Journal of Modern Applied Statistical Methods

Volume 7 | Issue 1 Article 9

5-1-2008

A Monte Carlo Power Analysis of Traditional Repeated Measures and Hierarchical Multivariate Linear Models in Longitudinal Data Analysis

Hua Fang
University of Nebraska-Lincoln, jfang2@unlnotes.unl.edu

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

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

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

Robert S. Barcikowski
Ohio University, barcikow@ohio.edu

Recommended Citation

Fang, Hua; Brooks, Gordon P.; Rizzo, Maria L.; Espy, Kimberly A.; and Barcikowski, Robert S. (2008) "A Monte Carlo Power Analysis of Traditional Repeated Measures and Hierarchical Multivariate Linear Models in Longitudinal Data Analysis," *Journal of Modern Applied Statistical Methods*: Vol. 7 : Iss. 1 , Article 9.
DOI: 10.22237/jmasm/1209614880
Available at: http://digitalcommons.wayne.edu/jmasm/vol7/iss1/9

Journal of Modern Applied Statistical Methods
May, 2008, Vol. 7, No. 1, 101-119

1538 – 9472/08/$95.00

A Monte Carlo Power Analysis of Traditional Repeated Measures and Hierarchical Multivariate Linear Models in Longitudinal Data Analysis

Hua Fang
University of Nebraska-Lincoln

Gordon P. Brooks
Ohio University

Maria L. Rizzo
Bowling Green State University

Kimberly A. Espy
University of Nebraska-Lincoln

Robert S. Barcikowski
Ohio University

The power properties of traditional repeated measures and hierarchical linear models have not been clearly determined in the balanced design for longitudinal studies in the current literature. A Monte Carlo power analysis of traditional repeated measures and hierarchical multivariate linear models are presented under three variance-covariance structures. Results suggest that traditional repeated measures have higher power than hierarchical linear models for main effects, but lower power for interaction effects. Significant power differences are also exhibited when power is compared across different covariance structures. Results also supplement more comprehensive empirical indexes for estimating model precision via bootstrap estimates and the approximate power for both main effects and interaction tests under standard model assumptions.

Key Words: Monte Carlo, power analysis, traditional repeated measures, hierarchical multivariate linear models, longitudinal study.

Introduction

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 employ adjusted univariate tests or traditional multivariate tests, which do not assume the variance-covariance (VC) structure (cf., Greenhouse & Geisser, 1959; Huynh & Feldt, 1976; Jennrich & Schluchter, 1986; Wolfinger & Chang, 1995). For the same longitudinal design, HMLM treat 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).

HMLM and TRM are essentially interrelated in their theoretical development, especially after advanced computational methods were developed to handle missing values and model the VC structures (Dempster, Laird & Rubin, 1977; Dempster, Rubin & Tsutakawa, 1981; Goldstein 1995; Jennrich & Schluchter, 1986; Littell, Milliken, Stroup &

Hua Fang is a Research Assistant Professor and Project Director in the Office of Research. E-mail: jfang2@unlnotes.unl.edu. Gordon P. Brooks is an Associate Professor of Research and Evaluation. E-mail: brooksg@ohio.edu. Maria L. Rizzo is an Assistant Professor. Email her at mrizzo@bgnet.bgsu.edu. Kimberly Andrews Espy is a Professor in the Department of Psychology. Robert S. Barcikowski is an emeritus professor of Research and Evaluation. E-mail: barcikow@ohio.edu.

Wolfinger, 2006; Little, 1995; Little & Rubin, 2002; Maas & Snijders, 2003; McCulloch & Searle, 2001; Raudenbush & Bryk, 2002; Van der Leenden, Vrijburg & de Leeuw, 1996). Jennrich and Schluchter were the first to model specific VC structures directly through maximum likelihood estimation based on traditional multivariate repeated measures approach whereas HMLM incorporates Jennrich and Schlutchter's multivariate repeated measures approach to longitudinal data analysis (Schluchter, 1988; Van der Leenden, 1998; Jennrich & Schlutchter, 1986; Raudenbush & Bryk, 2002). In literature, HMLM is simply called hierarchical linear models, more generally known as multilevel models, growth mixture models or generalized latent variable models (e.g., Goldstein, 1994, 1995; Hox, 2002; Maas & Snijders, 2003; Muthén, 2002, 2004; Muthén & Muthén, 2006; Raudenbush & Bryk, 2002; 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 et al., 2005; Snijders, 2005). As the current analytical power approximations are not comprehensive or necessarily accurate (Littell et al., 2006) and the power properties of TRM and HMLM have not yet been clearly compared in the balanced design, using Monte Carlo (MC) simulation approach would be efficient to examine 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); (c) and first-order autoregressive (AR(1)). Additionally, the bootstrap estimates for the treatment effect were compared for TRM and the three VC structures.

Two-level HMLM model

The hypotheses tested in this simulation assumed 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:

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

$$\text{Level 2: } \pi_{0i} = \beta_{00} + \beta_{01} * \text{TREATMENT} + u_{0i}$$
$$\pi_{1i} = \beta_{10} + \beta_{11} * \text{TREATMENT} + u_{1i} \quad (2)$$

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); e_{ti} refers to the random error at the t^{th} time point of the i^{th} person at level-1.

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

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

where the residual term, $r_{ti} = u_{0i} + u_{1i} * TIME + e_{ti}$, includes the leve-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 parenthesis and random effects embodied in the residual term r_{ti}.

TRM and three covariance structures under study

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

$$Y = X\beta + r \quad (4)$$

where Y is a $t_i \times 1$ response vector for subject i, and t represents the number of time points and i = 1, …, n; X is a $t_i \times a$ design matrix for fixed effect β, where a is the number of fixed effects (i.e., the 3 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, $r \sim N(0, \Sigma)$. 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:

$$e_{ti} \sim N(0, \sigma^2), \text{ and}$$

$$\begin{pmatrix} u_{0i} \\ u_{1i} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{pmatrix}\right].$$

The variance of level-1 error term (e_{ti}) is homogeneous and the covariance structure of level-2 random errors (u_{0i} and u_{1i}) is arbitrary. For Model (3), only 4 variance-covariance 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 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 unknown, one could fit an UN covariance matrix. The UN matrix for each level-2 subject with 3 time points can be expressed as

$$\begin{pmatrix} \sigma^2_{11} & \sigma^2_{12} & \sigma^2_{13} \\ \sigma^2_{21} & \sigma^2_{22} & \sigma^2_{23} \\ \sigma^2_{31} & \sigma^2_{32} & \sigma^2_{33} \end{pmatrix} \tag{5}$$

and requires the estimation of 3 variance parameters and 3 covariance parameters. When more time points are involved, UN can require an exorbitant number of parameters.

First Order Auto-Regressive (AR(1))

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

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

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

$$\sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{pmatrix} \tag{7}$$

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 (τ_{00}, τ_{11} and τ_0) which is usually assumed unstructured (cf. Littell, Henry, & Ammerman, 1998, pp. 1229-1230; McCulloch & Searle, 2001; Raudenbush & Bryk, 2002; Singer & Willett, 2003; Wolfinger, 1993). Thus, 5 variance-covariance parameters need to be estimated for AR(1) of Model (3).

Methodology

Monte Carlo Design

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

$$y_{ti} = (\beta_{01} \times \text{TREATMENT} + \beta_{10} \times \text{TIME} + \beta_{11} \times \text{TREATMENT} \times \text{TIME}) + r_{ti} \quad (8)$$

A stacked SAS macro was written by the author (the author, 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 5000, and the nominal alpha (α) for each sample test was .10 considering the relatively small number of iterations (e.g., compared to 10,000 iterations).

