

## Paper 138-31

**%hlmdata and %hlmrmppower: Traditional Repeated Measures vs. HLM for Multilevel Longitudinal Data Analysis—Power and Type I Error Rate Comparison**

Hua Fang, Ohio University

**ABSTRACT**

This paper proposes a stacked macro called %hlmdata and %hlmrmppower to generate multilevel longitudinal data and to compare power and Type I error rate of multivariate repeated measures (MRM) and hierarchical linear model (HLM) approaches. The comparison was conducted on the evaluation of a fixed between-group treatment effect in a balanced design through a Monte Carlo study. By varying correlation in the covariance matrix, effect size, reliability and ratio of level-2 sample size to level-1 time points, the power of the two models was examined and presented in tables and figures. Meanwhile, the bootstrap estimates for the treatment mean differences were calculated, such as bootstrap standard errors, bias and confidence intervals. The conclusion summarizes the Type I error rates and power patterns for the two models and discusses the utility of the macro.

**INTRODUCTION**

Longitudinal data analysis is used to analyze the data collected on each subject on two or more occasions (McCulloch & Searle, 2001). The balanced design in longitudinal data analysis assumes a complete data set with an equal number of measurements over time for each subject, while the unbalanced design has incomplete data without equal time intervals or time points for each subject (Maas & Snijders, 2003; Raudenbush, Bryk & Congdon, 2004). Literature shows that HLM and MRM can be employed in longitudinal data analysis (e.g., Hox, 2002; Kreft, 1996; Maas & Snijders, 2003; Raudenbush & Bryk, 2002; Van der Leenden, Vrijburg, & de Leeuw, 1996).

Traditionally, the application of repeated measures (RM) in longitudinal data analysis includes univariate and multivariate approaches. When sphericity is met, that is, under the Type H assumption (Huynh & Feldt, 1970), the univariate tests are appropriate because they will usually be more powerful than multivariate tests for the within-subject effect (Wolfinger & Chang, 1995). When sphericity is not met, we can employ the adjusted univariate tests, such as G-G (Greenhouse & Geisser, 1959) and H-F (Huynh & Feldt, 1976), or traditional multivariate tests with no assumption of covariance matrix, such as Wilks' Lambda, Pillai's Trace, Hotelling-Lawley Trace, and Roy's Greatest Root (Wolfinger & Chang, 1995). Sphericity is not a focus in this study because of two facets. First, this paper concerns the between-group effect test instead of within-group effect test. Second, in the theoretical development of MRM, the variance-covariance structure is not limited to the sphericity form (i.e., the univariate approach) or has no assumption of the covariance structure (i.e., the multivariate approach). Instead, the variance-covariance structure can be extended to various forms which can be modeled or specified (Jennrich & Schluchter, 1986; Raudenbush & Bryk, 2002; Singer & Willett, 2003; Wolfinger, 1993). The specification of variance-covariance matrix does affect the precision of estimates of the fixed effects, and will therefore impact hypothesis testing and confidence interval estimates (Singer & Willett, 2003). To simplify the comparison, the author concentrates on the traditional multivariate repeated measures (MRM) without assumption of covariance structure of the error term.

When hierarchical linear model (HLM) is applied to longitudinal data analysis, the level-1 units are the repeated measures for each subject and the level-2 units consist of subjects. The repeated measures are conceived as nested within each subject. The standard HLM possesses constant variance at level 1 and arbitrary covariance structure at level 2 (cf., Raudenbush & Bryk, 2002; Raudenbush, Spybrook, Liu & Congdon, 2005; Singer, 1998; Singer & Willett, 2003). If we intend to examine a two-group treatment effect on individuals' scores over time, a simple standard 2-level HLM with repeated measures at level1 and two-group treatment effect at level 2 can be expressed as follows:

$$\text{Level1: } y_{ti} = \pi_{0i} + \pi_{1i} * \text{TIME} + \varepsilon_{ti} \quad (1)$$

$$\begin{aligned} \text{Level2: } \pi_{0i} &= \beta_{00} + \beta_{01} * \text{TREATMENT} + u_{0i} \\ \pi_{1i} &= \beta_{10} + \beta_{11} * \text{TREATMENT} + u_{1i} \end{aligned} \quad (2)$$

$y_{ti}$ : Individual score

$\pi_{0i}$ : Individual's true score at time 0;

$\pi_{1i}$ : Individual's true slope (rate of change with respect to time)

$\beta_{00}$ : The average overall true initial score at time 0;

$\beta_{01}$ : The hypothesized difference in average true status from the effect of treatment;

$\beta_{10}$ : The average overall true annual rate of change at the second level;

$\beta_{11}$ : The hypothesized difference in average true annual rate of change from the effect of treatment;

$u_{0i}$ : The random effect for intercepts;

$u_{1i}$ : The random effect for slopes;

$\varepsilon_{it}$ : The random error with the  $t^{\text{th}}$  time point in the  $i^{\text{th}}$  individual.

