



Published in final edited form as:

*Rehabil Psychol.* 2008 August ; 53(3): 370–386. doi:10.1037/a0012765.

## Analyzing Longitudinal Data with Multilevel Models: An Example with Individuals Living with Lower Extremity Intra-articular Fractures

**Oi-Man Kwok,**

Department of Educational Psychology at Texas A&M University

**Andrea T. Underhill,**

Injury Control Research Center at the University of Alabama at Birmingham

**Jack W. Berry,**

Injury Control Research Center at the University of Alabama at Birmingham

**Wen Luo,**

Department of Educational Psychology at University of Wisconsin-Milwaukee

**Timothy R. Elliott, and**

Department of Educational Psychology at Texas A&M University

**Myeongsun Yoon**

Department of Educational Psychology at Texas A&M University

### Abstract

The use and quality of longitudinal research designs has increased over the past two decades, and new approaches for analyzing longitudinal data, including multi-level modeling (MLM) and latent growth modeling (LGM), have been developed. The purpose of this paper is to demonstrate the use of MLM and its advantages in analyzing longitudinal data. Data from a sample of individuals with intra-articular fractures of the lower extremity from the University of Alabama at Birmingham's Injury Control Research Center is analyzed using both SAS PROC MIXED and SPSS MIXED. We start our presentation with a discussion of data preparation for MLM analyses. We then provide example analyses of different growth models, including a simple linear growth model and a model with a time-invariant covariate, with interpretation for all the parameters in the models. More complicated growth models with different between- and within-individual covariance structures and nonlinear models are discussed. Finally, information related to MLM analysis such as online resources is provided at the end of the paper.

### Keywords

multilevel model; growth model; trajectory analysis; hierarchical linear model; rehabilitation

---

Correspondence concerning this article should be addressed to Oi-Man Kwok, Department of Educational Psychology, 4225 TAMU, College Station, TX 77843-4225, or via email: omkwok@neo.tamu.edu.

**Publisher's Disclaimer:** The following manuscript is the final accepted manuscript. It has not been subjected to the final copyediting, fact-checking, and proofreading required for formal publication. It is not the definitive, publisher-authenticated version. The American Psychological Association and its Council of Editors disclaim any responsibility or liabilities for errors or omissions of this manuscript version, any version derived from this manuscript by NIH, or other third parties. The published version is available at <http://www.apa.org/journals/rep>

Longitudinal designs have recently received more attention in a variety of different disciplines of psychology including clinical, developmental, personality and health psychology (West, Biesanz, & Kwok, 2003). In some areas, such as developmental psychology and personality psychology, a substantial number of recently published studies have been longitudinal (Biesanz, West, & Kwok, 2003; Khoo, West, Wu, & Kwok, 2006). For example, Khoo et al. (2006) found that slightly more than one third of articles published in *Developmental Psychology* in 2002 included at least one longitudinal study, defined as having at least two measurement occasions. This proportion is double the proportion of longitudinal studies published in the same journal in 1990. Furthermore, more than 70% of the longitudinal studies published in *Developmental Psychology* in 2002 included three or more measurement waves. In this paper, we will focus on the analyses of multiwave longitudinal data, where multiwave is defined as more than two waves.

With the growing use of longitudinal research, a number of methodological and statistical sources on the analysis of multiwave longitudinal data have appeared in the past decade (e.g., Bollen & Curran, 2006; Collins & Sayer, 2001; Singer & Willet, 2003), including discussions of traditional approaches such as repeated-measures Univariate Analysis of Variance (UANOVA) and Multivariate Analysis of Variance (MANOVA). Multi-Level Models (MLM), also known as Hierarchical Linear Models (HLM, Raudenbush & Bryk, 2002), random coefficient models (Longford, 1993), and mixed-effect models (Littell, et al., 2006), have become an increasingly important approach for analyzing multiwave longitudinal data.

Although multilevel models have been widely adopted in educational research for more than two decades (Raudenbush, 1988), these models are still relatively new to researchers in rehabilitation psychology. Growth models first appeared in the rehabilitation psychology literature over a decade ago: Clay et al. (1995) reported a compelling (yet circumscribed) demonstration of using growth modeling of the development of emotional distress and behavioral problems of children with juvenile arthritis and others with juvenile diabetes, and of children without diagnosed health problems. An expanded report from this database appeared three years later in the *Journal of Consulting and Clinical Psychology* (Frank, et al., 1998). Subsequent applications of hierarchical linear modeling (HLM) examined the dynamic trajectory of adjustment among family members over the first year of caregiving for a person with a spinal cord injury (Shewchuk, Elliott, & Richards, 1998), and later identifying salient predictors of these trajectories (Elliott, Shewchuk, & Richards, 2001). Another study used HLM to study the growth curve trajectory of functional abilities of persons receiving inpatient spinal cord injury rehabilitation (Warschausky, Kay, & Kewman, 2001). A study of problem-solving training with caregivers of stroke survivors was among the first to use HLM in a randomized clinical trial (Grant, et al., 2002), which heightened expectations about the utility of MLM in intervention research. The initial enthusiasm for MLM was expressed by one author who opined that these techniques could have "...an immense impact on program development and evaluation" and help the field identify "...who responds best to what and why, who is at risk, and who responds optimally regardless of treatment options" (Elliott, 2002, p. 138).

Unfortunately, very few studies utilizing MLM have recently appeared in the rehabilitation psychology literature, implying that the field has yet to realize the potential and possibilities of these approaches. This is particularly unfortunate in light of the informative and extensive longitudinal databases that have been collected over the years to help us understand certain high-cost, high-impact disabilities (e.g., spinal cord injury, traumatic brain injuries in the Model Systems projects).

Therefore, the present paper is intended to be an elementary introduction to multilevel modeling for researchers in rehabilitation psychology who have interest in and access to longitudinal data that could be analyzed with these techniques.

## Comparison of MLM with Repeated-Measures Univariate ANOVA (UANOVA)

Most researchers in rehabilitation psychology will be familiar with the use of repeated-measures univariate analysis of variance (UANOVA) for analyzing multiwave data. In this section, therefore, we will provide an overview of some of the similarities and differences between repeated-measures ANOVA and MLM. We provide a summary of our comparison in Table 1. In the simplest multiwave design, a purely within-subjects design, the researcher will have collected repeated measurements on the same sample of research participants over time, and the primary question is whether there is within-subjects change in the sample on a particular outcome variable. With repeated-measures UANOVA, time is thought of as a *categorical* factor, and the researcher can conduct a variety of contrasts to compare differences among time periods. For example, each post-baseline time period can be compared to baseline, or all adjacent time periods can be compared. An alternative strategy is polynomial trend analysis. In ANOVA trend analysis, the question is whether there is an average linear, quadratic, or higher-order polynomial trend in mean levels of the outcome variable over time.

The repeated-measures ANOVA-based analyses can be viewed as special cases of multi-level models (Kwok, West, & Green, 2007). Hence, MLM can employ these same analytic strategies for simple within-subjects designs, but, as we will describe in more detail below, MLM can provide several advantages over ANOVA in terms of handling missing data and flexible modeling of variance-covariance structures. MLM also offers a unique data analytic strategy for within-subjects designs that is not possible using in UANOVA. Namely, MLM can be used to model *individual-level trends over time*, in which polynomial trends can be estimated for each participant (rather than simply average trends). This approach is referred to as *individual growth models*. In UANOVA, individual growth models are not estimated; rather, an average growth model is estimated in a single analysis of all participants, and individual variation around the average model is treated as “unexplained” error. In MLM, regression parameters from all the individual growth models including intercepts, slopes, or both can be treated as random effects for estimation. Advantages of modeling these random effects have been reviewed in Kwok et al (2007).

There are different estimation methods for UANOVA and MLM. In repeated-measures UANOVA, least squares (LS) estimation is generally used, whereas in MLM, maximum likelihood (ML) is one of the commonly used estimation methods. A detailed discussion of LS and ML methods is beyond the scope of this paper. One additional note on ML is given because there are two commonly used ML methods<sup>1</sup> available in several major MLM programs (e.g., HLM, SPSS, SAS, and STATA) when the outcome variable is continuous and normally distributed. The default estimation method in both SPSS (MIXED) and SAS (PROC MIXED) for analyzing multilevel data with continuous outcomes is the restricted maximum likelihood (REML). The alternative estimation is the full information maximum likelihood (FIML). REML can provide more accurate results when sample size (especially the number of higher level units) is small (Hox, 2002;Raudenbush & Bryk, 2002). On the other hand, FIML can compare the goodness of fit for both fixed and random parts between nested models using likelihood ratio tests, whereas REML can only compare the goodness of fit for the random part between nested models. The results presented in this paper are based on the default REML estimation method in both SAS and SPSS. More detailed information on estimation can be found in the text by Raudenbush and Bryk (2002).

<sup>1</sup>The Bayesian methods (e.g., the empirical Bayesian and fully Bayesian estimators) are also widely used estimation method in MLM. The discussion of these methods is beyond the focus of this paper. More information on these Bayesian estimation methods can be found in Raudenbush and Bryk's (2002) text.

Another major difference between UANOVA and MLM is in the treatment of the time predictor. In individual growth models, time can be treated as a continuous variable in MLM. Because of this, MLM can accommodate unequal spacing between time intervals and unbalanced data. Observations may be collected at unequally spaced intervals (e.g., measurements collected 0 months, 3 months, 6 months, 1 year, 5 years following treatment). Observations may also be collected at different time points for different participants (e.g., for the first participant 0, 3, 6, 9 months following treatment; for the second participant 1, 5, 10, 12 months following treatment<sup>2</sup>). Such patterns of observations may occur because of practical problems in implementing the original data collection design. Unbalanced data and unequal spacing conditions can be flexibly handled under MLM through adequate specification of the time predictor. On the other hand, all participants are assumed to have the same number of assessments (balanced data) and the intervals between time periods are assumed to be equal (equal spacing) when using UANOVA.

In the same vein, missing data can be handled flexibly in MLM but not in UANOVA. Missing data can arise for many reasons in multiwave longitudinal research: missed appointments, participant incapacity, dropout, or loss to follow-up for a variety of reasons. However, only complete cases can be included in an analysis when using UANOVA. If a research participant is missing for even a single time period, all of the participant's data are removed from the analysis. The capacity of MLM (using likelihood-based estimation) to incorporate all available data in an analysis can be especially useful in conducting intention-to-treat (ITT) analyses in controlled clinical trials. Recent reviews of ITT studies have suggested that inappropriate handling of missing data is the chief problem with published reports of ITT-based clinical trials (Gravel, Opatrny, & Shapiro, 2007; Hollis & Campbell, 1999). The requirement of complete data in UANOVA can lead to substantial losses of statistical power and precision in longitudinal research. We discuss the possibility of imputing missing data values in a later section of this paper.

An advantage of MLM is that it can make use of all available data in the estimation of model parameters due to its flexible treatment of the time predictor. A research participant with only baseline data can be included in an analysis and contribute to the estimation of model parameters. The validity of using all available data does depend on whether missing data are missing completely at random (MCAR) (or missing at random (MAR) which is a less restrictive missing data assumption), and methods of assessing this requirement are available. Additionally, the treatment of time as a continuous instead of discrete variable in MLM can increase the statistical power for detecting the growth effects (Muthén & Curran, 1997). For these reasons, MLM is a preferred option for ITT analyses in clinical trials and other intervention studies, particularly when the theoretical model anticipates a gradual response to the intervention over time (as typified in most psychological theories of therapeutic response).

UANOVA and MLM also differ on the statistical assumptions related to the variance-covariance structure when analyzing longitudinal data. In repeated-measures UANOVA, the variance-covariance matrix of observations taken over time is assumed to meet the requirements of sphericity in which compound symmetry is a sufficient condition for fulfilling the sphericity assumption. Compound symmetry implies that the variances of measures at each time period are equal, and also that the covariances between all pairs of time periods are equal. This is a strong assumption, and is likely to be unrealistic for many (if not most) longitudinal studies (Kwok et al., 2007). For example, in some studies, research participants might show greater variability in an outcome measure over time, whereas in others there might be growing convergence over time. It is also possible that covariances between variables over time will be

---

<sup>2</sup>One alternative way to handle the unequal space issue in UANOVA is to create a covariate capturing the unequal space and include the covariate in the analysis (i.e., univariate analysis of covariance or UANCOVA).

smaller with greater distances between time intervals. Violations of the assumption of sphericity can lead to incorrect decisions in ANOVA-based analyses. In MLM, there is great flexibility in specifying the variance-covariance structure of longitudinal data (Chi & Reinsel, 1988; Diggle, 1988; Laird & Ware, 1982; Jones & Boadi-Boateng, 1991; Wolfinger, 1993). The most flexible option in MLM analysis is a general “unstructured” variance-covariance assumption in which every variance and covariance is free to be estimated from the data. Other, more restrictive assumptions include autocorrelated structures, in which covariances are a function of distances between time periods, and variances can be modeled as either homogeneous or variable.

One more difference between repeated-measures UANOVA and MLM concerns the use of covariates in statistical analyses. Covariates can be used in within-subjects research for many reasons: to reduce error variance, to statistically “equalize” participants on some variable of interest, or to find mediators of the relationship between time and an outcome variable. In repeated-measures UANOVA, all covariates in a model must be time-invariant. In other words, individual measures on the covariates do not change with time and therefore have a constant effect across all measurement occasions. Examples of such time-invariant covariates would be the age of a participant at the beginning of a study or standing on stable trait variables. Using MLM, in contrast, the researcher can include time-varying covariates in an analysis. Time-varying covariates are often assessed concurrently with major outcome variables and can change over time for each participant. The inclusion of time-varying covariates can provide a much more sensitive and realistic assessment of covariate effects for unstable, state-like variables that might influence primary outcome variables.

