An Improved Model for Evaluating Change in

Randomized Pretest, Posttest, Follow-up Designs

Constance A. Mara, Robert A. Cribbie, David B. Flora,

Cathy LaBrish, Laura Mills, & Lisa Fiksenbaum

Department of Psychology

York University

Abstract

Randomized pretest-posttest-follow-up (RPPF) designs are often used for evaluating the effectiveness of an intervention. These designs typically address two primary research questions: 1) Do the treatment and control groups differ in the amount of change from pretest to posttest? and 2) Do the treatment and control groups differ in the amount of change from posttest to follow-up? This study presents a model for answering these questions and compares it to recently proposed models for analyzing RPPF designs due to Mun, von Eye, & White (2009) using Monte Carlo simulation. The proposed model provides increased power over previous models for evaluating group differences in RPPF designs.

Keywords: *latent change modeling, longitudinal data, randomized pretest-posttest-follow-up designs.*

**An Improved Model for Evaluating Change in**

**Randomized Pretest, Posttest, Follow-up Designs**

Change is a process that is inevitably of interest to behavioral scientists, and interest in change is even more salient when it results (or is expected to result) from an intervention. Given this interest in the analysis of change, appropriate statistical techniques are necessary to accurately estimate the amount of change within a group or compare the amount of change across groups. The measurement of change has had a long, and often vibrant, history (e.g., Bryk & Raudenbush, 1987; Cronbach & Furby, 1970; Lord, 1967; Rogosa & Willett, 1983; Thurstone, 1930; Williams & Zimmerman, 1977), with modern researchers understanding that the appropriate statistical technique needed for evaluating change depends on many factors, including the nature of the research question, the research design, and the number of time points with observed data.

A common longitudinal design in intervention research is the randomized pretest, posttest, follow-up (RPPF) design (see Rausch, Maxwell & Kelley, 2003). In this design, participants are randomly assigned to treatment and control conditions, where only participants in the treatment group receive the intervention. All participants are measured prior to the intervention (pretest), immediately following (or shortly after) the intervention (posttest), and at some time following the termination of the intervention (follow-up). Although it is well known that more time points allows a better estimate of the pattern of growth over time (e.g., Willett, 1989), intervention studies rarely offer the opportunity to obtain multiple measurements between pretest and posttest or between posttest and follow-up. Further, these intermediate time points are often not relevant to the research questions addressed. For example, the two most common

research questions addressed by researchers utilizing RPPF designs are: 1) Do the groups differ in the amount of change from pretest to posttest? and 2) Do the groups differ in the amount of change from posttest to follow-up? In other words, clinicians are usually interested in both whether the intervention is effective as well as whether the treatment effects (if there are any) are sustained (or even accentuated) over time.

Many statistical approaches have been proposed for assessing change over time, or for investigating correlates of change over time, including paired-samples $t$-tests, analysis of variance (ANOVA), analysis of covariance (ANCOVA), multivariate analysis of covariance (MANCOVA), and multiple regression (Curran & Hussong, 2000). Structural equation modeling (SEM) and hierarchical linear modeling (HLM; also known as multilevel modeling) have also provided novel (and often improved) approaches to the analysis of change (e.g., Bryk & Raudenbush, 1987; Meredith & Tisak, 1990). Specific to RPPF designs, there has been considerable interest lately in latent change models (LCMs), which allow researchers to address questions related to group differences in changes across the three time points. In this paper we outline the statistical issues involved in measuring change and comparing changes across groups, discuss current LCMs for RPPF designs, and propose a simplified LCM for evaluating the primary research questions of researchers using RPPF designs.

**Traditional Approaches to Assessing Change over Time**

Traditionally, there have been two popular methods for comparing the amount of change across multiple groups: ANCOVA and ANOVA on change (posttest-pretest) scores. Change scores are also commonly referred to as 'difference scores' or 'gain scores'. Although most of the literature has focused on two time-point designs, this discussion is easily extended to three

time-point RPPF designs. More specifically, in RPPF designs the researcher typically

investigates group differences in change across each pair of time points separately (i.e., pretest to

posttest; posttest to follow-up), and thus, the only difference from a two time-point design is that

the specific time points utilized differ across analyses.

When participants are randomly assigned to groups, comparing the groups on the

outcome after covarying for the pretest will provide a more powerful test than will an ANOVA

on change scores (Huck & Maclean, 1975). Covarying for the pretest helps remove chance

variations between the groups, which also helps address issues due to regression to the mean and

floor or ceiling effects (Campbell & Kenny, 1999). However, when subjects are not randomly

assigned to groups (e.g., naturally occurring groups), covarying for the pretest is not appropriate

because differences between the groups occur not only because of chance variation, but also

because of systematic differences between the groups (e.g., Huitema, 1980; Lord, 1967; Rogosa,

1995; Schafer, 1992). An ANOVA on difference scores does not involve regressing the posttest

on the pretest (i.e., covarying for the pretest), and therefore it provides an unbiased estimate of

the difference in change across groups when the groups differ at baseline (Wright, 2006). More

simply put, when comparing the amount of change over two time points across groups, an

ANCOVA (with the pretest as covariate) is recommended when subjects are randomly assigned

to groups, and an ANOVA on difference scores is recommended when subjects are not randomly

assigned to groups.

