University of Nebraska - Lincoln

[DigitalCommons@University of Nebraska - Lincoln](https://digitalcommons.unl.edu/)

2009

Power of Models in Longitudinal Study: Findings From a Full-Crossed Simulation Design

Hua Fang University of Nebraska-Lincoln

Gordon P. Brooks Ohio University, brooksg@ohio.edu

Maria L. Rizzo Bowling Green State University, mrizzo@bgsu.edu

Kimberly A. Espy University of Nebraska-Lincoln, kespy2@unl.edu

Robert S. Barcikowski Ohio University

Follow this and additional works at: [https://digitalcommons.unl.edu/dcnlfacpub](https://digitalcommons.unl.edu/dcnlfacpub?utm_source=digitalcommons.unl.edu%2Fdcnlfacpub%2F48&utm_medium=PDF&utm_campaign=PDFCoverPages)

Part of the Neurosciences Commons

Fang, Hua; Brooks, Gordon P.; Rizzo, Maria L.; Espy, Kimberly A.; and Barcikowski, Robert S., "Power of Models in Longitudinal Study: Findings From a Full- Crossed Simulation Design" (2009). Developmental Cognitive Neuroscience Laboratory - Faculty and Staff Publications. 48. [https://digitalcommons.unl.edu/dcnlfacpub/48](https://digitalcommons.unl.edu/dcnlfacpub/48?utm_source=digitalcommons.unl.edu%2Fdcnlfacpub%2F48&utm_medium=PDF&utm_campaign=PDFCoverPages)

This Article is brought to you for free and open access by the Developmental Cognitive Neuroscience Laboratory at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Developmental Cognitive Neuroscience Laboratory - Faculty and Staff Publications by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

NIH Public Access

Author Manuscript

J Exp Educ. Author manuscript; available in PMC 2009 November 26.

Published in final edited form as:

J Exp Educ. 2009 April 1; 77(3): 215–254. doi:10.3200/JEXE.77.3.215-254.

Power of Models in Longitudinal Study: Findings From a Full-Crossed Simulation Design

HUA FANG,

University of Nebraska–Lincoln

GORDON P. BROOKS, Ohio University

MARIA L. RIZZO, Bowling Green State University

KIMBERLY ANDREWS ESPY, and University of Nebraska–Lincoln

ROBERT S. BARCIKOWSKI Ohio University

Abstract

Because the power properties of traditional repeated measures and hierarchical multivariate linear models have not been clearly determined in the balanced design for longitudinal studies in the literature, the authors present a power comparison study of traditional repeated measures and hierarchical multivariate linear models under 3 variance-covariance structures. The results from a full-crossed simulation design suggest that traditional repeated measures have significantly higher power than do hierarchical multivariate linear models for main effects, but they have significantly lower power for interaction effects in most situations. Significant power differences are also exhibited when power is compared across different covariance structures.

Keywords

covariance structure; hierarchical multivariate linear models; longitudinal study; power analysis; traditional repeated measures

> In longitudinal studies, both traditional repeated measures (TRM) and hierarchical multivariate linear models (HMLM) can be applied for a balanced design when the focus is testing fixed main effects. The balanced design assumes an equal number and spacing of measurements over time for each subject. TRM can be used for this design with univariate or multivariate approaches. When sphericity is met, the univariate tests are appropriate; when sphericity is not met, we can use adjusted univariate tests or traditional multivariate tests, which do not assume a given variance-covariance (VC) structure (cf. Greenhouse & Geisser, 1959; Huynh & Feldt, 1976; Jennrich & Schluchter, 1986; Wolfinger & Chang, 1995). For the same longitudinal design, HMLM treats the repeated observations nested within the subjects; that is, repeated measures at Level 1 and subjects at Level 2. A third or higher level of HMLM can be introduced to represent the contextual effects on the subjects' growth (Raudenbush & Bryk, 2002).

Address correspondence to Hua Fang, University of Nebraska–Lincoln, Room 102, 501 Building, Lincoln, NE 68588-0206, USA. jfang2@unl.edu.

(2)

HMLM and TRM are essentially interrelated in their theoretical development, especially after advanced computational methods were developed to handle missing values and estimate the VC structures (Dempster, Laird, & Rubin, 1977; Dempster, Rubin, & Tsutakawa, 1981; Goldstein 1995; Jennrich & Schluchter, 1986; Littell, Milliken, Stroup, & Wolfinger, 2006; Little, 1995; Little & Rubin, 2002; Maas & Snijders, 2003; McCulloch & Searle, 2001; Raudenbush & Bryk, 2002; Van der Leeden, Vrijburg, & de Leeuw, 1996). Jennrich and Schluchter were to first model specific VC structures directly through maximum likelihood estimation on the basis of traditional multivariate repeated measures approach, whereas HMLM incorporates Jennrich and Schlutchter's multivariate repeated measures approach to longitudinal data analysis (Raudenbush & Bryk; Jennrich & Schluchter; Schluchter, 1988; Van der Leenden, 1998). In the literature, HMLM is simply called *hierarchical linear models*, *multilevel models*, or more generally known as *generalized latent variable models* (e.g., Goldstein, 1994, 1995; Hox, 2002; Maas & Snijders, 2003; B. Muthén, 2002, 2004; L. Muthén & Muthén, 2006; Raudenbush & Bryk; Singer & Willett, 2003; Skrondal & Rabe-Hesketh, 2004).

The power analysis in longitudinal studies has been an active area but a uniform and standard criterion has not been established, especially based on the VC structures (cf. Hedeker, Gibbons, & Waternaux, 1999; Littell et al., 2006; Raudenbush, Spybrook, Liu, & Congdon, 2005; Snijders, 2005). Because the current analytical power approximations are not comprehensive or necessarily accurate (Littell et al.) and the power properties of TRM and HMLM have not yet been clearly compared in the balanced design, using a Monte Carlo (MC) simulation approach is efficient for examining their power properties simultaneously.

For parsimonious and exploratory purposes, TRM and three common VC structures were examined with the longitudinal data generated from a 2-level HMLM in this study. The three VC structures were (a) random slope with homogeneous Level 1 variance (RC), (b) unstructured (UN), and (c) first-order autoregressive (AR(1)).

Two-Level HMLM Model

The hypotheses tested in this simulation assumed the hypothesis of no fixed effects on the individuals' scores over time. The fixed effects were the two-group treatment effect (β_{01}) , time effect (β_{10}), and interaction (β_{11}). The underlying mathematical model for this simulation is as follows:

$$
Level 1: y_{ti} = \pi_{0i} + \pi_{1i}^* TIME + e_{ti}
$$
\n⁽¹⁾

Level 2:
$$
\pi_{0i} = \beta_{00} + \beta_{01}^*
$$
 TREATMENT+ u_{0i}
 $\pi_{1i} = \beta_{10} + \beta_{11}^*$ TREATMENT+ u_{1i} ,

where y_{ti} represents the score of person *i* at time *t*; π_{0i} is the score of person *i* at time 0; π_{1i} refers to the slope of person *i* (i.e., rate of change with respect to time); β_{00} is the average overall initial score at time 0; β_{01} stands for the hypothesized difference in average status from the effect of treatment; β_{10} is the average overall annual rate of change at Level 2; β_{11} represents the hypothesized difference in average annual rate of change from the effect of treatment; u_{0i} is the random effect for intercepts (i.e., random error of intercepts at Level 2); u_{1i} is the random effect for slopes (i.e., random error of slopes at Level 2); *eti* refers to the random error at the t^{th} time point of the i^{th} person at Level 1.

