5-1-2014

A Flexible Method for Conducting Power Analysis for Two- and Three-Level Hierarchical Linear Models in R

Yi Pan
University of North Carolina - Chapel Hill, yi.pan@unc.edu

Matthew T. McBee
East Tennessee State University, mcbeem@etsu.edu

JMASM Algorithms and Code:
A Flexible Method for Conducting Power Analysis for Two- and Three-Level Hierarchical Linear Models in R

Yi Pan
University of North Carolina – Chapel Hill
Chapel Hill, NC

Matthew T. McBee
East Tennessee State University
Johnson City, Tennessee

A general approach for conducting power analysis in two- and three-level hierarchical linear models (HLMs) is described. The method can be used to perform power analysis to detect fixed effects at any level of a HLM with dichotomous or continuous covariates. It can easily be extended to perform power analysis for functions of parameters. Important steps in the derivation of this approach are illustrated and numerical examples are provided. Sample code implementing this approach is provided using the free program *R*.

Keywords: power analysis, hierarchical linear model, mixed model, R, power analysis for hierarchical linear model

Hierarchical linear modeling (HLM) is widely used in various areas of social science (Singer, 1998; Raudenbush & Bryk, 2002). As with any quantitative method, it is frequently important to perform power analysis in order to determine the necessary sample size to achieve a given level of power, to describe the minimum detectable effect size, or to describe the level of precision in the estimation of effects that is achievable by a given study design and sample size.

Power analysis in the general linear model context is straightforward. Many empirical researchers are trained in the methods of performing power analysis for linear models and several excellent pieces of software, such as *GPower* and *SAS PROC GLMPOWER*, are widely available (Thomas & Krebs, 1997; Lewis, 2006). The penetration of HLM into the mainstream of a variety of social science disciplines has created a need for convenient tools to perform power analysis for HLMs. Several software applications are currently available for HLM power

Dr. Pan is a statistician at UNC-Chapel Hill's Frank Porter Graham Child Development Institute. Email him at yi.pan@unc.edu. Dr. McBee is an Assistant Professor in the Department of Psychology. Email him at: mcbeem@etsu.edu.

analysis. *Optimal Design* (Raudenbush, et al., 2004) is a widely used HLM power analysis software in social sciences, and allows researchers conduct power analysis on difference between treatment and control group in a number of cluster data analysis scenarios. However, it lacks the functionality of conducting power analysis for continuous predictors. *Power Analysis in Two-Level Designs* (PinT; Snijders & Bosker, 1993; Bosker, Snijders, & Guldemond, 1999) accommodates power analysis for continuous variables, but is limited to 2-level HLM's. Simulation-based power analysis software, like *MLPowSim* (Browne, Golalizadeh & Parker, 2009) and *ML-Des* (Cools, Van den Noortgate & Onghena, 2008), offer more flexibility, but it takes a much longer time to conduct simulation-based power analysis, and they do not allow unbalanced design.

This article provides insights about how to conduct power analysis in HLM studies and introduce ways to increase flexibility in power analysis previously mentioned pre-packaged software are lacking. Some reader familiarity with the basics of power analysis in a linear models framework is assumed; readers are referred to Cohen (Cohen, 1988, 1992) for a review of the fundamentals. A general strategy is put forth for performing power analysis in HLMs and the calculation of the covariance matrix of parameter estimators for models of various complexities, which is the critical component to calculate power, is illustrated. Also illustrated is how to use the equations derived to perform power calculations using *R*, although they could be performed in any software that performs matrix calculations. The goal is to provide a flexible and general approach that can be used for different scenarios, many of which may not be implemented in existing software.

Review of Power Analysis

Performing a power analysis involves calculating standard errors for estimators of parameters of interest. Once armed with an effect size and a standard error, a researcher can produce a test statistic that may then be compared against a chi-square, *T*, or *F* distribution (Cohen, 1998) to estimate approximate power. This paper focuses on the process of appropriately obtaining the standard error of a parameter estimator in HLM, which is the square root of the variance estimate of the parameter estimator. The actual power calculation using an assumed effect size and standard error is shown in numeric examples.

Statistical Power in HLM

The process of power analysis for multilevel models differs depending on whether one wishes to calculate power for a continuous variable or a dichotomous variable. This article will show that the dichotomous case is much simpler. In fact, an explicit analytical result is derived; however, the starting point is the more general continuous case. In the continuous case, the variance of the parameter estimator of interest depends on the sample data which researchers may not have when they conduct their power analysis. Therefore, additional information about unknown sample data must be assumed. In addition, the inclusion of covariates as well as whether the model contains random slopes will impact the power analysis. Although analytical solutions could be derived for some special cases, slightly different models could end up with very different analytical forms. Therefore a general numerical approach that may be used with a variety of models will be illustrated.

The goal is to calculate a test statistic, whose approximate distribution is known, that can be used to estimate the power of a statistical test of a parameter. Given certain assumptions regarding the model, parameter values, and sample data, the variance-covariance matrix of all the fixed effect parameters in the model can be approximated. This implies that the power to detect any fixed effect can be easily calculated. Furthermore, it will be demonstrated that the power to detect functions of parameters (e.g., contrasts) can also be calculated once the variance-covariance matrix is obtained. For maximum generality matrix notation is used to describe the model.