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_{1i}), 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 $X_{intercepts}$ and X_{slopes}, $G = \begin{pmatrix} \sigma_{00}^2 & \sigma_{01}^2 \\ \sigma_{10}^2 & \sigma_{11}^2 \end{pmatrix}$, and random mean vector, $\begin{pmatrix} \mu_0 \\ \mu_1 \end{pmatrix}$, were specified to simulate correlated bivariate normal data for $X_{intercepts}$ and X_{slopes}. 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 Inc., 2003). An index matrix was created for time, treatment and individual IDs. Two treatment groups were coded by 0 and 1, respectively. Individuals (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. Based on Model (4), a univariate response vector of y_{ti} was created. For example, each subject might have had 3 time points and each treatment group had 25 subjects.

The data generator (author, 2006) was validated with parameter estimates from Potthoff and Roy's data (1964). The results are shown in Appendix A.

Power comparison

Holding other factors constant, the power comparison was implemented by changing the levels of one of the four factors, respectively: (1) Correlation in G matrix (G), (2) reliability of level-1 coefficients (λ), (3) effect size (β) and (4) ratio of group sample size to time points (n/t) under specified conditions (see Table 1).

Power comparison by G matrix

This study used Cohen's indices (1988) for correlation, $\rho \in \{.1 \;\; .3 \;\; .5\}$; correspondingly, G matrix (G) for random intercepts (u_{0i}) and slopes (u_{1i}) was specified as $\begin{pmatrix} 1 & .1 \\ .1 & 1 \end{pmatrix}$, $\begin{pmatrix} 1 & .3 \\ .3 & 1 \end{pmatrix}$, or $\begin{pmatrix} 1 & .5 \\ .5 & 1 \end{pmatrix}$. To show the power pattern by varying correlation in G matrix, the MC design incorporated a moderate sample size (n = 75) and fixed time points (t = 3) (i.e., ratio of group sample size to time points, n/t = 75/3), effect size (β_{01} = .5, β_{10} = .5 or β_{11} = .5), and moderate reliability λ = .5, to simulate a specific situation and compare power at each G matrix. Based on the design, a general power pattern of TRM, RC, UN and AR(1) were presented at each of the three G matrices in the respective three tests, treatment effect (β_{01}) test, time effect (β_{10}) test and interaction (β_{11}) test (i.e., 4 × 3 × 3 = 36 cells).

Power comparison by reliability

By changing the averaged reliability of level-1 coefficients, $\lambda \in \{.01, .25, .5, .75, .1\}$, (Raudenbush & Bryk, 2002), the power pattern of RC, UN and AR(1) was compared by setting $G = \begin{pmatrix} 1 & .3 \\ .3 & 1 \end{pmatrix}$, n/t = 75/3, and β = .5. The power pattern of TRM, RC, UN and AR(1) was presented at each of the five reliability indexes in the respective three tests, treatment effect (β_{01}), time effect (β_{10}) and interaction (β_{11}) (i.e., 4 × 5 × 3 = 60 cells).

Power comparison by effect size

Cohen's indexes were also used for two-group treatment effect (β_{01}), time effect (β_{10}) and interaction (β_{11}), $\beta \in \{.2\ \ .5\ \ .8\}$. The MC design simulated a moderate situation where n = 75, $C_x = \begin{pmatrix} 1 & .3 \\ .3 & 1 \end{pmatrix}$, λ = .5 and t = 3 to compare power at three effect sizes, $\beta \in \{.2\ \ .5\ \ .8\}$, of the three fixed effects for the four models, TRM, RC, AR(1) and UN (i.e., 4 × 3 × 3 = 36 cells).

Power comparison by sample size ratio

For exploratory purpose, this study fixed the time points (t) at 3. As the maximum likelihood estimation requires relatively large sample sizes, the sample size per treatment group ($m=2$) was changed from 25 to 200 by an increase of 25 ($n \in \{25\ \ 75\ \ 100\ \ 125\ \ 150\ \ 175\ \ 200\}$), that is, the total sample size $N \in \{150\ \ 300\ \ 450\ \ 600\ \ 750\ \ 900\ \ 1050\ \ 1200\}$ ($N = m \times n \times t$). To compare the power by varying the sample size ratio, the condition was specified as β = .5, $G = \begin{pmatrix} 1 & .3 \\ .3 & 1 \end{pmatrix}$, λ = 3 and t = 3. For each specified condition, the power patterns for TRM and the three VC structures were presented at eight sample sizes, for the three fixed effects (i.e., 4 × 8 × 3 = 96 cells).

Monte Carlo Analysis

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

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

where p = .5 and n = 5000. If the pairwise differences are twice the upper bound of SE (i.e., SE_{upper} = .20), then the differences are labeled as significant. The power patterns are illustrated in tables and graphs. In addition to the power analysis, bootstrap standard CI, estimates, bias and standard errors for the estimates of the treatment mean difference (β_{01}) were calculated to compare the model precision.

Results

Empirical Power by G Matrix

The results (see Figure 1) indicated two general patterns when varying the *G matrix* under the specific circumstance in all three tests, treatment effect (β_{01}), time effect (β_{10}) and interaction (β_{11}). The first 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 matrices seems to be minimal.

The pairwise power tests (see Table 2) showed that TRM power was significantly higher than the other three in the treatment and time tests, but significantly lower than the other three in the interaction test by varying G matrices. Among the three VC structures, UN had significantly higher power than RC and AR(1) in both treatment and interaction tests, whereas AR(1) has significantly higher power than RC in the same two tests. As to the time test, UN power was significantly higher than RC across the three G matrices but was not significantly higher than AR(1) at all three G matrices.

Table 1. MC Design for Power Analysis of TRM, RC, AR(1) and UN by G Matrix, Effect Size and Sample Size of 5000 MC Samples at $\alpha = .10$

Factors (Cells) [a]	Conditions [b]
G Matrix (G) (36) $\{\begin{pmatrix}1 & .1\\ .1 & 1\end{pmatrix}\begin{pmatrix}1 & .3\\ .3 & 1\end{pmatrix}\begin{pmatrix}1 & .5\\ .5 & 1\end{pmatrix}\}$	Fixed [c]: $n = 75$, $t = 3$ and $\lambda = 3$ 1. $\beta_{01} = .5, \beta_{10} = 0, \beta_{11} = 0$ [d] 2. $\beta_{01} = 0, \beta_{10} = .5, \beta_{11} = 0$ [d] 3. $\beta_{01} = 0, \beta_{10} = 0, \beta_{11} = .5$ [d]
Reliability (λ) (60) {.01 .25 .5 .75 1}	Fixed [c]: $t = 3$, $\lambda = 3$ and $G = \begin{pmatrix}1 & .3\\ .3 & 1\end{pmatrix}$ 1. $\beta_{01} = .5, \beta_{10} = 0, \beta_{11} = 0$ [d] 2. $\beta_{01} = 0, \beta_{10} = .5, \beta_{11} = 0$ [d] 3. $\beta_{01} = 0, \beta_{10} = 0, \beta_{11} = .5$ [d]
Effect Size (β) (36) {.2 .5 .8}	Fixed [c]: $n = 75$, $t = 3$, $\lambda = 3$ and $G = \begin{pmatrix}1 & .3\\ .3 & 1\end{pmatrix}$ 1. $\beta_{10} = 0, \beta_{11} = 0$ [d] 2. $\beta_{01} = 0, \beta_{11} = 0$ [d] 3. $\beta_{01} = 0, \beta_{10} = 0$ [d]
Sample Size per Treatment Group (n) (96) {25 50 75 100 125 150 175 200}	Fixed [c]: $t = 3$, $\lambda = 3$ and $G = \begin{pmatrix}1 & .3\\ .3 & 1\end{pmatrix}$ 1. $\beta_{01} = .5, \beta_{10} = 0, \beta_{11} = 0$ [d] 2. $\beta_{01} = 0, \beta_{10} = .5, \beta_{11} = 0$ [d] 3. $\beta_{01} = 0, \beta_{10} = 0, \beta_{11} = .5$ [d]