The aforementioned two-level model can be reduced to a single-level model by substituting Equation (2) into Equation (1):

$$
y_{ti} = (\beta_{00} + \beta_{01} \times \text{TREATMENT} + \beta_{10} \times \text{TIME} + \beta_{11} \times \text{TREATMENT} \times \text{TIME}) + r_{ti},
$$
\n(3)

where the residual term, $r_{ti} = u_{0i} + u_{Ii} * TIME + e_{ti}$ includes the Level-1 random error (e_{ti}) and Level-2 random effects $(u_{0i}$ and u_{1i}); β_{00} , β_{01} , β_{10} , β_{11} , u_{0i} , u_{1i} , and e_{ti} are the same as those in Equation (1) and Equation (2). Hence, the HMLM model is also expressed as the mixed effect model with a mix of fixed effects in the parentheses and random effects embodied in the residual term *rti*.

TRM and Three Covariance Structures Under Study

The TRM approach to Equation (3) can be simply expressed in a matrix form:

$$
Y = X\beta + r,\tag{4}
$$

where *Y* is a $t_i \times 1$ response vector for subject *i*, *t* represents the number of time points, and *i* $= 1, ..., n$; *X* is a *t*_{*i*} × *a* design matrix for fixed effect β, where *a* is the number of fixed effects (i.e., the three parameters, β_{01} , β_{10} , and β_{11} in this study), and β is an $a \times 1$ vector; residual *r* is independently and normally distributed with a mean vector of 0 and variance of Σ , $r \sim N(0,$ Σ). The parameter estimates in the traditional approach are obtained using the method of moments (McCulloch & Searle, 2001; Montgomery, 2005; Wolfinger & Chang, 1995).

Random Slope With Homogeneous Level-1 Variance (RC)

Random slope with homogeneous Level-1 variance is often described as the covariance structure for standard MLM, also known as *standard hierarchical linear model* (HLM) or *random coefficient model* (RC; Raudenbush & Bryk, 2002, p. 191; Raudenbush, Bryk, & Congdon, 2004; Singer & Willett, 2003, p. 244–245, 251–265; Kreft, 1996). For convenience, RC is used for this covariance structure hereafter. The RC covariance structure of Model (3) residual r_{ti} , Σ_r , is expressed as two components:

$$
e_{ti} \sim N(0, \sigma^2)
$$
, and $\begin{pmatrix} u_{0i} \\ u_{1i} \end{pmatrix} \sim N \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{pmatrix}.$

The variance of Level-1 error term (e_i) is homogeneous and the covariance structure of Level-2 random errors $(u_{0i}$ and u_{1i}) is arbitrary. For Model (3), only 4 VC parameters need to be estimated; that is, σ^2 , τ_{00} , τ_{11} , and τ_{01} . Level-1 variance, σ^2 , is independent of Level-2 variance, τ.

Unstructured Covariance Matrix (UN)

The unstructured covariance matrix (also called *unrestricted structure* in the literature) places no restrictions on the structure of covariance matrix, Σ_r , and there is redundancy in mathematical formulation of this covariance structure (Littell, Henry, & Ammerman, 1998, pp. 1229–1230; Raudenbush et al., 2004). If the covariance structure of Σ*^r* is assumed to be unknown, one could fit a UN covariance matrix. The UN matrix for each Level-2 subject with 3 time points can be expressed as

$$
\begin{pmatrix}\n\sigma_{11}^2 & \sigma_{12}^2 & \sigma_{13}^2 \\
\sigma_{21}^2 & \sigma_{22}^2 & \sigma_{23}^2 \\
\sigma_{31}^2 & \sigma_{32}^2 & \sigma_{33}^2\n\end{pmatrix}
$$
\n(5)

and requires the estimation of three variance parameters and three covariance parameters. In general, the number of parameters to estimate in the UN approach is of order d^2 for a *d* by *d* covariance matrix, which may be computationally excessive when a large number of time points are involved.

First-Order Auto-Regressive (AR(1))

For Model (3), AR(1) can be written as follows:

$$
Var(r_{ti}) = \tau + \sigma^2
$$

\n
$$
Cov(r_{ti}, r_{t'i}) = \tau + \sigma^2 \rho^{|t-t'|},
$$
\n(6)

where τ stands for the Level-2 variance and $|t-t'|$ is the lag between two time points; ρ is the auto-correlation and σ^2 is the Level-1 variance at each time point. AR(1) allows the Level-1 errors to be correlated under Markov assumptions and Level-1 covariance structure is expressed as

$$
\sigma^2 \left(\begin{array}{ccc} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{array} \right) \tag{7}
$$

As the redundancy is not in the mathematical formulation of $AR(1)$, the covariance structure of Level-2 random effects $(u_{0i}$ and u_{1i}) must be specified to estimate the Level-2 variance τ $(\tau_{00}, \tau_{11}, \text{ and } \tau_{01})$, which is usually assumed unstructured (cf. Littell et al., 1998, pp. 1229– 1230; McCulloch & Searle, 2001; Raudenbush & Bryk, 2002; Singer & Willett, 2003; Wolfinger, 1993). Thus, five VC parameters need to be estimated for AR(1) of Model (3).

Method

MC Design

This research used a MC study to compare the empirical power of RC, UN, and AR(1). To make the results applicable over many possible situations (Raudenbush & Liu, 2000), a standardized model, Model (4), was employed in this simulation where the grand mean in Model (3) was set to zero (i.e., $\beta_{00} = 0$),

> $y_{ti} = (\beta_{01} \times \text{TREATMENT} + \beta_{10} \times \text{TIME} + \beta_{11})$ \times TREATMENT \times TIME)+ r_{ti} ,

(8)

A stacked SAS macro was written by Fang (2006) to generate the two-level repeated measures data with the RC covariance structure and to calculate the power for RC, UN, and AR(1). The number of iterations for this MC study was 10,000 and the nominal alpha for each sample test was .05. RC was selected as the underlying covariance structure mainly in view of the efficiency and complexity of this MC design and future expansion of the data generator. For example,

more parameters would be required to specify for UN and AR(1) than RC for the same data generation. Also, the redundancy problem of UN mentioned earlier would increase the complexity for the design of the data generator. In fact, choosing either of the three covariance structures would mimic the real application situation in which researchers would usually not know the real underlying covariance structure in advance but would need to fit one for their hierarchical data during analyses. Thus the choice of an appropriate covariance structure is at issue, and this power comparison study naturally becomes necessary.

