Published in final edited form as:
Comput Methods Programs Biomed. 2008 August ; 91(2): 122–127. doi:10.1016/j.cmpb.2008.03.001.

Sample Size and Power Calculations Based on Generalized Linear Mixed Models with Correlated Binary Outcomes

Qianyu Dang[a,b,*], **Sati Mazumdar**[b,c], and **Patricia R. Houck**[c]

[a]Department of Medicine, University of Pittsburgh, Pittsburgh, PA 15213, United States

[b]Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA 15213, United States

[c]Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA 15213, United States

Abstract

The generalized linear mixed model (GLIMMIX) provides a powerful technique to model correlated outcomes with different types of distributions. The model can now be easily implemented with SAS PROC GLIMMIX in version 9.1. For binary outcomes, linearization methods of penalized quasilikelihood (PQL) or marginal quasi-likelihood (MQL) provide relatively accurate variance estimates for fixed effects. Using GLIMMIX based on these linearization methods, we derived formulas for power and sample size calculations for longitudinal designs with attrition over time. We found that the power and sample size estimates depend on the within-subject correlation and the size of random effects. In this article, we present tables of minimum sample sizes commonly used to test hypotheses for longitudinal studies. A simulation study was used to compare the results. We also provide a Web link to the SAS macro that we developed to compute power and sample sizes for correlated binary outcomes.

Keywords

Clinical trials; GLIMMIX; Penalized quasi-likelihood; Marginal quasi-likelihood

1. Introduction

The increasing use of longitudinal study designs in applied clinical research has been accompanied by an increasing demand for power calculations and sample size determinations during the study planning stage. For each study, the power estimates should be based on the proposed model for the analysis of primary outcomes. The most popular statistical models for longitudinal data with repeated measures are marginal models, such as generalized estimating equation (GEE) models, and mixed effects models in which subjects are treated as random effects.

For GEE models, Liu and Liang [1] developed an approximate method based on the generalized score test with an asymptotically noncentral chi-square distribution of the test statistic under the alternative hypothesis. Shih [2] and Pan [3] discussed alternative

*Corresponding author. Tel: 412-647-4780; fax: 412-246-6954. E-mail address: dangq@upmc.edu.

approaches for sample size calculations based on the *z*-test and a robust variance estimator. Pan [3] also derived the formulas for different structures for the working correlation matrix. To calculate sample sizes, Rochon [4] developed a computer program (GEESIZE) that accounts for unequal allocation, staggered entry, and loss to follow-up.

For linear mixed models with normal outcomes, Snijders and Bosker [5] provided an approximate method to estimate the standard errors of regression coefficients. These standard errors can be used for power calculations for explanatory variables (e.g., group effects) in study designs with repeated measures for each subject. Hedeker and his colleagues [6] provided formulas for comparisons of two groups across the study time points in a two-group repeated-measures design with attritions over time based on asymptotical *z*-tests. They further expanded these formulas to accommodate unbalanced design and comparisons at specific time points during the study, and they used the Web-based program RMASS [7] to implement their algorithms. However, RMASS does not handle mixed models with non-Gaussian outcomes. For these models, the alternative approach is to use Monte Carlo simulation [8]. The main advantage of this approach is that it allows researchers to compute power for a relatively complex study for which no exact methods are available. The disadvantages are that the results usually cannot be generalized to other study designs and that the simulation experiments can be time consuming.

Recently, the generalized linear mixed model (GLIMMIX) has become a popular procedure in SAS version 9.1 (SAS Institute, Cary, North Carolina). This model can handle different types of outcomes, such as binary, Poisson, and log normal outcomes. In addition, GLIMMIX has a more lenient assumption than GEE models regarding mechanisms to handle attrition and missing data. In most clinical longitudinal studies, subjects can drop out for various reasons. The missing data can be categorized as missing completely at random (MCAR), missing at random (MAR), or missing not at random (MNAR). GLIMMIX assumes that the missing data are MAR, which means that they may depend on the observed values but not on the unobserved values. In contrast, GEE models assume that the missing data are MCAR, which requires that they not depend on either observed or unobserved values. In the formulas for sample size and statistical power calculations for longitudinal designs that we describe here, we allowed for subject attrition under the MAR assumption.

