Analyzing Change: A Primer on Multilevel Models with Applications to Nephrology

Jocelyn E. Holden[a]   Ken Kelley[a]   Rajiv Agarwal[b]

[a]Inquiry Methodology Program, Indiana University-Bloomington, Bloomington, Ind., [b]Department of Medicine, Indiana University, Indianapolis, Ind., USA

Abstract

The analysis of change is central to the study of kidney research. In the past 25 years, newer and more sophisticated methods for the analysis of change have been developed; however, as of yet these newer methods are underutilized in the field of kidney research. Repeated measures ANOVA is the traditional model that is easy to understand and simpler to interpret, but it may not be valid in complex real-world situations. Problems with the assumption of sphericity, unit of analysis, lack of consideration for different types of change, and missing data, in the repeated measures ANOVA context are often encountered. Multilevel modeling, a newer and more sophisticated method for the analysis of change, overcomes these limitations and provides a better framework for understanding the true nature of change. The present article provides a primer on the use of multilevel modeling to study change. An example from a clinical study is detailed and the method for implementation in SAS is provided.

Copyright © 2008 S. Karger AG, Basel

Analyzing Change in Nephrological Research: A Primer on Multilevel Models

Analysis of change is central to the study of kidney disease. For many reasons, obtaining repeated measurements on a single individual can be valuable. Decline of glomerular filtration rate (GFR), increase in proteinuria, circadian variation in blood pressure (BP), progression of fibrosis are some examples of change. Older and less sophisticated models for addressing questions of change over time sacrifice information; modern methods allow a richer set of research questions to be addressed. This primer on multilevel models for analyzing change is intended for researchers interested in studying phenomena that change over time, with examples provided from renal research.

Change may not be linear. In other words, it is important not to conceptualize change as only monotonically increasing or decreasing. The trajectory over time may show growth until a particular time point is reached, and then begin to atrophy. Two different types of questions can be asked regarding the analysis of change. Change in some attribute of the patient (e.g. BP, GFR) is *intraindividual* change. Examination of intraindividual change allows questions to be answered such as: (a) Is there a change? (b) What is the pattern of change: linear, quadratic, asymptotic? or (c) Are there factors which correlate (observationally) or influence (experimentally) how individuals change over time?

Rajiv Agarwal, MD
Professor of Medicine, Division of Nephrology, Department of Medicine
Indiana University and RLR VA Medical Center
1481 West 10th Street, 111N, Indianapolis, IN 46202 (USA)
Tel. +1 317 554 0000, ext. 82241, Fax +1 317 988 2171, E-Mail ragarwal@iupui.edu

Intraindividual change is distinct from *intraindividual variability*. Whereas intraindividual change typically reflects a process that is relatively stable (e.g. trait), intraindividual variability refers to fluctuations and are generally reversible (e.g. state). It is important to consider intraindividual variability when studying intraindividual change because intraindividual variability can be a confounding factor in understanding intraindividual change. Minor fluctuations in the outcome variable should not be mistaken for or mask the true nature of change over time [1].

How change differs between individuals is often referred to as *interindividual* change. Interindividual change can provide information about the similarity (or lack thereof) of change within a group of individuals, or between groups of individuals. Examination of interindividual change includes such questions as: (a) Do all individuals in a group change in the same manner or at the same rate? (b) Is there a difference in the rate of trajectories for groups? (c) Is there a difference in the type of change groups follow?

Both intraindividual change and interindividual change provide different but important information about the nature of change. Traditional statistical procedures (e.g. those in the general linear model framework) used to assess change have only been able to study one type of change at a time: either intraindividual or interindividual. Multilevel modeling (MLM) is a more powerful statistical tool because both types of change can be assessed simultaneously in a single model [2].