Power Analysis for Two-Level Models with Continuous Variables

According to the Gauss-Markov theorem, when errors are independently identically normally distributed with mean of zero and a constant variance in a simple linear regression model, the ordinary least squares estimator (OLSE) is the best linear unbiased estimator (BLUE; Hayashi, 2000). However, the assumption of independently identically distributed (i.i.d.) errors is not realistic for multilevel data. The variance-covariance matrix of random errors in response variables can be assumed to be $\sigma^2\Omega$ as opposed to $\sigma^2 I$, where I is an identity matrix according to the conditions specified in the Gauss-Markov theorem. As a result, the OLSE can be generalized to obtain a generalized least squares estimator

(GLSE) $\left(X'\Omega^{-1}X \right)^{-1} X'\Omega^{-1}Y$. Note that when $\Omega = I$, GLSE is OLSE. Under the assumption that $\sigma^2\Omega$ is specified correctly, the GLSE is also BLUE (Aitken, 1935). Suppose a researcher conducts a study in which she enrolls J groups of participants and each group consists of n individuals. There are all together m level-one predictors. The level-one equation in matrix form is $Y = X\beta + e$, where Y is a $nJ*1$ vector, X is a $nJ*mJ$ diagonal block matrix, β is $mJ*1$, and e is a $nJ*1$ vector.

The level-one equation in matrix form is

$$
\begin{pmatrix} y_{11} \\ \vdots \\ y_{n1} \\ y_{12} \\ \vdots \\ y_{nJ} \end{pmatrix}
=
\begin{pmatrix}
\begin{matrix} x_{111} & x_{112} & \cdots & x_{11m} \\ x_{211} & x_{212} & \cdots & x_{21m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n11} & x_{n12} & \cdots & x_{n1m} \end{matrix} & & \mathbf{0} & & \cdots & & \mathbf{0} \\
& \begin{matrix} x_{121} & x_{122} & \cdots & x_{12m} \\ x_{221} & x_{222} & \cdots & x_{22m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n21} & x_{n22} & \cdots & x_{n2m} \end{matrix} & & \cdots & & \mathbf{0} \\
\mathbf{0} & & \vdots & & \ddots & & \vdots \\
& & & & & \begin{matrix} x_{1J1} & x_{1J2} & \cdots & x_{1Jm} \\ x_{2J1} & x_{2J2} & \cdots & x_{2Jm} \\ \vdots & \vdots & \ddots & \vdots \\ x_{nJ1} & x_{nJ2} & \cdots & x_{nJm} \end{matrix} \\
\mathbf{0} & & \cdots & & \mathbf{0} &
\end{pmatrix}
\begin{pmatrix} \beta_{11} \\ \vdots \\ \beta_{1m} \\ \beta_{21} \\ \vdots \\ \beta_{Jm} \end{pmatrix}
+
\begin{pmatrix} e_{11} \\ \vdots \\ e_{n1} \\ e_{12} \\ \vdots \\ e_{nJ} \end{pmatrix}
$$

The elements of one column in each block in X are all 1 if the level-one model has an intercept. The intercept can also be considered as a slope when x is always equal to 1; no further distinction will be given between intercepts and slopes in the remainder of this article.

The level-two equation is

$$
\begin{pmatrix} \beta_{11} \\ \vdots \\ \beta_{1m} \\ \beta_{21} \\ \vdots \\ \beta_{Jm} \end{pmatrix}
=
\begin{pmatrix}
\begin{matrix} z_{111} & \cdots & z_{11p} \\ & \mathbf{0} & \\ & \vdots & \\ & \mathbf{0} & \\ z_{121} & \cdots & z_{12p} \\ & \mathbf{0} & \\ \vdots & & \\ & \mathbf{0} & \\ z_{1J1} & \cdots & z_{1Jp} \\ & \mathbf{0} & \\ & \vdots & \\ & \mathbf{0} & \end{matrix}
&
\begin{matrix} & \mathbf{0} & \\ z_{211} & \cdots & z_{21p} \\ \vdots & & \vdots \\ & \mathbf{0} & \\ & \mathbf{0} & \\ z_{221} & \cdots & z_{22p} \\ & & \\ & \mathbf{0} & \\ & \mathbf{0} & \\ z_{2J1} & \cdots & z_{2Jp} \\ \vdots & & \vdots \\ & \mathbf{0} & \end{matrix}
&
\begin{matrix} \cdots \\ \cdots \\ \ddots \\ \mathbf{0} \\ \cdots \\ \cdots \\ \ddots \\ \mathbf{0} \\ \cdots \\ \cdots \\ \ddots \\ \mathbf{0} \end{matrix}
&
\begin{matrix} & \mathbf{0} & \\ & \mathbf{0} & \\ & \vdots & \\ z_{m11} & \cdots & z_{m1p} \\ & \mathbf{0} & \\ & \mathbf{0} & \\ & \vdots & \\ z_{m21} & \cdots & z_{m2p} \\ & \mathbf{0} & \\ & \mathbf{0} & \\ & \vdots & \\ z_{mJ1} & \cdots & z_{mJp} \end{matrix}
\end{pmatrix}
\begin{pmatrix} \gamma_{11} \\ \vdots \\ \gamma_{1p} \\ \gamma_{21} \\ \vdots \\ \gamma_{mp} \end{pmatrix}
+
\begin{pmatrix} u_{11} \\ \vdots \\ u_{1m} \\ u_{21} \\ \vdots \\ u_{Jm} \end{pmatrix}
$$

where z_{ijk} indicates the kth second level predictor that is for the jth group and has an effect on i^{th} level-one variable.

