edu (A.C. Leon).

The objective of this manuscript is to examine the sample size required to detect a 2×2 interaction of two binary fixed effects in mixed-effects linear regression analyses. The model, described in detail in Section 2, also incorporates a time-varying covariate, but that covariate does not interact with group membership. We sought to determine if, as with the fixed-effects factorial ANOVA, the sample size needed to detect an interaction in a repeated measures design is fourfold that of a main effect. A formula for the sample size required to detect an interaction is presented below. A simulation study then examines the statistical power of the resulting sample sizes to detect interactions of various magnitudes in a 2×2 factorial design with repeated measures of a continuous outcome.

2. Mixed-effects linear regression model and sample size determination

A mixed-effects linear regression model of repeated measures of a continuous dependent variable, y_{ij}, is specified as:

$$y_{ij} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i} x_{2i} + \beta_4 t_j + \upsilon_i + \varepsilon_{ij} \tag{1}$$

for subject i ($i = 1, \ldots, N$), at time j ($j = 1, \ldots, k$), where β_0 is the intercept term, x_1, represents the treatment contrast ($x_1 = -1/2$ if placebo; $x_1 = 1/2$ if investigational treatment), x_2 represents the moderator contrast ($x_2 = -1/2$ if effect moderator is absent; $x_2 = 1/2$ if effect moderator is present), $x_1 x_2$ represents the treatment by moderator interaction. As defined by Kraemer et al. (2002), "... moderators identify on whom and under what circumstances treatments have different effects". Randomization to treatment assignment is stratified by the moderator. Note that N is the *total sample size*. Therefore $N/2$ subjects are randomized to each treatment and the sample size per cell is $N/4$ for the balanced design with two binary factors, which we consider here. The coefficients, β_1 to β_3, represent the magnitude of the corresponding main effects and interaction, t_j represents the time point of the j-th assessment and its coefficient β_4 represents the slope over time. This model assumes parallel slopes across treatment groups and that the slopes do not vary as a function of the moderator. These assumptions could be relaxed if either a treatment by time interaction or a treatment by moderator by time interaction were included in the model. However, here we have chosen to focus on the treatment by moderator interaction. Therefore, model (1) is an extension of the factorial fixed-effects ANOVA model, and can be described as a 2×2 factorial random intercept ANCOVA model with t_j as a time-varying covariate.

The subject-specific random intercept υ_i is assumed to be distributed $N(0, \sigma_\upsilon^2)$, and the conditional distribution of error term ε_{ij} for a given υ_i is assumed to be independent and identical with $N(0, \sigma_\varepsilon^2)$ across time points j within the i-th subject. The marginal distributions of υ_i and ε_{ij} are assumed to be mutually independent, that is $\text{Cov}(\varepsilon, \upsilon) = 0$. It follows from those conditional and mutual independence assumptions that $\text{Var}(Y_{ij}) \equiv \sigma^2 = \sigma_\upsilon^2 + \sigma_\varepsilon^2$ and $\text{corr}(Y_{ij}, Y_{ij'}) \equiv \rho = \sigma_\upsilon^2 / \sigma^2 = \sigma_\upsilon^2 / (\sigma_\upsilon^2 + \sigma_\varepsilon^2)$, the intraclass correlation coefficient (ICC), for $j \neq j'$. The standardized effects of β_1 to β_3 can be quantified as $\Delta_m = \beta_m / \sigma$, $m = 1, 2, 3$.

The variance of the estimated interaction is four times that of estimated main effect in the factorial fixed-effects ANOVA (Section 4.2 in Fleiss (1986)). That relation also holds for the 2×2 factorial random intercept ANCOVA model (1) that we are considering here, since neither $\text{Var}(Y_{ij}) \equiv \sigma^2$ nor the correlation, ρ, depends on subject i or time point j. Specifically, the following holds:

$$\text{Var}(\hat{\beta}_1) = \text{Var}(\hat{\beta}_2) = \text{Var}(\hat{\beta}_3)/4$$

and therefore

$$\text{Var}(\hat{\Delta}_1) = \text{Var}(\hat{\Delta}_2) = \text{Var}(\hat{\Delta}_3)/4, \tag{2}$$

where $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$ are corresponding maximum likelihood estimates of β_1, β_2, and β_3. It follows that the sample size needed to detect an interaction effect will be four times that for detecting a main effect of the identical magnitude because the sample size is a linear function of the variance of an effect estimate.

The total number of subjects, say $N(\Delta_1)$, required to detect a main effect with power $1 - \beta$ (where β is the level of type II error) was presented elsewhere (Donner et al., 1981; Donner and Klar, 2000; Diggle et al., 2002):

$$N(\Delta_1) = \frac{4(z_{\alpha/2} + z_\beta)^2 (1 + (k-1)\rho)\sigma^2}{k\beta_1^2} = \frac{4(z_{\alpha/2} + z_\beta)^2 (1 + (k-1)\rho)}{k\Delta_1^2}. \tag{3}$$

It follows that $N(\Delta_1) = N(\Delta_2)$ for $\Delta_1 = \Delta_2$. However, for effects of the same magnitude, $\Delta_1 = \Delta_3$, the total number of subjects, say $N(\Delta_3)$, required to detect an interaction effect with power $1 - \beta$ can then be expressed as fourfold that of the main effect. Finally, combining the sample size determination (3) for the main effect with the fourfold increase in the variance of the mle of the interaction effect of interest (2), we propose the following for sample size determination for detecting the interaction:

$$N(\Delta_3) = \frac{16(z_{\alpha/2} + z_\beta)^2 (1 + (k-1)\rho)\sigma^2}{k\beta_3^2} = \frac{16(z_{\alpha/2} + z_\beta)^2 (1 + (k-1)\rho)}{k\Delta_3^2} = 4N(\Delta_1). \tag{4}$$