(cf., Raudenbush & Bryk, 2002; Singer & Willett, 2003)

Substitute Equation (2) into (1), we can reduce the 2-level model into a single-level model:

$$y_{it} = (\beta_{00} + \beta_{01} * \text{TREATMENT} + \beta_{10} * \text{TIME} + \beta_{11} * \text{TREATMENT} * \text{TIME}) + (u_{0i} + u_{1i} * \text{TIME} + \varepsilon_{it}) \quad (3)$$

where  $\beta_{00}$ ,  $\beta_{01}$ ,  $\beta_{10}$ ,  $\beta_{11}$ ,  $u_{0i}$ ,  $u_{1i}$ , and  $\varepsilon_{it}$  are the same as those in Equation (1) and (2).

For Model (3), the standard HLM approach to variance-covariance structure of the random part ( $u_{0i} + u_{1i} * \text{TIME} + \varepsilon_{it}$ )

is that  $\varepsilon_{it} \sim N(0, \sigma^2)$  and  $\begin{bmatrix} u_{0i} \\ u_{1i} \end{bmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{00}^2 & \sigma_{01}^2 \\ \sigma_{10}^2 & \sigma_{11}^2 \end{pmatrix}\right]$ . For the same model, traditional MRM does not assume

the covariance matrix  $\Sigma_r$  of the error term  $r$  and  $r = u_{0i} + u_{1i} * \text{TIME} + \varepsilon_{it}$ .

To compare the two approaches, this study considers an empirical power analysis. The stacked macro, called %hlmdata and %hlmrmpower, has been written to generate multilevel data and implement power comparison respectively. This macro incorporated Base SAS (e.g., data step), SAS/STAT (e.g., PROC MIXED), SAS/IML, and MACRO facility into a unit, inheriting the existing functions in the SAS samples, for instance, Cholesky decomposition to generate correlated normal data. Monte Carlo methods, such as, parametric bootstrap standard confidence intervals, bias and standard errors, were also used to evaluate the model precision in this implementation.

## MONTE CARLO IMPLEMENTATION

The hypothesized test in this simulation using %hlmdata and %hlmrmpower is to detect the fixed two-group treatment effect on the individuals' scores over time for a balanced design. The MC study was conducted in SAS9.1 environment (SAS Institute Inc., 2003). SAS PROC MIXED was employed to conduct the estimation procedure for the two approaches. The underlying 2-level model for this simulation is presented in Equation (3). To make the results applicable over many possible applications, the standardized model was employed (cf. Raudenbush & Liu, 2000) where the grand mean was set to zero (i.e.,  $\beta_{00} = 0$ ) in Equation (3). Since time was assumed to have random slopes, it is reasonable to assume  $\beta_{10} = 0$  in (3). Thus, the standardized model in this simulation is as follows:

$$y_{it} = \beta_{01} * \text{TREATMENT} + \beta_{11} * \text{TREATMENT} * \text{TIME} + u_{0i} + u_{1i} * \text{TIME} + \varepsilon_{it} \quad (4)$$

To conduct the comparison in detecting the true treatment mean difference, the interaction between treatment and time is not considered (i.e.,  $\beta_{11} = 0$ ) for simplicity in this simulation. The corresponding traditional MRM and standard HLM model specification in PROC MIXED are shown in Appendix A.

## DATA GENERATION

Standard HLM has constant level-1 variance and arbitrary level-2 G matrix for random intercepts ( $u_{0i}$ ) and slopes ( $u_{1i}$ ). %hlmdata was used to generate data from Model (4). The data generation procedure was carried out as follows:

### 1. Level-1 Data

The error term at level-1 (i.e.,  $\varepsilon_{it}$ ) is assumed to be independent of the level-2 random effects (i.e.,  $u_{0i}$  and  $u_{1i}$ ), that is,  $\text{cov}(u_i, \varepsilon_{it}) = 0$ . The level-1 error term follows a normal distribution,  $\varepsilon_{it} \sim N(0, \sigma^2)$ , where,  $\sigma^2$  is determined by the reliability average across the level-1 coefficients (cf. Raudenbush & Bryk, 2002).

### 2. Level-2 Data

The random intercepts  $u_{0i}$  represented by  $X_{\text{intercept}}$  and slopes  $u_{1i}$  represented by  $X_{\text{slope}}$  follow a standardized bivariate normal distribution. A standardized G matrix (i.e., level-2 covariance matrix) for  $X_{\text{intercept}}$  and  $X_{\text{slope}}$ ,

$$C_X = \begin{pmatrix} \sigma_{00}^2 & \sigma_{01}^2 \\ \sigma_{10}^2 & \sigma_{11}^2 \end{pmatrix} \text{ or } C_X = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \text{ and random mean vector, } \begin{pmatrix} u_{0i} \\ u_{1i} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \text{ were specified to simulate correlated}$$