Note. [a] The factors are not crossed
[b] Conditions are specified for testing each fixed effect (β_{01}, β_{10}, and β_{11}) within each factor
[c] "Fixed" indicates the fixed parameters in the design within each factor
[d] Settings for testing the three fixed effects

Empirical Power by Reliability

Two power patterns were displayed by varying reliability indexes in the specified condition. First, with the increase of reliability, the power of TRM, RC, AR(1) and UN increased in all three tests. TRM had a higher power-increasing rate than the other three below the reliability of .5. Above the reliability of .5, all four seemed to increase power at a decreasing rate. The power of all four approached to the asymptote of 1 as the reliability reached 1 (see Figure 2)..80. In the time test, all four had power above .80. Yet, in the interaction test, TRM power was significant in all three tests, that is, TRM had significantly higher power than the other three, reliability reached 1 (see Figure 2).

The second pattern showed that TRM gained the highest power in the treatment and time tests but had the lowest power than the other three in the interaction test across the reliability indexes (see Table 3). Among RC, AR(1) and UN, UN power ranked the highest, AR(1) the second and RC the lowest across all reliability indexes in all three tests. At the reliability of .75, TRM power was above .80 in the treatment test whereas the power of all three VC structures seemed to be above .60 but below .80. In the time test, all four had power above .80. Yet, in the interaction test, TRM power was

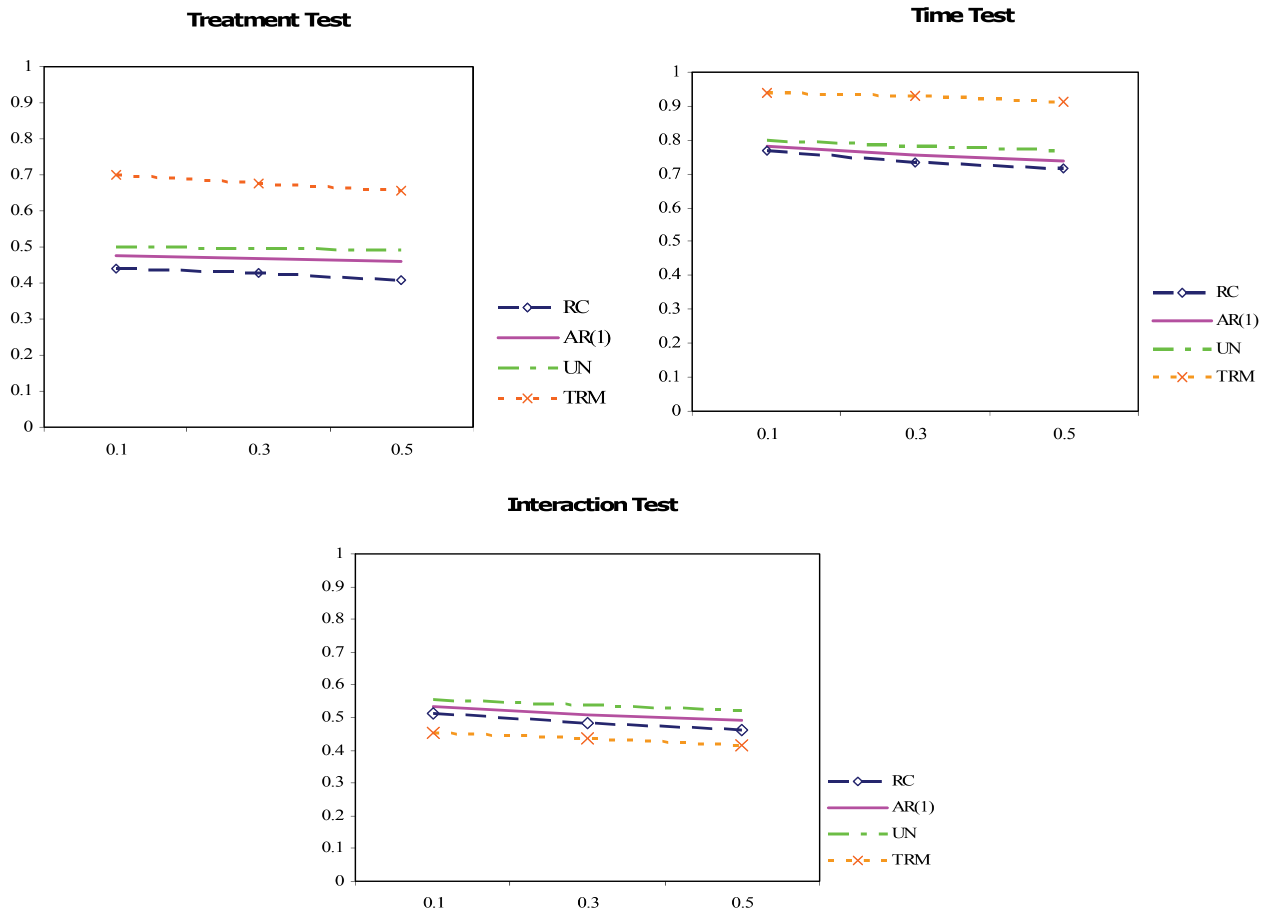


Figure 1. Power pattern of HLM, AR(1), UN and TRM by *G* matrix in treatment, time and interaction tests when $n = 75$, $t = 3$ and $\lambda = .5$ of 5000 MC samples at $\alpha = .10$.

below .60 while the power of the other three was at or above .60.

Table 4 indicated that when the reliability was between .5 and 1, the pairwise power differences among the four were significant in all three tests, that is, TRM had significantly higher power than the other three, respectively, in the treatment and time tests, but significantly lower power than the three in the interaction test. Generally, among the three VC structures, UN power was statistically higher than AR(1) and RC while AR(1) power was significantly higher than RC under the specified condition. Below the reliability of .5, the pairwise power differences among the four were not all significant across the three tests.

Empirical Power by Effect Size

As effect sizes increased, the power of the four models was enhanced in the treatment, time and interaction tests. Also, it seemed that the power of the four has a higher increasing rate from the small to the medium effect size than from the medium to the large effect size (Fig. 3).