Data Generation

The data generation procedure based on Model (3) was carried out as follows:

Level-1 Data—The error term at Level 1 (i.e., e_{ti}) was assumed to be independent of the Level-2 random effects (i.e., u_{0i} and u_{Ii}), that is, $cov(u_i, e_{ti}) = 0$. The Level-1 error term followed a normal distribution, $e_{ti} \sim N(0, \sigma^2)$.

Level-2 Data—The random intercepts u_{0i} ($X_{intercepts}$) and slopes u_{1i} (X_{slopes}) assumed a standard bivariate normal distribution. A standardized *G* matrix for *Xintercepts* and *Xslopes*,

$$
G=\left(\begin{array}{cc}\sigma_{00}^2 & \sigma_{01}^2\\ \sigma_{10}^2 & \sigma_{11}^2\end{array}\right),\text{ and random mean vector, }\left(\begin{array}{c}\mu_0\\ \mu_1\end{array}\right),
$$

were specified to simulate correlated bivariate normal data for *Xintercepts* and *Xslopes*. The Cholesky decomposition method was utilized to generate the correlated Level-2 normal data. This simulation was accomplished by multiplying the normal data by *L*, which is the Cholesky decomposition of *G*. The estimated variables were $\hat{X}_{intercepts}$ and \hat{X}_{slopes} .

Complete Data—Data were generated in the appropriate format required by PROC MIXED (SAS Institute, 2003). An index matrix was created for time, treatment, and individuals (IDs). Two treatment groups were coded by 0 and 1, respectively. IDs were considered nested within each treatment group; for instance, IDs ranged from 1 to 25 for Group 1, and 26 to 50 for Group 2. Time started from 0 and extended to the maximum specified for each study condition. On the basis of Model (4), a univariate response vector of y_{ti} was created. For example, each subject might have had three time points and each treatment group 25 subjects.

The data generator (Fang, 2006; Fang, Brooks, Rizzo, & Barcikowski, 2006) was validated with parameter estimates from Potthoff and Roy's (1964) data. The results (see the Appendix) show that the estimates from simulated data are slightly different in the third or fourth decimal from the true parameters. Also, as suggested by Littell et al. (2006), under the null hypothesis of no treatment effect, the *p* values based on the simulated data have an approximate uniform distribution, which further validates the data generator.

Power Comparison

When the time points are fixed, this study implemented a full-crossed design for power comparison; that is, the power was compared across the levels of the four factors: (a) correlation in *G* matrix (*G*), (b) reliability of Level-1 coefficients (λ), (c) effect size (β), and (d) sample size per treatment group (n) . The decision was made to fix the time points at 3 because of three main facets. First, three time points are practically the base requirement for longitudinal studies, although theoretically two time points would suffice (McCulloch & Searle, 2001). Second, a full-crossed design based on three time points lays the foundation for the second step where we fixed the total sample size while varying the sample size per treatment group and the number

of time points (i.e., the ratio of *n* to *t*). Third, in many real-world studies, it is simply not feasible to collect data at more than three time points due to costs, particularly in applied settings such as educational studies, clinical trials, and epidemiological research found in other social and physical science areas.

As shown in Figure 1, we used Cohen's indexes (1988) for correlation, $\rho \in \{.1, .3, .5\}$; correspondingly, G matrix (G) for random intercepts (u_{0i}) and slopes (u_{1i}) was specified as

 $\left(\begin{array}{cc} 1 & .1 \\ .1 & 1 \end{array}\right)$, $\left(\begin{array}{cc} 1 & .3 \\ .3 & 1 \end{array}\right)$, or $\left(\begin{array}{cc} 1 & .5 \\ .5 & 1 \end{array}\right)$;

the averaged reliability of Level-1 coefficients was set to vary from .01 to 1 by .25, $\lambda \in \{.01$. 25 .5 .75 1} (Raudenbush & Bryk, 2002; Raudenbush et al., 2004); the effect sizes also used Cohen's indexes for the two-group treatment effect (β_{01}), time effect (β_{10}), and interaction $(\beta_{11}), \beta \in \{.2, .5, .8\};$ the sample size per treatment group changed from 25 to 200 by 25 (*n* \in {25 75 100 125 150 175 200}). On the basis of this design, a general power pattern of TRM, RC, UN, and AR(1) were presented at each level of the four factors for the three tests (i.e., 4 \times 3 \times 5 \times 3 \times 7 \times 3 = 3,780 cells).

For exploratory purpose, we also compared the power of the four models by holding the total sample size (*N*) at 800 and varying the ratio of the treatment group size to the number of time points under a specified condition in which the reliability was fixed at .50. The sample size ratio was set at 100/4, 80/5, and 40/10, respectively. The power of the four models was examined across different levels of the correlation in *G* matrix and effect sizes for the three tests (i.e., $4 \times 3 \times 3 \times 3 \times 3 = 324$ cells; see Figure 1).

MC Analysis

We used the following function to calculate the upper bound of standard errors for pairwise empirical power (i.e., the standard error for the difference in proportions).

$$
SE_{upper} = \sqrt{2 \times \frac{p \times (1-p)}{n}},
$$
\n(9)

where $p = .5$ and $n = 10,000$. If the pairwise differences are twice the upper bound of *SE* (i.e., $SE_{\text{upper}} \cong .007$, then the differences are labeled as significant. The power patterns are illustrated in tables and graphs.

Results

The results shown in Tables 1–3 indicate a general pattern when varying the *G* matrix in all three tests, treatment effect (β_{01}), time effect (β_{10}) and interaction (β_{11}). The pattern showed that as the correlation in *G* matrix increases, the power of TRM, RC, UN, and AR(1) decreased slightly, which may imply that the lower the correlation between intercepts and slopes, the higher the power we can obtain. But it should be noted that the power change across *G* matrixes seems to be visually minimal. Because of this small difference among the four models, Figures 2–4 merely present the power pattern at *G* = .1, whereas Tables 1–3 provide the entire power information by varying *G* matrixes.

Empirical Power in Treatment Test

By fixing the number of time points at 3, this power study indicated that in the treatment test, the TRM power was significantly higher than the three HMLM models across different levels of sample sizes, *G* matrixes, reliability, and effect sizes, except when the reliability is very low, around .01, or when the reliability is perfect at 1 and effect size at .8 (see Table 1a). As displayed in Figure 2, the power of all four models is below .20 at the reliability of .01 and the power difference among the four is not statistically different based on the criteria discussed earlier (i.e., twice the upper bound of SE , $2 \times SE = .014$), even when the sample size changes to 200 and the effect size becomes .8 (see Table 1b).

Although the pairwise power difference among the four models seems to be visually indistinguishable at the reliability of 1, the power of four models is significantly different from each other across the sample sizes; correlation in the *G* matrix and effect sizes, except the difference between TRM and UN vanishes at the reliability of 1 (see Figure 2 and Table 1b). The pairwise power difference also displays that the power of UN, AR(1), and RC generally keeps the same descending order but is not significantly different across all levels of the four factors, especially at the reliability of .01 (see Figure 2a and Table 1b).