All methods for power and sample size calculations for models with correlated data require assumptions about the within-subject correlation structures. The common choices for these structures are independent, compound symmetry (CS), first-order autocorrelation (AR(1)), and unstructured (UN). Unlike a GEE method, which has the attractive property of using a robust variance estimator to determine regression coefficients, the GLIMMIX model relies on the correct specification for the within-subject correlation structure for its power estimation.

In the work reported here, we derived the formulas with AR(1) and CS correlation structures. We focused on the power calculation for the GLIMMIX model with correlated binary outcomes and used an approach similar to that of Pan [3]. We considered the situation in which treatment is given at the subject level and continued over time and in which repeated measures from the same subject are taken across a period of time. We calculated the power for a two-group comparison involving a treatment group and a control group. We implemented the methods in a SAS program.

2. Statistical Framework

2.1. GLIMMIX model with binary outcomes

The general form of the GLIMMIX model with a link function g can be written as

$$g(E(Y_{ij}|\gamma_i))=X_i\beta+Z_i\gamma_i \quad (1)$$

or as

$$E(Y_{ij}|\gamma_i)=g^{-1}(X_i\beta+Z_i\gamma_i)=g^{-1}(\eta)=\mu_i, \quad (2)$$

where Y_{ij} denotes the outcome for subject i at time j, where X_i denotes the design matrix for fixed effects, and where Z_i denotes the design matrix for random effects.

We assume that the random effects γ_i, which are also called G-side random effects, are distributed as $iid\ N(0, G)$. We also assume that β is the vector of regression coefficients for fixed effects. In particular, for binary outcomes Y_{ij}, we assume that $\mu_i = prob(Y_i) = 1|\gamma_i = E(Y_{ij}|\gamma_i)$. The conditional variance estimates are $Var(Y_{ij}|\gamma_i)=A_i^{1/2}RA_i^{1/2}$, where A_i is a diagonal matrix containing the variance function $\mu_i(1-\mu_i)$ and where the R matrix (R-side error structure) is an assumed covariance error structure such as CS or AR(1). For subject i with n_i observations, the within-subject CS covariance structure is defined as

$$R_i=\begin{Bmatrix} 1 & \rho & \rho & \dots & \rho \\ \rho & 1 & \rho & \dots & \rho \\ \rho & \rho & 1 & \dots & \rho \\ \dots & & & \dots & \\ \rho & \rho & \rho & \dots & 1 \end{Bmatrix}, \quad (3)$$

and the AR(1) covariance structure is defined as

$$R_i=\begin{Bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n_i-1} \\ \rho & 1 & \rho & \dots & \rho^{n_i-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{n_i-3} \\ \dots & & & \dots & \\ \rho^{n_i-1} & \rho^{n_i-2} & \rho^{n_i-3} & \dots & 1 \end{Bmatrix}. \quad (4)$$

For both structures, the correlation coefficient ρ is the only parameter. The correlation between the two binary outcomes y_i, y_j is defined as

$$\rho(y_i, y_j)=\frac{p_{ij}-p_ip_j}{\sqrt{p_i(1-p_i)p_j(1-p_j)}}, \quad (5)$$

where the marginal rates p_i = Prob(y_i = 1), p_j = Prob(y_j = 1) and p_{ij} = Prob(y_i = 1, y_j = 1).

Most of the solutions for generalized mixed models rely on some form of likelihood function. The marginal likelihood for GLIMMIX with the scale parameter ϕ can be obtained by integrating out the random effects γ_i through the conditional distribution $f(\gamma_i|G)$. The likelihood contribution for subject i is a function of ϕ, regression coefficients β, and random effects distribution G:

$$f(Y_i|\beta, G, \phi)=\int\prod_{j=1}^{n_i}f_{ij}(Y_{ij}|\gamma_i, \beta, \phi)f(\gamma_i|G)d\gamma_i, \quad (6)$$

and the total likelihood [9] of N subjects is

$$L(\beta, G, \phi)=\prod_{i=1}^{N} \int \prod_{j=1}^{n_i} f_{ij}(Y_{ij}|\gamma_i, \beta, \phi) f(\gamma_i|G) d\gamma_i. \quad (7)$$

Note that for binary outcomes with logit link function, ϕ is equal to 1. The difficulty in maximizing the above function to obtain the maximum likelihood estimate is that there are N integrals over the q-dimensional random effects γ_i. The integrals can only be solved analytically in a few models, such as linear mixed models with normally distributed responses. Otherwise, different techniques of numerical approximation are required.