These recommendations apply directly to analyses related to RPPF designs. When

comparing the amount of change from pretest to posttest, covarying for the pretest is

recommended because subjects are randomly assigned to groups and therefore any differences

between the groups can be attributed to chance. However, covarying for the posttest when comparing the amount of change from posttest to follow-up would be inappropriate because differences between the groups at posttest are the result of not only chance variation, but also systematic differences between the groups (i.e., posttest scores hopefully differ because one group received an intervention while the other did not; Lord, 1967). In this case, comparing the group change scores would be most appropriate. These points will play an important role in the development of an improved model for analyzing RPPF designs presented later in the paper.

**Latent Change Models**

SEM is a sophisticated statistical technique that allows researchers to assess the relationships among latent constructs estimated from observed measures (Bollen & Curran, 2006). LCMs, which are a specific type of SEM, are used to assess hypotheses based on the amount of change across multiple time points. Over the past couple decades there has been a substantial amount of research on LCMs, including LCMS for identifying predictors of change, and identifying latent classes that differ in their growth trajectories (Bollen & Curran; Duncan, Duncan, Strycker, Li & Alpert, 2006; Matschinger, Kilian, & Angermeyer, 2006; Raykov, 1993; for a review of longitudinal models of change see Rosel and Plewis, 2008). More specifically, LCMs offer information not only on average change (e.g, how much did the experimental group change from pretest to posttest), but also the extent, magnitude and variability of individual changes. In this paper we use the term 'latent change model' or 'LCM' to represent a wide variety of structural equation models that assess change over time, such as latent growth curve models, latent curve analysis models, and latent difference score models. In some instances we refer to a specific type of change model and in those instances we use the more specific labels.

LCMs use the repeated measures to estimate the amount of change across time. In a typical latent growth curve model, the slope and the intercept are the latent factors, where the intercept represents the reference point (typically the pretest) and the slope represents the rate of change over time. The rate of change is often expressed in units similar to a difference score, but this depends on the exact parameterization of the model (see Cribbie, Mills, & Jamieson, in press). Unlike SEM, which typically only incorporates information about the covariance structure of the observed measures, LCMs often incorporate information about the covariance structure as well as the mean structure of the observed measures. Recently, LCMs have become a popular technique because they offer unique advantages to addressing the complex research questions associated with longitudinal data (Curran, 2000). For example, LCMs permit the estimation of the measurement errors associated with the repeated measures. LCMs also have advantages over traditional methods in terms of how the relationships among the variables are modeled. For example, growth factors may be outcomes that are predicted from other variables (e.g., controlling for pretest or other covariates) or may be predictors of other latent or observed variables (e.g., change in alcohol use may predict academic achievement). Further, LCMs also make it very easy to have multiple indicators of growth at any time point, to analyze multiple growth processes in a single analysis (e.g., growth in alcohol use and growth in violence, and to model the relationships among the different growth processes), to include multiple groups with potentially different covariance matrices, etc. A typical conditional growth curve model is presented in Figure 1, where the intercept and slope are expected to be predicted by the two-level grouping variable.

Two recent papers present novel LCMs for comparing changes in groups over time in a

RPPF design. An important characteristic of these models is that they incorporate a latent

difference score approach (McArdle & Hamagami, 2001), where change from pretest to posttest

to follow-up is estimated using multiple difference score estimates of the amount of change

across each of the time points (for an excellent summary of latent difference score models, and

related models, see McArdle, 2009). This approach is different from a latent growth model, in

which the measurement of change focuses on the average amount of change across all time

points. The latent growth model approach is typically not appropriate for RPPF designs because

researchers are interested in isolating the specific changes that occur across each of the time

points (Mun, von Eye, & White, 2009). The first paper we highlight is from Willoughby,

Vandergrift, Blair, and Granger (2007), who modify the typical LCM to accommodate RPPF

designs. A conditional version of the Willoughby et al. model is presented in Figure 2, where the

grouping variable is hypothesized to relate to the intercept and change variables. In their model,

two change (slope) factors are used to represent the amount of change from pretest to posttest

(Slope 1) and from pretest to follow-up (Slope 2). However, there are three limitations of the

Willoughby et al. model for a RPPF design. First, there is no parameter to represent the posttest

to follow-up change, and therefore no way to compare the groups on the amount of change from

posttest to follow-up. Second, when estimating the amount of change from pretest to posttest, it

is more appropriate (i.e., more powerful) to covary for the pretest score because participants are

randomly assigned to groups. Third, the model is just-identified so no goodness of fit statistics

are available.