When the total sample size is fixed at 800 and reliability at .50, the results show that the power of the four models decreases as the ratio of treatment group size to the number of time points decreases (see Figure 2b). In other words, it seems that when the total sample size is held constant, the higher power is easily achieved by increasing the treatment group size rather than the number of time points in detecting the treatment effect (i.e., increasing the sample size at a higher level rather than the lower level, using the terminology of HMLM). Table 3 indicates that in the treatment test, the power at 100/4 is at least 1.5 times the power at 80/5 and at least twice the power at 40/10 for all four models. Under this setting, TRM power is still significantly higher than the HMLM models across the sample sizes, effect sizes, and correlation in the *G* matrix, except that TRM and UN power seem to be similar at the sample size ratio of 40/10 or when the effect size is small. Among the three HMLM models, UN power is still higher than the other two, and AR(1) is higher than RC, except at the ratio of 40/10.

Empirical Power in Interaction Test

The interaction test with three fixed time points indicates that the TRM power is generally significantly lower than HMLM models at the medium or large effect sizes with the reliability between .25 and .75 (see Table 2). At the reliability of 1, RC power is not always significantly higher than TRM across the samples sizes, effect sizes, or the correlation in the *G* matrix, whereas UN and AR(1) power are generally significantly higher than TRM. At a very low reliability of .01, the power of all four models is below .20 across the power factors and the pairwise power differences seem not to be identifiable (see Figure 3a).

Among the three HMLM models, UN power ranks the highest, AR(1) the second, and RC the lowest across the factors. At the medium or large effect sizes, the pairwise power difference among the three VC structures is always significant, except when the reliability is at or below . 25. At the small effect size, the pairwise differences among the three VC structures are significant, except when the reliability is at or below .50.

Holding the total sample size at 800 and the reliability at .50, the power pattern between TRM and HMLM models in the interaction test seems to be magnified (see Figure 3b). The power of UN, AR(1), and RC is still higher than TRM across the sample sizes, effect sizes, and the correlation in the *G* matrix. The pairwise differences among the four models are generally significant except with the power difference between RC and TRM at the small effect size and with the pairwise differences among the three HMLM models at a large effect size of .8 and

small correlation of .1 in the *G* matrix, as their power is all close to the asymptote of 1 (see Table 3).

Empirical Power in Time Test

As shown in Figure 4a, TRM and the three HMLM models are still not comparable at the reliability of .01. At the small or medium effect size, TRM power is significantly higher than the three HMLM models across sample sizes and the correlation indexes in *G* matrix when the reliability is at or above .25; among the three VC structures, UN is significantly higher than the other two, and AR(1) is higher than RC when the reliability is at or above .50. At the large effect size, as the power of all four models are close to the asymptote of 1, TRM power seems to be higher than the three HMLM models only when the sample size is below 75 and the reliability above .25.

Figure 4b indicates that while holding the total sample size at 800 and the reliability of .50, TRM power is higher than the other three models in the time test across the sample size ratios at the small or medium effect size, except at the ratio of 40/10 where UN is higher than TRM. At the large effect size, the power of all four models is close to 1 across the specified sample size ratios, and no significant pairwise power difference is found.

Discussion and Conclusions

In the present MC study, we were primarily concerned with the empirical power of TRM and HMLM under three VC structures in the longitudinal study. Specifically, we compared the power of TRM, RC, AR(1), and UN in three tests, two-group treatment effect (β_{01}), time effect (β_{10}) , and time-by-treatment interaction (β_{11}) , under the balanced design in longitudinal studies, assuming the underlying covariance structure is RC. The four factors in this power study are the *G* matrix (*G*), reliability (λ), effect size (β), and ratio of sample size per treatment group (*n*) to the number of time points (*t*).

Generally, the findings from this full-crossed design based on the fixed time points or the fixed total sample size show that TRM, in the treatment or time test, is more powerful than the HMLM under the three VC structures except for the situation in which the reliability is close to 0 or perfect at 1. Yet, in reality, the two extreme conditions are rare. This study also indicates that in the treatment test, it is easier to achieve the higher power by increasing the treatment group size rather than the number of time points, which corresponds to the discussion in the literature (e.g., Kreft, 1996; Raudenbush & Bryk, 2002). For example, the increase of time points typically would comparatively incur more research cost by following up on the subjects, and it would probably confront more participants' mobility, dropout, and other attrition problems and likely result in greater missing data, which would further complicate the research.

As for the interaction test, researchers have questioned what the power is to detect the interactions when they exist in the HMLM data. They expected HMLM to perform better than traditional models, but they did not find such proof (Davison, Kwak, Seo, & Choi, 2002; Kreft, 1996; Raudenbush, 1995). This study provided empirical power estimates in the interaction test for both TRM and HMLM. One of the interesting findings in this power study indicates that TRM has significantly lower power than the other three HMLM models—RC, AR(1), and UN—in the interaction test, except in the two extreme reliability conditions (close to 0 or 1).

From this study, we noticed that the power can be significantly different among different VC structures when using the HMLM models in the longitudinal study. In addition to referring to the model fit statistics (Akaike, 1973; Littell et al., 2006; Pinheiro & Bates, 2000; Schwarz, 1978; Singer & Willett, 2003), the empirical power results from this study could be a reference source when applying HMLM models. Also from these empirical results, practitioners may

estimate the sample sizes, the reliability, effect size, or the correlation in the *G* matrix for their studies if scenarios are similar to this study. On the basis of this study, TRM could be the choice if researchers are further interested in main effect tests and the practical situation is most similar to this research in which the balanced design is assumed and fixed effects are primarily the concern. If researchers are more concerned with interaction tests (i.e., the group difference in the change rate over time using the HMLM terminology), this study recommends that UN, AR (1), or RC be the method of choice. UN could be the best among the three HMLM models if the practical situation is most similar to this research and if there is a need to try an exploratory analysis when the VC structure is assumed unknown. In practice, if the candidate models are appropriate for a given analysis, practitioners tend to choose the more powerful model, which is reasonable. However, we encourage the thoughtful and responsible application of TRM and HMLM in recommended tests similar to this balanced design in order to properly contextualize the value of our findings.

Because no uniform agreement was made regarding the typical values for the settings, such as the effect sizes and the correlation among random effects in different applied research areas (e.g., education, psychology, clinical trials, and biology, etc.), we used the traditional Cohen's indexes, which are at least widely accepted in social science research. On the basis of our post hoc demonstration, the change of the sign of these indexes would not affect the power pattern among the four models compared in this study, although the magnitude of the power of these models changed slightly. However, we would recommend that practitioners closely collaborate with their project colleagues to make full sense of the results generated from this study. If there is a strong rationale that the expected indexes differ from the present study, the values should be changed to generate corresponding output using our macro.