Inadequacies of the Traditional Methods for Assessing Change

Traditionally, the most common ways to address change involve special cases of the general linear model. In particular, the most common method for analyzing change is the use of repeated measures ANOVA, for it allows for the testing of mean differences across time and trends over multiple time points. Repeated measures ANOVA, however, is not ideal for analyzing change in many situations. Repeated measures ANOVA makes the assumption of sphericity. For the assumption of sphericity to hold, there must be equal variance at each measurement occasion and an equal covariance between all pairs of time points [3]. This is a very restrictive assumption, because there may be many situations where there is greater individual variability as time passes, or the contrary, where there is less individual variability as time passes.

The assumption of sphericity can be avoided, however, by use of other general linear modeling (GLM) techniques such as multivariate analysis of variance (MANOVA) (or the paired samples t test if only two measurement occasions exist). Unfortunately, there can be problems when using MANOVA approaches for measuring change. MANOVA techniques provide some evidence about whether or not change is present and can test various polynomial trends. However, in general we learn little about individual or group growth trajectories and little about what influences the trajectories. Use of MANOVA can be particularly problematic in situations where there is a great amount of fluctuation in scores over time. In these situations, change can be masked: the mean change over time (which MANOVA detects) may in fact be zero even though there is large variability from time point to time point. Although MANOVA and MLMs would both show a mean change of zero, the MLM would also yield parameter estimates for the variability in individual change and potentially the correlates of change.

Difficulties also arise when nested levels are used in MANOVA. It can become difficult to determine whether data should be analyzed at the individual level or at the group level. For example, if individuals are nested within different treatment clinics, is it appropriate to analyze the change across the individuals? Or to analyze whether there was change across clinics? Such issues are referred to as the unit of analysis problem – in this situation neither level is optimal: individuals nested within clinic are not independent and thus violate the independence assumption, whereas aggregated scores across individuals for mean clinic scores reduce statistical power and precision and potentially lead to aggregation bias[1]. This type of question can be addressed by use of a nested MANOVA model; however, models of this type can quickly become complex and can be difficult to interpret. Such nested structures (individuals nested within clinic) are easy to deal with in the MLM framework.

Finally, MANOVA, ANOVA, t test, and repeated measures ANOVA techniques all encounter problems with missing data. Use of these techniques requires that each

---

[1] In some situations the level at which to nest is not clear – that is, how far should the nesting go? On the extreme, one could nest within individuals over time, within clinics, within physician, within examination room, within day of the week, etc. Obviously, such a complex nesting structure is not necessary or advised. When the nesting structure has theoretical or empirical support, it should be considered. As with many models in statistics, the effect of including an additional level of nesting is a model modification and those models are nested, which implies they can be empirically evaluated. In our experience, two- or three-level models are adequate for typical medical applications.

individual included in the sample have data for each measurement occasion. The general recommendation is for any individual or case missing data on one or more observations to be deleted from analysis in a listwise/casewise manner. In many studies, especially longitudinal studies, there are problems with subject attrition, which would, then, necessitate the deleting of all cases with missing data if such an analysis were approached from a general linear model perspective. Such a strategy can be detrimental to the sample size of a study and wastes information that is collected in the study. There are advanced statistical ways to deal with missing data in the general linear model framework; however, these advanced techniques are many times difficult to deal with and are unnecessary due to the benefits provided by the MLM. In situations with missing data, MLMs can be very useful because they use all available information and weight estimates based on the amount of data available [4, 5].

Moving outside of GLM techniques, another traditional approach to the analysis of change is time series analysis. One potential benefit of time series analysis is that unlike the GLM techniques previously discussed, although the residual variances at each time point are assumed uncorrelated after the model has been fit, the effect of correlations between scores at each time point is included as a parameter in the model. Thus nonindependence of observations is less of an issue. However, traditional time series analysis generally requires a data structure where relatively few individuals have many measurement occasions. In traditional longitudinal situations, which can be considered time series in and of themselves, the structure of the data is generally relatively few measurement occasions for a large numbers of individuals. Long time series can themselves be nested, where the unit of analysis issues noted for GLM techniques can be an issue. The convergence of many time series models and nested structures can often be addressed in an MLM context, where many individuals each have many measurement occasions, further illustrating the generality of the MLM. MLMs can use any number of time points (although three or more is preferred), and as was previously discussed, can easily deal with missing data and nested structure.