We have thus far described potential advantages of using MLM in the analysis of strictly within-subjects research designs. However, a common use of multiwave research combines within-subjects repeated measures with one or more between-subjects variables. An example of such a design is the randomized clinical trial with multiple outcome assessments over time. In such clinical trials, treatment condition is a between-subjects factor and time (assessment occasion) is a within-subjects factor. In using ANOVA to assess clinical trial data, the focus is on the statistical significance of the Treatment by Time interaction. If the interaction is significant, simple main effects can be conducted to determine the nature of the differential change between treatment conditions. Differences in average time contrasts and polynomial trends can be employed for this purpose.

The same strategies can be used in MLM, but MLM can also combine the advantages of individual growth-curve analysis with the examination of interactions of treatment with time. Specifically, because MLM separates the random effects into two parts (between-subject random effects and within-subject random errors), MLM allows for the examination of new effects of interest such as *cross-level* interaction effects. For example, researchers can examine how treatment condition (and other between-subject level predictor) influences the individual growth trajectories (within-subjects repeated measures) of research participants over time. Even if individual growth curves are not estimated, the MLM approach to assessing treatment effects over time offers clear advantages over the repeated-measures approach. For example, in a recent randomized controlled trial of problem-solving training for family caregivers of a loved one with a traumatic brain injury (Rivera, Elliott, Berry, & Grant, 2007), caregivers were assessed on a variety of psychological and health-related outcomes over four time periods. In this study, individual growth-curves were not estimated (because of convergence problems due to relatively small sample sizes in the control and treatment groups). Nonetheless, the MLM analysis permitted the inclusion of data from all available caregivers, allowed for the estimation of an unstructured variance-covariance structure, and provided the opportunity to include potentially mediating time-varying covariates in the models (such as problem-solving abilities at each assessment).



The major focus of this paper is to demonstrate how to analyze longitudinal rehabilitation data using MLM. We have already provided a sketch of some of the advantages of MLM for such data. In what follows, we provide a more detailed discussion of both the theoretical underpinnings of MLM and some practical guidance for how to conduct MLM. Our analyses will highlight the use of individual growth models, since these models are uniquely provided by the MLM approach. Data from a sample of individuals with intra-articular fractures of the lower extremity from the University of Alabama at Birmingham's Injury Control Research Center (UAB-ICRC) are analyzed using both SAS PROC MIXED and SPSS MIXED, and the corresponding annotated SAS and SPSS syntaxes for different growth model are presented. For more information on how to analyze longitudinal data using SAS and SPSS, readers can consult Singer's (1998) article on using SAS PROC MIXED, and Peugh and Enders' (2005) article on using SPSS MIXED. We start our presentation with data preparation, followed by the analyses of different growth models, including a simple linear growth model and the model with time-invariant covariate, with interpretation for all the parameters in the models.

## Data Description

A sample of individuals with intra-articular fractures (IAF) of the lower extremity from a large scale, ongoing project, *A Longitudinal Study of Rehabilitation Outcomes*, by the University of Alabama at Birmingham's Injury Control Research Center (UAB-ICRC) will be used for the demonstration. The UAB-ICRC longitudinal study included persons with at least one of four potentially disabling injuries (i.e., spinal cord injury, traumatic brain injury, severe burns, or IAF of the lower extremity) who were discharged from a sample of hospitals representing a cross-section of individuals in north-central Alabama. The criteria for inclusion in the longitudinal study were: (1) acute care length of stay of three or more days; (2) resided and injured in Alabama; (3) discharged alive from an acute care hospital between October 1, 1989 and September 30, 1992; (4) was more than 17 years old when injured; and, (5) can be contacted at pre-specified intervals after discharge. Data have been collected approximately annually from 12 months post-discharge to 180 months post discharge. For the purposes of this manuscript, data collected at 12, 24, 48, and 60 months post-discharge were utilized (no data was collected 36 months post-discharge). The details of the data collection procedure can be found elsewhere (Underhill et al., 2003; Underhill et al., 2004).

There were 251 individuals with IAF who had data for at least one time point. Among these 251 individuals, 131 had complete data for all four time points (i.e., 12, 24, 48, and 60-months post-acute care). The descriptive statistics for these two samples ( $N = 131$  and  $N = 251$ ) are presented in Table 2. We conducted attrition analyses by comparing the two samples on all demographic variables and found no significant differences between the two samples on any of the demographic variables. For pedagogical reasons we start the demonstration with the 131 participants with complete data. We will discuss the analyses with the full dataset (with  $N = 251$ ) in the later section.

We used the physical domain of the Functional Independence Measure (FIM) as the outcome variable for the demonstration. FIM is a widely-used self-report scale of functional status (Keith, Granger, Hamilton, & Sherwin, 1987) which contains 18 seven-point Likert-type items with responses ranging from (1) "total assistance" to (7) "complete independence". The FIM has been shown to have adequate reliability and validity (Putzke et al., 2004). In this study, the reliabilities of the FIM scale at different time points ranged from .91 to .98. The FIM scale can be further divided into two sub-domains, namely, a physical (or motor) domain (13 items) and a cognitive domain (5 items; Greenspan et al., 1996; Hall, Hamilton, Gordon, & Zasler, 1993; Heinemann et al., 1993). The physical domain includes items such as eating, grooming, bathing, dressing, toileting, bladder and bowel management, walking, and transferring to bed/toilet/tub. The cognitive domain contains items such as comprehension, expression, social

interaction, problem solving, and memory. Higher scores on these two domains indicate higher functional ability. In our demonstration, we focus on the change in the physical domain (FIM\_P) of the participants over time. The means and standard deviations of the four FIM\_P scores for the two samples (i.e.,  $N = 131$  and  $N = 251$ ) are presented in Table 3. Based on the information from Table 3, a potential negative trend of the FIM\_P scores and an increase in score variability over time may be found in both samples.

## Data Preparation

Generally, data are in multivariate (sometimes called “wide”) format as shown in Figure 1a; that is, each row represents a participant and each column represents a specific variable. To analyze the data using MLM, we need to transform the multivariate data (Figure 1a) into the univariate (sometimes called “long”) format (Figure 1b) in which each row presents a specific time point rather than a participant. Table 4 presents the corresponding annotated SAS and SPSS syntaxes for converting data from multivariate format to univariate format. In SPSS, menu-driven assistance is also available using “Restructure” under the main Data menu.

In multivariate format, each row represents a participant and each column represents a variable. All time-varying variables (i.e., allowing different variable value at different time point such as the FIM\_P score) and time-invariant variables (i.e., no change in the variable value over time such as gender or age at the first time measure) are presented in the columns of the data. In other words, each time measure is a variable in the multivariate format data (e.g., the four different time measures for FIM\_P are represented by four different variables, namely, FIM\_P12, FIM\_P24, FIM\_P48, and FIM\_P60). On the other hand, in the univariate format data, each row represents a specific time observation and each individual can have multiple rows of observations. For participants with complete data ( $N = 131$ ), each individual has four rows of data lines to represent the four different time measures (i.e., the 12-months, 24-months, 48-months and 60-months post charge measures). A new variable, Time, called an “index” variable, is created as the indicator for each row of time measures. Additionally, researchers should also screen their data to check for potential errors and outliers through the steps recommended by Fidell and Tabachnick (Fidell & Tabachnick 2003; Tabachnick & Fidell, 2006). Plotting individual-level data (e.g., spaghetti plots as shown in Figure 2) before conducting any data analyses is always recommended. These plots can be useful for determining possible polynomial trends that might fit the data, and can also be used to flag unusual levels or trajectories for individual participants.

## Model Specification and Analysis

In this section, we first start with a simple random intercept model (for calculating ICC) and a simple linear growth model, followed by a model with a time-invariant covariate. We then move beyond the default growth models and discuss more complicated models, including models with different between- and within-individual covariance structures and models with nonlinear growth patterns.

### A. Random Intercept Model and Simple Linear Growth Model

In multilevel models for longitudinal data, the lowest level of data is the specific measurement at a particular time. This lowest level is referred to as “Level-1” data. Each Level-1 measurement is nested within a particular research participant. The individual, then, constitutes Level-2 data. If there is another level of nesting (e.g., if participants are Analyzing Longitudinal Data 16 nested within schools), this would be Level-3 data, and so on. The simplest two-level model with a total of 131 participants and 4 repeated measures of FIM\_P per individual over time can be presented as below:

Level-1 (repeated-measures level) model:

$$\text{FIM\_P}_{ti} = \beta_{0i} + e_{ti} \quad (1)$$

where  $t$  represents the four different measurement occasions (i.e., 12, 24, 48, and 60-months follow-up) and  $i$  represents the 131 participants (i.e.,  $i = 1 \dots 131$ ).  $\beta_{0i}$  is the estimated average FIM\_P score (over the 4 FIM\_P scores) for the  $i$ -th individual.  $e_{ti}$  is the within-individual random error which captures the difference between the observed FIM\_P score at time  $t$  and the predicted (average) score of the  $i$ -th participant.  $e_{ti}$  is generally assumed to be normally distributed with variance equal  $\sigma^2$  (i.e.,  $e_{ti} \sim N(0, \sigma^2)$ ) which captures the within-individual variation. Equation (1) shows the average FIM\_P score for the  $i$ -th patient. In this example, we have 131 patients so that we have 131 average FIM\_P scores. We can further summarize these 131 average FIM\_P scores by the following equation:

Level-2 (individual level) models:

$$\beta_{0i} = \gamma_{00} + U_{0i} \quad (2)$$

where  $\gamma_{00}$  is the grand mean of the 131 average FIM\_P scores and  $U_{0i}$  is the difference between the  $i$ -th average FIM\_P score and the grand mean. As shown in Table 6 (Model ICC), the grand mean FIM\_P score was 85.954.  $U_{0i}$  is assumed to be normally distributed with variance equals  $\tau_{00}$  (i.e.,  $U_{0i} \sim N(0, \tau_{00})$ ). Additionally, the within-individual random errors ( $e_{ti}$ ) are assumed to be *independent* from the between-individual random effects ( $U_{0i}$ ). The combination of equations (1) and (2) is named as “random-intercept model” in which no predictor is included in these two equations. Intra-class correlation (ICC) which measures the magnitude of dependency between observations can be calculated by using the within- and between-individual variances (i.e.,  $\sigma^2$  and  $\tau_{00}$ ) from equations (1) and (2)

$$ICC = \frac{\tau_{00}}{\tau_{00} + \sigma^2} \quad (3)$$

The ICC for this example was:  $43.915 / (43.915 + 47.753) = .479$  based on the variance estimates presented in Table 6 (Model ICC). ICC is the proportion of the between-individual variance to the sum of the between- and within-individual variances of an outcome variable and generally ranges between 0 and 1. Hox (2002) interpreted ICC as “the proportion of the variance explained by the grouping structure in the population” (p.15). ICC can also be (roughly) viewed as the average relation between any pair of observations (i.e., the FIM\_P scores) within a cluster (i.e., a patient in our example). Barcikowski (1981) showed that the type I error rate could be inflated (e.g., from the nominal .05 level to .06) when a very small ICC (e.g., .01) occurred. ICC in educational research with cross-sectional design generally ranges between .05 and .20 (Snijders & Bosker, 1999). The relative high ICC from this example (.479) is probably due to the longitudinal nature of the data given that the exact same measure was assessed repeatedly from the same patient over time.

A simple linear growth model can be presented by the following equation:

Level-1 (repeated-measures level) model:

$$\text{FIM\_P}_{ti} = \beta_{0i} + \beta_{1i} \text{Time}_{ti} + e_{ti} \quad (4)$$



To simplify the illustration, we coded  $Time_{ti}$  as 0 for the first (i.e., 12 months) follow-up,  $Time_{ti}$  as 12 for the second (i.e., 24 months) follow-up,  $Time_{ti}$  as 36 for the third (i.e., 48 months) follow-up, and  $Time_{ti}$  as 48 for the fourth (i.e., 60 months) follow-up. Unlike in (equation 1) of the random intercept model,  $\beta_{0i}$  in (equation 4) is the estimated FIM\_P score for the  $i$ -th individual at the first (i.e., 12 months) follow-up when  $Time_{ti}$  is equal to 0.  $\beta_{1i}$  is the average monthly change in FIM\_P score for the  $i$ -th individual over time.  $e_{ti}$  is still the within-individual random error with variance equal  $\sigma^2$  which captures the within-individual variation (i.e.,  $e_{ti} \sim N(0, \sigma^2)$ ). More discussion on modeling the within-individual variation will be given in the later section. (Equation 4) shows the regression model based on the four FIM\_P scores for the  $i$ -th participant. Indeed, we can fit the same model to the 131 participants separately. Hence, we can have 131 different sets of regression coefficients (i.e., the intercept ( $\beta_0$ ) and the average monthly change ( $\beta_1$ )). We can summarize these 131 sets of parameter estimates by the following two equations:

Level-2 (individual level) models:

$$\beta_{0i} = \gamma_{00} + U_{0i} \quad (5)$$

$$\beta_{1i} = \gamma_{10} + U_{1i} \quad (6)$$

where  $\gamma_{00}$  is the average score of FIM\_P at the initial time point (i.e.,  $Time_{ti} = 0$ ) and  $\gamma_{10}$  is the average monthly change in FIM\_P over the 131 participants. Both  $U_{0i}$  and  $U_{1i}$  are the between-individual random effects and are assumed to be normally distributed (i.e.,

$\begin{bmatrix} U_{0i} \\ U_{1i} \end{bmatrix} \sim N(\underline{0}, T)$ , where  $T = \begin{bmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{bmatrix}$ ).  $U_{0i}$  captures the difference between the intercept ( $\beta_{0i}$ ) of the  $i$ -th participant from the average intercept  $\gamma_{00}$ , and  $U_{1i}$  captures the difference between the estimated monthly change in FIM\_P ( $\beta_{1i}$ ) of the  $i$ -th participant from the average monthly change in FIM\_P ( $\gamma_{10}$ ) across the 131 participants. The variances of  $U_{0i}$  and  $U_{1i}$  are  $\tau_{00}$  and  $\tau_{11}$  respectively which capture the between-individual variation.