bivariate normal data. The Cholesky decomposition method was used to generate the correlated level-2 normal data. This simulation was accomplished by multiplying the normal data by Cholesky root  $L$  which is the Cholesky decomposition of  $C_X$ . The estimated variables were  $\hat{X}_{\text{intercept}}$  and  $\hat{X}_{\text{slope}}$ . The fixed treatment effect has a mean vector

$$\text{of} \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}.$$

### 3. Complete Data

To conform to the data format required by PROC MIXED, an index matrix was created for time, treatment and individual IDs. Two treatment groups were coded of 0 and 1, respectively. IDs were nested within each treatment group, for instance, IDs were from 1 to 25 for Group 1 and 26 to 50 for Group 2. Time started from 0 and extended to the value as specified. Based on Model (4), a univariate response vector of  $y_{ti}$  was created. Although traditional MRM has multivariate response data, that is, individuals have multi-dimensional responses; it has to be transformed when using PROC MIXED.

### TEST STATISTICS COMPUTING

To compute the power, the test statistic, F-statistics, and its corresponding p-values for the treatment effect from HLM and MRM were calculated by SAS PROC MIXED procedure and drawn from ODS tables. To be comparable, the study used the same type of F-statistic for standard HLM and traditional MRM. Type III F-statistic is the default for the fixed effect in PROC MIXED and was used in this study.

### MONTE CARLO DESIGN

This MC study assumes linearity and normality. %hlmrmppower was used to calculate the Type I error rates and power for the two models. The simulation ran 2000 replications and a nominal alpha of .10 was adopted considering the relative small number of replications.

#### 1. Type I error rates

The Type I error rates were evaluated when the treatment group means are assumed truly equal,  $\mu_1 = \mu_2 =$

$$0, \begin{bmatrix} u_{0i} \\ u_{1i} \end{bmatrix} \sim N \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right], \varepsilon_{it} \sim N(0, \sigma^2) \text{ and } \text{cov}(u_{it}, \varepsilon_{it}) = 0. \text{ For each replication, group sample size (n)}$$

increases from 25 to 200 by 25 with fixed time points (t) of 3. To acquire the type I error rates, the number of p-values for F statistic less than .10 were counted and divided by the number of replications.

#### 2. Power comparison

Power comparison was conducted by varying correlation of G matrix, effect size, reliability, and the ratio of the number of individuals to the number of time points. This study used Cohen's indices for effect size (d) of the two-group treatment effect (cf. Raudenbush & Liu, 2000; Raudenbush et al., 2005),  $d \in \{.2 \ .5 \ .8\}$ , and Cohen's indices for correlation, that is,  $\rho \in \{.1 \ .3 \ .5\}$ . Correspondingly, G correlation matrix ( $C_x$ ) for random intercepts and slopes was

specified as  $\begin{pmatrix} 1 & 0.1 \\ 0.1 & 1 \end{pmatrix}$ ,  $\begin{pmatrix} 1 & 0.3 \\ 0.3 & 1 \end{pmatrix}$  or  $\begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$ . The range of reliability varied from .01 to 1,

specifically,  $\lambda \in \{.01 \ .25 \ .5 \ .75 \ 1\}$ . The ratio of sample sizes to time points was categorized into two groups: The first group assumed the number of time points (t) was fixed at 3 and n increased from 25 to 200 by 25, ratio  $\theta_1 \in \{25/3 \ 50/3 \ 75/3 \ 100/3 \ 125/3 \ 150/3 \ 175/3 \ 200/3\}$ ; the second group assumed the total number N (n\*t) was fixed and ratio  $\theta_2 \in \{20/20 \ 40/10 \ 80/5 \ 100/4\}$ . Each of the four factors was examined by holding the other three constant so as to optimize the comparison and present general power patterns of HLM and MRM under specified situations, which is illustrated in following tables.

#### a. MC design for power comparison by G matrix

To show the power pattern by varying correlation in G matrix, the MC design chose a moderate sample size ( $n = 75$ ), effect size ( $d = .5$ ), reliability ( $\lambda = .5$ ) and fixed time points ( $t = 3$ ) to simulate a specific situation and compare power under each correlation G matrix, which is present as below:

Table 1

MC Design for Power Pattern of HLM and MRM by correlation in G Matrix

Condition: $n = 75$ , $d = .5$ , $\lambda = .5$ and $t = 3$			
G matrix ( $C_x$ )	$\begin{pmatrix} 1 & 0.1 \\ 0.1 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.3 \\ 0.3 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$

Based on the design, a general power pattern of HLM an MRM was represented under each G matrix in this specified situation.

b. MC design by effect sizes

Similarly, the MC design simulated a moderate situation where  $n = 75$ ,  $C_x = \begin{pmatrix} 1 & 0.3 \\ 0.3 & 1 \end{pmatrix}$ ,  $\lambda = .5$  and  $t = 3$  to compare power for each effect size, which is present as below:

Table 2

*MC Design for Power Pattern of HLM and MRM by Effect Sizes*

Condition: $n = 75$ , $C_x = \begin{pmatrix} 1 & 0.3 \\ 0.3 & 1 \end{pmatrix}$ , $\lambda = .5$ and $t = 3$			
$d$	0.2	0.5	0.8