Table 4 showed the significant pairwise power differences among TRM, RC, AR(1) and UN at the medium effect size in all three tests: TRM power was significantly higher than the other three VC structures in treatment and time tests, but had significantly lower power than the other three in the interaction test. Still at the medium effect size, UN was significantly higher than RC and AR(1) while AR(1) was

Table 3. Power and Pairwise Power Difference of TRM, RC, AR(1) and UN by Reliability when n = 75, t = 3 and G = $\begin{pmatrix} 1 & .3 \\ .3 & 1 \end{pmatrix}$ of 5000 MC Samples at α = .10

	Power				Power Difference					
λ	RC	AR(1)	UN	TRM	UN vs. RC	UN vs. AR(1)	AR(1) vs.RC	TRM vs. RC	TRM vs. AR(1)	TRM vs.UN
Two-group Treatment Effect (β_{01} = .5, β_{10} = 0, β_{11} = 0)										
.01	.1050	.1092	.1166	0.1148	.0116	.0074	.0042	.0098	.0056	-.0018
.25	*.2436*	*.2650*	*.2726*	*0.4472*	*.0290**	*.0076*	*.0214**	*.2036**	*.1822**	*.1746**
.50	.4264	.4688	.4946	0.6762	.0682*	.0258*	.0424*	.2498*	.2074*	.1816*
.75	.6054	.6646	.7292	0.8208	.1238*	.0646*	.0592*	.2154*	.1562*	.0916*
1	.7830	.8374	.9334	0.8920	.1504*	.0960*	.0544*	.1090*	.0546*	-.0414*
Time Effect (β_{01} = 0, β_{10} = .5, β_{11} = 0)										
.01	.1162	.1200	.1216	.1292	.0054	.0016	.0038	.0130	.0092	.0076
.25	.5210	.5338	.5418	.7050	.0208*	.0080	.0128	.1840*	.1712*	.1632*
.50	.7354	.7576	.7828	.9284	.0474*	.0252*	.0222*	.1930*	.1708*	.1456*
.75	.8586	.8776	.9208	.9838	.0622*	.0432*	.0190*	.1252*	.1062*	.0630*
1	.9138	.9330	.9764	.9962	.0626*	.0434*	.0192*	.0824*	.0632*	.0198*
Interaction (β_{01} = 0, β_{10} = 0, β_{11} = 0.5)										
.01	.1092	.1146	.1194	.1092	.0102	.0048	.0054	.0000	-.0054	-.0102
.25	.3222	.3336	.3416	.2796	.0194*	.0080	.0114	-.0426*	-.0540*	-.0620*
.50	.4844	.5092	.5390	.4368	.0546*	.0298*	.0248*	-.0476*	-.0724*	-.1022*
.75	.5972	.6348	.6952	.5576	.0980*	.0604*	.0376*	-.0396*	-.0772*	-.1376*
1	.6726	.7258	.8220	.6638	.1494*	.0962*	.0532*	-.0088	-.0620*	-.1582*

Note: * indicates the difference is significant, that is, twice the upper bound of standard error for empirical power (SE = $\sqrt{2\times\frac{p\times(1-p)}{n}} = .01$ where p = .5 and n =5000), 2×SE = .02.

significantly higher than RC. At the small or large effect size, the pairwise power differences among the three VC structures were not all significant across the three tests.

Empirical Power by Sample Size

With the increase of the sample size, the power of TRM, RC AR(1) and UN increased (see Figure 4). It seemed that below the sample size of 100 (i.e., N = 600), the four models increased power at an increasing rate and above the sample size of 100, the four enhanced their power at a decreasing rate, approaching to the asymptote of 1.

When the sample size was small, n = 25 (i.e., N = 150), the power of all four models was low (e.g., around or below .3 in the treatment and interaction tests). At the sample size of 100 (i.e., N = 600), TRM gained power around .75 and .95 in the treatment and time tests, respectively, but below .55 in the interaction test; whereas all three VC structures obtained power merely above .50 and .80 in the treatment and time tests, respectively, but above .60 in the

Table 2. Power Patten and Pairwise Power Difference of TRM, RC, AR(1) and UN by G Matrix when n = 75, t = 3 and λ = .5 of 5000 MC Samples at α = .10

	Power				Power Difference					
G	RC	AR(1)	UN	TRM	UN vs. RC	UN vs. AR(1)	AR(1) vs.RC	TRM vs. RC	TRM vs. AR(1)	TRM vs.UN
Two-group Treatment Effect (β_{01} = .5, β_{10} = 0, β_{11} = 0)										
$\begin{pmatrix} 1 & .1 \\ .1 & 1 \end{pmatrix}$	.4382	.4756	.4982	.7020	.0600*	.0226*	.0374*	.2638*	.2264*	.2038*
$\begin{pmatrix} 1 & .3 \\ .3 & 1 \end{pmatrix}$	.4264	.4688	.4946	.6762	.0682*	.0258*	.0424*	.2498*	.2074*	.1816*
$\begin{pmatrix} 1 & .5 \\ .5 & 1 \end{pmatrix}$	.4070	.4610	.4918	.6562	.0848*	.0308*	.0540*	.2492*	.1952*	.1644*
Time Effect (β_{01} = 0, β_{10} = .5, β_{11} = 0)										
$\begin{pmatrix} 1 & .1 \\ .1 & 1 \end{pmatrix}$	.7686	.7814	.8002	.9396	.0316*	.0188	.0128	.1710*	.1582*	.1394*
$\begin{pmatrix} 1 & .3 \\ .3 & 1 \end{pmatrix}$	.7354	.7576	.7828	.9284	.0474*	.0252*	.0222*	.1930*	.1708*	.1456*
$\begin{pmatrix} 1 & .5 \\ .5 & 1 \end{pmatrix}$	.7144	.7376	.768	.9146	.0536*	.0304*	.0232*	.2002*	.1770*	.1466*
Interaction (β_{01} = 0, β_{10} = 0, β_{11} = 0.5)										
$\begin{pmatrix} 1 & .1 \\ .1 & 1 \end{pmatrix}$	.5110	.5318	.5570	.4554	.0460*	.0252*	.0208*	-.0556*	-.0764*	-.1016*
$\begin{pmatrix} 1 & .3 \\ .3 & 1 \end{pmatrix}$	.4844	.5092	.5390	.4368	.0546*	.0298*	.0248*	-.0476*	-.0724*	-.1022*
$\begin{pmatrix} 1 & .5 \\ .5 & 1 \end{pmatrix}$	.4618	.4930	.5210	.4132	.0592*	.0280*	.0312*	-.0486*	-.0798*	-.1078*