The meaning of  $\tau_{00}$ ,  $\tau_{11}$ , and the covariance,  $\tau_{01}$  (or  $\tau_{10}$ ) between the two random effects (i.e.,  $U_{0i}$  and  $U_{1i}$ ) can be further explained through visualization. As shown in Figure 3a, the average model is the bolded straight line with the average intercept  $\gamma_{00}$  (i.e., the average FIM\_P score at the first time measure) and the average monthly change  $\gamma_{10}$ , and the other straight lines are the individual predicted models for different participants. The variation between the 131 intercepts (of the 131 regression models) and the average intercept  $\gamma_{00}$  is captured by  $\tau_{00}$  and the variation between the 131 monthly changes and the average monthly change  $\gamma_{10}$  is captured by  $\tau_{11}$ . When  $\tau_{00}$  is equal to zero as shown in Figure 3b, all the intercepts of the 131 regression models are the same as the average intercept  $\gamma_{00}$  and all 131 regression models pass through  $\gamma_{00}$ . On the other hand, when  $\tau_{11}$  is equal to zero, all the predicted monthly changes of the 131 regression models are the same as the average monthly change  $\gamma_{10}$  and all 131 regression models are parallel to the average model (i.e., the bolded straight line) as shown in Figure 3c. Figure 3d shows a possible look for a positive covariance  $\tau_{01}$ . That is, individuals who have a higher FIM\_P score at the first time measure are more likely to have larger predicted monthly change in FIM\_P than those who score lower on the FIM\_P at the first time point. The annotated SPSS and SAS syntaxes for fitting all the growth models in this paper are presented in Table 5.

Consistent with the mean pattern as shown in Table 3, the FIM\_P score decreased over time (see Model A in Table 6). The average FIM\_P score at the first time measure (i.e., 12-months follow-up) was 87.351, and it decreased .058 point per month (or .70 point per year). On

average, the patients reported worse functional abilities as time passed. Except for the covariance  $\tau_{01}$ , the variances of the two random effects (i.e.,  $\tau_{00}$  and  $\tau_{11}$ ) were statistically significant, which indicated a significant amount of variation between the 131 individual regression models and the average model. The significance of these two random-effect variances also implied that some potential individual-related variables might be able to explain/account for the variation between individual regression models and the average model. The significant variance of the within-individual random error,  $\sigma^2$ , indicated a significant amount of variation between the observations at different time points and the individual regression model within each person.

## B. Model with Time-Invariant Covariate

Because of the two significant random-effect variances, we can further examine some potential individual-related predictors which may be able to account for the variation between individual regression models and the average model. Many individual-related variables are time-invariant variables because the values of these variables are the same over time, such as participant's gender and ethnicity. We use the grand-meancentered participant's age at the first time measure (i.e.,  $c\_Age_i = Age_i - m\_Age$ ) as the person-related predictor for the demonstration.  $Age_i$  is the initial age or age at the first time measure of the  $i$ -th participant and  $m\_Age$  is the mean age at the first time measure over the 131 participants. For example, as shown in Figures 1a and 1b, the initial age for the first participant (i.e.,  $participant\_id = 1100067$ ) was 22 and the mean-centered age for this individual was  $-23.82$  given that the mean initial age for the 131 participants was 45.82 (i.e.,  $-23.82 = 22 - 45.82$ ). A major reason for using the centered age is to have meaningful interpretation for the intercept. Biesanz et al (2004) discussed the centering issue (especially on centering the time variable) in longitudinal analysis. More discussions on centering in the general MLM framework can be found in Kreft, de Leeuw and Aiken (1995) and Enders and Tofghi (2007).

In our example, the level-1 (repeated-measures level) model is the same as shown in equation (4), whereas the level-2 (individual level) models are:

$$\beta_{0i} = \gamma_{00} + \gamma_{01} c\_Age_i + U_{0i} \quad (7)$$

$$\beta_{1i} = \gamma_{10} + \gamma_{11} c\_Age_i + U_{1i} \quad (8)$$

By substituting equations (7) and (8) back into equation (4), we have:

$$FIM\_P_{it} = \gamma_{00} + \gamma_{01} c\_Age_i + \gamma_{10} Time_{it} + \gamma_{11} c\_Age_i * Time_{it} + U_{0i} + U_{1i} * Time_{it} + e_{it} \quad (9)$$

, where the first four terms (i.e.,  $\gamma_{00}$ ,  $\gamma_{01} c\_Age_i$ ,  $\gamma_{10} Time_{it}$ , and  $\gamma_{11} c\_Age_i * Time_{it}$ ) are the fixed-effects which capture the average model. The last three terms (i.e.,  $U_{0i}$ ,  $U_{1i} * Time_{it}$ , and  $e_{it}$ ) are the random effects which capture the variation between individual regression models and the average model (i.e.,  $U_{0i}$  and  $U_{1i} * Time_{it}$ ) and the variation between individual observations and the regression model within each person (i.e.,  $e_{it}$ ). The results are presented in Table 6 (Model B).

Because of the added cross-level interaction effect (i.e.,  $c\_Age_i * Time_{it}$  in (equation 9)), the regression coefficients of the lower order terms (i.e.,  $\gamma_{00}$  of the intercept,  $\gamma_{01}$  of  $c\_Age_i$ , and  $\gamma_{10}$  of  $Time_{it}$ ) are conditional terms and have to be interpreted along with the interaction term. For example, the regression coefficient of  $Time_{it}$ ,  $-.058$ , was not the average monthly change

in FIM\_P score for all 131 participants but for those whose initial age or age at the first time measure equal to 45.82 because the values of  $c\_Age_i$  for these individuals were equal to zero. Similarly, the intercept, 87.351, was not the average FIM\_P score at the first time measure for all 131 participants but for those whose initial age equal to 45.82. The non-significant coefficient of  $c\_Age_i$ ,  $\gamma_{01}$ , indicated that there was no relation between the mean-centered age and the FIM\_P score at  $Time_{ti} = 0$ .

To understand the meaning of the interaction effect,  $c\_Age_i * Time_{ti}$ , we can use the steps suggested by Aiken and West (1991) to decompose the interaction effect. The general idea of the Aiken and West (1991) procedure is using a 2-dimensional figure to present a 3-dimensional relationship (i.e., two predictors,  $c\_Age_i$  and  $Time_{ti}$ , and the outcome variable,  $FIM\_P_{ti}$ ). It is quite straightforward to decompose the interaction effect in longitudinal analysis, in which the Y-axis is always the outcome variable ( $FIM\_P_{ti}$ ) and the X-axis is always the “ $Time_{ti}$ ” predictor. Hence we only need to substitute some meaningful values for the second predictor (i.e., the predictor other than  $Time_{ti}$  which is the time invariant covariate,  $c\_Age_i$ ). The three commonly used values for substitution are: the mean of the predictor, one standard deviation (*SD*) above the mean value of the predictor, and one *SD* below the mean value of the predictor. Hence, the three values of  $c\_Age_i$  we used for decomposing and plotting the interaction effect were:  $-17.02$  (1SD below the mean of  $c\_Age_i$ ),  $0$  (mean of  $c\_Age_i$ ) and  $17.02$  (1SD above the mean of  $c\_Age_i$ ). The corresponding predicted model for each specific  $c\_Age_i$  value is shown below:

Younger participants (with  $c\_Age_i = -17.02$  or original  $Age_i = 45.82 - 17.02 = 28.80$  years old):

$$\begin{aligned}\widehat{FIM\_P}_{ti} &= 87.351 - .058 Time_{ti} - .028 c\_Age_i - .003 Time_{ti} * c\_Age_i \\ &= 87.351 - .058 Time_{ti} - .028 (-17.02) - .003 Time_{ti} * (-17.02) \\ &= 87.828 - .007 Time_{ti}\end{aligned}$$

Mean age participants (with  $c\_Age_i = 0$  or  $Age_i = 45.82$  years old):

$$\begin{aligned}\widehat{FIM\_P}_{ti} &= 87.351 - .058 Time_{ti} - .028 (0) - .003 Time_{ti} * (0) \\ &= 87.351 - .058 Time_{ti}\end{aligned}$$

Older participants (with  $c\_Age_i = 17.02$  or  $Age_i = 45.82 + 17.02 = 62.84$  years old):

$$\begin{aligned}\widehat{FIM\_P}_{ti} &= 87.351 - .058 Time_{ti} - .028 (17.02) - .003 Time_{ti} * (17.02) \\ &= 86.874 - .109 Time_{ti}\end{aligned}$$

Figure 4 presents the three predicted models for each of the three age groups. In general, the younger participants had slightly higher FIM\_P scores at the first time measure, even though it was not significantly different from those in other age groups. On the other hand, the declination rate of the FIM\_P scores was significantly slower in the younger-participant group (i.e.,  $-.007$  point/month) than the older-participant group (i.e.,  $-.109$  point/month). In other words, older individuals experienced a steeper rate of decline in functional abilities over time than younger individuals.

We can also evaluate the effectiveness of the time-invariant covariate,  $c\_Age_i$ , on explaining the between-individual variation by using the Pseudo- $R^2$  statistic (Raudenbush & Bryk, 2002; Singer & Willet, 2003) as shown below:

$$Pseudo\_R^2_{\tau_{00}} = \frac{\tau_{00\_Unconditional} - \tau_{00\_Conditional}}{\tau_{00\_Unconditional}}$$

and

$$Pseudo\_R^2_{\tau_{11}} = \frac{\tau_{11\_Unconditional} - \tau_{11\_Conditional}}{\tau_{11\_Unconditional}}$$

for  $\tau_{00}$  and  $\tau_{11}$  respectively.  $\tau_{00\_Unconditional}$  and  $\tau_{11\_Unconditional}$  are the variances of the random effects for the model *without* the time invariant covariate  $c\_Age_i$ , whereas  $\tau_{00\_Conditional}$  and  $\tau_{11\_Conditional}$  are the variances of the random effects for the model *with*  $c\_Age_i$ . Hence, the *Pseudo-R<sup>2</sup>* statistic is the proportion of explained variance in the random effect by the time invariant covariate. Based on the information in Table 6 models A and B, the *Pseudo-R<sup>2</sup>* statistics for  $\tau_{00}$  and  $\tau_{11}$  are:

$$Pseudo\_R^2_{\tau_{00}} = \frac{\tau_{00\_Unconditional} - \tau_{00\_Conditional}}{\tau_{00\_Unconditional}} = \frac{15.253 - 15.231}{15.253} = .001 (or .1\%)$$

and

$$Pseudo\_R^2_{\tau_{11}} = \frac{\tau_{11\_Unconditional} - \tau_{11\_Conditional}}{\tau_{11\_Unconditional}} = \frac{.061 - .059}{.061} = (or 3.3\%).$$

That is, the initial age or age at the first follow-up measure (i.e.,  $c\_Age_i$ ) could only explain .1% of the variance in  $\tau_{00}$  (i.e., the variation of the FIM\_P score over the 131 participants at the first time point) but 3.3% of the variance in  $\tau_{11}$  (i.e., the variation of the monthly change in FIM\_P score over the 131 participants). Given that the explained variance is the analog of the R-square change in the OLS regression, we can adopt Cohen's (1988) guideline (i.e., .02, .13 and .26 in R-square change representing small, medium and large effect respectively) and conclude that the initial age has no effect on predicting the functional abilities at the initial time point (i.e., 12-months follow-up) but a small effect on predicting the linear rate of change in the functional abilities over time. These findings are consistent with the tests of significance for the individual parameter estimates as shown in Table 6 model B (i.e., the  $p$ -value of  $\gamma_{01}$  was .29, whereas the  $p$ -value of  $\gamma_{11}$  was less than .05). Additionally, these small explained variances imply the omission of other important time-invariant covariates in the model and further examination of the model is needed.

The advantage of using *Pseudo-R<sup>2</sup>* is that it provides an easy to use and understandable measure of effect size. However, unlike the regular  $R^2$ , negative *Pseudo-R<sup>2</sup>* can be obtained especially when level-1 (repeated-measures/within-individual level) predictors only contribute to the within-individual variation but not the between-individual variation. This can also be explained by the compensatory relation between the within-individual variance and between-individual variance (Kwok, et al., 2007; Snijders & Bosker, 1994; 1999). Snijders and Bosker (1994, 1999) provided more discussion on the use of *Pseudo-R<sup>2</sup>* and suggested an alternative way<sup>3</sup> for calculating the  $R^2$  for different level to prevent the negative explained variance.