The two linearization methods implemented in the SAS PROC GLIMMIX procedure are the penalized quasi-likelihood (PQL) method and the marginal quasi-likelihood (MQL) method [9, 10].

The PQL is based on the Taylor expansion of Y_{ij} around current estimates of fixed effects $\hat{\beta}$ and random effects $\widehat{\gamma_i}$. The expansion around both fixed and random effects in equation (1) yields:

$$Y_{ij} \approx g(X_{ij}\widehat{\beta}+Z_{ij}\widehat{\gamma_i})+g'(X_{ij}\widehat{\beta}+Z_{ij}\widehat{\gamma_i})X_{ij}(\beta-\widehat{\beta})+g'(X_{ij}\widehat{\beta}+Z_{ij}\widehat{\gamma_i})Z_{ij}(\gamma_i-\widehat{\gamma_i})+\varepsilon_{ij} \quad (8)$$

The first term on the right side equals the current predictor for the conditional mean $\widehat{\mu_i}=E(Y_{ij}|\gamma_i)$. Rewriting everything in vector notation and letting $\tilde{\Delta}_i=g'(X_i\widehat{\beta}+Z_i\widehat{\gamma_i})$, we have

$$Y_i \approx \widehat{\mu_i}+\tilde{\Delta}_i X_i(\beta-\widehat{\beta})+\tilde{\Delta}_i Z_i(\gamma_i-\widehat{\gamma_i})+\varepsilon_i, \quad (9)$$

where $\tilde{\Delta}$ is actually the diagonal matrix of derivatives of the conditional mean μ_i evaluated at the expansion locus of $\hat{\beta}$ and $\widehat{\gamma_i}$. Rearranging the above terms, we obtain

$$Y_i^* \equiv \tilde{\Delta}_i^{-1}(Y_i-\widehat{\mu_i})+X_i\widehat{\beta}+Z_i\widehat{\gamma_i} \approx X_i\beta+Z_i\gamma+\varepsilon_i^* \quad (10)$$

and

$$VAR[Y_i^*|\gamma_i]=\tilde{\Delta}_i^{-1}A_i^{1/2}RA_i^{1/2}\tilde{\Delta}_i^{-1}. \quad (11)$$

Equation (10) can be viewed as a linear mixed model with the new continuous pseudo-observation Y_i^*, fixed effects β, random effects γ, and error term $\varepsilon_i^*=\tilde{\Delta}_i^{-1}\varepsilon_i$. Given starting values for β and G, the estimates for these unknown parameters are obtained by solving this approximate linear mixed model. Then the pseudodata are updated with these estimates, and the procedure is repeated iteratively until convergence is reached.

The MQL is similar to the PQL except that the Taylor expansion of Y_{ij} is around $\hat{\beta}$ and that γ_i is equal to 0. Rodriguez and Goldman [11] found that the point estimates can be biased for binary outcomes for both PQL and MQL, but the variance estimates are acceptable.

2.2. Sample size and power calculations

Both sample size and power estimates for any types of mixed models depend on the variance estimates of the fixed effect

$$VAR(\widehat{\beta})=(\sum_{i=1}^{N}X_i'V_i^{-1}X_i)^{-1}, \quad (12)$$

where V_i is the unconditional variance of Y_i. This variance consists of two components: the G-side random effects and the R-side errors. The variance is formulated as

$$V_i=var(Y_i)=Z_iGZ_i'+\tilde{\Delta}_i^{-1}A_i^{1/2}R_iA_i^{1/2}\tilde{\Delta}_i^{-1}. \quad (13)$$

This can be rewritten as

$$V_i=\tilde{\Delta}_i^{-1}(Z_iG^*Z_i'+A_i^{1/2}R_iA_i^{1/2})\tilde{\Delta}_i^{-1} \quad (14)$$

by putting the first term into the sandwich-like error structure of $\tilde{\Delta}_i^{-1}A_i^{1/2}RA_i^{1/2}\tilde{\Delta}_i^{-1}$ so that G becomes transformed G*. We can further simplify equation (14) by extracting a combined variance parameter σ_i^2 so the transformed R* matrix contains only the correlation parameters ρ*. The result is

$$Z_iG^*Z_i'+A_i^{1/2}R_iA_i^{1/2}=\sigma_i^2(A_i^{1/2}R^*{}_iA_i^{1/2}) \quad (15)$$

Note that the component of ρ* in the R* matrix is not equal to the original within-subject correlation coefficient in the R matrix. Instead, ρ* depends on both the G matrix and the R matrix. It is easy to see that ρ* becomes larger as the size of the random effects and the size of the fixed effect coefficient $\hat{\beta}$ increase.