A general power pattern of HLM an MRM was shown for each effect size in this specified situation.

c. MC design for power comparison by reliability

Holding a moderate situation where  $d=.5$ ,  $C_x =$ ,  $t = 3$ , and  $n = 75$ , a MC design for power comparison by reliability is described below:

Table 3

*MC Design for Power Pattern of HLM and MRM by Reliability*

Condition: $d=.5$ , $C_x = \begin{pmatrix} 1 & 0.3 \\ 0.3 & 1 \end{pmatrix}$ , $t = 3$ , and $n = 75$					
$\lambda$	.01	.25	.50	.75	1

In this specified situation, the power comparison was undertaken for each reliability index and a general power pattern for MRM and HLM was illustrated.

d. MC design for power comparison by ratio

For Group I ratio,  $\theta_1$ , the condition was specified as  $d=.5$ ,  $C_x = \begin{pmatrix} 1 & 0.3 \\ 0.3 & 1 \end{pmatrix}$ ,  $t = 3$  and  $\lambda = .5$ . For Group II ratio,  $\theta_2$ , the

condition was set as total  $N = 200$ ,  $d = .5$ ,  $C_x = \begin{pmatrix} 1 & 0.3 \\ 0.3 & 1 \end{pmatrix}$ ,  $\lambda = .5$ . Table 5 shows the design for the two groups.

Table 5

*MC Design for Power Pattern of HLM and MRM by Ratio*

Group I condition: $d=.5$ , $C_x = \begin{pmatrix} 1 & 0.3 \\ 0.3 & 1 \end{pmatrix}$ , $t = 3$ and $\lambda = .5$								
$\theta_1$	25/3	50/3	75/3	100/3	125/3	150/3	175/3	200/3
Group II condition: $N = 800$ , $d = .5$ , $C_x = \begin{pmatrix} 1 & 0.3 \\ 0.3 & 1 \end{pmatrix}$ , $\lambda = .5$								
$\theta_2$	20/20		40/10		80/5		100/4	

For each specified condition, power pattern for HLM and MRM was presented by different ratios of number of samples sizes to number of time points.

In addition to the power analysis, bootstrap standard CI, bias and standard errors for the estimates of the treatment mean difference were calculated for each design.

### MONTE CARLO RESULTS

The empirical results for the MC designs are summarized as follows:

1. Empirical Type I error rates

Table 6

*Type I Error Rates of HLM and MRM by Group Sample Sizes at the Nominal Alpha of .10*

<i>n</i>	25	50	75	100	125	150	175	200
HLM	0.1020	0.0805	0.0995	0.0945	0.0970	0.1020	0.1035	0.0900
MRM	0.1055	0.0935	0.0920	0.0965	0.0950	0.1045	0.1080	0.0995

As indicated in Table 6 and Figure 1, the type I error rates for HLM and MRM are, apparently, close to .1 across sample sizes. (For graph, see Appendix B).

## 2. Empirical power

The empirical power of HLM and MRM for detecting the fixed two-group treatment effect is presented for the four designs along with bootstrap estimates of treatment mean difference, bias, standard errors and standard confidence intervals.

### a. Power pattern of HLM and MRM by correlation in G matrix

Table 7

*Power Pattern of HLM and MRM by correlation in G Matrix under Condition of  $n = 75$ ,  $d = .5$ ,  $\lambda = .5$  and  $t = 3$* 

G matrix ( $C_x$ )	$\begin{pmatrix} 1 & 0.1 \\ 0.1 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.3 \\ 0.3 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$
HLM	0.4315	0.4215	0.4010
MRM	0.6995	0.6760	0.6530

Under this specified circumstance, the power of HLM and MRM decreases by approximately 2% on average as the correlation in G matrix increases by 20%. Traditional MRM seems to have an average of 26% higher power than standard HLM across the G matrices. (For graph, see Appendix B).

### b. Power pattern of HLM and MRM by effect sizes

Table 8

*Power Pattern of HLM and MRM by Effect Sizes under Condition of  $n = 75$ ,  $C_x = \begin{pmatrix} 1 & 0.3 \\ 0.3 & 1 \end{pmatrix}$ ,  $\lambda = .5$  and  $t = 3$* 

<i>d</i>	0.2	0.5	0.8
HLM	0.1315	0.4215	0.7700
MRM	0.2120	0.6760	0.9650

In this specific condition, the power of HLM and MRM increases approximately on average by 32% and 38% respectively as the effect sizes raise by 30%. Standard HLM seems to have an 18% lower power on average than traditional MRM across the effect sizes. (For graph, see Appendix B).