In addition to the explained variance, researchers can compare different multilevel models using the information criteria such as Akaike information criterion (AIC) and Bayesian information criterion (BIC). The general guideline for using these information criteria is to

select the model with the smallest value on either AIC or BIC. Additional guideline on comparing models using BIC is available (Raftery, 1996). There is no substantial difference between two models if the BIC difference is less than 2. On the other hand, there is a substantial difference between two models if the difference between the two BIC values is larger than 10. Nevertheless, a few studies showed that the effectiveness of using these information criteria on model selection (especially on selecting the correct variance-covariance matrix of the random effects and errors) was relatively low (Keselman, Algina, Kowalchuk, & Wolfinger, 1998). Hence, these information criteria should be used with caution and should not be used as the sole criterion on selecting models without considering both theoretical explanation and application of the model.

### C. Beyond the Default Models

In previous sections we discussed models which are the common/default models in most of statistical programs. These models always come with some default assumptions, specifically on the random-effect part of the model. For example, a specific covariance structure, an identity (ID) structure (i.e.,  $\sigma^2\mathbf{I}$ ), is always assumed for the covariance structure of the within-individual variation (i.e.,  $e_{it}$ ), which may not be applicable, especially for longitudinal analysis (Kwok, et al, 2007). The misspecification of random effect covariance structure may result in biased estimation of the variances of the random effects, which in turn may affect the estimation of the standard errors and the test of significance of the fixed effects.

One of the advantages of using MLM for analyzing longitudinal data is that MLM offers great flexibility in modeling the covariance structure for both between-individual random effects and within-individual random errors. Researchers can search for the optimal covariance structure, which theoretically results in the highest statistical power and increases the precision of estimates of the fixed effects (Davis, 2002; Diggle, Heagerty, Liang & Zeger, 2002; Keselman, Algina, & Kowalchuk, 2001; Singer & Willet, 2003; Wolfinger, 1996).

#### Modeling the Covariance Structure for the Between-Individual Random Effects

—Recall that we examined two models, a simple linear growth model (model A) and a model with a time invariant covariate,  $c\_Age_i$  (model B). In these two models, we estimated three elements (i.e.,  $\tau_{00}$ ,  $\tau_{11}$ , and  $\tau_{01}$ ) in the covariance structure for the between-individual random effects. As shown in Table 6 (models A and B), the covariance,  $\tau^{01}$ , was not significant in either one of the models. The covariance structure used for these two models is called unstructured (UN) in which all unique elements in the covariance structure are free for estimation:

$$V \begin{bmatrix} U_{0i} \\ U_{1i} \end{bmatrix} = \begin{bmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{bmatrix}.$$

Because of the non-significance of the covariance (i.e.,  $\tau_{01}$  not significantly different from zero), we have fitted the same two models (i.e., models C1 and C2 in Table 6) with a simpler covariance structure, namely, UN(1) structure (in SAS) or DIAG structure (in SPSS) as presented below:

<sup>3</sup>By fitting a multilevel model with only random effect associated with the intercept (i.e., no random effect associated any coefficients other than the intercept), Snijders and Bosker (1994, 1999) provided an alternative way to calculate the explained variance by redefining

the within-individual variance as  $\sigma^2$  and the between-individual variance as  $\frac{\sigma^2}{n} + \tau_{00}$  in which  $n$  is the cluster size (i.e., the number of observation per cluster). For unbalanced design,  $n$  can be the harmonic mean of the cluster size across all clusters.



$$V \begin{bmatrix} U_{0i} \\ U_{1i} \end{bmatrix} = \begin{bmatrix} \tau_{00} & 0 \\ 0 & \tau_{11} \end{bmatrix},$$

in which only the variances but not covariance of the random effects are estimated. In other words,  $\tau_{01}$  is constrained to zero which implies that there is no relation between the functional abilities at the initial time point (i.e., the 12 months follow-up) and the rate of change in the functional abilities over time across individuals.

By using the likelihood ratio test, we can compare models A and C1 with their -2LL values. Because the only difference between these two models is the covariance  $\tau_{01}$  which has been estimated in model A but not in model C1 (i.e.,  $\tau_{01}=0$  in model C1), the difference in the -2LL values of these two *nested* models follows a chi-square distribution with 1 degree of freedom. The likelihood ratio test,  $\chi^2(1) = -2LL_{\text{Model A}} - (-2LL_{\text{Model C1}}) = 3449.133 - 3449.097 = .036$ , is not statistically significant ( $p = .850$ ) which means that  $\tau_{01}$  is not different from zero. The estimated  $\tau_{00}$  and  $\tau_{11}$  in model C1 are slightly larger than the ones in model A due to the redistribution of the variance between the random effects (Luo & Kwok, 2006; Meyers & Beretvas, 2006). Then, we added the time-invariant covariate,  $c\_Age_i$ , back in model C1 and the results of this model (model C2) are presented in Table 6. Basically, the only difference on the model specification between models B and C2 is that  $\tau_{01}$  is estimated in model B but constrained to zero in model C2. We can use the same equation presented previously to obtain the Pseudo- $R^2$  statistics for the changes in  $\tau_{00}$  and  $\tau_{11}$  after including  $c\_Age_i$  in the model. By constraining  $\tau_{01}$  to zero in models C1 and C2, the Pseudo- $R^2$  statistics for both  $\tau_{00}$  and  $\tau_{11}$  are larger (i.e.,  $R^2_{\tau_{00}} = .015$  and  $R^2_{\tau_{11}} = .048$  than the ones based on models A and B (i.e.,  $R^2_{\tau_{00}} = .001$  and  $R^2_{\tau_{11}} = .033$ ). Instead of using the default setting, modeling the variance-covariance matrix of the between-individual random effects may result in higher explained variances.

### Modeling the Covariance Structure for the Within-Individual Random Errors—

As Kwok et al (2007) pointed out, when analyzing longitudinal data under the MLM framework, researchers typically assume the *within-individual errors* to be independently and identically distributed (i.i.d.) with mean zero and homogenous variance  $\sigma^2$  for all participants (i.e.,  $\underline{e} \sim N(0, \sigma^2 \mathbf{I})$ ). The simplification of the within-individual covariance structure (i.e.,  $\sigma^2 \mathbf{I}$ , also named identity structure) may bias the estimation of the standard errors of the fixed effects, which, in turn, may lead to incorrect statistical inferences for the fixed effects. Singer and Willet (2003) also emphasized that obtaining an adequate within-individual covariance structure is a key element in estimating the proper effect size and properly accounting for missing values in multilevel data. Thus, choosing the optimal error structure is an important task in MLM. As Campbell and Kenny (1999) stated, “the correlational structure of longitudinal data almost always has a proximally autocorrelated structure: adjacent waves of measurement correlate more highly than nonadjacent waves, and more remote in time, the lower the correlation (Campbell & Reichardt, 1991; Kenny & Campbell, 1989)...except for data that are highly cyclical (Warner, 1998), proximal autocorrelation is the norm” (p.121).

Given that the proximal autocorrelation is a common phenomenon in longitudinal data, the first-order autoregression (AR(1)) structure was also fitted to the within-individual covariance of our example data. AR(1) is one of the commonly used covariance structures when analyzing longitudinal data and it has been widely applied in the latent growth models (Bollen & Curran, 2004; 2006; Curran & Bollen, 2001). AR(1) contains two parameters (the error variance  $\sigma^2$  and the autocorrelation coefficient  $\rho$ ). An example of AR(1) with four repeated measures is shown below (compared with the default identity (ID) structure in model C2):

$$AR(1)_{Model, C3} = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix} \text{ vs. } ID_{Model, C2} = \sigma^2 \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The results based on the AR(1) structure are shown in Table 6 (model C3). The only difference between models C2 and C3 is in the **within-individual covariance** where model C2 has the default identity structure (i.e.,  $\sigma^2\mathbf{I}$ ) while model C3 has the AR(1) structure. The difference between these two models on fitting the within-individual covariance matrix can be examined by the likelihood ratio test given that the default identity structure is nested within the AR(1) structure. The significant difference between the -2LL values of these two models (i.e.,  $\chi^2(1) = 3459.544 - 3445.827 = 13.717, p < .001$ ) indicated that the AR(1) structure (model C3) fitted the within-individual covariance matrix better than the default identity structure (model C2). Nevertheless, most of the parameter estimates of model C3 were very similar to the ones in model C2. One noticeable difference between the two models is that the change in the  $p$ -value of the Time\*c\_Age interaction effect (i.e., from  $p = .042$  in model C2 reduced to  $p = .037$  in model C3) implicitly showed the increment in the statistical power after modeling the within-individual covariance as the AR(1) structure. Additionally, the negative autocorrelation coefficient ( $\rho = -.450$ ) indicated a tendency of the FIM\_P score to oscillate within patients over time (after controlling for the average linear growth trend). That is, a negative relation ( $\rho = -.450$ ) between the adjacent time points was found while a smaller positive relation ( $\rho^2 = .202$ ) between the non-adjacent time points (i.e., the 1<sup>st</sup> and 3<sup>rd</sup> time points, and the 2<sup>nd</sup> and 4<sup>th</sup> time points) was presented.

#### D. Handling Missing Data

Missing data may occur in which individuals are absent from one or more data collection occasions. MLM addresses this issue by taking all observations into account regardless of the design of the study. In our example dataset, there were 251 individuals who had at least one response in the four time measures. Among these 251 individuals, 27 of them only responded to one time point, 39 of them responded to two time points, 54 of them responded to three time points, and 131 of them responded to all four time points. Because of the required multivariate data format (see Figure 1a) for a repeated-measures UANOVA analysis, listwise deletion would be adopted and only the 131 individuals with complete data could be included<sup>4</sup>. On the other hand, all 251 individuals can be included HLM, SAS in the analysis when MLM is used. The results based on all 251 individuals are presented in Table 6 model D. Models C3 and D are exactly the same in the model setup except the sample size (i.e., model C3 only included 131 individuals whereas model D included 251 individuals).

Compared with model C3, a very similar pattern of results is found in model D. There are two major differences, however, between the two models: 1) all regression coefficients are significant in model D, and 2) the variance of the random intercept  $\tau_{00}$  is substantially larger in model D than in model C3. The significance of all regression coefficients is the result of the increased sample size (i.e., increased statistical power). The increment in  $\tau_{00}$  is the result of the inclusion of more individuals, especially the ones with only a single response which could

<sup>4</sup>An alternative way to analyze incomplete data using UANOVA is to incorporate the multiple imputation procedure (Schafer, 1997) if the data are missing completely at random (MCAR) or missing at random (MAR, Little & Rubin, 2002). Multilevel missing data can be imputed using either the PAN routine provided by Schafer (<http://www.stat.psu.edu/~jls/misoftwa.html>), or the SAS PROC MI routine (i.e., to impute the multivariate data then convert the imputed data to univariate format for analysis). HLM, SAS (PROC MIANALYZE) and SPSS (MVA module) have the missing data routine which can analyze the imputed data. Nevertheless, as pointed out by Twisk (2006), "it has even been shown that applying multilevel analysis to an incomplete dataset is even better than applying imputation methods (Twisk & de Vente, 2002; Twisk, 2003)" (p.107).

only contribute to the estimation of between-individual parameters and variations (Muthen, 2002, February 20).

### E. Required Sample Size for Longitudinal Analysis using MLMs

There are many rules of thumb on the required sample size for multilevel models (e.g., 15 units per cluster by Bryk & Raudenbush (1992); 30 clusters/30 units per cluster rule by Kreft (1996), and the 50 clusters/20 units per cluster rule for detecting cross-level interaction effect by Hox (1998)). However, none of these rules of thumb can provide an accurate estimation of the sample size (in terms of a desired level of statistical power along with a specific size of the target effect). Moreover, all these rules of thumb require a relatively large number of level-1 units (i.e., the number of repeated measures) which may not be applicable for longitudinal studies with educational or psychological data given that these data often contain a fewer number of waves/repeated measures. Generally, more reliable estimates for the individual growth models can be obtained with a relative large number of measurement waves (e.g., 8 or more). Moreover, a larger number of higher level units (i.e., the number of patients in our example) can increase the statistical power for detecting the effects of the higher level predictors and the cross-level interaction effects between the within- and between-individual predictors. More accurate sample size estimation for longitudinal analysis in MLM can be obtained through freeware such as PINT (version 2.1; Bosker, Snijders, & Guldemond, 2003) which can be downloaded from: <http://stat.gamma.rug.nl/snijders/>, and Optimal Design (version 1.76; Spybrook, Raudenbush, Liu, Congdon, & Martinez, 2008) which can be downloaded from: [http://sitemaker.umich.edu/group-based/optimal\\_design\\_software](http://sitemaker.umich.edu/group-based/optimal_design_software). Additionally, A freeware application for determining sample sizes for MLM for two-group repeated measures designs (e.g., clinical trials) can be obtained from <http://tigger.uic.edu/~hedeker/ml.html>. This application (RMASS2) is based on the work of Hedeker, Gibbons, and Waternaux (1999), and allows for attrition, both fixed and random effects models, and several variance-covariance structures for repeated measures.