The variance estimator of the fixed effects can thus be formulated as

$$VAR(\widehat{\beta})=(\sum_{i=1}^{N}X_i'V_i^{-1}X_i)^{-1}=\sum_{i=1}^{N}\{X_i'\tilde{\Delta}_i(Z_iG^*Z_i'+A_i^{1/2}R_iA_i^{1/2})^{-1}\tilde{\Delta}_iX_i\}^{-1}=\sum_{i=1}^{N}\{X_i'\tilde{\Delta}_i\sigma_i^2(A_i^{1/2}R_i^*A_i^{1/2})^{-1}\tilde{\Delta}_iX_i\}^{-1} \quad (16)$$

For a mixed model with binary outcomes, all the formulas above are based on the conditional mean μ_i instead of the marginal average probability of event $p= E(Y_{ij})$. The conditional mean and the marginal average probability are not equal [12]:

$$\mu_i=E(Y_{ij}|\gamma_i)=\frac{\exp(X_i\beta+Z_i\gamma_{ij})}{1+\exp(X_i\beta+Z_i\gamma_{ij})}\neq E(Y_{ij}). \quad (17)$$

However, at the planning stage of a study, we can assume that μ_i is an approximation of $E(Y_{ij})$.

In the example we present here, we assume the treatment group to be the fixed effect and each subject within the group to be the random effect. The purpose is to test if the binary responses of the treatment group are better than those of the control group. The design matrices for fixed and random effects for subject i are:

Control group: $X_i = [1_{ni}\, 0_{ni}]$, $Z = 1_{ni}$;

Treatment group: $X_i = [1_{ni}\, 1_{ni}]$, $Z = 1_{ni}$.

The null hypothesis is H_0: β = 0, and the alternative hypothesis is H_a: β = b > 0. The test for a hypothesis such as this is a Wald-type test based on asymptotically normal distributions as

in the marginal model ([2], [3]). Thus, for given sample size N, the power $1 - \Delta$ can be obtained by

$$1 - \Delta \approx 1 - \Phi\left(Z_{\alpha/2} - b\sqrt{\frac{N}{VAR(\widehat{\beta})}}\right), \quad (18)$$

and the required sample size N for a given power $1 - \Delta$ is

$$N = \frac{VAR(\widehat{\beta})(Z_{\alpha/2} - Z_{1-\Delta})^2}{b^2}. \quad (19)$$

We denoted $\boldsymbol{\pi}$ as the sample allocation ratio and denoted p_0 and p_1 as the marginal average rates for control and treatment groups and then applied the same derivation procedure used by Shih [2] and Pan [3] to rewrite the variance estimator (16) as a function of $\boldsymbol{\pi}$, p_0, p_1, R^* and σ_i^2:

$$VAR(\widehat{\beta}) = \left(\sum_{i=1}^{N} \frac{1_{n_i}' R^{*-1} 1_{n_i} M}{\sigma_i^2}\right)^{-1} = \left(\sum_{i=1}^{N} \frac{1_{n_i}' R^{*-1} 1_{n_i}}{\sigma_i^2}\right)^{-1} M^{-1}, \quad (20)$$

where

$$M^{-1} = \frac{1}{\pi p_0(1-p_0)} \begin{bmatrix} 1 & -1 \\ -1 & 1 + \frac{\pi p_0(1-p_0)}{(1-\pi)p_1(1-p_1)} \end{bmatrix}. \quad (21)$$

When VAR($\hat{\beta}$) becomes available, we can plug it into equation (18) or (19) to obtain the estimated power or sample size.