Future studies may consider extending this MC study by comparing power at different total sample sizes across more variable ratios of sample size per group to the number of time points to examine whether the power differences could diminish or increase at the low ratio. Also, the data generator could be modified to further investigate the power differentiation under other underlying covariance structures. Instead of reliability, interclass correlation could be considered in the power analysis. Although the magnitude of power difference and power decreasing or increasing rates can vary, the general power patterns among TRM and the three VC structures are expected to be similar to this study. The HMLM data generator and power comparison macro (Fang, 2006; Fang et al., 2006) could be expanded to generate missing data or nonnormal longitudinal data in order to be more practical and to examine the statistical properties and power of more complex growth models.

Acknowledgments

This research was supported in part by R01 DA014661 from the National Institute on Drug Abuse (Espy, PI), R01 MH065668 from the National Institute of Mental Health (Espy, PI), and R01 HD050309 (Taylor, PI) from the National Institute of Child Health and Development. The previous support for the first author by the Voinovich School of Leadership and Public Affairs of Ohio University is also gratefully acknowledged.

References

- Akaike, H. Information theory as an extension of the maximum likelihood principle. In: Petrov, BN.; Csaki, F., editors. Second international symposium on information theory. Budapest, Hungary: Akademiai Kiado; 1973. p. 267-228.
- Cohen, J. Statistical power analysis for the behavioral sciences. Mahwah, NJ: Erlbaum; 1988.
- Davison ML, Kwak N, Seo YS, Choi J. Using hierarchical linear models to examine moderator effects: Person-by-organization interactions. Organizational Research Methods 2002;5:231–254.
- Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society 1977;39(B):1–8.

- Dempster AP, Rubin DB, Tsutakawa RK. Estimation in covariance components models. Journal of the American Statistical Association 1981;76:341–353.
- Fang, H. %hlmdata and %hlmrmpower: Traditional repeated measures vs. HLM for multilevel longitudinal data analysis: Power and type I error rate comparison. Proceedings of the Thirty-First Annual SAS Users Group Conference; Cary, NC: SAS Institute; 2006.
- Fang, H.; Brooks, GP.; Rizzo, ML.; Barcikowski, RS. An empirical power analysis of multilevel linear model under three covariance structures in longitudinal data analysis. 2006 Proceedings of the American Statistical Association, Section on Quality and Productivity; Alexandria, VA: American Statistical Association; 2006. [CD-ROM]

Goldstein H. Multilevel cross-classified models. Sociological Methods & Research 1994;22:364–376. Goldstein, H. Multilevel statistical models. Vol. 2. London: Edward Arnold; 1995 .

- Greenhouse SW, Geisser S. On methods in the analysis of profile data. Psychometrika 1959;32:95–112.
- Hedeker D, Gibbons RD, 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(1):70–93.
- Hox, JJ. Multilevel analysis techniques and applications. Mahwah, NJ: Erlbaum; 2002.
- Huynh H, Feldt LS. Estimation of the box correction for degrees of freedom from sample data in the randomized block and split plot designs. Journal of Educational Statistics 1976;1:69–82.
- Jennrich RI, Schluchter MD. Unbalanced repeated-measures models with structured covariance matrices. Biometrics 1986;42:805–820. [PubMed: 3814725]
- Kreft, IGG. Are multilevel techniques necessary? An overview, including simulation studies. 1996. Retrieved August 13, 2005, from<http://www.calstatela.edu/faculty/ikreft/quarterly.html>
- Littell RC, Henry PR, Ammerman CB. Statistical analysis of repeated measures data using SAS procedures. Journal of Animal Science 1998;76:1216–1231. [PubMed: 9581947]
- Littell, RC.; Milliken, GA.; Stroup, WW.; Wolfinger, RD. SAS system for mixed models. Vol. 2. Cary, NC: SAS Institute; 2006.
- Little RJA. Modeling the dropout mechanism in repeated-measures studies. Journal of the American Statistical Association 1995;90:1112–1121.
- Little, RJ.; Rubin, DB. Statistical analysis with missing data. Vol. 2. New York: John Wiley; 2002.
- Maas CJM, Snijders TB. The multilevel approach to repeated measures for complete and incomplete data. Quality & Quantity 2003;37:71–89.
- McCulloch, CE.; Searle, SR. Generalized, linear, and mixed models. New York: Wiley; 2001.
- Montgomery, DC. Design and analysis of experiments. Hoboken, NJ: Wiley; 2005.
- Muthén B. Beyond SEM: General latent variable modeling. Behaviormetrika 2002;29:81–117.
- Muthén, B. Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In: Kaplan, D., editor. Handbook of quantitative methodology for the social sciences. Newbury Park, CA: Sage; 2004. p. 345-368.
- Muthén, L.; Muthén, B. Mplus user's guide. Vol. 4. Los Angeles: Author; 2006.
- Pinheiro, JC.; Bates, DM. Statistics and computing: Mixed-effects models in S and S-PLUS. New York: Springer; 2000.
- Potthoff RF, Roy SN. A generalized multivariate analysis of variance model useful especially for growth curve problems. Biometrika 1964;51:313–326.
- Raudenbush SW. Maximum likelihood estimation for unbalanced multilevel covariance structure models via the EM algorithm. British Journal of Mathematical and Statistical Psychology 1995;48:359–370.
- Raudenbush, SW.; Bryk, AS. Hierarchical linear models: Applications and data analysis methods. London: Sage; 2002.
- Raudenbush, SW.; Bryk, AS.; Congdon, R. HLM6: Hierarchical linear and nonlinear modeling. Chicago: Scientific Software International; 2004.
- Raudenbush SW, Liu X. Statistical power and optimal design for multisite randomized trials. Psychological Methods 2000;5:199–213. [PubMed: 10937329]

NIH-PA Author Manuscript NIH-PA Author Manuscript

- Raudenbush, SW.; Spybrook, J.; Liu, X.; Congdon, R. Optimal design for longitudinal and multilevel research: Documentation for the "Optimal Design" software. 2005. Retrieved September 20, 2006, from http://sitemaker.umich.edu/group-based/optimal_design_software
- SAS Institute. SAS/STAT user's guide, version 9.1. Cary, NC: Author; 2003.
- Schluchter, MD. Unbalanced repeated measures models with structured covariance matrices. In: Dixon, WJ., editor. BMDP statistical software manual. Vol. 2. Berkeley, CA: University of California Press; 1988. p. 1081-1114.
- Schwarz G. Estimating the dimensions of a model. Annals of Statistics 1978;6:461–464.
- Singer, JD.; Willett, JB. Applied longitudinal data analysis. New York: Oxford University Press; 2003.
- Skrondal, A.; Rabe-Hesketh, S. Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models. New York: Chapman & Hall/CRC; 2004.
- Snijders, TAB. Power and sample size in multilevel linear models. In: Everitt, BS.; Howell, DC., editors. Encyclopedia of statistics in behavioral science. Vol. 3. Chicester, England: Wiley; 2005. p. 1570-1573.
- Van der Leeden R, Vrijburg K, de Leeuw J. A review of two different approaches for the analysis of growth data using longitudinal mixed linear models: Comparing hierarchical linear regression (ML3, HLM) and repeated measures design with structured covariance matrices (BMDP-5V). Computational Statistics and Data Analysis 1996;21:583–605.
- Van der Leeden R. Multilevel analysis of repeated measures data. Quality & Quantity 1998;32:15–29.
- Wolfinger RD. Covariance structure selection in general mixed models. Communications in Statistics, Simulation and Computation 1993;22:1079–1106.
- Wolfinger, RD.; Chang, M. Comparing the SAS GLM and MIXED procedures for repeated measures. SUGI Proceedings; Cary, NC: SAS Institute; 1995.