### c. Power pattern of HLM and MRM by reliability

Table 9

*Power Pattern of HLM and MRM by Reliabilities under Condition of  $n = 75$ ,  $d = .5$ ,  $C_x = \begin{pmatrix} 1 & 0.3 \\ 0.3 & 1 \end{pmatrix}$  and  $t = 3$* 

$\lambda$	.01	.25	.50	.75	1
HLM	0.0995	0.2385	0.4220	0.6050	0.7755
MRM	0.1125	0.4425	0.6775	0.8210	0.8970

Table 9 indicates that power of HLM and MRM increases as reliability enhances. Standard HLM raises power at an approximate average rate of 17%. Traditional MRM seems to increase power at a decreasing rate from initial rate of approximate 33% between reliability .01 and .25 to the rate of 8% between .75 to 1. On average, MRM has a higher power of 16% than HLM across the range of reliability. (For graph, see Appendix B).

### d. Power pattern of HLM and MRM by ratios

Table 10

*Power Pattern of HLM and MRM by Group I and Group II Ratios of Sample Sizes to Time Points*

Group I condition: $d = .5$ , $C_x = \begin{pmatrix} 1 & 0.3 \\ 0.3 & 1 \end{pmatrix}$ , $t = 3$ and $\lambda = .5$								
$\theta_1$	25/3	50/3	75/3	100/3	125/3	150/3	175/3	200/3
HLM	0.1885	0.3010	0.4215	0.5135	0.5900	0.6810	0.7430	0.8070
MRM	0.3470	0.5255	0.6760	0.7935	0.8565	0.9095	0.9375	0.9695
Group II condition: $N = 800$ , $d = .5$ , $C_x = \begin{pmatrix} 1 & 0.3 \\ 0.3 & 1 \end{pmatrix}$ , $\lambda = .5$								
$\theta_2$	20/20		40/10		80/5		100/4	
HLM	0.0205		0.0725		0.2850		0.4285	
MRM	0.1645		0.2900		0.6195		0.7435	

Under Group I condition where time points are fixed at 3, the power of HLM and MRM increases as the group sample size expands. Both have a relatively higher rate of increase when sample size rises from 25 to 75, HLM with an average increase of 12% and MRM of 16%. When sample size increases from 75 to 200, HLM has an average power-increasing rate of 8% while the power increase of MRM slows down from 12% to 3%. Across Group I ratios, traditional MRM seem to have an average of 22% higher power than standard HLM. (For graph, see Appendix B).

When holding the total sample size  $N$  constant,  $N=800$  in this situation, both HLM and MRM have obvious power increasing rates as the ratio of sample size to time points increases (i.e., power increases as the ratio of the level-2 sample size to the number of level-1 repeated measures increases). Under each ratio category, traditional MRM has a higher power than standard HLM and the power difference becomes larger as the ratio increases, for instance, MRM is about 14% higher power than HLM at the ratio of 20/20 and about 32% at the ratio of 100/4, which is shown in Figure 3.

Along with the power comparison, bootstrap bias, standard error and standard confidence intervals (Martinez and Martinez, 2002) were calculated for the estimates of treatment mean difference in the above designs. Generally, the results show that standard HLM and traditional MRM are similarly unbiased for the mean estimates while MRM has smaller standard errors and narrower confidence intervals than HLM. (See table in Appendix B).

## CONCLUSION

Using %hlmdata and %hlmrmppower, this research examines the empirical power of traditional MRM and standard HLM in detecting the two-group fixed treatment effect on the individuals' scores over time in the balanced longitudinal analysis. The traditional MRM does not assume the covariance structure of the random part while standard HLM assumes homogeneous level-1 variance and arbitrary level-2 covariance structure (G matrix). The two approaches were compared based on the data generated from a simple 2-level standard HLM model (i.e., repeated measures at level-1 and individuals at level-2 with random intercepts and slopes).

The results show that the two models have the similar empirical Type I error rates under specified condition. The power of standard HLM and traditional MRM was compared by varying one of four factors and holding constant the other three, correlation in G matrix, effect sizes, reliability and the ratio of sample size to time points (by fixing the time points or the total size). The overall results show that standard HLM is more conservative than traditional MRM. When correlation in G matrix increases, the power of HLM and MRM decreases slightly by 2% in this study. As effect size increases, both power increases on average by approximately the amount of the increase of effect size, for example, HLM power increase by 32% and MRM by 36% when effect size rises by 30%. When reliability is low, both have a lower power. With increasing reliability, HLM power increases by 17% at an average rate over the specified range while MRM power increases at a decreasing rate under the specified condition. With fixed time points and increasing sample size, both power rise but the power of MRM increases faster than HLM. When fixing the total  $N$  and the ratio of sample size to time points increases, the power of the two models improves at a higher rate, which indicates that the increase of the level-2 size is more important than level-1 size. This has been discussed in the literature (e.g., Hox, 2002; Raudenbush et al., 2005).

The general pattern of traditional MRM and standard HLM can be extended to other specified conditions as long as the repeated measures data is generated from the standard HLM (i.e., with standard HLM covariance structure and data structure) and the research is confined to the balanced longitudinal analysis. In addition to traditional MRM,

Jennrich and Schluchter's MRM model (1986) with structured covariance matrix can also be examined using this macro. As an example, the model specification with the covariance matrix of first-order autoregressive (AR(1)) using PROC MIXED is included in Appendix C and the results are compared with those from the HMLM module in HLM6.0 (Raudenbush, Bryk & Congdon, 2004).