A recent paper by Mun et al. (2009) addresses the first and third limitations of the

Willoughby et al. model; specifically, they propose multiple over-identified models in which the

reference point changes in each model. Conditional versions of two of the models proposed by Mun et al. are presented in Figure 3. Specifically, the first model (Model a) enables an evaluation of group differences on the change from pretest to posttest (i.e., the 'Group' to 'Slope 1' path), and posttest to follow-up (i.e., the 'Group' to 'Slope 2' path). The second model (Model b) enables an evaluation of group differences from pretest to follow-up (i.e., the 'Group' to 'Slope 1' path) and from posttest to follow-up (i.e., the 'Group' to 'Slope 2' path). Further, the regression of the intercept factor on the grouping variable is a test of group differences at posttest (Model a) or follow-up (Model b). However, the models proposed by Mun et al. have one important limitation with respect to analyzing change in RPPF designs; namely that they use a difference score based approach to controlling for the pretest. In other words, a more powerful test of whether the groups differ on the amount of change from pretest to posttest would be obtained if the model covaried for the pretest score since the participants are randomly assigned to groups.

**Proposed Latent Change Model**

In order to address the limitations of the Willoughby et al. (2007) and Mun et al. (2009) models, we derived a model that covaries for the pretest to achieve greater power for detecting changes from pretest to posttest and posttest to follow-up. This model expands on the two time-point 'change-regression' model (McArdle, 2009). It is important to recall that it is not appropriate to covary posttest scores when evaluating group differences in change from posttest to follow-up because posttest group differences (other than just chance variation) are expected due to the impact of the intervention. The proposed LCM is presented in Figure 4. In this model, the differences at pretest are covaried from both the posttest and follow-up variables, rendering

the interpretation of these variables as the amount of change from pretest to posttest (posttest variable) and from pretest to follow-up (follow-up). The change score based latent portion of the model is then used to estimate the amount of change from posttest to follow-up ('Slope 2'). The regression of 'Slope 1' on 'Group' can be interpreted as a test of group differences in change from pretest and posttest, and the regression of 'Slope 2' on 'Group' can be interpreted as a test of group differences in change from posttest to follow-up.

## Monte Carlo Study

A simulation study was used to evaluate the proposed model and compare the Type I error control and power of the proposed model to that of the Mun et al. models for evaluating group differences in change from pretest to posttest and posttest to follow-up. Two sample sizes were used, $N = 100$ and $N = 500$. Subjects ($i = 1, ... , N$) were randomly assigned to groups ($g = 1, 2$). Pretest scores were a function of $X$ (a true score on the outcome construct of interest; $\mu = 0$, $\sigma = 1$) and error, $\varepsilon_{pre}$ ($\mu = 0$, $\sigma = 1$):

$$\text{Pretest}_i = X_i + \varepsilon_{pre, i}.$$

Two variables were then used to represent the amount of change from pretest to posttest ($C_{pp}$), and from posttest to follow-up ($C_{pf}$), each measured with error ($\mu = 0$, $\sigma = 1$):

$$C_{pp,i} = C_{pp,g} + \varepsilon_{pp,i}$$

$$C_{pf,i} = C_{pf,g} + \varepsilon_{pf, i}.$$

Note that $C_{pp,g}$ and $C_{pf,g}$ are constants that represent the amount of change from pretest to posttest, and posttest to follow-up, respectively, for the group that the $i$th participant was randomly assigned to. Posttest scores were then a function of $X$, $C_{pp}$, and error, $\varepsilon_{post}$ ($\mu = 0$, $\sigma = 1$):

$$\text{Posttest}_i = X_i + C_{pp,i} + \varepsilon_{\text{post},\,i}\,,$$

and follow-up scores were a function of $X$, $C_{pp}$, $C_{pf}$, and error, $\varepsilon_f$ ($\mu = 0$, $\sigma = 1$):

$$\text{Follow-up}_i = X_i + C_{pp,i} + C_{pf,i} + \varepsilon_{f,\,i}\,.$$

Two specific research questions were addressed for each of the Mun et al. and the proposed models: 1) Do the groups differ on the amount of change from pretest to posttest? and 2) Do the groups differ on the amount of change from posttest to follow-up? For the Mun et al. procedure, Model a was used to evaluate these research questions. To evaluate whether the groups differed on the amount of change from pretest to posttest, the coefficient from regressing 'Slope 1' on 'Group' was evaluated, and to evaluate whether the groups differed on the amount of change from posttest to follow-up, the coefficient from regressing 'Slope 2' on 'Group' was evaluated. In the proposed model, the coefficient from regressing 'Slope 1' on 'Group' evaluated the question of whether the groups differed on the amount of change from pretest to posttest, and the coefficient from regressing 'Slope 2' on 'Group' evaluated the question of whether the groups differed on the amount of change from posttest to follow-up.