Biographies

Hua Fang is a research assistant professor at the University of Nebraska–Lincoln. Her research interests are hierarchical linear models, structural equation models, growth mixture models, power analysis, missing data analysis, computational statistics, cluster analysis, and classification.

Gordon P. Brooks is an associate professor of educational research and evaluation and Chair of the Department of Educational Studies at Ohio University. His primary research interests are statistics education, power and sample size analysis, and Monte Carlo programming.

Maria L. Rizzo is an assistant professor and director of actuarial science in the Department of Mathematics and Statistics at Bowling Green State University. Her research interests are multivariate analysis, goodness-of-fit tests, cluster analysis and classification, and computational statistics.

Kimberly Andrews Espy is a professor at the University of Nebraska–Lincoln. Trained as a neuroscientist, she studies how young children and infants develop regulatory skills and how that process goes awry in various medical conditions.

Robert S. Barcikowski is a professor emeritus of educational research and evaluation at Ohio University. His primary research interests hierarchical linear models, multivariate statistics, and Monte Carlo programming. Fang, Brooks, Rizzo, Espy, & Barcikowski 253

FIGURE 1.

Monte Carlo (MC) design for power analysis of TRM, RC, AR(1), and UN by *G* matrix, effect size, and sample size of 10,000 MC samples at $\alpha = .05$. TRM = traditional repeated measures; $RC =$ random slope with homogeneous level-1 variance; $AR(1) =$ first-order autoregressive; $UN =$ unstructured. ^aThe factors are crossed. ^bConditions are specified for testing each fixed effect (β_{01} , β_{10} , and β_{11}) across levels of factors. ^{ca} Fixed" indicates the fixed parameters in the design. dSettings for testing the three fixed effects.

FIGURE 2.

Power patterns of variance-covariance structures, random slope with homogeneous Level-1 variance (RC), first-order autoregressive (AR(1)), unstructured (UN), and traditional repeated measures (TRM) at α = .05 of 10,000 MC samples in treatment test. For Figure 2a, $t = 3.0$ and *G* = 0.1 in treatment test. For Figure 2b, $N = 800$ and $\lambda = .50$ in treatment test.

 $\mathbf b$

FIGURE 3.

Power patterns of variance-covariance structures, random slope with homogeneous Level-1 variance (RC), first-order autoregressive (AR(1)), unstructured (UN), and traditional repeated

measures (TRM) at α = .05 of 10,000 MC samples in the interaction test. For Figure 3a, $t = 3.0$ and $G = 0.1$ in interaction test. For Figure 3b, $N = 800$ and $\lambda = .50$ in interaction test.

a

FIGURE 4.

Power patterns of variance-covariance structures, random slope with homogeneous Level-1 variance (RC), first-order autoregressive (AR(1)), unstructured (UN), and traditional repeated measures (TRM) at α = .05 of 10,000 MC samples in time test. For Figure 4a, $t = 3.0$ and $G =$ 0.1 in time test. For Figure 4b, $N = 800$ and $\lambda = .50$ in time test.

FANG et al. Page 18

TABLE 1a. Power Pattern of TRM, RC, AR(1), and UN When $t = 3$ of 10,000 MC Samples at $a = .05$ in Treatment Test **TABLE 1a. Power Pattern of TRM, RC, AR(1), and UN When** *t* **= 3 of 10,000 MC Samples at α = .05 in Treatment Test**

FANG et al. Page 19

NIH-PA Author Manuscript NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

 NIH-PA Author Manuscript NIH-PA Author Manuscript TABLE 1a. Power Pattern of TRM, RC, AR(1), and UN When $t = 3$ of 10,000 MC Samples at $a = .05$ in Treatment Test **TABLE 1a. Power Pattern of TRM, RC, AR(1), and UN When** *t* **= 3 of 10,000 MC Samples at α = .05 in Treatment Test**

FANG et al. Page 20

 NIH-PA Author ManuscriptNIH-PA Author Manuscript

 NIH-PA Author Manuscript NIH-PA Author Manuscript

> TABLE 1b. Power Difference in Treatment Test **TABLE 1b. Power Difference in Treatment Test**

NIH-PA Author Manuscript NIH-PA Author Manuscript

 NIH-PA Author ManuscriptNIH-PA Author Manuscript

 NIH-PA Author Manuscript NIH-PA Author Manuscript

> TABLE 1b. Power Difference in Treatment Test **TABLE 1b. Power Difference in Treatment Test**

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Note. Italicized number indicates the difference is significant; that is, twice the upper bound of standard error for empirical power (λ D = $\sqrt{2}$ X $\frac{m}{n}$ \approx \approx $\sqrt{2}$ X $\sqrt{2}$ and $n = 100000$), $2 \times \text{SE} =$ *G* = Correlation in *Note.* Italicized number indicates the difference is significant; that is, twice the upper bound of standard error for empirical power ($SE = \sqrt{2 \times \frac{p \times (1-p)}{n}} \approx .0007$, where $p = .5$ and $n = 10000$), $2 \times SE = .014$. $\lambda =$ rel matrix, i.e., between growth parameter estimates, EZ = effect size of treatment. TRM = traditional repeated measures; RC = random slope with homogeneous Level-1 variance; AR(1) = first-order autoregressive; UN = unstructur matrix, i.e., between growth parameter estimates, EZ = effect size of treatment. IEM = traditional repeated measures; RC = random slope with homogeneous Level-1 variance; AR(1) = first-outler grossive; DIN = unstructured; structures.