The size of the level-two predictor design matrix Z is $mJ * mp$. The size of the level-two parameter vector γ is $mp * 1$, and the random slope U is a $mj * 1$ vector. Note that the above equation assumes that all level two predictors have effects on all β's. In practice, the design matrix of level two predictors should be constructed according to the actual model of interest. Also, researchers may specify some level-one parameters to have random effects. A level-two HLM equation can also be written in the following fashion:

$$Y = X\left(Z\gamma + U\right) + e \tag{1}$$

By distributing X :

$$Y = XZ\gamma + XU + e \tag{2}$$

Because γ can be considered as a vector of fixed effects, only $\left(XU + e\right)$ is random in **Y**.

$$
\begin{aligned}
V = Var\left(Y\right) &= Var\left(XU + e\right) \\
&= X\tau_k^2 X' + \pi^2 I = \sigma^2 \Omega
\end{aligned}
\tag{3}
$$

As can be observed, the variance components are divided into multiple parts, and the number depends on how many level-one predictors have random effects. Directly following generalized linear model theory, results in

$$Var\left(\hat{\gamma}\right) = \left(\left(XZ\right)'\left(\sigma^2\Omega\right)^{-1}\left(XZ\right)\right)^{-1} \tag{4}$$

(Snijders & Bosker, 1993). Using a conclusion from De Leeuw and Kreft (1986, p. 25) that

$$\left(\sigma^2\Omega\right)^{-1} = \begin{pmatrix} \pi^{-2}I - \pi^{-2}X\left(X'X\right)^{-1}X' \\ +X\left(X'X\right)^{-1}\left(\pi^2\left(X'X\right)^{-1}+\tau^2\right)^{-1}\left(X'X\right)^{-1}X' \end{pmatrix} \qquad (5)$$

As a result, if the $\sigma^2, \tau^2, X, Z,$ and V matrices are known, the variance-covariance matrix of $\hat{\gamma}$ can be calculated. If this information and the assumed effect size of $\hat{\gamma}$ are determined before conducting power analysis, power can be estimated using the χ^2 distribution. However, one problem remains: Prospective power analysis takes place before the study begins, so some of the needed information may be unavailable. In order to proceed with the power calculation, it is necessary to make some assumptions about the values of level one and level two covariates. One obvious option is to gather information on the distribution of the covariates, draw random variates from the distributions, and use that information in the calculations.

A general strategy to estimate power for the effects of covariates in two-level HLM is now presented; the procedure is as follows: First, assume values for the following: the effect size for the parameter of interest, the level one residual variance π^2, the level-two random effects' variance-covariance matrix τ^2, means and variance-covariance matrix of the level-one covariates, X, and the means of the level-two covariates, Z. Second, write down their specific models in matrix form and get detailed expressions for $X, Z,$ and Ω. Third, perform the matrix calculation and describe and obtain an estimation of the variance-covariance matrix of all the fixed effects' parameter estimators. Finally, the assumed effect size and the variance of the parameter estimator of interest can be used to construct a χ^2 statistic to obtain the estimated power.

Example Power Analysis for Two-Level Model Where Covariate Values are Known

Consider an example of a growth model with ten time points, a random intercept and a random slope for time. The model may be written as:

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + \varepsilon_{ij}$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01}z_j + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}z_j + u_{1j} \tag{6}$$

$$y_{ij} = \gamma_{00} + \gamma_{01}z_j + \gamma_{10}x_{ij} + \gamma_{11}x_{ij}z_j + u_{0j} + u_{1j}t_{ij} + \varepsilon_{ij}$$

$$u_j \sim N\left(0, \tau^2\right), \varepsilon_{ij} \sim N\left(0, \sigma^2\right)$$

When considering the values of the covariates x_{ij} and z_j, researchers may face two situations. One is that x_{ij} and z_j are completely or partially unknown prior to data collection. In this case, to conduct power analysis, the researcher will have to assume the first and second moments of the covariates. The second situation is that x_{ij} and z_j are known. For example, if z_j represents different levels of treatment, the number of levels and the number of individuals assigned to each is known in advance of data collection. For this example, assume that x_{ij} represents the coding of ten equally-spaced time points, z_j represents five levels of treatment, and the model assumes linear effects of x_{ij} and z_j.

Step one: Assume necessary values.

Assume that the effect size, δ, of γ_{01} is 1.0, The level-one error variance, π^2, is 10. The variance-covariance matrix of the level-two random components u_j is $\begin{pmatrix} 5.0 & 1.0 \\ 1.0 & 4.0 \end{pmatrix}$. The number of clusters, j, is 50. The number of repeated measures per cluster, n, is 10. The total sample size is 500; input these values into R by creating variables to hold them.

```
pisq <- 10
tausq <- array(c(5,1,1,4), dim=c(2,2))
delta <- 1
n <- 10
j <- 50
```

Step two: Write out matrix forms of X and Z

The matrix format of the reduced form equation is:

$$\begin{pmatrix} y_{11} \\ \vdots \\ y_{n1} \\ y_{12} \\ \vdots \\ y_{nJ} \end{pmatrix} = \begin{pmatrix} \begin{matrix} 1 & x_{11} \\ 1 & x_{21} \\ \vdots & \vdots \\ 1 & x_{n1} \end{matrix} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \begin{matrix} 1 & x_{12} \\ 1 & x_{22} \\ \vdots & \vdots \\ 1 & x_{n2} \end{matrix} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \begin{matrix} 1 & x_{1J} \\ 1 & x_{2J} \\ \vdots & \vdots \\ 1 & x_{nJ} \end{matrix} \end{pmatrix} \left[\begin{pmatrix} 1 & z_1 & 0 \\ 0 & 1 & z_1 \\ 1 & z_2 & 0 \\ 0 & 1 & z_2 \\ \vdots & \vdots & \\ 1 & z_j & 0 \\ 0 & 1 & z_j \end{pmatrix} \begin{pmatrix} \gamma_{00} \\ \gamma_{01} \\ \gamma_{10} \\ \gamma_{11} \end{pmatrix} + \begin{pmatrix} u_{01} \\ u_{11} \\ u_{02} \\ u_{12} \\ \vdots \\ u_{0J} \\ u_{1J} \end{pmatrix} \right] + \begin{pmatrix} e_{11} \\ \vdots \\ e_{n1} \\ e_{12} \\ \vdots \\ e_{nJ} \end{pmatrix}$$