To evaluate Type I error rates for each model for the question of whether the groups differ on the amount of change from pretest to posttest, we set $C_{pp,1} = C_{pp,2} = 0$, and to evaluate power, we set $C_{pp,1} = 0$ and $C_{pp,2} = .3$. To evaluate Type I error rates for each model for the question of whether the groups differ on the amount of change from posttest to follow-up, we set $C_{pf,1} = C_{pf,2} = 0$, and to evaluate power, we set $C_{pf,1} = 0$ and $C_{pf,2} = .3$. Combinations of these conditions resulted in three different scenarios:

Scenario 1: No difference between the groups in the amount of change from pretest to posttest or from posttest to follow-up ($C_{pp,1} = C_{pp,2} = C_{pf,1} = C_{pf,2} = 0$)

Scenario 2: A difference between the groups in the amount of change from pretest to posttest, but no difference between the groups in the amount of change from posttest to follow-up ($C_{pp,1} = 0$; $C_{pp,2} = .3$; $C_{pf,1} = C_{pf,2} = 0$)

Scenario 3: A difference between the groups in the amount of change from pretest to posttest and from posttest to follow-up ($C_{pp,1} = 0$; $C_{pp,2} = .3$; $C_{pf,1} = 0$; $C_{pf,2} = .3$).

Five thousand simulations were conducted for each condition with a nominal alpha level of .05. The simulations were conducted in R (R Development Core Team, 2009), using the 'sem' package (Fox & Cramer, 2009) with maximum likelihood estimation.

## Results

### Type I Error Control and Power

Type I error control and power results for the Mun et al. LCM and the proposed LCM are presented in Table 1.

**Scenario 1 ($C_{pp,1} = C_{pp,2} = C_{pf,1} = C_{pf,2} = 0$).** For this situation, Type I error control is evaluated for both research questions because there is no change in either group across any of the time points. For the Mun et al. model, Type I error rates for evaluating whether the groups differ on the amount of change from pretest to posttest were slightly elevated (.061 for $N = 100$ and .058 for $N = 500$), while Type I error rates for evaluating whether the groups differ on the amount of change from posttest to follow-up were deflated (.026 for $N = 100$ and .025 for N = 500). For the proposed model, Type I error rates for evaluating whether the groups differ on change from pretest to posttest, and from posttest to follow-up, for both sample sizes, were very accurate, ranging from .048 to .055.

**Scenario 2 ($C_{pp,1} = 0$, $C_{pp,2} = .3$; $C_{pf,1} = C_{pf,2} = 0$).** Evaluating the difference between the groups in change from pretest to posttest is a power condition because the scores for group two change while the scores for group one do not change, while evaluating the difference between the groups in change from posttest to follow-up is a Type I error condition because there is no change for either groups across this time period. With respect to power for detecting the different amounts of change across the groups from pretest to posttest, the proposed model had greater power than the Mun et al. model, with a larger difference (about 5%) observed in the $N = 500$ condition. With respect to Type I error control, as in Scenario 1, the Type I error rates for the Mun et al. model were slightly deflated, while the Type I error rates for the proposed model were accurate.

**Scenario 3 ($C_{pp,1} = 0$, $C_{pp,2} = .3$; $C_{pf,1} = 0$, $C_{pf,2} = .3$).** The amount of change from pretest to posttest, and the amount of change from posttest to follow-up, differ across groups and therefore both conditions are investigating power. Results for this scenario are also in Table 1. For both research questions there is greater power for the proposed model than for the Mun et al. model, with a greater difference observed with the larger sample size. For example, with $N = 500$, there is approximately 76% power for detecting the difference in change between the two groups from pretest to posttest with the proposed model, and about 72% for the Mun et al. model, whereas there is approximately 68% power for detecting the difference in change between the two groups from posttest to follow-up with the proposed model, and about 57% for the Mun et al. model.

## Discussion

Researchers interested in evaluating the effectiveness of an intervention will often utilize

a RPPF design. Although there are numerous strategies for analyzing the data from RPPF designs, LCMs have become extremely popular and allow researchers the opportunity to model change across all three time points simultaneously while also including predictors or outcomes of change. This study discussed previously proposed models for comparing the amount of change across groups in RPPF designs and compared the Type I error control and power of a novel LCM for investigating group differences in change across pretest, posttest, and follow-up to models recently proposed by Mun et al. (2009). The proposed model uses a regression-based approach for controlling for group differences in change at pretest (i.e., covarying for the pretest), whereas the Mun et al. model uses a latent difference score based approach. The regression-based strategy is more appropriate than the change score approach when the subjects are randomly assigned to groups and therefore any differences between the groups can be considered error.

The results of the current study indicate that the proposed model provides more precise Type I error control and higher power than the models proposed by Mun et al. for evaluating group differences in change from pretest to posttest and posttest to follow-up. The proposed model can also easily be extended to designs with more than two treatment groups by utilizing a multiple group structural equation model, to designs with more than one change process (e.g., measuring change in both depression and anxiety), or to models with multiple predictors or outcomes of change. Recent research on Bayesian models may also present researchers with the opportunity to incorporate prior information into the modeling process and to allow researchers to fit these types of models with smaller samples than what would be required with a maximum likelihood solution (Zhang, Hamagami, Wang, Nesselroade, & Grimm, 2007).

To summarize, researchers utilizing RPPF designs are often interested in two specific

research questions, whether the treatment and control groups differ in change from pretest to posttest and whether the treatment and control groups differ in change from posttest to follow-up. The proposed latent change model provides a very simple and effective strategy for evaluating these research questions.