Generalized estimating equations (GEE) also provide an alternative to MLM. Similar to time series analysis, GEE can model correlations between the individual time points, thus again nonindependence of time points is less of an issue. GEE are also well suited to handle missing data as encountered by longitudinal studies. One potential drawback to GEE, however, is that the variance-covariance matrix is practically ignored. Thus, all information about the variability between scores, individuals or groups is lost. MLMs not only model longitudinal data over time but allow for analysis of the variance-covariance matrix providing the researcher with more information. Also, like the GLM and time series analyses, issues of nested structure may arise which the MLM can deal with easily [6].

As can be seen, traditional methods for detecting change are largely inferior to modern methods for many research purposes. As an alternative to the procedures mentioned above, MLMs provide a method for modeling change, which explicitly accounts for interindividual and intraindividual change simultaneously in a single model. MLMs, also known in different contexts as hierarchical (non)linear models, mixed effects models, random effects models, random coefficient models, covariance components models, or value added models offer a method which can fit individual change equations for each subject or case in the study, and then model those individual change coefficients at another level of the model in an attempt to explain interindividual differences in change.

MLM is so called because it analyzes variables at several different levels of analysis using nested models. By nested models, we refer to situations where individuals, or cases more generally, are contained within groups, and those groups may be further contained in other groups. For example, BP readings are obtained from each arm. These readings are nested within patients. These patients might be nested within a treatment group versus a control group. In addition to the analysis of both inter- and intraindividual change, MLMs can also include information on factors that affect the change.

The Model

The simplest model of change follows a straight line. The overall model is often decomposed into two parts: level 1 and level 2.

Level 1
Level 1 of the longitudinal MLM measures intraindividual change [7]. Level 1 of each MLM generally consists of N prediction equations, where N represents the number of participants. The straight line change model can be described mathematically as:

$$y_{it} = \pi_{0i} + \pi_{1i}(Time_{it}) + e_{it}, \tag{1}$$

where y_{it} is the criterion variable for the i-th individual ($i = 1, \ldots, N$) at the t-th time point ($t = 1, \ldots, T$), π_{0i} is the intercept for the i-th individual, π_{1i} is the slope for the i-th individual, and e_{it} is the

error in predicting the i-th individual at the t-th time point. In this model, $Time_{it}$ is the only explanatory variable. When time is used as an explanatory variable in the level 1 model, the model can be conceptualized as a longitudinal model [8].

More specifically, in longitudinal models, the criterion variable y_{it} is the variable in which we are looking for change. The regression intercept for the i-th individual, π_{0i}, is the predicted value of y_i when $Time_{it}$ is zero for the i-th individual. Similarly, the regression slope π_{1i} is the expected change in the criterion with a unit increase in $Time$ for the i-th individual. In the case of the straight line change model described in equation 1, π_{1i} refers to the rate of change in y_i for a 1 unit change in $Time$. The final component of the regression equation, the error denoted e_{it}, refers to the regression error left unexplained by the level 1 regression model.

Many times researchers will want to rescale or recenter the time variable in order to make results more interpretable. At times, this constant is the sample mean of the time variable, though the constant can change from situation to situation depending on the nature and metric of the data collected [6]. For example, if we are studying change in BP after dialysis, time zero should be the end of dialysis, instead of midnight or clock time. Rescaling of the time variable is important to consider, because it can make interpretation of the level 2 parameters more intuitive and clear.

The level 1 longitudinal model need not be limited to using $Time$ as the only explanatory variable. Sometimes, as in multiple regression models, we want to explain some criterion by a combination of several variables. Models of this type typically take the form of:

$$y_{it} = \pi_{0i} + \pi_{1i}(Time_{it}) + \pi_{2i}(X_{1t}) + \ldots + \pi_{ji}(X_{jt}) + e_{it}, \quad (2)$$

where $\pi_{2i} \ldots \pi_{ji}$ are the regression weights for explanatory variables $X_{1t} \ldots X_{jt}$. As can be seen, this model is identical to the model discussed above except it has multiple slope parameters for the covariates included in the model. These slope coefficients can be interpreted in exactly the same way as the simple example above.