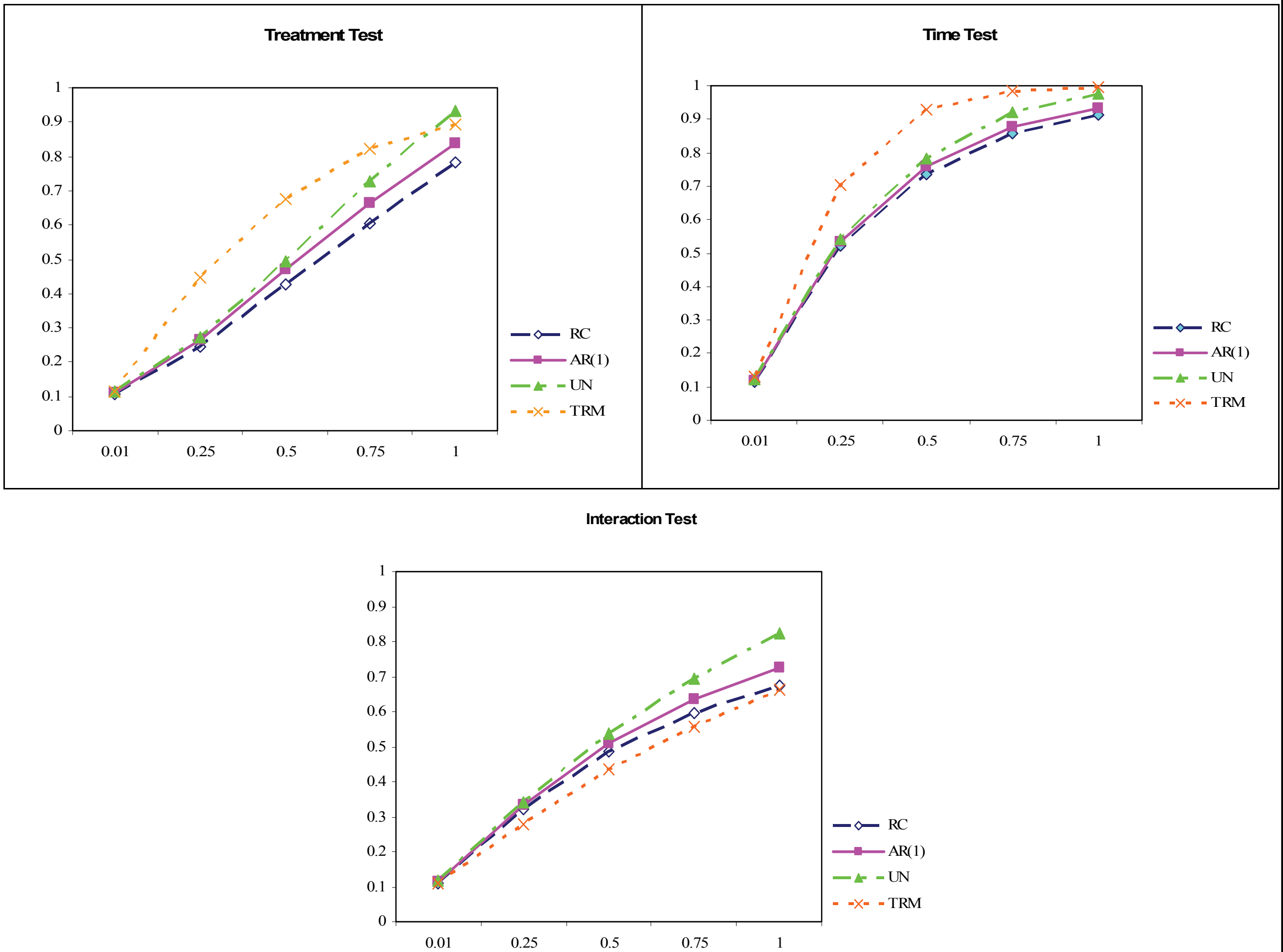


Figure 2. Power pattern of HLM, AR(1), UN and TRM by reliability in treatment, time and interaction tests when $n = 75$, $t = 3$ and $G = \begin{pmatrix} 1 & .3 \\ .3 & 1 \end{pmatrix}$ of 5000 MC samples at $\alpha = .10$

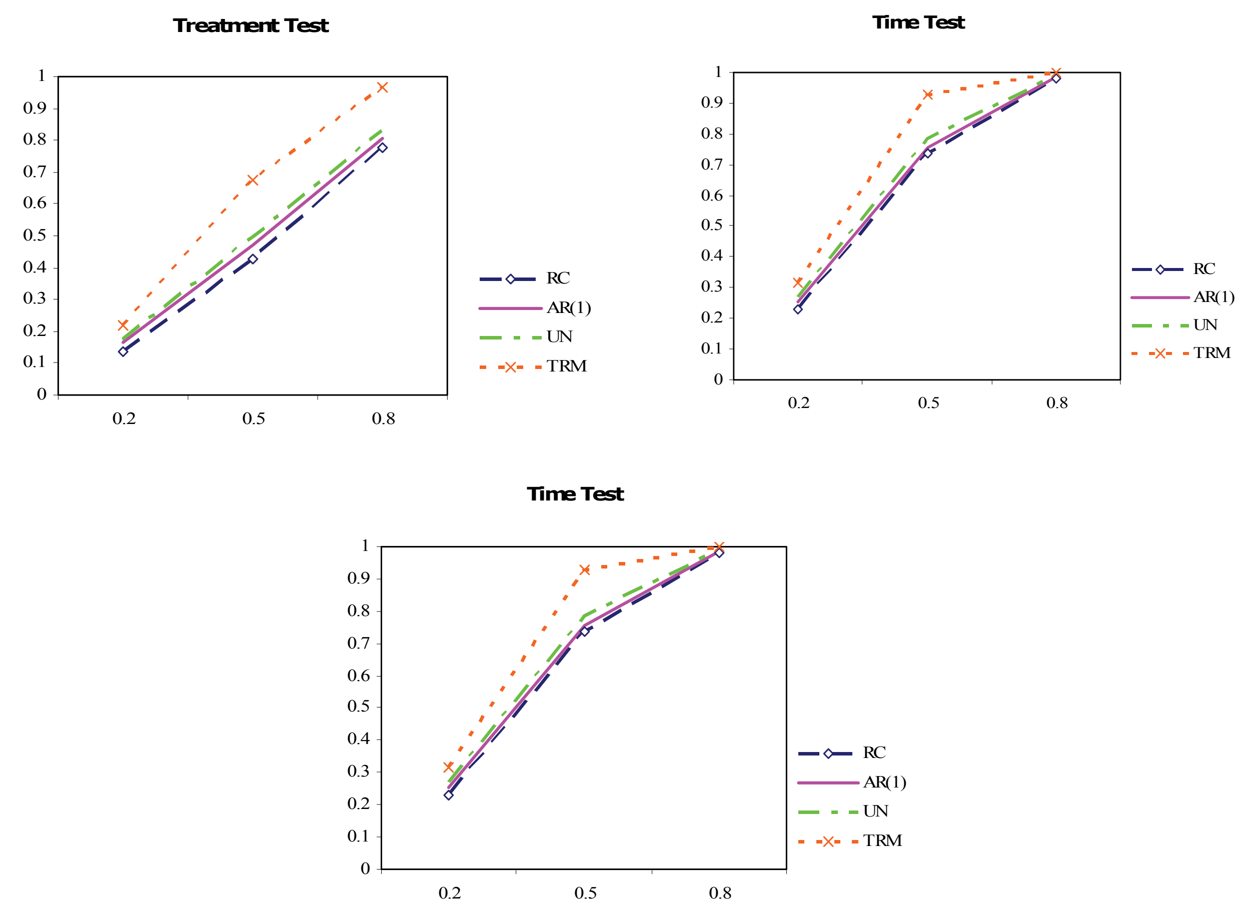


Figure 3: Power pattern of HLM, AR(1), UN and TRM by effect size in treatment, time and interaction tests when $n = 75$, $t = 3$, $\lambda = .5$ and $G = \begin{pmatrix} 1 & .3 \\ .3 & 1 \end{pmatrix}$ of 5000 MC samples at $\alpha = .1$

Table 4 Power Pattern and Pairwise Power Difference of RC, AR(1) and UN by Effect Size when n = 75, t = 3, λ = .5 and G = $\begin{pmatrix} 1 & .3 \\ .3 & 1 \end{pmatrix}$ of 5000 MC Samples at α = .10