# References

Bollen, K. A., & Curran, P. J. (2006). Latent curve models: A structural equation perspective. Hoboken, NJ: Wiley.

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

Campbell, D. T. & Kenny, D. A. (1999). *A primer on regression artifacts.* New York: Guilford.

Cribbie, R. A., Jamieson, J. & Mills, L. (in press). *Evaluating change over two time points*. In J. Jaccard & W. Silverman (eds.) Analytic methods for clinical research: A primer of concepts and techniques. New York: Lawrence Erlbaum Associates.

Cronbach, L. J., & Furby, L. (1970). How should we measure "change" - or should we? *Psychological Bulletin, 74*, 68-80.

Curran, P.J. (2000).  A latent curve framework for the study of developmental trajectories in adolescent substance abuse. In Rose, J.S., Chassin, L., Presson, C.C., & Sherman, S.J. (Eds.), *Multivariate Applications in Substance Use Research: New Methods for New Questions.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Curran, P.J. & Hussong, A.M. (2002). Structural equation modeling of repeated measures data: Latent curve analysis. In Moskovitz, D.R. & Hershberger, S.L. (Eds.), *Modeling Intraindividual Variablity with Repeated Measures Data: Methods and Applications,* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Duncan, T. E., Duncan, S. C., Strycker, L. A., Li, F., & Alpert, A. (2006). *An introduction to latent variable growth curve modeling: Concepts, issues, and applications.* Second edition. Mahwah, NJ: Lawrence Erlbaum.

Elliott, M. R., Gallo, J. J., Ten Have, T. R., Bogner, H. R., & Katz, I. R. (2005). Using a

    Bayesian latent growth curve model to identify trajectories of positive affect and negative

    events following myocardial infarction. Biostatistics, 6, 119-143.

Fox, J. & Kramer, A. (2009). sem: Structural Equation Models. R package version 0.9-17. URL:

    http://CRAN.R-project.org/package=sem

Huck, S. W., & McLean, R. A. (1975). Using a repeated measures ANOVA to analyze the data

    from a pretest–posttest design: A potentially confusing task. *Psychological Bulletin, 82,*

    511–518.

Huitema, B. E., (1980). *The analysis of covariance and alternatives.* New York: Wiley.

Lord, F.M. (1967). A paradox in the interpretation of group comparisons. *Psychological Bulletin,*

    *68,* 304-305.

Matschinger, H., Kilian, R. & Angermeyer, M. (2006). Detecting specific patterns of

    change for two outcome scores in a mental health study by means of linear mixture

    models. *Methodology, 2,* 113-123.

McArdle, J. J., & Hamagami, F. (2001). Latent difference score structural models for linear

    dynamic analyses with incomplete longitudinal data. In L. M. Collins & A. G. Sayer

    (Eds.), *New methods for the analysis of change* (pp. 139–175). Washington, DC:

    American Psychological Association.

McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal

    data. *Annual Review of Psychology, 60*, 577-605.

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika, 55*, 107-122.

Mun, E.Y., von Eye, A., & White, H.R. (2009). An SEM approach for the evaluation of

intervention effects using pre-post-post designs. *Structural Equation Modeling: A Multidisciplinary Journal, 16*, 315- 337.

R Development Core Team (2009). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org.

Rausch, J.R., Maxwell, S.E., & Kelley, K. (2003). Analytic methods for questions pertaining to a randomized pretest, posttest, follow-up design. *Journal of Clinical Child & Adolescent Psychology, 32,* 467 – 486.

Rogosa, D. R. (1995). Myths and methods: "Myths about longitudinal research" plus supplemental questions. In J. M. Gottman (Ed.), *The analysis of change*. Mahwah, NJ, Lawrence Erlbaum Associates, Inc.

Rogosa, D. R., & Willett, J. B. (1983). Demonstrating the reliability of the difference score in the measurement of change. *Journal of Educational Measurement*, *20*, 335-343.

Rosel, J., & Plewis, I. (2008). Longitudinal data analysis with structural equations. *Methodology, 4*, 37–50.

Schafer, W. D. (1992). Analysis of pretest-posttest designs. *Measurement and Evaluation in Counseling and Development, 25*, 2-4.

Thurstone, L. L. (1930). The learning function. *Journal of General Psychology, 3,* 469–493.

Tomarken, A. J. & Waller, N. G. (2005). Structural equation modeling: Strengths, limitations, and misconceptions. *Annual Review of Clinical Psychology, 1,* 31-65.

Willett, J.B. (1989). Some results on reliability for the longitudinal measurement of change: Implications for the design of studies of individual growth. *Educational and Psychological Measurement, 49,* 587-602.

Williams, R.H., & Zimmerman, D.W. (1977). The reliability of difference scores when errors are correlated. *Educational and Psychological Measurement, 37,* 679-689.

Willoughby, M., Vandergrift, N., Blair, C., & Granger, D.A. (2007). A structural equation modeling approach for the analysis of cortisol data collected using pre-post-post designs. *Structural Equation Modeling: A Multidisciplinary Journal, 14*, 125 – 145.