NIH-PA Author Manuscript NIH-PA Author Manuscript

Power Pattern and Pairwise Power Difference of TRM, RC, AR(1), and UN When $t = 3$ of 10,000 MC Samples at $\alpha = .05$ in the Interaction Test $\alpha = 0.05$ in the Interaction Test Power Pattern and Pairwise Power Difference of TRM, RC, AR(1), and UN When *t* = 3 of 10,000 MC Samples at **TABLE 2**

NIH-PA Author Manuscript NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

3,0641
0,0593
0,0628 TRM 175 0.9388 0.9571 0.9869 0.9427 0.9011 0.9302 0.9761 0.9082 0.8606 0.8983 0.9651 0.8723 200 0.9644 0.9740 0.9931 0.9681 0.9370 0.9555 0.9880 0.9445 0.9051 0.9330 0.9811 0.9163 .75 25 0.2121 0.2421 0.2688 0.1908 0.1911 0.2250 0.2561 0.1766 0.1713 0.2077 0.2427 0.1653 50 0.3644 0.4037 0.4508 0.3284 0.3290 0.3738 0.4231 0.2981 0.2988 0.3453 0.4047 0.2776 75 0.5146 0.5522 0.6077 0.4741 0.4685 0.5108 0.5772 0.4354 0.4305 0.4767 0.5495 0.4048 100 0.6409 0.6787 0.7335 0.6020 0.5899 0.6360 0.7047 0.5575 0.5483 0.5929 0.6743 0.5140 125 0.7441 0.7738 0.8253 0.7050 0.6931 0.7334 0.7962 0.6568 0.6462 0.6932 0.7710 0.6125 150 0.8175 0.8401 0.8828 0.7895 0.7757 0.8076 0.8596 0.7472 0.7330 0.7719 0.8343 0.7090 175 0.8741 0.8916 0.9280 0.8545 0.8360 0.8621 0.9082 0.8144 0.7965 0.8289 0.8895 0.7756 200 0.9148 0.9353 0.9368 0.9346 0.9246 0.9446 0.9446 0.9462 0.8377 0.8370 0.8514 0.8716 0.92610 0.9285 0.92353 .50 25 0.1613 0.1808 0.1905 0.1382 0.1510 0.1706 0.1836 0.1317 0.1418 0.1631 0.1789 0.1267 50 0.2704 0.2947 0.3101 0.2277 0.2533 0.2790 0.2968 0.2174 0.2390 0.2677 0.2863 0.2082 75 0.3890 0.4135 0.4337 0.3363 0.3614 0.3900 0.4162 0.3154 0.3404 0.3719 0.4018 0.2987 100 0.4969 0.5230 0.5417 0.4353 0.4675 0.4947 0.5203 0.4074 0.4373 0.4655 0.5001 0.3863 125 0.5843 0.6102 0.6375 0.5155 0.5495 0.5821 0.6150 0.4872 0.5220 0.5525 0.5927 0.4624 150 0.6645 0.6863 0.7081 0.6049 0.6313 0.6587 0.6882 0.5767 0.6014 0.6314 0.6670 0.5484 175 0.7298 0.7506 0.7734 0.6689 0.6973 0.7228 0.7511 0.6379 0.6681 0.6959 0.7328 0.6118 200 0.7881 0.8068 0.8292 0.7334 0.7598 0.7811 0.8070 0.7039 0.7306 0.7552 0.7865 0.6768 .25 25 0.1040 0.1144 0.1250 0.0926 0.1010 0.1121 0.1220 0.0905 0.0999 0.1121 0.1204 0.0876 50 0.1617 0.1733 0.1778 0.1329 0.1571 0.1682 0.1750 0.1303 0.1527 0.1647 0.1715 0.1271 75 0.2308 0.2424 0.2480 0.1883 0.2254 0.2380 0.2442 0.1844 0.2174 0.2302 0.2385 0.1790 100 0.2931 0.3066 0.3101 0.2388 0.2855 0.2969 0.3031 0.2324 0.2747 0.2899 0.2952 0.2263 125 0.3480 0.3611 0.3658 0.2824 0.3370 0.3515 0.3598 0.2747 0.3262 0.3410 0.3531 0.2681 150 0.4077 0.4199 0.4288 0.3393 0.3965 0.4084 0.4182 0.3295 0.3838 0.4005 0.4100 0.3202 175 0.4662 0.4790 0.4830 0.3833 0.4520 0.4672 0.4719 0.3708 0.4341 0.4521 0.4616 0.3641 200 0.5110 0.5236 0.5236 0.4327 0.4943 0.5161 0.5161 0.5161 0.4202 0.5161 0.5221 0.5161 0.5074 0.5074 0.5074 0
0.4120 0.5021 0.5021 0.5237 0.4327 0.4321 0.5161 0.421 .01 25 0.0472 0.0539 0.0539 0.0516 0.0473 0.0512 0.0516 0.0516 0.0516 0.0473 0.0539 0.0473 0.0539 0.0539 0.053 50 0.0510 0.0551 0.0576 0.0480 0.0504 0.0552 0.0576 0.0481 0.0502 0.0549 0.0575 0.0476 75 0.0586 0.0618 0.0644 0.0592 0.0586 0.0618 0.0647 0.0589 0.0596 0.0614 0.0645 0.0581 100 0.0615 0.0639 0.0669 0.0596 0.0538 0.0638 0.0667 0.0671 0.0613 0.0645 0.0645 0.0645 0.0645 0.0645 0.0671 0
2001 0.0615 0.0639 0.0669 0.0596 0.0598 0.0588 0.0589 125 0.0580 0.0605 0.0629 0.0576 0.0576 0.0606 0.0632 0.0573 0.0573 0.0607 0.0624 0.0569 150 0.0643 0.0663 0.0640 0.0640 0.0640 0.0668 0.0668 0.0668 0.0676 0.0676 0.0643 0.0643 0.0643 0.068 0.0679 0.
0.0641 0.0643 0.0668 0.0643 0.0648 0.0648 0.0648 0.0648 0.0648 0.0648 0.0648 0.0648 0.0648 0.0679 0.0648 0.07 175 0.0685 0.0685 0.0685 0.0698 0.0690 0.0691 0.0696 0.0594 0.0594 0.0699 0.0685 0.0685 0.0694 0.0694 0.0694 0
Casa 0.0685 0.0685 0.0698 0.0691 0.0690 0.0691 0.0690 0.0690 0.0699 0.069 22000 0.06900 0.089010 0.09010 0.09010 0.069010 0.09010 0.09010 0.07010 0.06010 0.08910 0.089010 0.0700 0.070
0.0702010 0.089010 0.09010 0.09010 0.09010 0.09010 0.09010 0.07010 0.06010 0.070 25 0.5400 0.6033 0.7025 0.5220 0.4802 0.5519 0.6605 0.4668 0.4262 0.5045 0.6202 0.4210 50 0.8394 0.8768 0.9404 0.8353 0.7813 0.8304 0.9163 0.7824 0.7265 0.7864 0.8882 0.7319 75 0.9542 0.9680 0.9899 0.9577 0.9291 0.9486 0.9838 0.9333 0.9009 0.9279 0.9744 0.9059 100 0.9902 0.9933 0.9992 0.9923 0.9780 0.9858 0.9973 0.9825 0.9634 0.9754 0.9941 0.9698 125 0.9973 0.9981 0.9984 0.9984 0.9984 0.9984 0.9986 0.99860 0.9988 0.99999 0.99955 0.99355 0.9923 0.9993 0.99
0.9985 0.9923 0.9936 0.9984 0.9984 0.9928 0.9928 0.9929 0.9993 0.9993 0.9993 0.9993 0.9923 0.9993 0.9993 0.10 160 16000 0.9996 0.9998 0.9993 0.9993 0.9993 0.9993 0.9993 0.9991 0.0001 0.9990 0.9980 0.9980 0.9980 0.9987 0.
0.9978 0.9997 0.9980 0.9981 0.9982 0.9980 0.9993 0.9993 0.9980 0.99820 0.99820 0.9980 0.9980 0.9980 0.9980 0.9 175 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 0.9998 0.9995 0.0000 0.0000 0.9995 0.9995 1
0.9995 0.9995 0.99980 0.9998 0.9999 0.9999 0.9999 0.9999 0.9995 0.9995 0.0000 0.0000 0.0000 0.0000 0.0000 0.00 200 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1
0.9999 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.000 .75 25 0.4506 0.4962 0.5457 0.4049 0.4122 0.4611 0.5188 0.3754 0.3786 0.4283 0.4920 0.3490 50 0.7522 0.7857 0.8318 0.7110 0.7043 0.7432 0.8028 0.6673 0.6596 0.7044 0.7785 0.6233 75 0.9026 0.9203 0.9460 0.8817 0.8706 0.8921 0.9308 0.8478 0.8354 0.8641 0.9142 0.8123 100 0.9623 0.9701 0.9817 0.9547 0.9448 0.9561 0.9753 0.9356 0.9251 0.9396 0.9677 0.9171 125 0.9896 0.9921 0.9949 0.9865 0.9865 0.9849 0.9920 0.9789 0.9758 0.9758 0.9758 0.987757 0.9891 0.9891 0.9875 150 0.9987 0.9983 0.9992 0.9929 0.9929 0.9929 0.9983 0.9983 0.9983 0.9983 0.9972 0.9972 0.9972 0.9972 0.9972 0
0.9972 0.9972 0.9920 0.9921 0.9920 0.9920 0.9920 0.9920 0.9920 0.9920 0.9920 175 0.9995 0.9995 0.9995 0.9993 0.9993 0.9993 0.9983 0.9983 0.9983 0.9983 0.9987 0.9967 0.9967 0.997
0.9980 0.9980 0.9996 0.9987 0.9992 0.9987 0.9987 0.9987 0.9983 0.9983 0.9983 0.9980 0.9980 0.9980 200 0.9996 0.9998 0.9998 0.9998 0.9999 0.9998 0.9991 0.9992 0.9998 0.9992 0.9992 0.9988 0.9988 0.9988 0.9988 0
2002 0.9998 0.9988 0.9982 0.9998 0.9998 0.9998 0.9998 0.9998 0.9982 0.9982 0.9982 .50 25 0.3384 0.3683 0.3872 0.2867 0.3192 0.3512 0.3742 0.2711 0.3009 0.3354 0.3605 0.2545 *n* **RC AR(1) UN TRM RC AR(1) UN TRM RC AR(1) UN TRM** Š *G* **= .5** AR(1) $\begin{smallmatrix} 6 & 6 & 7 & 8 & 8 & 6 & 7 & 8 & 8 & 7 & 8 & 7 & 8 & 8 & 7 & 8 & 7 & 8 & 8 & 7 & 8 & 7 & 8 & 8 & 7 & 8 & 7 & 8 & 8 & 7 & 8 & 7 & 8 & 8 & 7 & 8 & 7 & 8 & 8 & 7 & 8 & 7 & 8 & 7 & 8 & 7 & 8 & 7 & 8 & 7 & 8 & 7 & 8 & 7 & 8 & 7 & 8 & 7 & 8 & 7 & 8 & 7 & 8 & 7 & 8 & 7 & 8 & 7 & 8 & 7 & 8 &$ ${\bf R}$ TRM Š *G* **= .3** AR(1) EZ=.8 $\mathbf{R}^{\mathbf{C}}$ TRM $\overline{5}$ **G = .1** AR(1) ${\bf R}$ \overline{a} Egnerggergnerggergnerggergnerggergger <u>ងនិងទិនីទីទីនានិងទិនីទីទីទីនាន</u> 75 $50\,$ 25 75 $50\,$ Ξ **λ**