It should be noted that explanatory variables in a MLM can be either time varying or time invariant. Covariates such as gender, ethnicity, or group status (treatment/control) are *time-invariant explanatory variables* because they need only be measured one time and do not change as time passes. Covariates such as age or weight can be regarded as *time-varying covariates*, because they potentially change over time and thus could be measured at each measurement occasion in order to be used in the model [7]. In equation 2, each X_{jt} is a time-varying covariate because it appears in the level 1 equation. Furthermore, quadratic, cubic, or other polynomial trends can be modeled simply by adding a slope coefficient for the term, for example $y_{it} = \pi_{0i} + \pi_{1i}(Time_{it}) + \pi_{2i}(Time_{it}^2) + e_{it}$. Models nonlinear in their parameters, for example inclusion of asymptotes or inflection points, can also be very useful [9, 10].

Level 2

The multilevel part of a MLMs comes in at the level 2 (and higher) parts of the model, for it is these higher levels that allow us to explain differences in the initial status and growth rates obtained in level 1 of the model. Until now, we have been modeling

a dependent variable with time and possibly time-varying covariates. At level 2, the individual change coefficients are modeled as the dependent variable with time-invariant predictors. For example, the intercept and slope of the N individuals might be modeled by gender or treatment group.

Consider the straight line change model from the first example. In that example, recall that there are two level 1 parameters, π_{0i} (intercept) and π_{1i} (slope) as well as an error term. This means that the level 2 model will consist of 2 equations: one for each of the level 1 change coefficients. Below is one possible level 2 model for this situation, which is termed an unconditional model because the individual change coefficients are not conditional on a time-invariant predictor:

$$\text{Level 2: } \pi_{0i} = \beta_{00} + r_{0i} \quad (3)$$

$$\pi_{1i} = \beta_{10} + r_{1i}, \quad (4)$$

where β_{00} and β_{01} are the intercepts (fixed effects) for the intercept and slope, respectively, and r_{0i} and r_{1i} are the unique effects for the i-th individual on the intercept and slope, respectively. Notice that the unique effects are simply the difference between the fixed effect (the estimate of the population value) and the unique (random) effect associated with the individual. As can be seen, we are now using regression equations to model the change coefficients from level 1. In the level 2 equations shown above, we are only attempting to model the slope and intercept with a horizontal line: an intercept + error term (fig. 1b). However, many times it is desirable to model the level 1 coefficients with time-invariant predictors. For example, theory might suggest that the initial status and trajectory over time of some variable is in part a function of gender. In other words, we want to test the hypothesis that there are group differences in initial status and growth between men and women. A suitable level 2 model to test this hypothesis is given below:

$$\text{Level 1: } y_{it} = \pi_{0i} + \pi_{1i}(Time_{it}) + e_{it} \quad (5)$$

$$\text{Level 2: } \pi_{0i} = \beta_{00} + \beta_{01}(Gender_i) + r_{0i} \quad (6)$$

$$\pi_{1i} = \beta_{10} + \beta_{11}(Gender_i) + r_{1i} \quad (7)$$

In this model, the intercept and slope from the level 1 model are each being predicted by regression lines with their own intercept and slope representing the influence of gender on each of these level 1 parameters (fig. 1d), where β_{11} represents the difference between males and females in the rate of change over time. The influence of gender does not, however, need to be tested for all parameters. If prior research supports the notion that gender has no effect on the rate of change, but might impact the initial status, a model could be constructed where the intercept is predicted by gender, but the slope is simply predicted by a means only model (i.e. a horizontal line). That is to say, each of the level 1 coefficients could potentially have their own unique prediction equation.