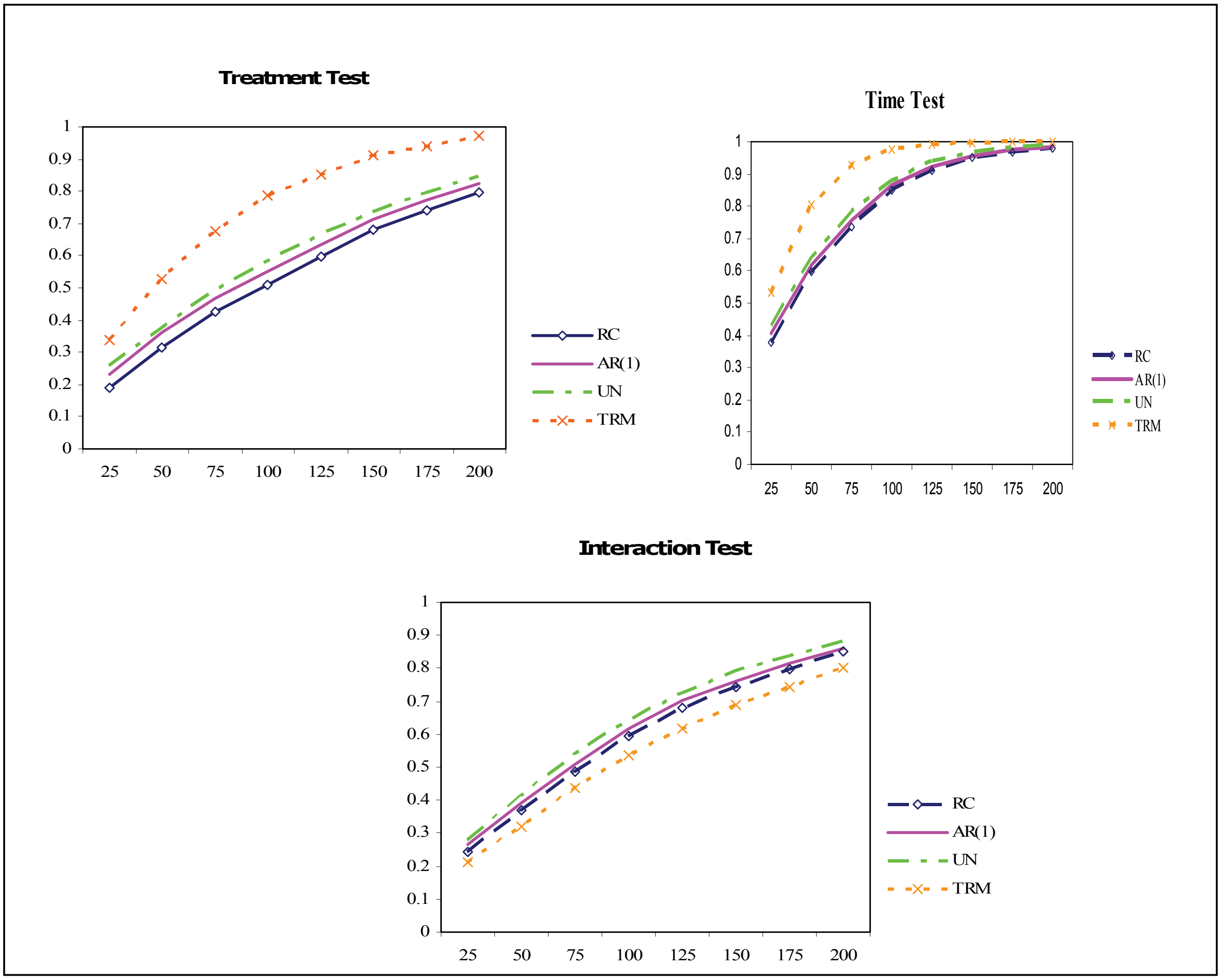
	Power				Power Difference					
G	RC	AR(1)	UN	TRM	UN vs. RC	UN vs. AR(1)	AR(1) vs.RC	TRM vs. RC	TRM vs. AR(1)	TRM vs.UN
Two-group Treatment Effect (β_{01} = .5, β_{10} = 0, β_{11} = 0)										
.20	.1372	.1650	.1746	.2166	.0374*	.0096	.0278*	.0794*	.0516*	.0420*
.50	.4264	.4688	.4946	.6762	.0682*	.0258*	.0424*	.2498*	.2074*	.1816*
.80	.7758	.8042	.8308	.9642	.0550*	.0266*	.0284*	.1884*	.1600*	.1334*
Time Effect (β_{01} = 0, β_{10} = .5, β_{11} = 0)										
.20	.2316	.2542	.2698	.3168	.0382*	.0156*	.0226*	.0852*	.0626*	.0470*
.50	.7354	.7576	.7828	.9284	.0474*	.0252*	.0222*	.1930*	.1708*	.1456*
.80	.9804	.9852	.9892	1.0000	.0088	.0040	.0048	.0196*	.0148*	.0108
Interaction (β_{01} = 0, β_{10} = 0, β_{11} = 0.5)										
.20	.1686	.1856	.1928	.1602	.0242*	.0072	.0170	-.0084	-.0254*	-.0326*
.50	.4844	.5092	.5390	.4368	.0546*	.0298*	.0248*	-.0476*	-.0724*	-.1022*
.80	.8342	.8498	.8680	.7864	.0338*	.0182	.0156	-.0478*	-.0634*	-.0816*

Note: * indicates the difference is significant, that is, twice the upper bound of standard error for empirical power ($SE = \sqrt{2\times\frac{p\times(1-p)}{n}} = .01$ where p = .5 and n =5000), $2\times SE$ = .02.

Table 5. Power Pattern and Pairwise Power Difference of HLM, AR(1), UN and TRM by Sample Size when t = 3, β =.5, λ = .5 and G = $\begin{pmatrix} 1 & .3 \\ .3 & 1 \end{pmatrix}$ of 5000 MC Samples at α = .10

	Power				Power Difference					
G	RC	AR(1)	UN	TRM	UN vs. RC	UN vs. AR(1)	AR(1) vs.RC	TRM vs. RC	TRM vs. AR(1)	TRM vs.UN
Two-group Treatment Effect (β_{01} = .5, β_{10} = 0, β_{11} = 0)										
25	.1900	.2334	.2572	.3368	.0672*	.0238*	.0434*	.1468*	.1034*	.0796*
50	.3166	.3590	.3754	.5270	.0588*	.0164	.0424*	.2104*	.1680*	.1516*
75	.4264	.4688	.4946	.6762	.0682*	.0258*	.0424*	.2498*	.2074*	.1816*
100	.5092	.5502	.5836	.7884	.0744*	.0334*	.0410*	.2792*	.2382*	.2048*
125	.5976	.6362	.6652	.8538	.0676*	.0290*	.0386*	.2562*	.2176*	.1886*
150	.6794	.7130	.7376	.9130	.0582*	.0246*	.0336*	.2336*	.2000*	.1754*
175	.7430	.7750	.7982	.9418	.0552*	.0232*	.0320*	.1988*	.1668*	.1436*
200	.7950	.8228	.8452	.9704	.0502*	.0224*	.0278*	.1754*	.1476*	.1252*
Time Effect (β_{01} = 0, β_{10} = .5, β_{11} = 0)										
25	.3790	.4050	.4306	.5320	.0516*	.0256*	.0260*	.1530*	.1270*	.1014*
50	.5970	.6192	.6416	.8058	.0446*	.0224*	.0222*	.2088*	.1866*	.1642*
75	.7354	.7576	.7828	.9284	.0474*	.0252*	.0222*	.1930*	.1708*	.1456*
100	.8504	.8644	.8786	.9740	.0282*	.0142*	.0140	.1236*	.1096*	.0954*
125	.9124	.9230	.9386	.9916	.0262*	.0156*	.0106	.0792*	.0686*	.0530*
150	.9496	.9548	.9658	.9976	.0162*	.0110	.0052	.0480*	.0428*	.0318*
175	.9686	.9736	.9806	.9996	.0120	.0070	.0050	.0310*	.0260*	.0190*
200	.9814	.9842	.986	.9996	.0046	.0018	.0028	.0182*	.0154*	.0136
Interaction (β_{01} = 0, β_{10} = 0, β_{11} = 0.5)										
25	.2428	.2638	.2800	.2126	.0372*	.0162	.0210*	-.0302*	-.0512*	-.0674*
50	.3692	.3902	.4166	.3208	.0474*	.0264*	.0210*	-.0484*	-.0694*	-.0958*
75	.4844	.5092	.5390	.4368	.0546*	.0298*	.0248*	-.0476*	-.0724*	-.1022*
100	.5956	.6162	.6422	.5340	.0466*	.0260*	.0206*	-.0616*	-.0822*	-.1082*
125	.6812	.7008	.7264	.6158	.0452*	.0256*	.0196	-.0654*	-.0850*	-.1106*
150	.7442	.7630	.7916	.6896	.0474*	.0286*	.0188	-.0546*	-.0734*	-.1020*
175	.7982	.8146	.8378	.7436	.0396*	.0232*	.0164	-.0546*	-.0710*	-.0942*
200	.8498	.8622	.8838	.8000	.0340*	.0216*	.0124	-.0498*	-.0622*	-.0838*

Note: * indicates the difference is significant, that is, twice the upper bound of standard error for empirical power ($SE = \sqrt{2\times\frac{p\times(1-p)}{n}} = .01$ where p = .5 and n =5000), $2\times SE$ = .02.