Wright, D. B. (2006). Comparing groups in a before-after design: When *t* test and ANCOVA produce different results. *British Journal of Educational Psychology, 76,* 663-675.

Zhang, Z., Hamagami, F., Wang, L., Nesselroade, J., & Grimm, K. J. (2007). Bayesian analysis of longitudinal data using growth curve models. *International Journal of Behavioral Development, 31*, 374–383.

Table 1.
*Type I error and Power Results for Investigating Group Differences in the Amount of Change from Pretest to Posttest and Posttest to Follow-up.*

| Scenario | Pretest to Posttest | | Posttest to Follow-up | |
|---|---|---|---|---|
| | Mun et al. | Proposed | Mun et al. | Proposed |
| **N = 100** | | | | |
| $C_{pp,1} = C_{pp,2} = 0$; $C_{pf,1} = C_{pf,2} = 0$ | .061 | .051 | .026 | .055 |
| $C_{pp,1} = 0, C_{pp,2} = .3$; $C_{pf,1} = C_{pf,2} = 0$ | .206[p] | .226[p] | .023 | .052 |
| $C_{pp,1} = 0, C_{pp,2} = .3$; $C_{pf,1} = 0, C_{pf,2} = .3$ | .214[p] | .226[p] | .132[p] | .205[p] |
| **N = 500** | | | | |
| $C_{pp,1} = C_{pp,2} = 0$; $C_{pf,1} = C_{pf,2} = 0$ | .058 | .048 | .025 | .050 |
| $C_{pp,1} = 0, C_{pp,2} = .3$; $C_{pf,1} = C_{pf,2} = 0$ | .711[p] | .765[p] | .024 | .051 |
| $C_{pp,1} = 0, C_{pp,2} = .3$; $C_{pf,1} = 0, C_{pf,2} = .3$ | .719[p] | .763[p] | .568[p] | .684[p] |