(7)

In this case, X is a block diagonal matrix. Each block contains a vector of ones to define the intercept and a second vector coding the time points according to the model. Z is stack of block diagonal matrices. Within each submatrix, the first row describes how the level-one intercept is a function of the level-two parameters, $\gamma_{00}, \gamma_{01}, \gamma_{10},$ and γ_{11}. The second row describes how the slope for the time parameters is a function of the same parameters. These matrices are specified in R by creating two operation matrices, A and B, and then their Kronecker product is calculated to obtain X. A is a j by j identity matrix and B is an n by 2 matrix containing a column vector of ones and column vector containing the coding of time.

```
A <- diag(j)
B <- array(c(1,1,1,1,1,1,1,1,1,1,1,2,3,4,5,6,7,8,9,10), dim=c(10,2))
X <- kronecker(A,B)
```

The Z matrix is created by the following code:

```
Zmean1 <- 1
Zmean2 <- 2
Zmean3 <- 3
Zmean4 <- 4
Zmean5 <- 5
B1 <- matrix(data=c(1,0,Zmean1,0,0,1,0,Zmean1), nrow=2, ncol=4)
A1 <- array(1, dim=c(j/5,1))
Z1 <- kronecker(A1,B1)
B2 <- matrix(data=c(1,0,Zmean2,0,0,1,0,Zmean2), nrow=2, ncol=4)
A2 <- array(1, dim=c(j/5,1))
Z2 <- kronecker(A2,B2)
B3 <- matrix(data=c(1,0,Zmean3,0,0,1,0,Zmean3), nrow=2, ncol=4)
A3 <- array(1, dim=c(j/5,1))
Z3 <- kronecker(A3,B3)
```

```
B4 <- matrix(data=c(1,0,Zmean4,0,0,1,0,Zmean4), nrow=2, ncol=4)
A4 <- array(1, dim=c(j/5,1))
Z4 <- kronecker(A4,B4)
B5 <- matrix(data=c(1,0,Zmean5,0,0,1,0,Zmean5), nrow=2, ncol=4)
A5 <- array(1, dim=c(j/5,1))
Z5 <- kronecker(A5,B5)
Z <- rbind(Z1,Z2,Z3,Z4,Z5)
```

In this example, a balanced design is assumed because there are equal numbers of time points assigned to each individual and equal number of individuals assigned to each level of treatment. However, researchers can conduct power analysis for unbalanced designs using this method by assigning varying numbers of time points to individuals or varying numbers of individuals across levels of treatment.

Step three: Obtain the approximate variance-covariance matrix

In order to simplify the syntax for calculating $\left(\sigma^2 * \Omega\right)^{-1}$, as shown in Equation 5, pre-define the identity matrix I and perform a calculation to obtain ψ, the block-diagonal matrix with J blocks of the 2 by 2 τ^2 matrix of variance components:

```
I <- diag(n*j)
I1 <- diag(j)
psi <- kronecker(I1,tausq)
```

Now $\left(\sigma^2 * \Omega\right)^{-1}$ can be obtained using the following code. The *a, b, c,* and *d* matrices correspond with the components of Equation 5. Use the solve command to perform matrix inversion.

```
a <- (pisq^-1)*I
b <- (pisq^-1)*(X %*%solve(t(X) %*% X)%*% t(X))
c <- (pisq)*(solve(t(X)%*%X))+ psi
d <- X%*%(solve(t(X)%*%X))%*%(solve(c))%*%(solve(t(X)%*%X))%*%t(X)
OmegaInv <- (a-b+d)
```

With this information, obtain the variance-covariance matrix of the parameter estimates using Equation 4.

```
e <- t(Z) %*% t(X) %*% OmegaInv %*% X %*% Z
Var_gamma <- solve(e)
```

After the covariance matrix has been obtained, the power estimate may be calculated by using the chi-square approximation. In the following code, compute

the non-centrality parameter $Z1$ by dividing the squared effect size by the relevant element from the covariance matrix and then obtaining the probability from the chi-square distribution with 1 degree of freedom. Interest lies in the power to detect the effect of γ_{01}, the parameter describing the effect of z_j on the outcome, which is the second of the four fixed effects. Its variance is represented by the (2, 2) entry of the variance-covariance matrix. It turns out the power to detect γ_{01} for an effect size of 1.0 under all the above assumptions is about 0.9. The following R code produces the power estimate:

```
Z1 <- (delta^2)/Var_gamma[2,2]
pchisq(3.841459, 1, Z1, lower.tail=FALSE)
```

Example Power Analysis for Two-Level Model where Covariate Values are Unknown

In the previous example, the values of the level-one and level-two covariates were known prior to data collection. The level-one covariate x in the growth model represented ten time points while the level-two covariate z represented five levels of treatment. Because the values were known, the X and Z design matrices could be constructed with the known values. However, in many cases the values of covariates are unknown prior to data collection. In this situation researchers will need to assume values for the means, variances, and covariances of the covariates in X and Z. The design matrices may then be constructed with values obtained from taking random draws from the appropriate univariate or multivariate distributions. In this section, power analysis for the model considered in the first example will be performed, but this time X and Z will contain continuous covariates with unknown values.