In this study, the two treatment groups are assumed to have an equal size while the stacked macro can allow the unequal size. However, if time variable needs to have a polynomial form or treatment groups require size expansion or the data needs to have other covariance structures (e.g., AR(1)), a small amount of modification is needed for the macro. This macro can be tailored to other specific HLM or mixed model designs. Under those circumstances users are encouraged to contact the author.

## APPENDIX A

Traditionally, multivariate repeated measures analysis is conducted in PROC GLM. PROC MIXED can mimic PROC GLM to give the same evaluation of the fixed between group effect. Take a typical example of longitudinal data from Pothoff and Roy (1964) for illustration. The data consist of dental measurements from the center of the pituitary to the pteryomaxillary fissure for 11 girls and 16 boys at ages 8, 10, 12, and 14. The subjects are the individual children, and there are four repeated measurements on each, and therefore this is a balanced design. For illustrative purpose, the original variable "gender" (similar to "treatment" in this study) with values of "F" and "M" is coded as 0 for female and 1 for male; the variable "age" (similar to "time" in this study) is coded as 0, 1, 2, 3, standing for the original four ages. The original data can be found in SAS documentation or Wolfinger and Chang's article (1995). Using traditional MRM approach, the PROC GLM statements and corresponding output are as follows:

```
proc glm data=forglm;
  class gender;
  model y1-y4=gender / nouni;
  repeated age / printe;
run;
```

The GLM Procedure					
Repeated Measures Analysis of Variance					
Tests of Hypotheses for Between Subjects Effects					
Source	DF	Type III SS	Mean Square	F Value	Pr > F
gender	1	140.4648569	140.4648569	9.29	0.0054
Error	25	377.9147727	15.1165909		

The following PROC MIXED statements mimics PROC GLM and generates the same evaluation of the between-group gender effect (cf. Wolfinger & Chang, 1995).

```
proc mixed data=formixed ;
  class gender age person;
  model y = gender age gender*age ;
  repeated age/ type=un sub=person;
  lsmeans gender / pdiff;
run;
```

(1)

The Mixed Procedure				
Type 3 Tests of Fixed Effects				
Effect	Num	Den	F Value	Pr > F
gender	DF	DF		
gender	1	25	9.29	0.0054
age	3	25	34.45	<.0001
gender*age	3	25	2.93	0.0532

The standard HLM specification in PROC MIXED is as follows:

```
proc mixed data=formixed method=ml;
  class person;
```



```

model y = gender age gender*age/s ddfm=bw;
random INTERCEPT age / SUBJECT =person TYPE = UN G;
run;

```

(cf. Singer, 1998) (2)

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr >  t
Intercept	21.2091	0.6110	25	34.71	<.0001
gender	<b>1.4065</b>	0.7937	25	1.77	0.0886
age	0.9591	0.1996	79	4.80	<.000
gender*age	0.6097	0.2593	79	2.35	0.0212

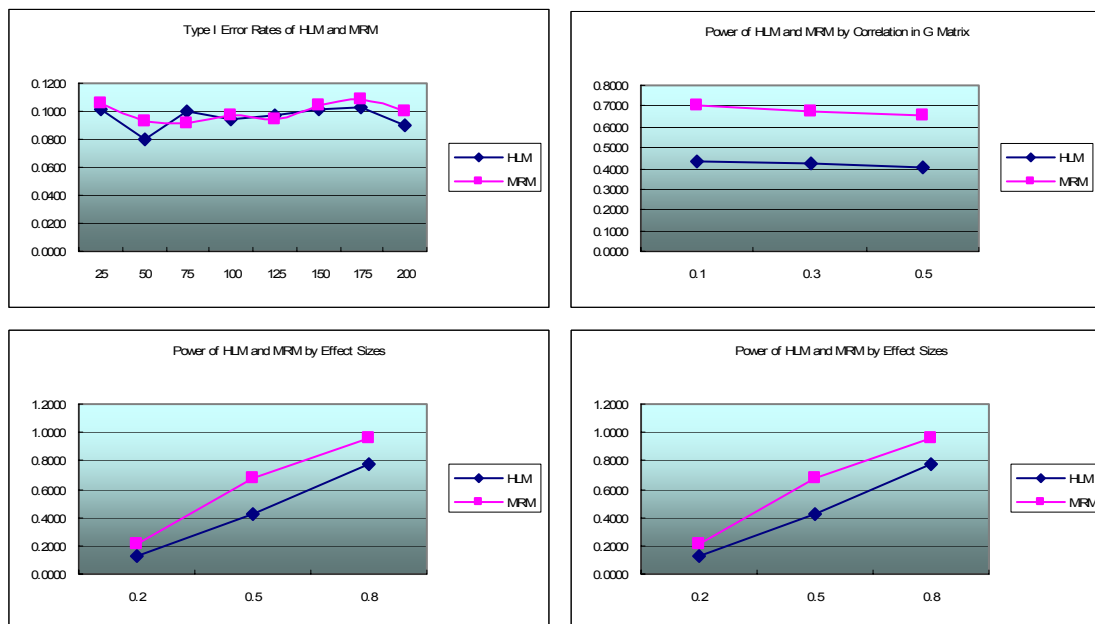
The same estimates were generated from HLM2 in HLM6.0 (Raudenbush, Bryk & Congdon, 2004)