Treatment Test
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
25
50
75
100
125
150
175
200
RC
AR(1)
UN
TRM
Time Test
RC
AR(1)
UN
TRM
Interaction Test
RC
AR(1)
UN
TRM

interaction test. As the sample size reached 200 (i.e., N = 1200), the power of all four was above .80 (see Table 5).

Table 5 also showed that the pairwise differences between TRM and the three VC structures were significant across the samples sizes in all three tests. Generally, TRM had significantly higher power than the three VC structures in the treatment and time tests but significantly lower power in the interaction test. It also appeared that the pair-wise differences shrunk as sample sizes increased.

Bootstrap Estimates

The bootstrap estimates, bias, standard errors and standard 95% confidence intervals of the treatment effect were examined within the three factors, G matrix, effect size and sample size per treatment group under specified conditions (see Appendix B). The results indicate that TRM, RC, AR(1) and UN generate unbiased and identical estimates of the treatment effect. TRM has slightly smaller bootstrap standard errors and hence slightly narrower confidence intervals. The bootstrap estimates of all three VC structures have similar patterns within each factor. As the correlation increases, the standard errors become slightly larger and therefore the confidence intervals are wider. As the reliability and sample sizes increase, the bootstrap standard errors decrease and confidence intervals become narrower.

Conclusion

This MC study primarily concerns the empirical power of TRM and HMLM under three variance-covariance (VC) structures in the longitudinal study. Specifically, this paper compared the power of TRM, 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. The three factors in this power study are the G matrix (G), reliability (λ), effect size (β) and sample size per treatment group (n).

Researchers have raised the question on what is the power to detect the interactions when they do exist in the HMLM data and expected HMLM perform better than traditional models but without proof (Davison, Kwak, Seo, and Choi, 2002; Kreft, 1996; Raudenbush, 1995). This study provided an 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, although it gains the significantly highest power in the main effects tests, treatment and time tests under the balanced design in the specified generic situations.

This study also supplements more comprehensive empirical indexes for estimating the model precision based on the bootstrap estimates and the approximate power for both main effects and interaction tests under more generic situations, including the empirical power indexes of HMLM under three different covariance structures which have not yet been specifically addressed in the literature. Based on this study, TRM could be the choice if researchers are more interested in main effect tests and the practical situation is most similar to this research where the balanced design is assumed and fixed effects are primarily the concern. If researchers are more concerned with interaction tests, this study recommends that UN, AR(1) or RC be the method of choice. When the number of repeated measures is 3, UN has the higher power than AR(1) or RC in the three tests within each factor. UN could be the choice if the practical situation is most similar to this research and if we need to try an exploratory analysis when the VC structure is assumed unknown.

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, the practitioners may estimate the sample sizes, the reliability, effect size or the correlation in G matrix for their studies if scenarios are similar to this study.

Future studies may consider extending this MC study by comparing power across factors instead of within each factor or fixing conditions and comparing the power by varying the sample size ratios between the number of subjects and time points while holding the total sample size. Instead of reliability, interclass correlation (ICC) 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 (the author, 2006) could be expanded to generate missing data or non-normal longitudinal data in order to be more practical and to examine the statistical properties and power of more complex growth models.

References

Akaike, H. (1973). Information theory as an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), *Second international symposium on information theory* (pp. 267-228). Budapest, Hungry: Akademiai Kiado.

Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin*. *101, 1*, 147-158.

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

Davison, M. L., Kwak, N., Seo, Y. S., & Choi, J. (2002). Using hierarchical linear models to examine moderator effects: Person-by-organization interactions. *Organizational Research Methods*, *5*, 231-254.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, 39(B)*, 1-8.

Dempster, A. P., Rubin, D. B., & Tsutakawa, R. K. (1981). Estimation in covariance components models. *Journal of the American Statistical Association*, *76*, 341-353.

Goldstein, H. (1994). Multilevel cross-classified models. *Sociological Methods & Research, 22*, 364-376.

Goldstein, H. (1995). *Multilevel statistical models* (2nd ed.). London: Arnold.

Greenhouse, S.W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, *32*, 95-112.

Hedeker, D., Gibbons, R. D., & Waternaux, C. (1999). Sample size estimation for longitudinal designs with attrition: Comparing time-related contrasts between two groups. *Journal of Educational and Behavioral Statistics*, *24(1)*, 70-93.

Huynh, H., & Feldt, L.S. (1976). Estimation of the Box Correction for Degrees of Freedom from Sample Data in the Randomized Block and Split Plot Designs. *Journal of Educational Statistics*, *1*, 69-82.

Hox, J. J. (2002). *Multilevel analysis techniques and applications*. Mahwah, N.J.: Lawrence Erlbaum Associates.

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

Kreft, I. G. G. (1996): Are multilevel techniques necessary? An overview, including simulation studies. Retrieved August 13, 2005, http://www.calstatela.edu/faculty/ikreft/quarterly/quarterly.html.

Littell, R. C., Henry, P. R., & Ammerman, C. B. (1998). Statistical analysis of repeated measures data using SAS procedures. *Journal of Animal Science*, 76, 1216-1231.

Littell, R. C., Milliken, G. A., Stroup, W. W., & Wolfinger, R. D. (2006). *SAS system for mixed models*. (2nd ed.). Cary, NC: SAS Institute Inc.

Little, R. J. A. (1995). Modeling the dropout mechanism in repeated-measures studies. *Journal of the American Statistical Association*, *90*, 1112-1121.

Little, R.J., & Rubin, D.B. (2002). *Statistical analysis with missing data (2nd edition)*. New York: John Wiley.

Maas, C. J. M., & Snijders, T. B. (2003). The multilevel approach to repeated measures for complete and incomplete data. *Quality & Quantity, 37*, 71–89.

McCulloch, C. E., & Searle, S. R. (2001). *Generalized, linear, and mixed models*. New York: John Wiley & Sons, Inc. Potthoff, R. F., & Roy, S. N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika*, *51*, 313-326.

Montgomery, D. C. (2005). *Design and analysis of experiments.* Hoboken, NJ : John Wiley & Sons, Inc.

Muthén, B. (2002). Beyond SEM: General latent variable modeling. *Behaviormetrika*, *29*, 81-117.

Muthén, B. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (ed.), *Handbook of quantitative methodology for the social sciences* (pp. 345-368). Newbury Park, CA: Sage Publications.

Muthén, L. & Muthén, B. (2006). *Mplus user's guide*. (4th ed.). Los Angeles: Muthén & Muthén.

Pinheiro, J. C., & Bates, D. M. (2000). *Statistics and computing: Mixed-effects models in S and S-PLUS*. New York: Springer.

Potthoff, R. F., & Roy, S. N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika*, *51*, 313-326.