Step one: Assume necessary values. The assumed values for all model parameters will be the same as the previous example, except $x \sim N(0,1)$ and $z \sim N(0,1)$. The effect size of γ_{01} is 1.0, the level-one error variance, π^2, is 10. The variance-covariance matrix of the level-two random components u_j is $\begin{pmatrix} 5.0 & 1.0 \\ 1.0 & 4.0 \end{pmatrix}$ as before. The number of clusters, j, is 50. The number of repeated measures per cluster, n, is 10. The total sample size is 500. The R code is identical to that provided for the first example, with the exception of the creation of the X and Z matrices. New variables, however, will be added to hold the means and standard deviations of the covariates. If the model included more than

one covariate in X or Z, additional variables would be needed to contain their pairwise covariances.

```
meanx <- 0
sdx <- 1
meanz <- 0
sdz <- 1
```

Step two: Write out matrix forms of X and Z

Readers are referred to Equations 6 and 7 for scalar and matrix representations of the model. The status of covariates as known or unknown does not affect the representation of the model. The issue is the creation of X and Z with randomly drawn values. The following code will perform this task:

```
library(Matrix)
B <- list()
set.seed(1234)
for (i in 1:j) {
Bx1 <- rep(1, times=n)
Bx2 <- rnorm(n, mean=meanx, sd=sdx)
B[[i]] <- cbind(Bx1, Bx2)
}
```

This code loads the Matrix library and defines the object B as a list. A loop creates a design matrix for each j by creating a vector of ones to code the intercept and then making n draws from the normal distribution to determine plausible values in X. These blocks are stored in objects named $B[[1]]$ to $B[[j]]$. Now these must be assembled into the overall design matrix X which has a block-diagonal structure as shown in Equation 7. The random number seed ensures that repeated runs of the code will produce identical pseudo-random draws for x.

```
C <- list()
for (i in 1:j) {
if (i == 1) {C[[i]] <- B[[1]]}
else {C[[i]] <- bdiag(C[[i-1]], B[[i]])}
}
X <- C[[j]]
```

This code assembles the X matrix by adding one block at a time using the bdiag command from the Matrix package. A similar procedure will be used to create the Z matrix.

```
D <- list()
set.seed(4321)
```

```
for (i in 1:j) {
zj <- rnorm(1, mean=meanz, sd=sdz)
Dz1 <- c(1, zj, 0, 0)
Dz2 <- c(0, 0, 1, zj)
D[[i]] <- rbind(Dz1, Dz2)
}
E <- list()
for (i in 1:j) {
if (i == 1) {E[[i]] <- D[[1]]}
else {E[[i]] <- rbind(E[[i-1]], D[[i]])}
}
Z <- E[[j]]
```

The first loop creates j design matrices, stored in objects $D[[1]$ to $D[[j]]$. Because the both the level-one intercept and slope are regressed on the same the same variable, a single draw for z is used in both rows of the "D" matrix. The second loop binds all j matrices together into the complete Z. A different random number seed should be specified here so the random draws that provide values for z are not identical to the first j draws of x.

Step three: Obtain the approximate variance-covariance matrix

After X and Z are specified the variance-covariance matrix of fixed effects parameter estimates may be obtained using the same code used in the previous example. The (2, 2) entry of this matrix provides the approximate variance of the fixed effect γ_{01}. The following code performs this calculation:

```
I <- diag(n*j)
I1 <- diag(j)
psi <- kronecker(I1,tausq)
a <- (pisq^-1)*I
b <- (pisq^-1)*(X %*%solve(t(X) %*% X)%*% t(X))
c <- (pisq)*(solve(t(X)%*%X))+ psi
d <- X%*%(solve(t(X)%*%X))%*%(solve(c))%*%(solve(t(X)%*%X))%*%t(X)
OmegaInv <- (a-b+d)
e <- t(Z) %*% t(X) %*% OmegaInv %*% X %*% Z
Var_gamma <- solve(e)
```

If the assumed effect size is 1.0, then the power estimate is obtained by the following code:

```
Z1 <- (delta^2)/Var_gamma[2,2]
pchisq(3.841459, 1, Z1, lower.tail=FALSE)
```

The power estimate is about 0.60. It is important to note that when this approach is used there may be considerable sampling variation across runs in the

draws of x and z . The amount of sampling variability in x is much smaller than in z because there are nj x's but only j z's. This may lead to some between-run variability in the power estimate. It is recommended that researchers run the program several times with different random number seeds and average the power estimates across runs.

Power Analysis for Three-Level Models with Continuous Variables

Next is an outline of how to perform power analysis for a three-level model using the same method. First a general matrix formulation of a three-level HLM is provided.

$$Y = X\lfloor Z(W\omega+V)+U \rfloor+e \tag{8}$$

$$Y = XZW\omega + XZV + XU + e \tag{9}$$

Only $(XZV + XU + e)$ is random in Y .

$$Var(Y) = Var(XZV + XU + e)$$
$$= XZv^2Z'X' + X\tau^2X' + \pi^2 I = \sigma^2\Omega \tag{10}$$

In Equation 10, v^2 is the variance-covariance matrix of level-three random components; the remaining terms are defined as previously. The variance-covariance matrix of $\hat{\gamma}$ can be calculated using:

$$Var(\hat{\gamma}) = \left((XZW)'(\sigma^2\Omega)^{-1}(XZW) \right)^{-1} \tag{11}$$

Through simple derivation:

$$(\sigma^2\Omega)^{-1} = \begin{pmatrix} \pi^{-2}I - \pi^{-2}X(X'X)^{-1}X' \\ +X(X'X)^{-1}\left[\pi^2(X'X)^{-1} + Zv^2Z' + \tau^2 \right]^{-1}(X'X)^{-1}X' \end{pmatrix} \tag{12}$$

As the following example will illustrate, the remainder of the procedure for the power analysis in a three-level HLM follows the same logic as that in a two-level HLM.