Final estimation of fixed effects (with robust standard errors)						
Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value	
For INTRCPT1, B0						
INTRCPT2, G00	21.209091	0.560431	37.844	25	0.000	
GENDER, G01	<b>1.406534</b>	0.773799	1.818	25	0.081	
For AGE slope, B1						
INTRCPT2, G10	0.959091	0.126265	7.596	25	0.000	
GENDER, G11	0.609659	0.233735	2.608	25	0.015	

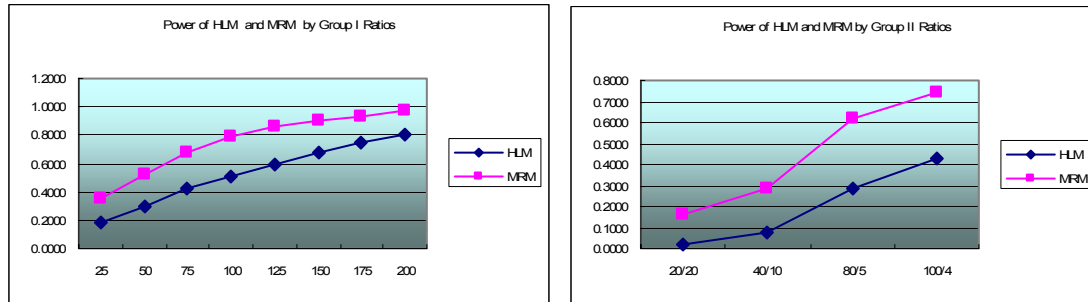
Therefore, PROC MIXED statement (1) and (2) are used for traditional MRM and standard HLM respectively

## APPENDIX B

The graphs for the four power comparison designs are as follows:







The bootstrap estimates are presented below:

	HLM					MRM				
	$\hat{d}$	Bias $\hat{s}$	S $\hat{E}$	CI_low	CI_high	$\hat{d}$	Bias $\hat{s}$	S $\hat{E}$	CI_low	CI_high
<b>G Matrix (<math>\rho</math>)</b>										
0.10	0.493	-0.007	0.309	-0.113	1.100	0.496	-0.004	0.225	0.055	0.938
0.30	0.493	-0.007	0.312	-0.119	1.105	0.496	-0.004	0.234	0.038	0.954
0.50	0.494	-0.006	0.315	-0.124	1.111	0.496	-0.004	0.242	0.021	0.970
<b>Effect Sizes (<math>d</math>)</b>										
0.20	0.193	-0.007	0.312	-0.419	0.805	0.196	-0.004	0.234	-0.262	0.654
0.50	0.493	-0.007	0.312	-0.119	1.105	0.496	-0.004	0.234	0.038	0.954
0.80	0.793	-0.007	0.312	0.181	1.405	0.796	-0.004	0.234	0.338	1.254
<b>Reliability (<math>\lambda</math>)</b>										
0.01	0.465	-0.035	2.583	-4.597	5.528	0.479	-0.021	1.602	-2.660	3.618
0.25	0.491	-0.009	0.481	-0.452	1.434	0.495	-0.005	0.325	-0.142	1.131
0.50	0.493	-0.007	0.312	-0.119	1.105	0.496	-0.004	0.234	0.038	0.953
0.75	0.495	-0.005	0.230	0.044	0.946	0.497	-0.003	0.194	0.116	0.877
1.00	0.497	-0.003	0.175	0.153	0.840	0.498	-0.002	0.172	0.161	0.834
<b>Group I Ratios (<math>\theta_1</math>)</b>										
25	0.502	0.002	0.548	-0.573	1.576	0.499	-0.001	0.406	-0.297	1.295
50	0.491	-0.009	0.370	-0.233	1.216	0.491	-0.009	0.283	-0.064	1.045
75	0.493	-0.007	0.312	-0.119	1.105	0.496	-0.004	0.234	0.038	0.954
100	0.499	-0.001	0.273	-0.036	1.033	0.500	0.000	0.200	0.109	0.891
125	0.497	-0.003	0.244	0.018	0.975	0.499	-0.001	0.182	0.141	0.857
150	0.500	0.000	0.228	0.054	0.946	0.500	0.000	0.166	0.174	0.826
175	0.501	0.001	0.209	0.092	0.910	0.504	0.004	0.158	0.195	0.813
200	0.501	0.001	0.189	0.131	0.871	0.498	-0.002	0.144	0.216	0.780
<b>Group II Ratios (<math>\theta_2</math>)</b>										
20/20	0.498	-0.002	1.127	-1.711	2.708	0.499	-0.001	0.855	-1.176	2.175
40/10	0.484	-0.016	0.606	-0.703	1.671	0.504	0.004	0.451	-0.379	1.387
80/5	0.493	-0.007	0.341	-0.174	1.161	0.502	0.002	0.249	0.015	0.990
100/4	0.500	0.000	0.285	-0.059	1.059	0.500	0.000	0.213	0.083	0.916