### F. Beyond Linear Growth Models

In this paper we only focused on linear growth models (i.e., the change in the outcome variable FIM\_P is in a linear fashion over time). In fact, the great majority of applications of growth models to date have used linear models. The advantages of using linear growth models include: 1) these models are simple and easy to interpret, and 2) they can adequately represent the growth process when the number of measurement waves is small or the study is short, or both. However, some phenomena such as the development of crystallized and fluid intelligence over the entire life span (Finkel, et al., 2003) cannot be adequately captured by linear growth models. Other forms of growth models such as a quadratic growth model (i.e., adding a quadratic term,  $time^2$ , into the model) or piecewise model (i.e., segmenting a nonlinear process into multiple linear growth models; Khoo, 2001; Raudenbush & Bryk, 2002) are commonly used for representing nonlinear growth processes. In addition to examining a single developmental process, parallel process models (PPM) provides the opportunity for studying the relation between multiple developmental processes simultaneously. More information on the setup and interpretation of PPM can be found in the studies by Cheong, MacKinnon, and Khoo (2003), and by Kwok, West and Sousa (2006).

In the current example we only examined a two-level model with repeated measures nested within patients. In some other settings, patients can also be nested within some higher level units/clusters such as different wards or hospitals. The variables associated with these higher level units are contextual variables (e.g., the total number of patients in a hospital, the ratio of number of doctors to number of patients, and the hospital type). These contextual variables may be able to further explain/account for the variation in the initial FIM\_P scores as well as the variation in the change in the FIM\_P scores over time across all patients. More information

on fitting a higher-level (e.g., three-level) models and interpreting the contextual effects can be found in both Raudenbush and Bryk's (2002) text and Snijders and Bosker's (1999) text.

## Conclusion

In this paper, we illustrated how to analyze longitudinal data under the MLM framework. There are two major phases for analyzing longitudinal data: preparation phase and analysis phase. In the preparation phase, researchers should carefully inspect their data to screen for errors or outliers in the data, and obtain some basic descriptive statistics for their data such as means, variances, skewness and kurtosis. As described in the data preparation section, data have to be converted into univariate format before analyzing them in MLM. Moreover, researchers are encouraged to plot their data with spaghetti plots to see whether there is any potential trend in the data over time.

In the analysis phase, Wallace and Green (2002) suggested four steps for analyzing longitudinal data in MLM: 1) review the past literatures to formulate the initial model; 2) examine the initial model and evaluate the fixed part of the model (i.e., shape of the average growth model plus covariates); 3) evaluate the random part of the model (i.e., the covariance structure of the random effects) with the same fixed part based on step 2; and 4) fine tune the fixed part of the model. An iterative process between steps 3 and 4 is recommended until a stable and interpretable model is obtained.

Similar steps have been adopted in our example. We first analyzed the data with a simple linear model (i.e., model A). We then added the time-invariant covariate ( $c\_Age_i$ ) to see whether it could account for the variations in both intercepts and slopes over the 131 participants (i.e., model B). After we confirmed the fixed part of the model (i.e., the shape of the average model plus covariates), we examined the random part of the model (i.e., the covariance structures of both between-individual variations and within-individual random errors). We first fitted a simpler between-individual covariance structure (i.e., without estimating  $\tau_{01}$ ; models C1 and C2). Then, we examined the AR(1) structure for the within-individual covariance structure (i.e., model C3), and the likelihood ratio test showed that the AR(1) structure fitted better to the data than the default identity structure.

In this paper, we provided only a simple overview of models and procedures for analyzing longitudinal data under the MLM framework. There are many texts such as Singer and Willet (2003), Hedeker and Gibbons (2006), and Weiss (2005) which have provided more in-depth treatments on analyzing longitudinal data in MLM. Readers can also find more information on using the latent growth model (LGM) to analyze longitudinal data in the texts by Duncan et al (1999) and Bollen and Curran (2006). Additionally, there are many useful resources from the internet, including the University of Bristol (UK) Centre for Multilevel Modeling (<http://www.cmm.bristol.ac.uk/index.shtml>) and the UCLA multilevel modeling portal (<http://statcomp.ats.ucla.edu/mlm/>) which contain useful information such as links to other MLM related online resources, and reviews and links to different MLM software including HLM, MLwinN, Mplus, and Stata.

## Acknowledgments

This study was supported in part by National Institute of Child Health and Human Development (NICHD) grant R01HD039367 awarded to the first author, and Grant No. R49-CE000191 from the U. S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Injury Prevention and Control to the University of Alabama at Birmingham, Injury Control Research Center. The contents of this study are solely the responsibility of the authors and do not necessarily represent the official views of the funding agencies.

## References

- Aiken, LS.; West, SG. Multiple regression: Testing and interpreting Interaction. Newbury Park, CA: Sage; 1991.
- Barcikowski RS. Statistical power with group mean as the unit of analysis. *Journal of Educational Statistics* 1981;6:267–285.
- Biesanz JC, West SG, Kwok O. Personality over time: Methodological approaches to the study of short-term and long-term development and change. *Journal of Personality* 2003;71:905–941. [PubMed: 14633053]
- Bollen KA, Curran PJ. Autoregressive latent trajectory (ALT) models: A synthesis of two traditions. *Sociological Methods and Research* 2004;32:336–383.
- Bollen, KA.; Curran, PJ. Latent curve models: A structural equation perspective. Hoboken, NJ: John Wiley & Sons; 2006.
- Bosker, RJ.; Snijders, TAB.; Guldmond, H. PINT (Power IN Two-level designs): Estimating standard errors of regression coefficients in hierarchical linear models for power calculations. 2003. Retrieved from: [http://stat.gamma.rug.nl/Pint21\\_UsersManual.pdf](http://stat.gamma.rug.nl/Pint21_UsersManual.pdf)
- Bryk, AS.; Raudenbush, SW. Hierarchical linear models: Applications and data analysis methods. Newbury Park, CA: Sage; 1992.
- Campbell, DT.; Kenny, DA. A primer on regression artifacts. New York, NY: Guilford; 1999.
- Campbell, DT.; Reichardt, CS. Problems in assuming the comparability of pretest and posttest in autoregressive and growth models. In: Snow, RE.; Wiley, E., editors. *Improving Inquiry in Social Science: A Volume in Honor Lee Cronbach*. Mahwah NJ:: Erlbaum; 1991. p. 201-219.
- Cheong J, MacKinnon DP, Khoo S. Investigation of mediational processes using latent growth curve modeling. *Structural Equation Modeling* 2003;10:238–262. [PubMed: 20157639]
- Chi EM, Reinsel GC. Models for longitudinal data with random effects and AR(1) errors. *Journal of the American Statistical Association* 1988;84:452–459.
- Clay DL, Wood PK, Frank RG, Hagglund KJ, Johnson JC. Examining systematic differences in adaptation to chronic illness: A growth modeling approach. *Rehabilitation Psychology* 1995;40:61–70.
- Cohen, J. Statistical power analysis for the behavioral sciences. 2nd Ed.. Mahwah, NJ: Erlbaum; 1988.
- Collins, LM.; Sayer, AG. New methods for the analysis of change. Washington, D.C: American Psychological Association; 2001.
- Curran, PJ.; Bollen, KA. The best of both worlds: combining autoregressive and latent curve models. In: Collins, LM.; Sayer, AG., editors. *New Methods for the Analysis of Change*. Washington, DC: American Psychological Association; 2001. p. 107-135.
- Davis, CS. Statistical methods for the analysis of repeated measurements. San Diego, CA: Springer; 2002.
- Diggle PJ. An approach to the analysis of repeated measurements. *Biometrics* 1988;44:959–971. [PubMed: 3233259]
- Diggle, P.J.; Heagerty, P.; Liang, K.; Zeger, S.L. Analysis of longitudinal data. 2nd Ed.. New York, NY: Oxford; 2002.
- Duncan, TE.; Duncan, SC.; Strycker, LA.; Li, F.; Alpert, A. An introduction to latent variable growth curve modeling: Concepts, issues, and applications. Mahwah, NJ: Erlbaum; 1999.
- Elliott T. Presidential address: Defining our common ground to reach new horizons. *Rehabilitation Psychology* 2002;47:131–143.
- Elliott T, Shewchuk R, Richards JS. Family caregiver social problem-solving abilities and adjustment during the initial year of the caregiving role. *Journal of Counseling Psychology* 2001;48:223–232.
- Enders CK, Tofighi D. Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods* 2007;12:121–138. [PubMed: 17563168]
- Fidell, LS.; Tabachnick, BG. Preparatory data analysis. In: Schinka, John A.; Velicer, Wayne F., editors. *Comprehensive Handbook of Psychology (Volume 2): Research Methods in Psychology*. New York, NY: Wiley; 2003. p. 115-141.



- Finkel D, Reynolds CA, McArdle JJ, Gatz M, Pedersen NL. Latent growth curve analyses of accelerating decline in cognitive abilities in adulthood. *Developmental Psychology* 2003;39:535–550. [PubMed: 12760521]
- Frank RG, Thayer JF, Hagglund KJ, Veith AZ, Schopp LH, Beck NC, Kashani JH, Goldstein DE, Cassidy JT, Clay DL, Chaney JM, Hewett JE, Johnson JC. Trajectories of adaptation in pediatric chronic illness: The importance of the individual. *Journal of Consulting and Clinical Psychology* 1998;66:521–532. [PubMed: 9642891]
- Grant JS, Elliott TR, Weaver M, Bartolucci AA, Giger JN. Telephone intervention with family caregivers of stroke survivors after rehabilitation. *Stroke* 2002;33:2060–2065. [PubMed: 12154263]
- Gravel J, Opatrny L, Shapiro S. The intention-to-treat approach in randomized controlled trials: Are authors saying what they do and doing what they say? *Clinical Trials* 2007;4:350–356. [PubMed: 17848496]
- Greenspan AI, Wrigley JM, Kresnow M, Branche-Dorsey CM, Fine PR. Factors influencing failure to return to work due to traumatic brain injury. *Brain Injury* 1996;10:207–218. [PubMed: 8777392]
- Hall KM, Hamilton BB, Gordon WA, Zasler ND. Characteristics and comparisons of functional assessment indices: Disability rating scale, functional independence measure, and functional assessment measure. *Journal of Head Trauma Rehabilitation* 1993;8:60–74.
- Hedeker, D.; Gibbons, RD. *Longitudinal data analysis*. Hoboken, NJ: Wiley; 2006.
- Hedeker D, Gibbons RD, Waternaux C. Sample size estimation for longitudinal designs with attrition: comparing time-related contrasts between two groups. *Journal of Educational and Behavioral Statistics* 1999;24:70–93.
- Heinemann AW, Linacre JM, Wright BD, Hamilton BB, Granger C. Relationships between impairment and physical disability as measured by the functional independence measure. *Archives of Physical Medicine and Rehabilitation* 1993;74:566–573. [PubMed: 8503745]
- Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomized controlled trials. *British Medical Journal* 1999;319:670–674. [PubMed: 10480822]
- Hox, J. Multilevel modeling: when and why. In: Balderjahn, I.; Mathar, R.; Schader, M., editors. *Classification, data analysis, and data highways*. New York, NY: Springer Verlag; 1998. p. 147-154.
- Hox, J. *Multilevel analysis: techniques and applications*. Mahwah, NJ: Erlbaum; 2002.
- Jones RH, Boadi-Boateng F. Unequally spaced longitudinal data with AR(1) serial correlation. *Biometrics* 1991;47:161–175. [PubMed: 2049497]
- Kenny DA, Campbell DT. On the measurement of stability in over-time data. *Journal of Personality* 1989;57:445–481. [PubMed: 2769563]
- Keselman HJ, Algina J, Kowalchuk RK. The analysis of repeated measures designs: A review. *British Journal of Mathematical and Statistical Psychology* 2001;54:1–20. [PubMed: 11393894]
- Keselman HJ, Algina J, Kowalchuk RK, Wolfinger RD. A comparison of two approaches for selecting covariance structures in the analysis of repeated measurements. *Communications in Statistics: Simulation and Computation* 1998;27(3):591–604.
- Keith RA, Granger CV, Hamilton BB, Sherwin FS. The functional independence measure: A new tool for rehabilitation. *Advances in Clinical Rehabilitation* 1987;1:6–18.
- Khoo S. Assessing program effects in the presence of treatment-baseline interactions: A latent curve approach. *Psychological Methods* 2001;6:234–257. [PubMed: 11570230]
- Khoo, S.; West, SG.; Wu, W.; Kwok, O. Longitudinal methods. In: Eid, M.; Diener, E., editors. *Handbook of Psychological Measurement: A Multimethod Perspective*. Washington, DC: APA; 2006. p. 301-317.
- Kreft, IGG. Are multilevel techniques necessary? An overview, including simulation studies. Unpublished manuscript. 1996. Retrieved from <http://www.stat.ucla.edu/~kreft/quarterly/quarterly.html>
- Kreft IGG, DeLeeuw J, Aiken LS. Variable centering in hierarchical linear models: Model parameterization, estimation, and interpretation. *Multivariate Behavioral Research* 1995;30:1–21.
- Kwok O, West SG, Green SB. The impact of misspecifying the within-subject covariance structure in multiwave longitudinal multilevel models: A monte carlo study. *Multivariate Behavioral Research* 2007;42:557–592.