In the special case in which the R matrix has a CS covariance structure with ρ_0^* and ρ_1^* as the components in R^* and with σ_0 and σ_1 as the combined variance parameter mentioned above for the control and treatment groups, the variance estimation formula is

$$VAR(\widehat{\beta}) = \left(\sum_{N_0} \frac{n_i}{1+(n_i-1)\rho_0^* {\sigma_0}^2} + \sum_{N_1} \frac{n_i}{1+(n_i-1)\rho_1^* {\sigma_1}^2}\right)^{-1} \left[\frac{1}{\pi\rho_0(1-p_0)} + \frac{1}{(1-\pi)p_1(1-p_1)}\right]. \quad (22)$$

But if the R matrix has an AR(1) covariance structure, the formula becomes

$$VAR(\widehat{\beta}) = \left(\sum_{N_0} \frac{n_i - (n_i-2)\rho_0^*}{(1+\rho_0^*){\sigma_0}^2} + \sum_{N_1} \frac{n_i - (n_i-2)\rho_1^*}{(1+\rho_1^*){\sigma_1}^2}\right)^{-1} \left[\frac{1}{\pi\rho_0(1-p_0)} + \frac{1}{(1-\pi)p_1(1-p_1)}\right]. \quad (23)$$

In both cases, N_0 and N_1 is the sample size in control and treatment group.

2.3. Longitudinal designs with attrition

All of the formulas described above assume that the sample size is constant across time. In practice, however, this assumption rarely holds true in longitudinal studies.

The “completers only” approach to the problem uses the minimum expected sample size at any time point in a study. But this conservative approach can lead to an underestimated power or an inflated sample size.

A better approach is to modify the formulas to accommodate expected attrition over time. To modify the formulas, we can begin by denoting N_{0k} and N_{1k} as the number of subjects who have k observations ($k = 1..$, n) in the control group and the treatment group, respectively. If the assumed covariance structure for R^* is CS, then the estimated variance will be

$$VAR(\widehat{\beta})=\left(\sum_k\sum_{N_{0k}}\frac{k}{1+(j-1)\rho_0^*{\sigma_0}^2}+\sum_k\sum_{N_{1k}}\frac{k}{1+(j-1)\rho_1^*{\sigma_1}^2}\right)^{-1}\left[\frac{1}{\pi\rho_0(1-\rho_0)}+\frac{1}{(1-\pi)\rho_1(1-\rho_1)}\right]. \quad (24)$$

If the covariance structure is AR(1), then the estimated variance will be

$$VAR(\widehat{\beta})=\left(\sum_k\sum_{N_{0k}}\frac{k-(k-2)\rho_0^*}{(1+\rho_0^*){\sigma_0}^2}+\sum_k\sum_{N_{1k}}\frac{k-(k-2)\rho_1^*}{(1+\rho_1^*){\sigma_1}^2}\right)^{-1}\left[\frac{1}{\pi\rho_0(1-\rho_0)}+\frac{1}{(1-\pi)\rho_1(1-\rho_1)}\right]. \quad (25)$$

2.4. Examples of calculations

Based on the above formulas, we developed an SAS program to compute the power and sample size for correlated binary outcomes. The program can be downloaded from our Web site at http://www.pitt.edu/~qidst1/abstract.htm.

For purposes of illustration, we present a hypothetical example in which 200 subjects are divided into two groups (a treatment group and a control group) and followed for 1 year. The binary responses are taken every 3 months after baseline. The assumed marginal rate of outcome is 0.1 for the treatment group and 0.2 for the control group. Each subject is considered to be a random effect. Based on the results of a pilot study, we estimate the variance of the random effect G to be 1 and the within-subject correlation ρ to be 0.7 for the total sample. We expect the overall attrition rate to be 20% for the total sample and the dropouts to be evenly distributed throughout the year. Thus, the number of subjects at 3, 6, 9, and 12 months will be 190, 180, 170, and 160. We assume all dropouts follow missing at random mechanism.

For comparison, we ran a simulation program similar to Moineddin [13] with CS within subject correlation structure to confirm the results. For the purpose of comparing of simulation and our formula, we assume the sample size is 200 and without any drop outs. Among 500 sets of randomly generated longitudinal data as described above, SAS GLIMMIX procedure converged on 480 of them with the p values of 358 estimated treatment effects to be less than 0.05. The power estimation of 74.6% is close to 71.8% computed from the formula we derived if we do not count those 20 sets of data which did not converge. The actual simulated power is somewhere between 71.6% and 75.6%, as these two boundaries are set by counting these 20 results to be all non-significant or all significant. Unfortunately, unlike CS, the AR(1) within subject correlation can not be written as the ratio of the within subject and total variance, so it is very difficult to implement power estimation through simulation for longitudinal binary data with auto-correlated within subject correlation structure.