J Exp Educ. Author manuscript; available in PMC 2009 November 26.

162.4270 0.6099'0 10.6"0 0.6559 1867'0 0.629'0 576.5"0 0.595"0 0.525"0 0.575"0 0.679"0 0.0929"0 0.109"0 0.6089

 NIH-PA Author ManuscriptNIH-PA Author Manuscript

($\mathcal{S}E = \sqrt{2 \times \frac{p \times (1-p)}{n}} \approx .007$, where $p = .5$ and $n = 10,000$, $2 \times \mathcal{S}E = .014$), then the differences are labeled as significant. The results for the time test will be provided upon request. λ = reliability, (λ $E = \sqrt{2 \times \frac{m}{m} E}$ \approx \sim 5 and *n* = 10,000, 2 × *SE* = .014), then the differences are labeled as significant. The results for the time test will be provided upon request. λ = reliability,

J Exp Educ. Author manuscript; available in PMC 2009 November 26.

 $n = sample size per group, G = correlation in G matrix, i.e., between growth parameters estimated, EZ = effect size of interaction. TRM = traditional repeated measures; RC = random slope with homogeneous
of the size per group.$ *G* matrix, i.e., between growth parameter estimates, EZ = effect size of interaction. TRM = traditional repeated measures; RC = random slope with homogeneous Level-1 variance; AR(1) = first-order autoregressive; UN = unstructured; all are power patterns of variance-covariance structures. Level-1 variance; AR(1) = first-order autoregressive; UN = unstructured; all are power patterns of variance-covariance structures. *G* = correlation in *n* = sample size per group,

Not

Power Pattern and Pairwise Power Difference of TRM, RC, AR(1), and UN When *t* = 3 and

Power Pattern and Pairwise Power Difference of TRM, RC, AR(1), and UN When $t = 3$ and $\lambda = .50$ of 10,000 MC Samples at $\alpha = .05$ in Three Tests

 $\lambda = .50$ of 10,000 MC Samples at

 $\alpha = .05$ in Three Tests

J Exp Educ. Author manuscript; available in PMC 2009 November 26.

(\triangle $\frac{p-2}{p}$ \triangle \triangle $\frac{p-1}{p}$ where $p = .5$ and $n = 10000$, $2 \times 5E = .014$), then the differences are labeled as significant. λ = reliability, $n =$ sample size per group,

between growth parameter estimates, EZ = effect size of treatment, interaction, or time. Ratio (*n*/*t*) = sample size per group (*n*) to the number of time points (*t*) holding total sample size (

between growth parameter estimates, $EZ =$ effect size of treatment, interaction, or time. Ratio (n/t) = sample size per group (n) to the number of time points (t) holding total sample size (N) constant. TRM = traditional repeated measures; RC = random slope with homogeneous Level-1 variance; AR(1) = first-order autoregressive; UN = unstructured; all are power patterns of variance-covariance

TRM = traditional repeated measures; RC = random slope with homogeneous Level-1 variance; AR(1) = first-order autoregressive; UN = unstructured; all are power patterns of variance-covariance

structures.

G = correlation in

G matrix, i.e.,

N) constant.

APPENDIX

Validation of Data Generator Using R. F. Potthoff and S. N. Roy's (1964) Data