Note: $C_{pp,1}$ = amount of change from pretest to posttest for the first group; $C_{pp,2}$ = amount of change from pretest to posttest for the second group; $C_{pf,1}$ = amount of change from posttest to follow-up for the first group; $C_{pf,2}$ = amount of change from posttest to follow-up for the second group; p = power condition.
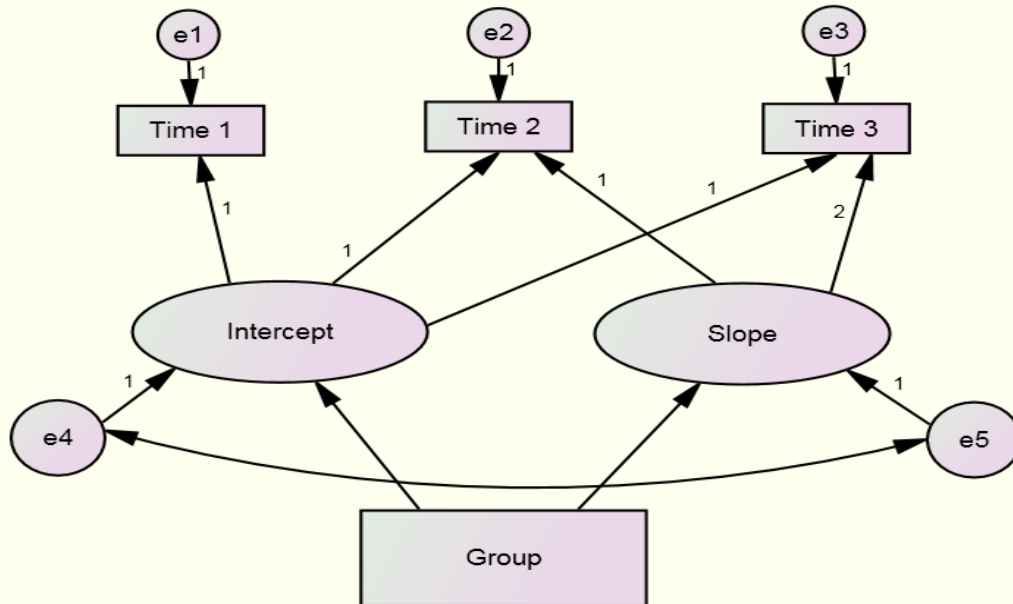
*Figure 1*. Conditional latent change model with three time points and a dichotomous grouping variable.
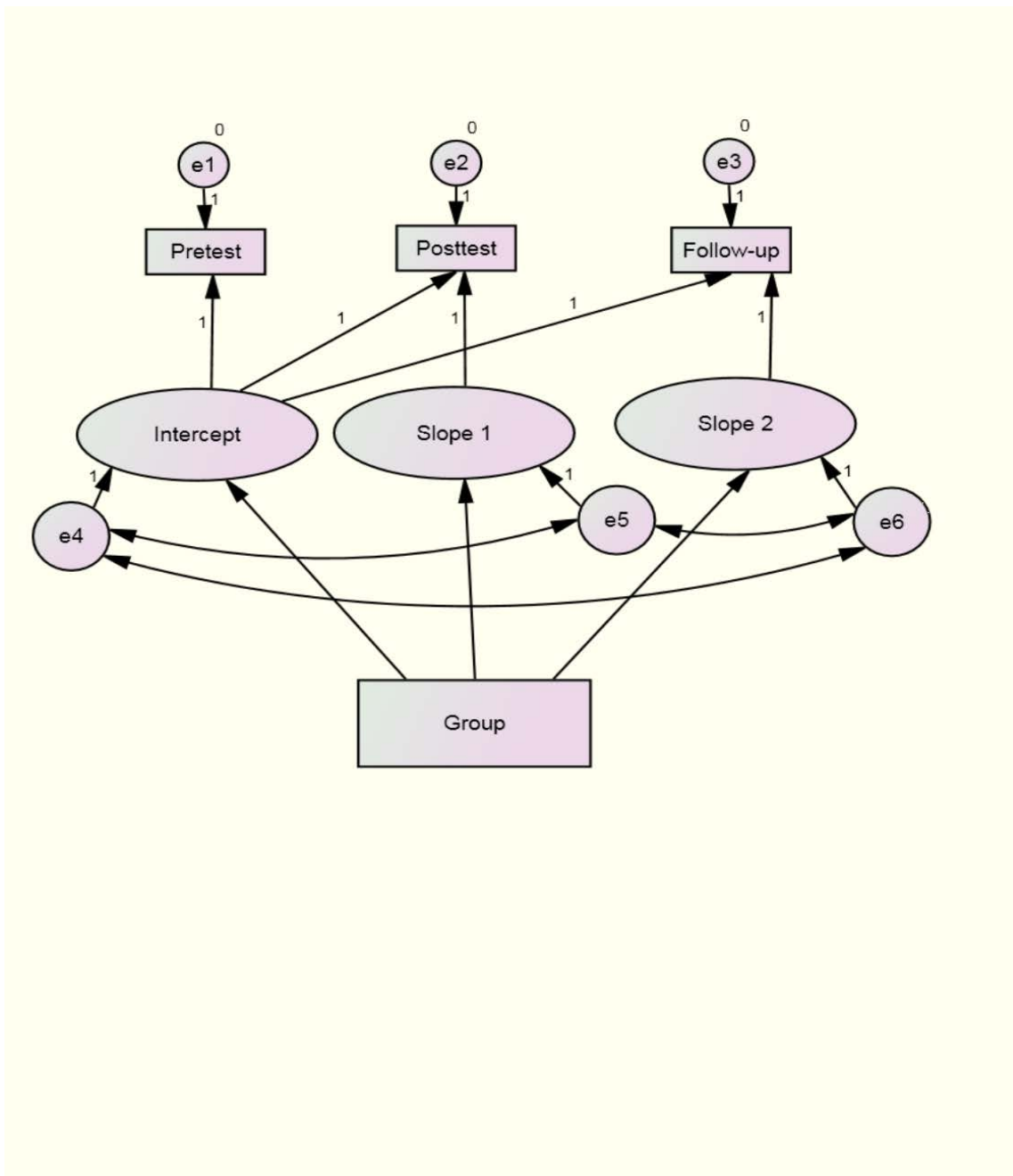
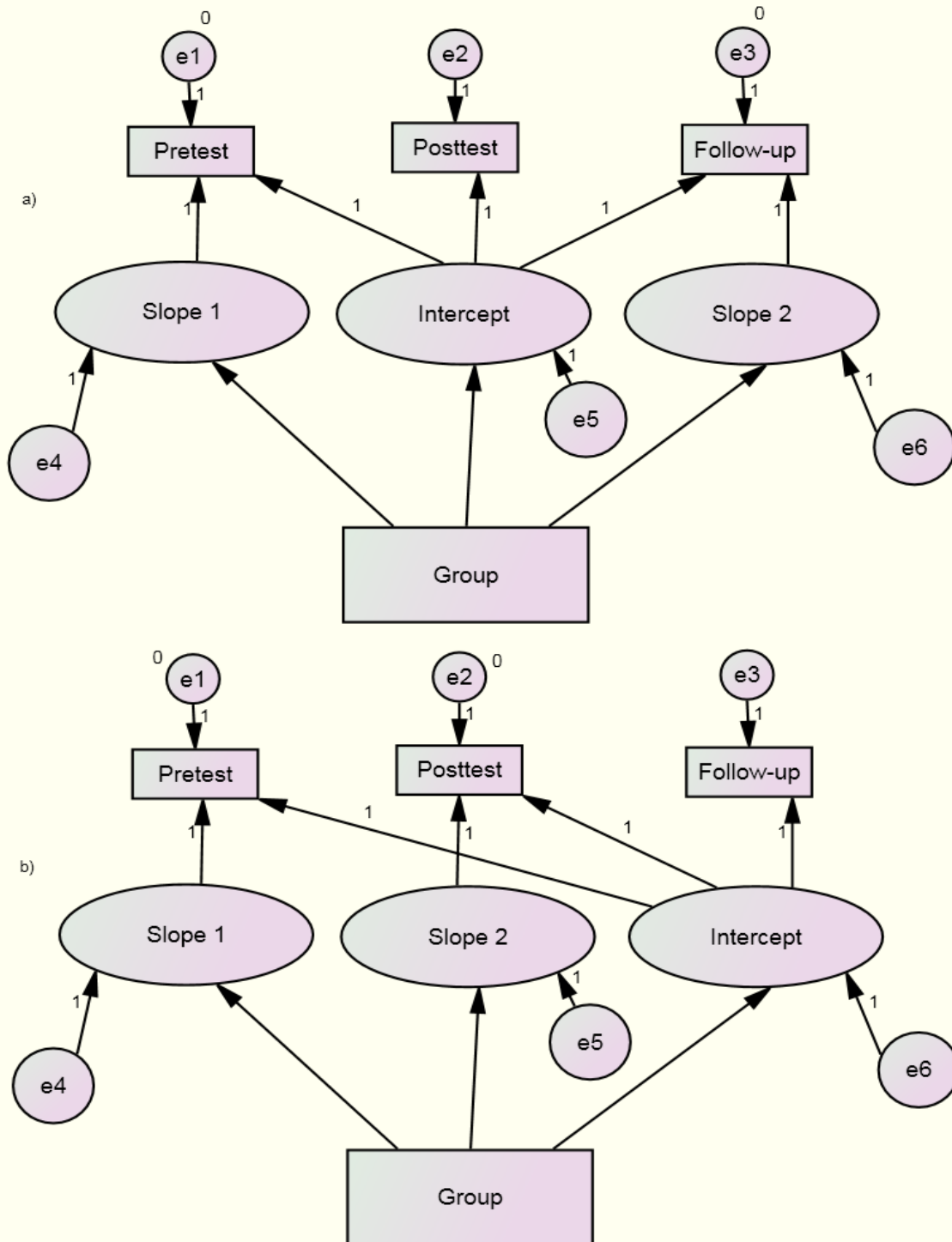*Figure 2*. Conditional latent change model due to Willoughby et al. (2007) for a RPPF design.