- Kwok, O.; West, SG.; Sousa, KH. Analyzing longitudinal parallel processing (LPP) model under structural equation modeling (SEM) framework: An example from the AIDS Time-oriented Health Outcome Study (ATHOS); Paper presented at 14th Annual Meeting of the Society for Prevention Research; San Antonio, TX. 2006.
- Laird N, Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982;38:963–974. [PubMed: 7168798]
- Littell, RC.; Milliken, GA.; Stroup, WW.; Wolfinger, RD.; Schabenberber, O. SAS System for Linear Mixed Models. 2nd Ed.. NC: SAS Institute; 2006.
- Little, RJA.; Rubin, DB. Statistical analysis with missing data. 2nd Ed.. New York, NY: Wiley; 2002.
- Longford, NT. Random coefficient models. New York, NY: Oxford University Press; 1993.
- Luo, W.; Kwok, O. Impacts of ignoring a crossed factor in analyzing cross-classified multilevel data: A Monte Carlo study; Paper presented at the 71st Annual Meeting of the Psychometric Society; Montreal, Canada. 2006.
- Meyers J, Beretvas SN. The impact of inappropriate modeling of cross-classified data structures. *Multivariate Behavioral Research* 2006;41:473–497.
- Muthén, B. Multilevel Data/Complex Sample: Cluster Size. 2006 Feb 2. Message posted to Mplus Discussion, archived at <http://www.statmodel.com/discussion/messages/12/164.html>
- Muthén B, Curran P. General longitudinal modeling of individual differences in experimental designs: A latent variable framework for analysis and power estimation. *Psychological Methods* 1997;2:371–402.
- Peugh JL, Enders CK. Using the SPSS Mixed procedure to fit hierarchical linear and growth trajectory models. *Educational and Psychological Measurement* 2005;65:811–835.
- Putzke JD, Barrett JJ, Richards JS, Underhill AT, LoBello SG. Life satisfaction following spinal cord injury: Long-term follow-up. *The Journal of Spinal Cord Medicine* 2004;27:106–110. [PubMed: 15162879]
- Raftery, AE. Bayesian model selection in social research. In: Marsden, PV., editor. *Sociological Methodology*. Oxford: Basil Blackwell; 1996. p. 111-163.
- Raudenbush SW. Educational applications of hierarchical models: A review. *Journal of Educational Statistics* 1988;13:85–116.
- Raudenbush, SW.; Bryk, AS. Hierarchical linear models: Applications and data analysis methods. 2nd Ed.. Thousand Oaks, CA: Sage Publications; 2002.
- Rivera P, Elliott T, Berry J, Grant J. Problem-solving training for family caregivers of persons with traumatic brain injuries: A randomized controlled trial. *Archives of Physical Medicine and Rehabilitation*. in press.
- Schafer, JL. Analysis of incomplete multivariate data. New York, NY: Chapman & Hall; 1997.
- Shewchuk R, Richards JS, Elliott T. Dynamic processes in health outcomes among caregivers of individuals with spinal cord injuries. *Health Psychology* 1998;17:125–129. [PubMed: 9548703]
- Singer JD. Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics* 1998;24:323–355.
- Singer, JD.; Willett, JB. Applied longitudinal data analysis: Modeling change and event occurrence. New York, NY: Oxford University Press; 2003.
- Snijders TAB, Bosker RJ. Modeled variance in two-level models. *Sociological Methods & Research* 1994;22:342–363.
- Snijders, TAB.; Bosker, RJ. Multilevel analysis: An introduction to basic and advanced multilevel modeling. CA: Sage Publications; 1999.
- Spybrook, J.; Raudenbush, SW.; Liu, X.; Congdon, R.; Martinez, A. Optimal design for longitudinal and multilevel research: Documentation for the “Optimal Design” software. 2008. Retrieved from: <http://sitemaker.umich.edu/group-based/files/od-manual-20080312-v176.pdf>
- Tabachnick, BG.; Fidell, LS. Using multivariate statistics. 5th ed.. Boston, MA: Allyn & Bacon; 2006.
- Twisk, JWR. Applied longitudinal data analysis for epidemiology: A practical guide. New York, NY: Cambridge University Press; 2003.
- Twisk, JWR. Applied multilevel analysis. New York, NY: Cambridge University Press; 2006.

- Twisk JWR, de Vente W. Attrition in longitudinal studies: How to deal with missing data. *Journal of Clinical Epidemiology* 2002;55:329–337. [PubMed: 11927199]
- Underhill AT, Lobello SG, Stroud TP, Terry KS, Devivo MJ, Fine PR. Depression and life satisfaction in patients with traumatic brain injury: A longitudinal study. *Brain Injury* 2003;17:973–982. [PubMed: 14514448]
- Underhill AT, Lobello SG, Fine PR. Reliability and validity of the family satisfaction scale with survivors of traumatic brain injury. *Journal of Rehabilitation Research and Development* 2004;41:603–610. [PubMed: 15558388]
- Wallace, D.; Green, SB. Analysis of repeated-measures designs with linear mixed models. In: Moskowitz, DS.; Hershberger, SL., editors. *Modeling Intraindividual Variability with Repeated Measures Data: Method and Applications*. Englewood Cliffs, NJ: Erlbaum; 2002. p. 103-134.
- Warschausky S, Kay J, Kewman D. Hierarchical linear modeling of FIM instrument growth curve characteristics after spinal cord injury. *Archives of Physical Medicine and Rehabilitation* 2001;82:329–334. [PubMed: 11245754]
- Warner, R. *Spectral analysis of time-series data*. New York, NY: Guilford; 1998.
- Weiss, RE. *Modeling longitudinal data*. New York, NY: Springer; 2005.
- West, SG.; Biesanz, JC.; Kwok, O. Within subject and longitudinal experiments: design and analysis issues. In: Sansone, C.; Morf, CC.; Panter, AT., editors. *Handbook of Methods in Social Psychology*. Thousand Oaks, CA: Sage; 2003. p. 287-312.
- Wolfinger RD. Covariance structure selection in general mixed models. *Communications in Statistics-Simulation and Computation* 1993;22:1079–1106.
- Wolfinger RD. Heterogeneous variance-covariance structures for repeated measures. *Journal of Agricultural, Biological, and Environmental Statistics* 1996;1:205–230.

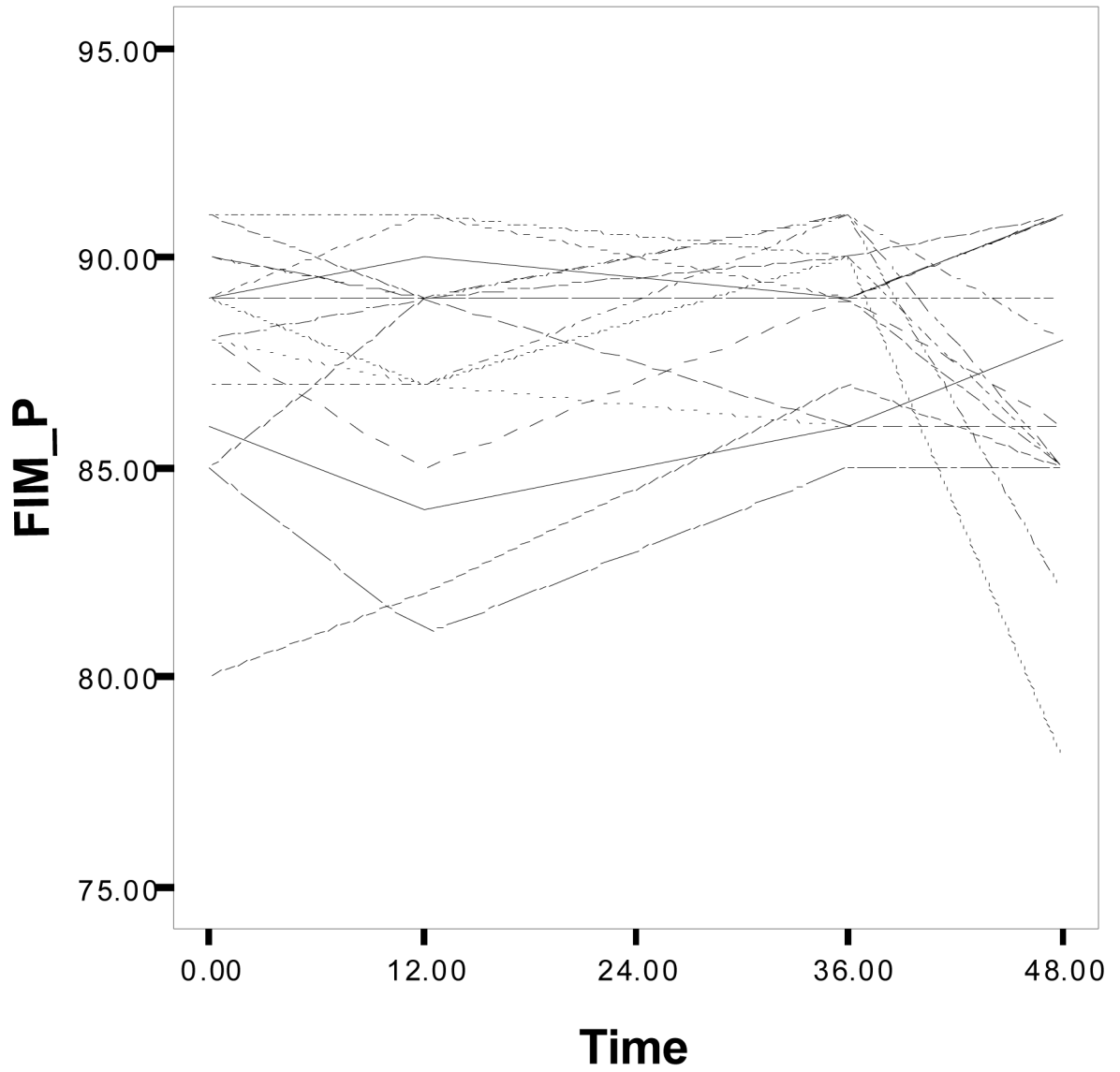
Figure 1\_multi [DataSet2] - SPSS Data Editor

	patient_id	age	c_age	FIM_P12	FIM_P24	FIM_P48	FIM_P60
1	1100067	22	-23.82	89.00	90.00	89.00	91.00
2	1100545	76	30.18	83.00	89.00	88.00	87.00
3	1130108	66	20.18	84.00	87.00	89.00	88.00