In the situation where the drop out rate was 20% over 12 months, the program showed that in order to achieve 80% power, we need to recruit 308 (with CS correlation structure) and 263 (with AR(1) correlation structure) patients using “completers only” approach. With the same power, the required number of patients is 262 with CS structure or 226 with the AR (1) structure if we can use the partial data of those 20% drop outs. It is clear that more subjects

are recruited than necessary by applying the conservative “completers only” approach in the sample size estimation.

Table 1 shows that the estimated power for a mixed effects model depends on assumptions about the variance of random effects and within-subject correlations. If the variance of random effects increase but the within-subject correlations remain the same, the estimated power increases. If the random effects remain the same but the within-subject correlations increase, the estimated power decreases. These trends are similar to mixed model with normal outcomes. Notice that the treatment effect coefficient beta is also a function of the random variance G, so the power change can also be seen as the results of effect size change. This is different from marginal models like GEE, in which the effect size is determined only by the marginal event rate of each group.

Table 2 shows that the estimated sample size required is affected by the number of time points, the expected marginal probabilities of outcome, and the within-subject structure and correlations. Overall, larger sample sizes are required for smaller numbers of repeated measures to obtain the same level of power. Therefore, to achieve the most cost-effective design, the researcher needs to find the optimal sample size and number of repeated measures for each subject.

3. Discussion and Summary

The GLIMMIX model provides a powerful technique to model correlated outcomes with different types of distributions, and this model can now be easily implemented with SAS PROC GLIMMIX version 9.1. For binary outcomes, linearization methods of PQL or MQL provide relatively accurate variance estimates for fixed effects. Using GLIMMIX based on these linearization methods, we derived formulas for power and sample size calculations for longitudinal designs with attrition over time.

The power can also be computed through computer simulation programs. But such approach is limited for GLIMMIX model. First, simulation is difficult for AR(1) and other complicated covariance structure other than CS since they can not be easily written as the ratio of random variance and total variance. The other problem is that the convergence rate of GLIMMIX models can be very low for data with medium to small sample sizes or small number of repeated measures. Such problem makes accurate power estimation to be difficult for most longitudinal studies as discussed by Moineddin [12]. The convergence rate is slightly better for SAS NLMIXED procedure using numerical integration instead of PQL, but NLMIXED lacks the flexibility for different within subject covariance structures.

Because GLIMMIX is a subject-specific model, it actually estimates each subject’s unique parameters of random effects in addition to estimating the fixed effect as a “common factor.” Power calculations need to be based on subject-specific information. Usually, however, it is not possible to obtain this information in the form of the conditional mean $\mu_i = E(Y_{ij}|\gamma_i)$ for each subject during the study design stage. If no pilot data for a study are available, we can use the marginal rates $p=E(Y_{ij})$ as an approximation. But if pilot data are available, we can instead to use the average of the estimated conditional means of all the subjects from the pilot study to reduce the possible bias caused by this approximation. However, our simulations show that the approximation gives results fairly close to the simulated one.

In summary, the sample size and power calculations for GLIMMIX are affected by prior information about random effects, within-subject correlations, and other values often difficult to obtain without pilot data. Statisticians should consult with clinicians to make reasonable estimates of this information or should base their assumptions on any preliminary

data that are available. They may then wish to use our SAS program (http://www.pitt.edu/~qidst1/abstract.htm) to compute their power and sample sizes for correlated binary outcomes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported in part by the Intervention Research Center and Advanced Center for Intervention and Services Research for Late Life Mood Disorders (P30 MH71944).