Example Power Analysis for Three-Level Model

An example is provided to perform power analysis for a simple three-level model.

$$
\begin{aligned}
y_{ij} &= \beta_{0jk} + \varepsilon_{ij} \\
\beta_{0jk} &= \gamma_{00} + u_{0jk} \\
\gamma_{00k} &= \omega_{000} + \omega_{001}\omega_{1k} + v_{00k} \\
y_{ijk} &= \omega_{000} + \omega_{001}\omega_{1k} + u_{0jk} + v_{00k} + \varepsilon_{ijk} \\
v_{00k} &\sim N\left(0, v_{00}^2\right), u_{0jk} \sim N\left(0, \tau_{00}^2\right), \varepsilon_{ij} \sim N\left(0, \sigma^2\right)
\end{aligned}
\tag{13}
$$

The model could represent students clustered within classrooms and classrooms clustered within schools. The model contains two fixed effects, a grand-mean intercept and a single level-three covariate, presumed to be continuous, w_{1k}. Like the previous example, assume that the levels of w_{1k} are known prior to data collection. Sample code is provided only where there are marked differences from the previous example.

Step one: Assume necessary values

Assume that the effect size of ω_{001} is .20. The outcome is standardized with a total variance, σ^2, of 1.0. The within-cluster variance, π^2 is .80. The level-two variance, τ_{00}^2 is .10. The level-three variance, v_{00}^2, is also .10. The number of level-two units per level-three unit, J, is 5. The number of level-three units, K, is 30. The number of individuals per level-two unit, n, is 10, yielding a total sample size of 1,500.

Step two: *X*, *Z*, and *W*

The structures of *X*, *Z*, and *W*, based on the model equations, follow similar logic to the previous example. They are structured as follows:

$$
\begin{pmatrix} y_{111} \\ \vdots \\ \vdots \\ y_{nJK} \end{pmatrix} = \begin{pmatrix} 1 & & & & \\ \vdots & 0 & \cdots & 0 \\ 1 & & & & \\ 0 & \begin{matrix}1\\\vdots\\1\end{matrix} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \begin{matrix}\vdots\\1\end{matrix} \end{pmatrix} \left[\begin{pmatrix} 1 & & & & \\ \vdots & 0 & \cdots & 0 \\ 1 & & & & \\ 0 & \begin{matrix}1\\\vdots\\1\end{matrix} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \begin{matrix}\vdots\\1\end{matrix} \end{pmatrix} \left[\begin{pmatrix} 1 & w_{11} \\ \vdots & \vdots \\ 1 & w_{1k} \end{pmatrix} \begin{pmatrix} \omega_{000} \\ \omega_{001} \end{pmatrix} + \begin{pmatrix} v_{001} \\ \vdots \\ v_{00K} \end{pmatrix} \right] + \begin{pmatrix} u_{011} \\ u_{021} \\ \vdots \\ u_{0JK} \end{pmatrix} \right] + \begin{pmatrix} e_{111} \\ \vdots \\ e_{nJK} \end{pmatrix}
$$

$$(14)$$

Step three: Obtain the approximate variance-covariance matrix

In order to calculate $\left(\sigma^2 * \Omega \right)^{-1}$, as shown in Equation 12, use the code:

```
I <- diag(n*j*k)
a <- (pisq^-1)*I
b <- (pisq^-1)*(X %*%solve(t(X) %*% X)%*% t(X))
c <- (pisq)*(solve(t(X)%*%X))+(Z%*%Tausqv%*%t(Z)+Tausqu)
d <- X%*%(solve(t(X)%*%X))%*%(solve(c))%*%(solve(t(X)%*%X))%*%t(X)
OmegaInv <- (sigmasq*(a-b+d))
```

The variance-covariance matrix of the parameter estimates is obtained using Equation 11.

```
e <- t(W) %*% t(Z) %*% t(X) %*% OmegaInv %*% X %*% Z %*% W
VarW <- solve(e) * sigmasq
```

Now that the covariance matrix is obtained, the power estimate may be calculated by using the chi-square approximation. Interest lies in the power to detect the effect of ω_{001}. The final power estimate result is 0.68; the power estimate is obtained using:

```
Z1 <- (delta^2)/VarW[2,2]
pchisq(3.841459, 1, Z1, lower.tail=FALSE)
```

Power Analysis for Models with Dichotomous Predictors

Raudenbush & Liu (2000) described a simplified method of calculating the power to detect the effect of a dichotomous predictor. For example, this method would conveniently apply to intervention studies with two levels of treatment. Consider the following simple multilevel model with only one dichotomous fixed level-two variable and a random intercept to illustrate some of the issues involved in power calculation. Suppose a researcher is interested in whether an intervention helps participants improve their outcome scores (y_{ij}). *J* groups are randomly enrolled to have the intervention as the experimental group and *J* groups are randomly chosen to be the control group ($z_{0j} = 1$ if j^{th} group receive the intervention, otherwise the value is 0). There are *n* students from each group enrolled in the study. The researcher is interested in estimating the main effect of intervention (γ_{01}) on participants' outcome scores. The model is:

$$
\begin{aligned}
y_{ij} &= \beta_0 + \varepsilon_{ij} \\
\beta_0 &= \gamma_{00} + \gamma_{01} * z_{0j} + u_{0j} \\
y_{ij} &= \gamma_{00} + \gamma_{01} * z_{0j} + u_{0j} + \varepsilon_{ij} \\
u_{0j} &\sim N\left(0, \tau_{00}^2\right), \varepsilon_{ij} \sim N\left(0, \sigma^2\right)
\end{aligned}
\tag{15}
$$