*Figure 3*. Conditional latent change models due to Mun et al. (2009) for a RPPF design. (a)

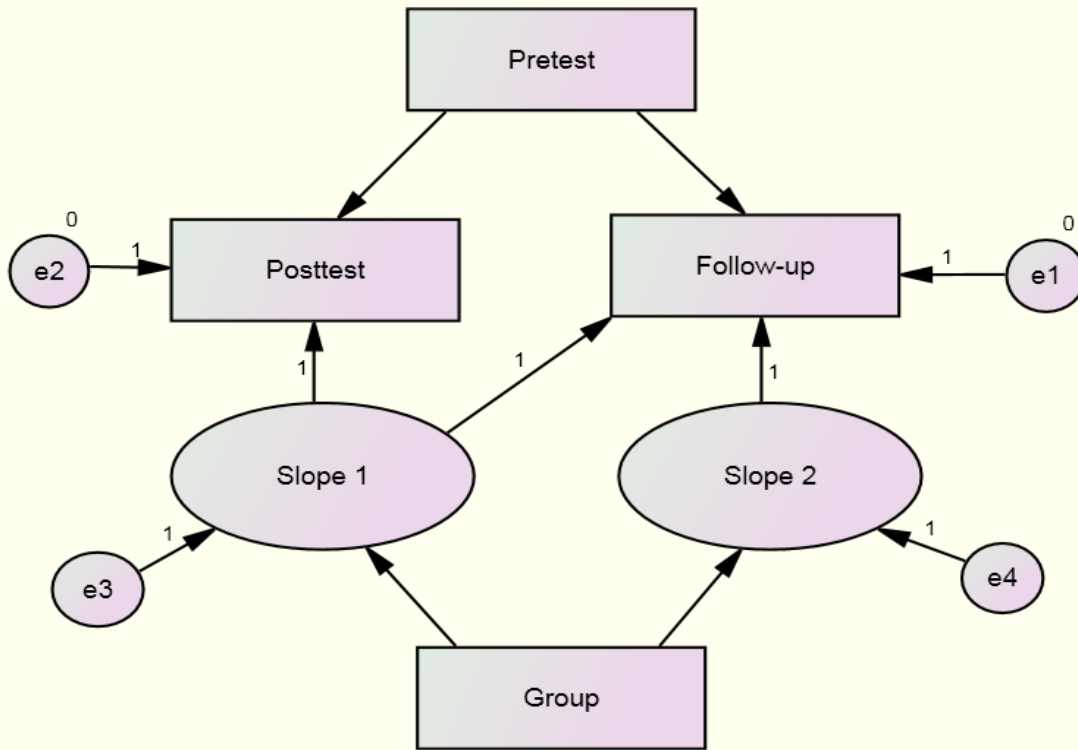Posttest as the referent variable; (b) Follow-up as the referent variable.

*Figure 4*. Proposed conditional latent change model for a RPPF design.