Figure 1\_uni [DataSet3] - SPSS Data Editor

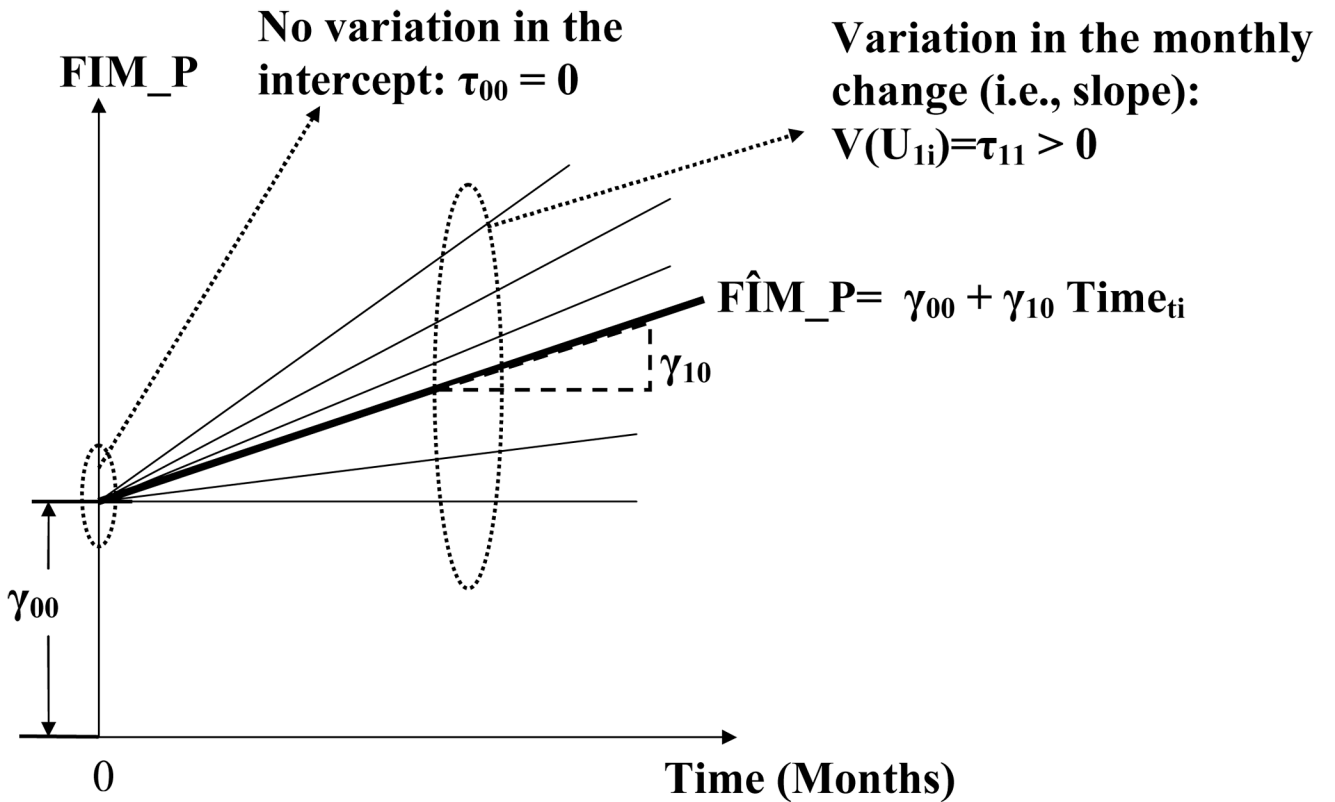
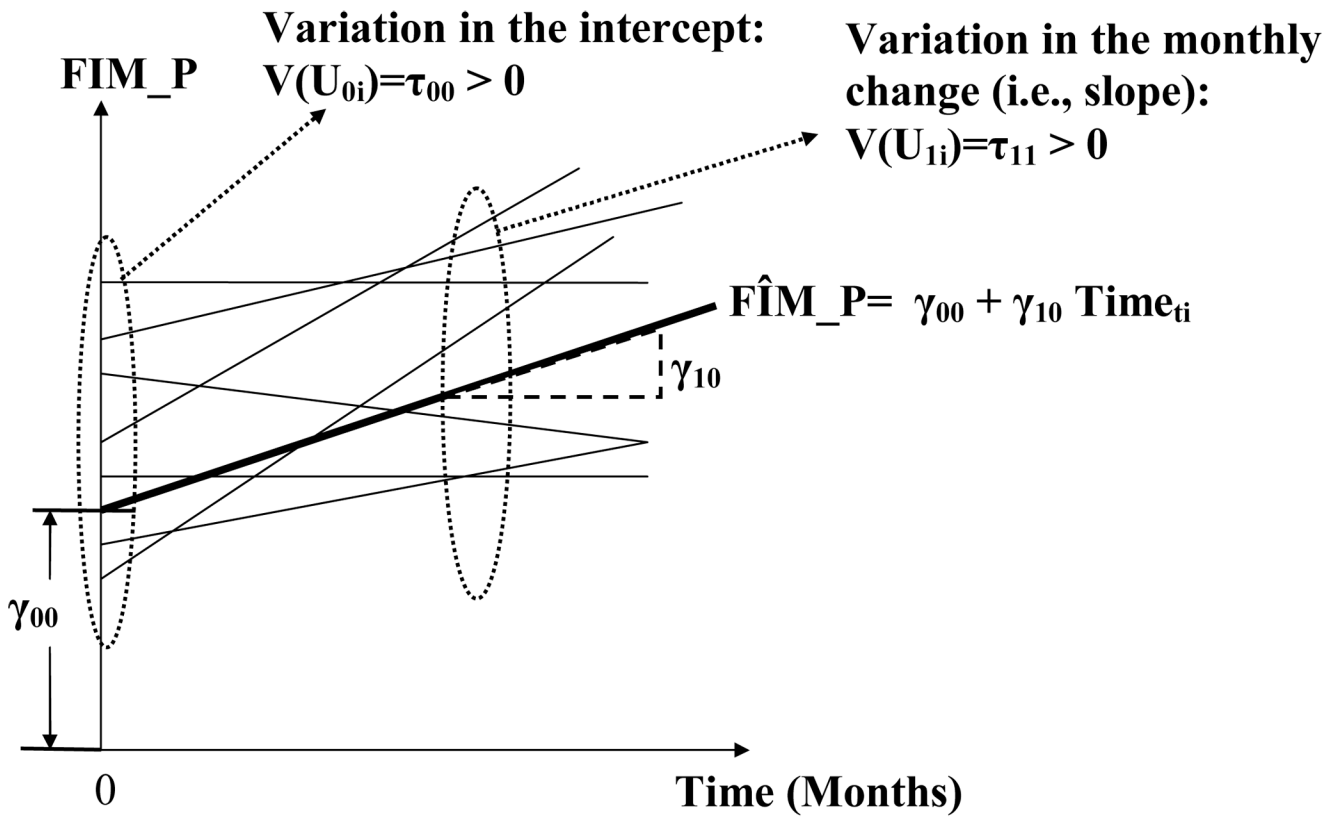
	patient_id	age	c_age	time	Index1	FIM_P
1	1100067	22	-23.82	.00	1	89.00
2	1100067	22	-23.82	12.00	2	90.00
3	1100067	22	-23.82	36.00	3	89.00
4	1100067	22	-23.82	48.00	4	91.00
5	1100545	76	30.18	.00	1	83.00
6	1100545	76	30.18	12.00	2	89.00
7	1100545	76	30.18	36.00	3	88.00
8	1100545	76	30.18	48.00	4	87.00
9	1130108	66	20.18	.00	1	84.00
10	1130108	66	20.18	12.00	2	87.00
11	1130108	66	20.18	36.00	3	89.00
12	1130108	66	20.18	48.00	4	88.00

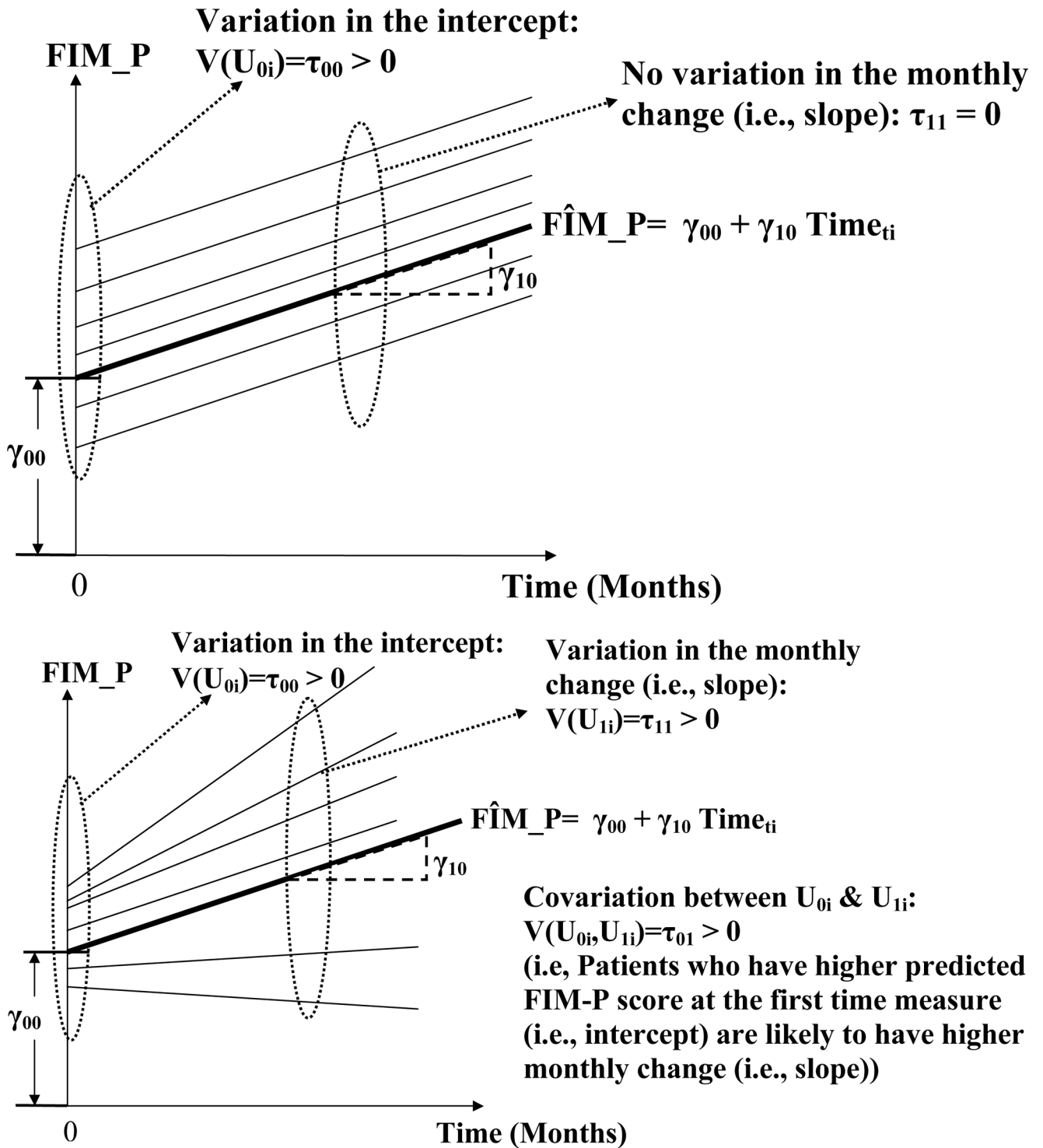
**Figure 1.**  
 Figure 1a. Data in Multivariate Format (SPSS)  
 Figure 1b. Data in Univariate Format (SPSS)



**Figure 2.**  
Spaghetti plots of a Random Sample with 20 Participants







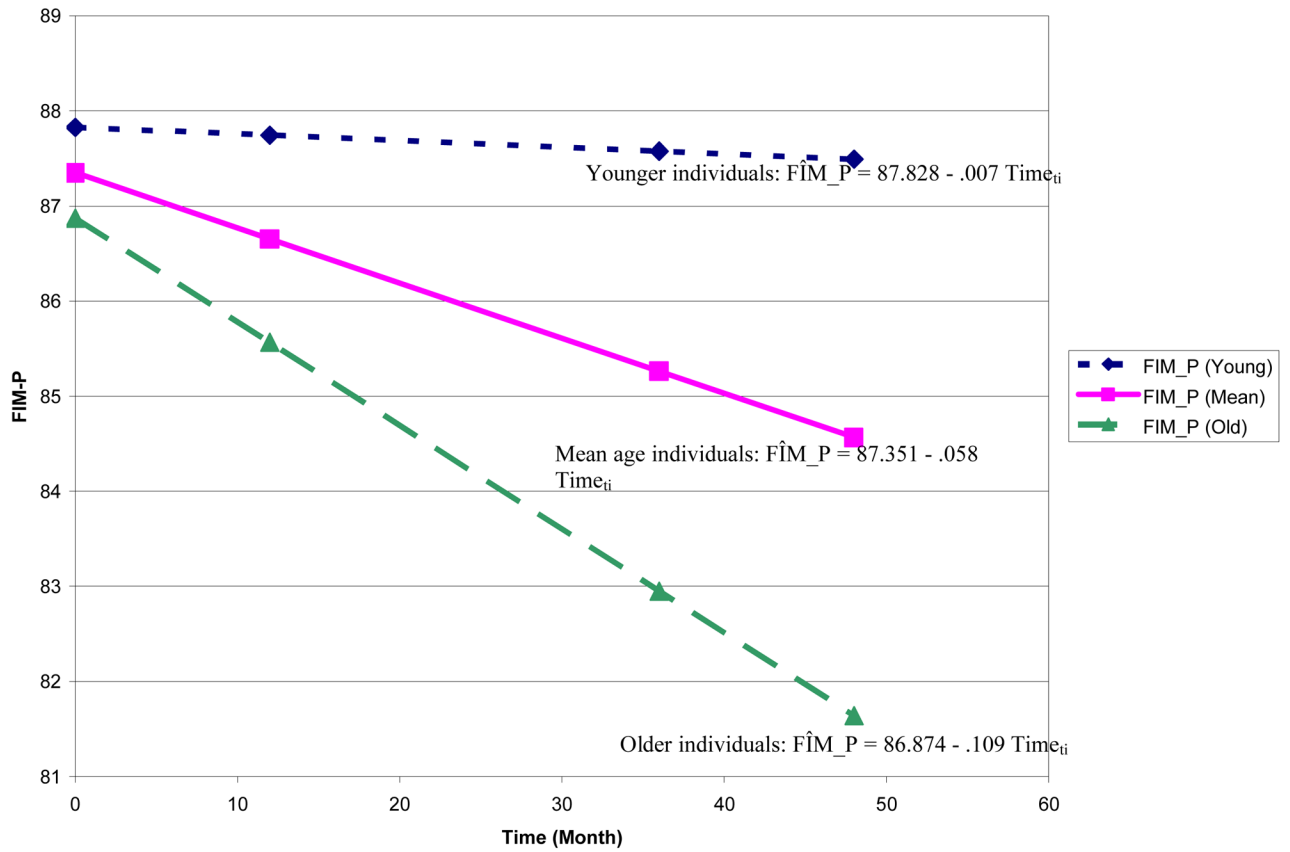
**Figure 3.**

*Figure 3a.* The Visual Demonstration of  $\tau_{00}$ ,  $\tau_{11}$ , and  $\tau_{01}$  ( $\tau_{00} > 0$  and  $\tau_{11} > 0$ )

*Figure 3b.* The Visual Demonstration of  $\tau_{00}$ ,  $\tau_{11}$ , and  $\tau_{01}$  ( $\tau_{00} = 0$  and  $\tau_{11} > 0$ )

*Figure 3c.* The Visual Demonstration of  $\tau_{00}$ ,  $\tau_{11}$ , and  $\tau_{01}$  ( $\tau_{00} > 0$  and  $\tau_{11} = 0$ )

*Figure 3d.* The Visual Demonstration of  $\tau_{00}$ ,  $\tau_{11}$ , and  $\tau_{01}$  ( $\tau_{00} > 0$ ,  $\tau_{11} > 0$ , and  $\tau_{10} = \tau_{01} > 0$ )



Overall model:  $\hat{FIM}_P = 87.351 - .058 Time_{ti} - .028 c\_Age_i - .003 Time_{ti} * c\_Age_i$