Because z_{0j} is dichotomous ($z_{0j} = 0$ for all participants in the control group and $z_{0j} = 1$ for the treatment group), all observations in treatment and control groups can be summed respectively to

$$
\overline{y}_{ij\ treat} = \gamma_{00} + \gamma_{01} + \frac{\sum_{j=1}^{J} n * u_{0j\ treat}}{J * n} + \frac{\sum_{j=1}^{J}\sum_{i=1}^{n} \varepsilon_{ij\ treat}}{J * n}
\tag{16}
$$

$$
\overline{y}_{ij\ control} = \gamma_{00} + \frac{\sum_{j=1}^{J} n * u_{0j\ control}}{J * n} + \frac{\sum_{j=1}^{J}\sum_{i=1}^{n} \varepsilon_{ij\ control}}{J * n}
\tag{17}
$$

Subtracting Equation 17 from Equation 16, it is possible to cancel out γ_{00} and get

$$\bar{y}_{ij\ treat} - \bar{y}_{ij\ control} = \gamma_{00} + \frac{\sum_{j=1}^{J} n*\left(u_{ij\ treat} - u_{ij\ control}\right)}{J} + \frac{\sum_{j=1}^{J}\sum_{i=1}^{n}\left(u_{ij\ treat} - u_{ij\ control}\right)}{J*n} \quad (18)$$

If the expectation of $\bar{y}_{ij\ treat} - \bar{y}_{ij\ control}$ is taken, all random intercepts and residuals drop out because their expectations are all 0 according to the assumption. Finally this results in

$$E\left(\bar{y}_{ij\ treat} - \bar{y}_{ij\ control}\right) = E\left(\gamma_{01}\right) = \gamma_{01} \quad (19)$$

The fact that z_{01} is either 0 or 1, and ordinary assumptions about random slopes and residuals allow a simple unbiased estimator of γ_{01} to be derived. Because observations from treatment and control group are independent of each other, the property:

$$Var\left(\hat{\gamma}_{01}\right) = Var\left(\bar{y}_{ij\ treat} - \bar{y}_{ij\ control}\right)$$
$$= Var\left(\bar{y}_{ij\ treat}\right) + Var\left(\bar{y}_{ij\ control}\right) \quad (20)$$

is observed. Therefore, using Equations 18 and 20, results in

$$Var\left(\hat{\gamma}_{01}\right) = \left| \frac{\sum_{j=1}^{J} n*\left(u_{0j\ treat} - u_{0j\ control}\right)}{J} + \frac{\sum_{j=1}^{J}\sum_{i=1}^{n}\left(\varepsilon_{ij\ treat} - \varepsilon_{ij\ control}\right)}{J*n} \right| \quad (21)$$

According to the assumptions, $Var\left(u_{0j}\right) = \tau_{00}^2$ for all u_{0j} 's, and $Var\left(\varepsilon_{ij} = \sigma^2\right)$ for all ε_{ij} 's, Equation 21 can be expressed as:

$$Var\left(\hat{\gamma}_{01}\right) = \frac{\sum_{j=1}^{J} n*\left(\tau_{0j\ treat}^2 - \tau_{0j\ control}^2\right)}{J^2} + \frac{\sum_{j=1}^{J}\sum_{i=1}^{n}\left(\sigma_{ij\ treat}^2 - \sigma_{ij\ control}^2\right)}{n^2 * J^2}$$

$$= \frac{2*\left(\tau_{00}^2 + \dfrac{\sigma^2}{n}\right)}{J} \tag{22}$$

Given the derivation in Equation 22, the non-centrality chi-square distribution parameter of random quantity $\dfrac{\left(\hat{\gamma}_{01} - \gamma_0\right)^2}{\mathrm{var}\left(\hat{\gamma}_{01}\right)}$ can be estimated, where γ_0 is the parameter under the null hypothesis. For A two-level HLM with only a random intercept, the intra-class correlation p is defined as $\dfrac{\tau_{00}^2}{\tau_{00}^2 + \sigma^2}$. As a result,

$$\lambda \approx \frac{\left(\hat{\gamma}_{01} - \gamma_0\right)^2}{\mathrm{var}\left(\hat{\gamma}_{01}\right)} = \frac{\dfrac{\left(\hat{\gamma}_{01} - \gamma_0\right)^2}{\tau_{00}^2 + \sigma^2}}{\dfrac{2*\left(\tau_{00}^2 + \dfrac{\sigma^2}{n}\right)}{J*\left(\tau_{00}^2 + \sigma^2\right)}} = \frac{J*\delta^2}{2*\left(p + \dfrac{1-p}{n}\right)} \tag{23}$$

Where $\delta^2 = \dfrac{\left(\hat{\gamma}_{01} - \gamma_0\right)^2}{\tau_{00}^2 + \sigma^2}$, which is the standardized effect size. The same result is provided in Raudenbush, et al. (2004). Because researchers will need to use results from similar previous studies to obtain an assumed effect size and measurements in various studies may be measured on different scales, it makes sense to consider a standardized response variable. After standardization, the variance of the response variable is 1, which means $\tau_{00}^2 + \sigma^2 = 1$.

Therefore, $\delta = \hat{\gamma}_{01} - \gamma_0$. Equation 23 can be used to generate the proposed test statistic for power calculation.