## APPENDIX C

Fitting the same data in Appendix A, the PROC MIXED statements for AR(1) covariance structure and output are as follows:

```
proc mixed data=formixed method = ml;
  class wave person;
  model y = gender|age/s ddfm=bw;
```

```

random INTERCEPT age / SUBJECT =person TYPE = UN G;
repeated wave/ type=AR(1) sub=person R;
run;
(cf. Singer, 1998)

```

Solution for Fixed Effects					
Effect	Estimate	Error	Standard DF	t Value	Pr >  t
Intercept	21.2226	0.5952	25	35.66	<.0001
gender	1.3142	0.7731	25	1.70	0.1016
age	0.9515	0.2020	79	4.71	<.0001
gender*age	0.6441	0.2624	79	2.45	0.0163

The above estimates are the same as those evaluated from HMLM in HLM6.0, standalone HLM software. See below:

Final estimation of fixed effects:					
Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1, P0					
INTRCPT2, B00	21.222636	0.595150	35.659	25	0.000
GENDER, B01	1.314197	0.773122	1.700	25	0.101
For AGE slope, P1					
INTRCPT2, B10	0.951497	0.201995	4.711	25	0.000
GENDER, B11	0.644101	0.262399	2.455	25	0.022

## REFERENCES

- Greenhouse, S.W. and S. Geisser. 1959. "On Methods in the Analysis of Profile Data," *Psychometrika*, 32, 95-112.
- Hox, J. J. 2002. *Multilevel Analysis Techniques and Applications*. Mahwah, N.J.: Lawrence Erlbaum Associates.
- Huynh, H., and Feldt, L. S. 1970, "Conditions under Which Mean Square Ratios in Repeated Measurements Designs Have Exact F-Distribution." *Journal of the American Statistical Association*, 65, 1582-1589.
- Huynh, H., and 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.
- Jennrich, R. I. & M. D. Schluchter. 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. <http://www.calstatela.edu/faculty/ikreft/quarterly/quarterly.html>
- Maas, C. J. M. & T. B. Snijders. 2003. "The Multilevel Approach to Repeated Measures for Complete and Incomplete Data". *Quality & Quantity* 37: 71-89.
- Martinez, W. L., and A. R. Martinez. 2002. *Computational Statistics Handbook with Matlab*. New York: Chapman & Hall/CRC.
- McCulloch, C. E., and S. R. Searle. 2001. *Generalized, linear, and mixed models*. New York: John Wiley & Sons, Inc.
- Raudenbush S. W., and A. S. Bryk. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*. London: Sage publications, Inc..
- Raudenbush, S.W., and X. Liu. 2000. "Statistical Power and Optimal Design for Multisite Randomized Trials." *Psychological Methods*, 5, 219-213.
- Raudenbush, S. W., A. S. Bryk, and R. Congdon. 2004. *HLM6: Hierarchical Linear and Nonlinear Modeling*. Chicago: Scientific Software International.
- Raudenbush, S. W., J. Spybrook, X. Liu and R. Congdon 2005. Optimal Design for Longitudinal and Multilevel Research: Documentation for the "Optimal Design" Software. <http://scholar.google.com>.
- Singer, J., 1998. "Using SAS PROC MIXED to Fit Multilevel Models, Hierarchical Models, and Individual Growth Models". *Journal of Educational and Behavioral Statistics*, 24, 4, 323-355.
- Singer, J. D., and J. B. Willett. 2003. *Applied Longitudinal Data Analysis*. New York: Oxford University Press, Inc..

- Van der Leeden, R., Vrijburg, K., and 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(23).
- Wolfinger, R. D. and M. Chang.1995. "Comparing the SAS GLM and MIXED Procedures for Repeated Measures". *SUGI Proceedings*. Cary, NC: SAS Institute Inc..
- Wolfinger, R.D. 1993. "Covariance Structure Selection in General Mixed Models." *Communications in Statistics, Simulation and Computation*, 22, 4, 1079 - 1106.

### ACKNOWLEDGMENTS

I am grateful to Gordon P. Brooks, Maria L. Rizzo, Robert S. Barcikowski and Nancy Brucken for valuable assistance in the preparation of this paper.

### CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Hua Fang  
Ohio University  
22 First Street  
Athens, Ohio 45701  
E-mail: hf101702@ohio.edu  
Web: <http://oak.cats.ohiou.edu/~hf101702/>

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.