**Figure 4.**  
Decomposing the  $c\_Age_i * Time_{ti}$  Interaction Effect

**Table 1**

## Comparisons between ANOVA-based Analyses and MLM Analyses

	Repeated Measures UANOVA	Multilevel Models
Analysis Strategies	<p><b>A.</b> Simple within-subjects designs: Time contrasts and average polynomial trends (time as a categorical factor; balanced data and equal time spacing assumed)</p> <p><b>B.</b> Mixed Within/Between Designs (e.g., randomized clinical trials): Interaction effects between time and other between-subjects factors.</p>	<p><b>A.</b> Simple within-subjects designs: Time contrasts and average polynomial trends (time as a categorical factor; balanced data and time spacing assumed)</p> <p><b>B.</b> Mixed Within/Between Designs (e.g., randomized clinical trials):</p> <ol style="list-style-type: none"> <li><b>1.</b> Interaction effects between time and other between-subjects factors.</li> <li><b>2.</b> <i>Cross-level interactions</i> effects of between-subjects factors/variables on individual growth trajectories).</li> </ol> <p><b>C.</b> Regression parameters from individual growth models including intercepts, slopes, both can be treated as random effects for estimation (time continuous variable; unbalanced data and unequal time spacing accommodated)</p>
Estimation Method	Least Squares	Maximum Likelihood
Missing Data	Complete cases only	All available data
Variance-Covariance Structure	Compound Symmetry (or Huynh-Feldt (H-F) which is a more general form of the compound symmetry structure) to meet the sphericity assumption	Flexible Structure
Covariates	Time invariant covariates only	Both time invariant and time-varying covariates

**Table 2**

## Descriptive Statistics of the Participants with IAF

	N = 131	N = 251
Age (Standard Deviation of Age)	45.82 (17.02)	43.89 (17.59)
Gender		
Men	75 (57.3%)	142 (56.6%)
Women	56 (42.7%)	109 (43.4%)
Ethnicity		
African American	38 (29.0%)	74 (29.5%)
Caucasian	93 (71.0%)	175 (69.7%)
Others	0 (0.0%)	2 (.8%)
Employment Status at the 12 Months Follow up:		
1 = Employed, full time	41 (31.3%)	72 (28.7%)
2 = Employed, part time	3 (2.3%)	9 (3.6%)
3 = Self-employed	4 (3.1%)	10 (4.0%)
4 = Unemployed	18 (13.7%)	35 (13.9%)
5 = Student	4 (3.1%)	6 (2.4%)
6 = Retired	20 (15.3%)	37 (14.7%)
7 = Not working because of a previous disability	35 (26.7%)	67 (26.7%)
8 = Other	5 (3.8%)	11 (4.4%)
9 = Unknown	1 (.8%)	4 (1.6%)
Marital Status at the 12 Months Follow up:		
1 = Single	28 (21.4%)	61 (24.3%)
2 = Married	74 (56.5%)	126 (50.2%)
3 = Divorced	16 (12.2%)	31 (12.4%)
4 = Separated	2 (1.5%)	8 (3.2%)
5 = Widowed	10 (7.6%)	21 (8.4%)
9 = Other/Unknown	1 (.8%)	4 (1.6%)
Education Status at the 12 Months Follow up:		
1 = 8 <sup>th</sup> grade or less	15 (11.5%)	23 (9.2%)
2 = 9 <sup>th</sup> through 11 <sup>th</sup> grade	27 (20.6%)	59 (23.5%)
3 = High school diploma/GED	48 (36.6%)	86 (34.3%)
4 = Trade school	2 (1.5%)	3 (1.2%)
5 = Some college, no degree	19 (14.5%)	41 (16.3%)
6 = Associate degree	1 (.8%)	4 (1.6%)
7 = Bachelor's degree	11 (8.4%)	21 (8.4%)
8 = Master's degree	6 (4.6%)	6 (2.4%)
9 = Doctorate	0 (0.0%)	1 (.4%)
10 = Other	2 (1.5%)	3 (1.2%)
99=Unknown	0 (0.0%)	4 (1.6%)

**Table 3**

Descriptive Statistics of the FIM\_P measure over time

	N = 131	N = 251
FIM_P at the first (12 Months) follow-up	87.21 <sup>†</sup> (5.21 <sup>††</sup> )	86.37 (8.37)
FIM_P at the first (24 Months) follow-up	86.79 (5.68)	86.42 (7.44)
FIM_P at the first (48 Months) follow-up	85.41 (11.52)	85.96 (10.03)
FIM_P at the first (60 Months) follow-up	84.40 (13.06)	84.64 (12.64)

<sup>†</sup> Mean<sup>††</sup> Standard Deviation



**Table 4**

SPSS and SAS Syntax for Transforming the Data from Multivariate Format to Univariate Format

	Syntax	Explanation
SPSS	<pre> VARSTOCASES /MAKE FIM_P FROM FIM_P12   FIM_P24 FIM_P48 FIM_P60 /INDEX = Time(4) /KEEP = id age c_age /NULL = KEEP. RECODE Time (1=0) (2=12) (3=36) (4=48).</pre>	<p><b>Varstocases:</b> call the data transformation procedure to convert the multivariate data format to univariate data format</p> <p><b>Make FIM_P from FIM_P12 FIM_P24 FIM_P48 FIM_P60:</b> create new (univariate) outcome variable FIM_P from FIM_P12 to FIM_P60</p> <p><b>Index=time(4):</b> create an index variable to represent different data line within each participant (e.g., individual). In this example, we create a new index variable (time) with 1,2,3,4 to represent the 4 different data lines within each individual in the univariate dataset</p> <p><b>Keep=id age c_age:</b> keep the time invariant variable (e.g., id and age at the first time measure)</p> <p><b>Null=keep</b> keep the missing data as a separate data line (e.g., if a individual has no data at both times 3 and 4, this individual will still have four data lines in the new converted univariate dataset with missing data shown for both 3<sup>rd</sup> and 4<sup>th</sup> data lines.</p> <p><b>RECODE Time (1=0) (2=12) (3=36) (4=48):</b> recode the values in the newly created Time variable (i.e., 1, 2, 3, &amp; 4) to the corresponding time values we used in the example (i.e., 0, 12, 36, &amp; 48 months).</p>
SAS	<pre> data uni; set mult; FIM_P=FIM_P12;time=0;output; FIM_P=FIM_P24;time=12;output; FIM_P=FIM_P48;time=36;output; FIM_P=FIM_P60;time=48;output; keep id age c_age time FIM_P;</pre>	<p><b>Data:</b> name the new converted univariate format data as “uni”</p> <p><b>Set:</b> read in the original multivariate format data named “mult”</p> <p><b>FIM_P=FIM_P12:</b> create new variable named FIM_P using the original variable FIM_P12</p> <p><b>Time=0:</b> create new variable “time” and set the value for the first time measure as 0</p> <p><b>Output:</b> output to the new converted univariate dataset (The same commands apply to the next three command lines for different time measures)</p> <p><b>Keep id c_age time FIM_P:</b> include all these variables (i.e., id c_age time FIM_P) in the new converted univariate dataset named “uni”</p>

Table 5

Syntaxes for analyzing different models in SAS and SPSS.

Model	SPSS syntax	SAS syntax
ICC	Mixed fim_p /fixed = intercept /random intercept   subject(individual_id) covtype(UN) /print = solution testcov.	proc mixed data=uni covtest; class individual_id; model fim_p= /solution; random intercept / type=un subject=individual_id;
A	Mixed fim_p with time /fixed = intercept time /random intercept time   subject(individual_id) covtype(UN) /print = solution testcov.	proc mixed data=uni covtest; class individual_id; model fim_p=time /solution; random intercept time / type=un subject=individual_id;
B	Mixed fim_p with time c_age /fixed = intercept time c_age time* c_age /random = intercept time   subject(individual_id) covtype(UN) covtype(UN) /print = solution testcov.	proc mixed data=uni covtest; class individual_id; model fim_p=time c_age time*c_age /solution; random intercept time/ type=un subject=individual_id;
C1	Mixed fim_p with time /fixed = intercept time /random = intercept time   subject(individual_id) covtype(diag) covtype(diag) /print = solution testcov.	proc mixed data=uni covtest; class individual_id; model fim_p=time /solution; random intercept time / type=un(1) subject=individual_id;
C2	Mixed fim_p with time c_age /fixed = intercept time c_age time* c_age /random = intercept time   subject(individual_id) covtype(diag) covtype(diag) /print = solution testcov.	proc mixed data=uni covtest; class individual_id; model fim_p=time c_age time*c_age /solution; random intercept time/ type=un(1) subject=individual_id;
C3	Mixed fim_p with time c_age by index1 /fixed = intercept time c_age time* c_age /random = intercept time   subject(individual_id) covtype(diag) /repeated=index1  subject(individual_id) covtype(ar1) /print = solution testcov.	proc mixed data=uni covtest; class individual_id; model fim_p=time c_age time*c_age /solution; random intercept time/ type=un(1) subject=individual_id; repeated / type=ar(1) subject=individual_id;
D	(Same as Model C3)	(Same as Model C3)

Note:

For SPSS syntax (model C3 as example):

**Mixed fim\_p with time c\_age by index1:** call the Mixed procedure in SPSS and identify the dependent variable (fim\_p) *with* the continuous predictors (time and c\_age) *by* the categorical predictor(s) (index1: label for each data line within each individual) ; **fixed:** specify the average growth (or fixed-effect) model; **random:** specify the random effects (or request for the estimation of the between-individual covariance matrix); **subject (individual\_id):** specify the level-2 cluster ID (i.e., individual id); **covtype(diag)** (in the “random” command line): specify the structure of the between-individual covariance matrix as diagonal (diag) structure; **repeated=index1:** request for the estimation of the within-individual covariance matrix; **covtype(ar1)** (in the “repeated” command line): specify the structure of the within-individual covariance matrix as the firstorder autoregressive

structure; **print=solution**: request for the growth parameter estimates (e.g.,  $\beta_0$  &  $\beta_1$ ) and their corresponding standard errors (e.g.,  $SE\beta_0$  &  $SE\beta_1$ ); **testcov**: request for the tests of the parameter estimates in the random effects and errors

For SAS syntax (model C3 as example):

**proc mixed**: call the proc mixed procedure in SAS; **data**: the dataset for the analysis; **covtest**: request for the tests of the parameter estimates in the random effects; **class**: specify the categorical variable individual\_id; **model**: specify the average growth (or fixed-effect) model; **solution**: request for the growth parameter estimates (e.g.,  $\beta_0$  &  $\beta_1$ ) and their corresponding standard errors (e.g.,  $SE\beta_0$  &  $SE\beta_1$ ); **random**: specify the random effects (or request for the estimation of the between-individual covariance matrix); **type=un(1)** (next to the “random” command): specify the structure of the between-individual covariance matrix as un(1) structure (same as the diagonal structure in SPSS); **subject**: specify the level-2 cluster ID (i.e., individual\_id); **repeated**: request for the estimation of the within-individual covariance matrix; **type=ar(1)** (next to the “repeated” command): specify the structure of the within-individual covariance matrix as the first-order autoregressive structure

**Table 6**

Parameter estimates for different model

	Model ICC <sup>†</sup>	Model A	Model B	Model C1	Model C2	Model C3	Model D
Fixed effects							
Intercept ( $\gamma_{00}$ )	85.954 (.653)	87.351 (.448)	87.351 (.448)	87.351 (.450)	87.351 (.448)	88.378 (.447)	86.476 (.492)
p-value	P<.001	p<.001	p<.001	p<.001	p<.001	p<.001	p<.001
Time ( $\gamma_{10}$ )	-.058 (.024)	-.058 (.024)	-.058 (.023)	-.058 (.024)	-.058 (.023)	-.058 (.023)	-.035 (.017)
p-value		.015	.014	.015	.014	.014	.040
c_Age ( $\gamma_{01}$ )	--	--	-.028 (.026)	--	-.028 (.026)	-.022 (.026)	-.116 (.028)
p-value			.294	--	.294	.411	p<.001
Time*c_Age ( $\gamma_{11}$ )	--	--	-.003 (.001)	--	-.003 (.001)	-.003 (.001)	-.003 (.001)
p-value			.043		.042	.037	.008
Random effects							
$\tau_{00}$	43.915	15.253	15.231	15.459	15.222	22.004	55.384
$\tau_{11}$		.061	.059	.062	.059	.067	.051
$\tau_{01}$		.024	-.001	--	--	--	--
$\sigma^2$	47.753	17.073	17.073	17.019	17.075	11.662	10.302
$\rho$		--	--	--	--	-.450	-.403
Overall Model Test							
-2LL	3713.001	3449.097	3459.544	3449.133	3459.544	3445.827	5295.452
AIC	3717.001	3457.097	3467.544	3455.133	3465.544	3453.827	5303.452
BIC	3725.520	3474.128	3484.559	3467.906	3478.306	3470.842	5322.125

<sup>†</sup> N = 131 for Models ICC, A, B, C1, C2, & C3; N=251 for Model D.

<sup>††</sup> Standard error in parentheses.

The ***bolded and italic values*** are NOT statistically significant ( $p > .05$ ). Model ICC contains no predictor; model A has time as predictor and model B has both time and  $c\_Agej$  as predictors. The model specification between models A and C1 is exactly the same except that  $\tau_0j$  is estimated in model A but not model C1. Similarly, the model specification between models B and C2 is exactly the same except that  $\tau_0j$  is estimated model B but not model C2. The only difference between models C2 and C3 is on the specification of the within-individual variance-covariance matrix in which C2 is specified with the default identity (ID) structure while C3 is specified as the first-order autoregressive (AR(1)) structure. Model specification for models C3 and C4 is exactly the same except model C3 including patients with all 4 repeated measures (N=131) whereas model C4 including all patients (N=251).