References

1. Liu G, Liang KY. Sample size calculations for studies with correlated observations. Biometrics. 1997; 53:937–947. [PubMed: 9290224]
2. Shih WJ. Sample size and power calculations for periodontal and other studies with clustered samples using the method of generalized estimating equations. Biometrical Journal. 1997; 39:899–908.
3. Pan W. Sample size and power calculations with correlated binary data. Controlled Clinical Trials. 2001; 22:211–227. [PubMed: 11384786]
4. Rochon J. Application of GEE procedures for sample size calculations in repeated measures. Statistics in Medicine. 1998; 17:1643–1658. [PubMed: 9699236]
5. Snijders TA, Bosker RJ. Standard errors and sample sizes for two-level research. Journal of Educational Statistics. 1993; 18:237–259.
6. Hedeker D, Gibbons R, Waternaux C. Sample size estimation for longitudinal designs with attrition: comparing time-related contrasts between two groups. Journal of Educational and Behavioral Statistics. 1999; 24:70–93.
7. RMASS. Center for Health Statistics, University of Illinois at Chicago; Available at: http://www.uic.edu/labs/biostat/projects.html. [Accessed August 16, 2007]
8. Gastanaga VM, McLaren CE, Delfino RJ. Power calculations for generalized linear models in observational longitudinal studies: a simulation approach in SAS. Computer Methods and Programs in Biomedicine. 2004; 84:27–33. [PubMed: 16982112]
9. Wolfinger R, O'Connell M. Generalized linear mixed models: a pseudo-likelihood approach. Journal of Statistical Computation and Simulation. 1993; 4:233–243.
10. SAS. GLIMMIX Procedure Documentation. Cary, North Carolina: SAS Institute Inc.; 2006 Jun.
11. Rodriguez G, Goldman N. An assessment of estimation for multilevel models with binary responses. Journal of the Royal Statistical Society A. 1995; 158:73–89.
12. Molenberghs, G.; Verbeke, G. Models for Discrete Longitudinal Data. New York: Springer-Verlag; 2005.
13. Moineddin R, Matheson F, Glazier R. A simulation study of sample size for multilevel logistic regression models. BMC Medical Research Methodology. 2007:7–34. [PubMed: 17291355]

Table 1

Power estimates based on the fixed effect, random effects, and withinsubject correlation[a]

Fixed effect (β)	Random effects (*G* side)	Within-subject correlation (ρ)	Power
0.982	0.1	0.6	74.0%
1.022	0.5	0.6	76.3%
1.138	1	0.6	83.5%
1.514	2	0.6	96.7%
1.989	3	0.6	99.9%
1.138	1	0.3	94.9%
1.138	1	0.4	91.4%
1.138	1	0.5	87.6%
1.138	1	0.7	83.5%
1.138	1	0.8	79.5%
1.138	1	0.9	72.0%

[a]Estimates are for a two-group longitudinal study with 100 subjects in the treatment group and 100 subjects in the control group and with 4 repeated measures for each subject. We assume CS within subject covariance structure and $\alpha = 0.05$.

Table 2

Minimum total sample size based on the number of time points, the expected marginal probabilities of the outcome, and the within-subject covariance structure and correlation [a]

Number of time points	Probabilities		Compound symmetry (CS) within-subject covariance structure and correlation			First-order autocorrelation (AR(1)) within-subject covariance structure and correlation		
	p_1	p_2	ρ = 0.2	ρ = 0.5	ρ = 0.8	ρ = 0.2	ρ = 0.5	ρ = 0.8
T = 3	0.1	0.2	152	214	276	140	192	260
		0.3	52	74	94	48	66	88
		0.4	30	42	52	28	38	50
		0.5	22	28	36	20	26	34
	0.2	0.3	224	314	402	206	282	380
		0.4	68	92	118	62	84	112
		0.5	34	48	60	32	42	56
	0.3	0.4	280	386	492	254	348	466
		0.5	76	106	134	70	96	126
	0.4	0.5	308	424	540	280	382	512
T = 6	0.1	0.2	110	188	264	76	122	206
		0.3	40	64	90	28	42	70
		0.4	24	38	52	16	26	40
		0.5	22	26	34	12	18	28
	0.2	0.3	166	276	386	114	180	302
		0.4	50	82	114	34	54	88
		0.5	26	42	58	18	28	46
	0.3	0.4	208	342	474	140	222	370
		0.5	58	94	130	40	62	102
	0.4	0.5	230	376	520	156	244	406

[a]Estimates are for the minimum total sample required in two groups, with a sample allocation ratio of 1:1, for comparison of event probabilities p_1 and p_1 assumed to be constant over time, with use of the logit link for various values T, p_1, p_1, and ρ. Estimates shown are for setting the random effects to $G = 1$, type I error to 0.05, and type II error to 0.20.