Power Analysis for Functions of Parameters in HLM

Sometimes researchers are interested in the power to detect functions of parameters. For example, if a study considers three levels of treatment, detecting

differences between each pair of treatments may be the primary research question. Because the entire variance-covariance matrix will result for the parameter estimators, the power of detecting linear combinations of parameters can be easily calculated. For example, to calculate power of detecting the effect of $a\beta_1 + b\beta_2$ for constant scalars a and b. It is easy calculate the standard error of the linear combination.

$$S.E.\left(a\hat{\beta}_1 + b\hat{\beta}_2\right) = \sqrt{a^2 \hat{V}ar\left(\hat{\beta}_1\right) + b^2 \hat{V}ar\left(\hat{\beta}_2\right) + 2ab\hat{C}ov\left(\hat{\beta}_1, \hat{\beta}_2\right)} \qquad (24)$$

Then the assumed effect size of the two parameters and the calculated standard error of the linear combination of interested parameter estimators can be used to estimate power. Unfortunately this procedure does not apply to the simplified analytical method for dichotomous predictors because the whole variance-covariance matrix is not obtained. Referring back to the first example of the two-level growth model, if it is desirable to conduct power analysis for when $\gamma_{00} + \gamma_{01}z_j$ when $z_j = 3$ with additional assumption that the estimated effect size of γ_{00} is 0.50, then by substituting each term in Equation 24 by its corresponding numeric value and results in the power to detect the effect of the linear combination is almost 1. The following code illustrates this calculation:

```
Z2 <- (3.5*3.5)/(1.063333+9*.09666667-6*.29)
pchisq(3.841459, 1, Z2, lower.tail=FALSE)
```

In the cases where interest lies in power of detecting a nonlinear function of parameters, the Taylor expansion can be used to obtain an approximation of the variance of the nonlinear function of parameter estimators.

Discussion

This article outlined a method for approximating power for a wide variety of HLMs. A theoretical foundation for performing power analysis in two- and three-level models was presented. Examples including R code were provided, though any software that can carry out matrix computations and generate random variates may be used. This approach is very flexible and can easily be carried out for models whose power cannot be estimated (or is inconvenient to estimate) using currently available software. The method outlined can perform power analysis for three-level models with many different types of covariates.

One limitation of the usability of this approach for applied researchers is the requirement of writing out their models in matrix form. This may be unfamiliar to many researchers and could prevent widespread adoption of this method. In the future, it is hoped that software will be created to automate this process to simplify the implementation of this method and to broaden its appeal.

A second limitation of the approach is the sensitivity to sampling variation when covariate values are unknown. This approach may be thought of as a hybrid of numerical approximation and simulation. The sensitivity increases as the projected sample sizes decrease, which becomes more severe at higher levels of the model. To obtain power estimates robust to sampling variability, it will perhaps be necessary to perform many repetitions of the procedure and obtain a power estimate averaged across repetitions. In most software it is easy to automate multiple repetitions of the procedure to produce the desired stability in the power estimate.

A final limitation is that this method calculates power assuming values for all relevant parameters, such as sample sizes and effect sizes. However, when planning studies, researchers are often interested in determining either the sample size required to reach a given level of power or in the minimum detectable effect size given required power and a fixed sample size. In this case it would be set power equal to some value and solve for the parameter of interest (i.e., sample size or minimum detectable effect size). This article did not directly address these scenarios, as focus was placed on the calculation of power given all other parameter values. However, it is easy to repeat the procedure described in this paper multiple times, specifying a range of values for the parameter of interest in order to find the value of the parameter leading to the desired power.

References

Aitken, A. (1935). On least squares and linear combinations of observations. In *Proceedings of the Royal Society of Edinburgh*, *55*: 42-28. Farmington, Pennsylvania, USA.

Bosker, R. J., Snijders, T. A. B., & Guldemond, H. (1999). *PINT (Power IN Two-level Designs): Estimating Standard Errors of Regression Coefficients in Hierarchial Linear Models for Power Calculations: User's Manual*. Retrieved from: http://www.stats.ox.ac.uk/~snijders/Pint21_UsersManual.pdf.

Browne, W. J., Golalizadeh, M., & Parker, R. (2009). *A guide to sample size calculations for random effect models via simulation and the mlpowsim software package*. University of Bristol.

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

Cohen, J. (1992). A power primer. *Psychological bulletin*, *112*(1): 155-159.

Cools, W., Van den Noortgate, W., & Onghena, P. (2008). ML-DEs: A program for designing efficient multilevel studies. *Behavioral Research Methods, 40*(1): 236-249.

De Leeuw, J., & Kreft, I. (1986). Random coefficient models for multilevel analysis. *Journal of Educational and Behavioral Statistics, 11*(1): 57-86.

Hayashi, F. (2000). *Econometrics*. Princeton, NJ. Princeton University Press Princeton.

Lewis, K. (2006). Statistical power, sample sizes, and software to calculate them easily. *BioScience, 56*(7): 607-612.

Raudenbush, S. & Bryk, A. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oakes, CA: SAGE Publications Inc.

Raudenbush, S. & Liu, X. (2000). Statistical power optimal design for multisite randomized trials. *Psychological Methods, 5*(2): 199-213.

Raudenbush, S., Spybrook, J., Liu X., Congdon, R. (2004). *Optimal design for longitudinal and multilevel research: Documentation for the Optimal Design software*. http://hlmsoft.net/od/.

Singer, J. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics, 23*(4): 323-355.

Snijders, T., & Bosker, R., (1993). Standard errors and sample sizes for two-levels research. *Journal of Educational and Behavioral Statistics, 18*(3): 237-259.

Thomas, L., & Krebs, C. J. (1997). A review of statistical power analysis software. *Bulletin of the Ecological Society of America*, *78*(2): 126-138.