Raudenbush, S. W. (1995). Maximum likelihood estimation for unbalanced multilevel covariance structure models via the EM algorithm. *British Journal of Mathematical and Statistical Psychology, 48* (2)*,* 359-370.

Raudenbush S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. London: Sage publications, Inc.

Raudenbush, S. W., Bryk, A. S., & Congdon, R. (2004). *HLM6: Hierarchical linear and nonlinear modeling.* Chicago: Scientific Software International.

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

Raudenbush, S. W., Spybrook, J., Liu, X., & Congdon, R. (2005). *Optimal design for longitudinal and multilevel research: Documentation for the "Optimal Design" software.* Retrieved September 20, 2006, from http://sitemaker.umich.edu/group-based/optimal_design_software

Rogosa, D. R., & Willett, B. (1985). Understanding correlates of changes by modeling individual differences in growth. *Psychometrica*, *50*, 203-228.

SAS Institute Inc. (2003). *SAS/STAT user's guide, version 9.1*. Cary, NC: SAS Institute Inc..

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis.* New York: Oxford University Press, Inc.

Skrondal, A. & Rabe-Hesketh, S. (2004). Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models. New York: Chapman & Hall/CRC.

Snijders, T. A. B. (2005). Power and sample size in multilevel linear models. In B.S. Everitt & D.C. Howell (Eds.), *Encyclopedia of Statistics in Behavioral Science*. Volume 3, 1570–1573. Chicester: Wiley.

Schluchter, M. D. (1988). Unbalanced repeated measures models with structured covariance matrices. In W. J. Dixon (Ed.), *BMDP statistical software manual*, *2*, 1081-1114. Berkeley, CA: University of California Press.

Schwarz, G. (1978). Estimating the dimensions of a model. *Annals of Statistics*, *6*, 461-464.

Van der Leeden, R., Vrijburg, K., & de Leeuw, J. (1996). 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, 21, 5*, 583-605.

Van der Leenden, R. (1998). Multilevel analysis of repeated measures data. *Quality & Quantity, 32*, 15–29.

Wolfinger, R. D. (1993). Covariance structure selection in general mixed models. *Communications in Statistics, Simulation and Computation*, *22, 4*, 1079 - 1106.

Wolfinger, R. D., & Chang, M. (1995). Comparing the SAS GLM and MIXED procedures for repeated measures. *SUGI Proceedings*. Cary, NC: SAS Institute Inc.

Acknowledgement: 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), 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.

Appendix A.

Table 6. Validation of Data Generator Using Potthoff and Roy's Data

	Potthoff and Roy's Data	Simulated Data
Intercept	21.2091	21.2063
Gender	1.4065	1.4065
Time	0.9591	0.9587
Gender*Time	0.6097	0.6115

Appendix B

Table 7. Bootstrap Estimates of Treatment Effect for RC, AR(1), UN and TRM by G Matrix, Effect Size, Reliability and Sample Size of 5000 MC Samples at α = .10. Table continued on next page.

(a) RC and RM

	Model									
	RC					RM				
	β_{01}	BIAS	SE	CI_low	CI_high	B01	BIAS	SE	CI_low	CI_high
G Matrix (ρ)										
0.10	0.50	0.00	0.31	-0.11	1.11	0.50	0.00	0.23	0.06	0.95
0.30	0.50	0.00	0.31	-0.12	1.12	0.50	0.00	0.23	0.04	0.96
0.50	0.50	0.00	0.32	-0.12	1.12	0.50	0.00	0.24	0.03	0.97
Effect size (d)										
0.20	0.20	0.00	0.31	-0.42	0.82	0.20	0.00	0.23	-0.26	0.66
0.50	0.50	0.00	0.31	-0.12	1.12	0.50	0.00	0.23	0.04	0.96
0.80	0.80	0.00	0.31	0.18	1.42	0.80	0.00	0.23	0.34	1.26
Reliability (λ)										
0.01	0.49	-0.01	2.59	-4.59	5.58	0.49	-0.01	1.62	-2.68	3.66
0.25	0.50	0.00	0.48	-0.45	1.45	0.50	0.00	0.33	-0.14	1.14
0.50	0.50	0.00	0.31	-0.12	1.12	0.50	0.00	0.23	0.04	0.96
0.75	0.50	0.00	0.23	0.05	0.95	0.50	0.00	0.19	0.12	0.88
1.00	0.50	0.00	0.17	0.16	0.84	0.50	0.00	0.17	0.16	0.84
Sample Size (n)										
25	0.50	0.00	0.54	-0.56	1.56	0.50	0.00	0.41	-0.31	1.31
50	0.50	0.00	0.38	-0.24	1.24	0.50	0.00	0.29	-0.06	1.06
75	0.50	0.00	0.31	-0.12	1.12	0.50	0.00	0.23	0.04	0.96
100	0.50	0.00	0.27	-0.04	1.03	0.50	0.00	0.20	0.10	0.90
125	0.50	0.00	0.25	0.02	0.98	0.50	0.00	0.18	0.14	0.86
150	0.50	0.00	0.23	0.06	0.94	0.50	0.00	0.17	0.17	0.83
175	0.50	0.00	0.20	0.10	0.90	0.50	0.00	0.16	0.19	0.81
200	0.50	0.00	0.19	0.12	0.88	0.50	0.00	0.15	0.22	0.79

(b) AR(1) and UN

	Model									
	AR(1)					UN				
	β_{01}	BIAS	SE	CI_low	CI_high	B01	BIAS	SE	CI_low	CI_high
G Matrix (ρ)										
0.10	0.50	0.00	0.31	-0.12	1.12	0.50	0.00	0.31	-0.11	1.11
0.30	0.50	0.00	0.32	-0.12	1.12	0.50	0.00	0.31	-0.11	1.11
0.50	0.50	0.00	0.32	-0.13	1.13	0.50	0.00	0.31	-0.11	1.11
Effect Sizes (d)										
0.20	0.20	0.00	0.32	-0.42	0.82	0.20	0.00	0.31	-0.41	0.81
0.50	0.50	0.00	0.32	-0.12	1.12	0.50	0.00	0.31	-0.11	1.11
0.80	0.80	0.00	0.32	0.18	1.42	0.80	0.00	0.31	0.19	1.41
Reliability (λ)										
0.01	0.49	-0.01	2.60	-4.60	5.58	0.49	-0.01	2.60	-4.60	5.59
0.25	0.50	0.00	0.49	-0.45	1.45	0.50	0.00	0.48	-0.45	1.45
0.50	0.50	0.00	0.32	-0.12	1.12	0.50	0.00	0.31	-0.11	1.11
0.75	0.50	0.00	0.24	0.04	0.96	0.50	0.00	0.22	0.06	0.93
1.00	0.50	0.00	0.18	0.14	0.86	0.50	0.00	0.16	0.19	0.81
Sample Size (n)										
25	0.50	0.00	0.55	-0.58	1.57	0.50	0.00	0.54	-0.56	1.56
50	0.50	0.00	0.38	-0.25	1.25	0.50	0.00	0.37	-0.24	1.23
75	0.50	0.00	0.32	-0.12	1.12	0.50	0.00	0.31	-0.11	1.11
100	0.50	0.00	0.28	-0.04	1.04	0.50	0.00	0.27	-0.03	1.02
125	0.50	0.00	0.25	0.01	0.98	0.50	0.00	0.24	0.02	0.97
150	0.50	0.00	0.23	0.05	0.95	0.50	0.00	0.22	0.06	0.93
175	0.50	0.00	0.21	0.10	0.90	0.50	0.00	0.20	0.11	0.89
200	0.50	0.00	0.19	0.12	0.88	0.50	0.00	0.19	0.13	0.87