As in multiple regression, use of a dichotomous explanatory variable, such as gender in our example, is theoretically testing whether the change coefficients for different groups are significantly different. For example, if our data are coded 0 = male and 1 = female, equation 6 is theoretically equivalent to testing the difference between the following two equations:

$$\text{Male: } \quad \pi_{0i} = \beta_{00} + \beta_{01}(0) + r_{0i} \quad (8)$$

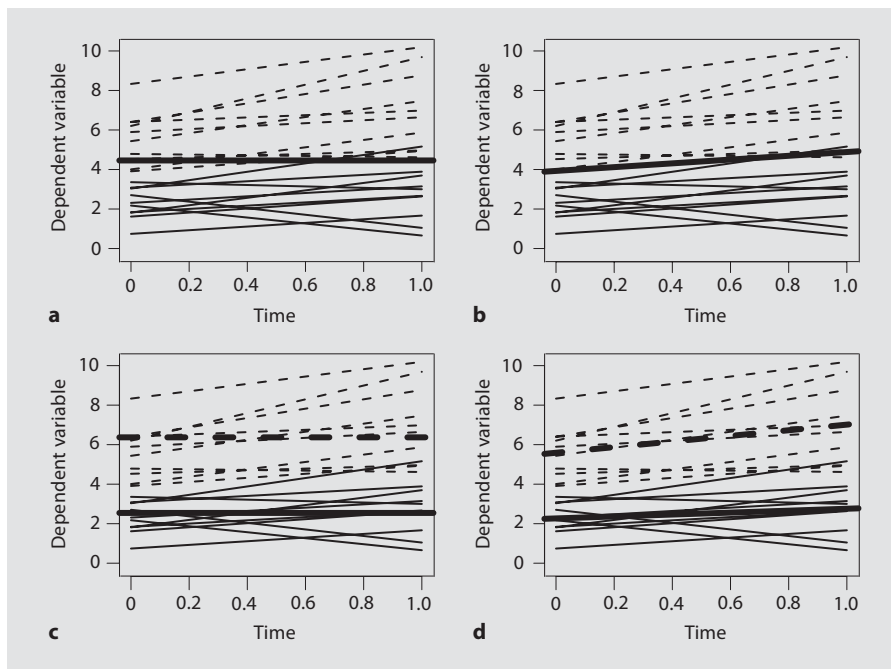$$\text{Female: } \pi_{0i} = \beta_{00} + \beta_{01}(1) + r_{0i} \quad (9)$$

Fig. 1. Four different models for the same hypothetical data. **a** Unconditional: intercept. **b** Unconditional: intercept +. **c** Conditional: intercept only. **d** Conditional: intercept +.

Because the equations are both predicting the intercept of the level 1 model, it follows that β_{01} is testing for a difference in the level one intercept between males and females. A similar construction can be obtained from equation 7 where β_{11} is the difference in the level one slopes between males and females.

At each level, multiple explanatory variables may be used. With the use of multiple explanatory variables, the possibility exists to also include the interaction of explanatory variables as an explanatory variable. For example, a second explanatory variable to the level 2 model:

$$\text{Level 1: } y_{it} = \pi_{0i} + \pi_{1i}(Time_{it}) + e_{it} \tag{10}$$

$$\text{Level 2: } \pi_{0i} = \beta_{00} + \beta_{01}(Gender_i) + \beta_{02}(Age) \\ + \beta_{03}(Gender_i \times Age) + r_{0i} \tag{11}$$

$$\pi_{1i} = \beta_{10} + \beta_{11}(Gender_i) + \beta_{12}(Age) \\ + \beta_{13}(Gender_i \times Age) + r_{1i} \tag{12}$$

This model uses two explanatory variables (gender and age) as well as the interaction of the two variables to model the level 1 parameters.

As can be seen, there are many possible level 2 models that exist for any level 1 model. Similar to level 1, we can also introduce linear models with polynomial trends, use models nonlinear in their parameters, or add multiple explanatory variables at level 2. Also, if the situation warrants it, we can add higher levels to the model in order to explain the level 2 parameters (i.e. a level 3 or potentially a level 4 model, etc.).

Estimation of Model Parameters
After the model is specified, the levels are combined into the full model so that model parameters can be estimated. The full model is formed by substituting the highest level (i.e. level 2 in this case) regression equations into the lower level regression equa-

Table 1. Example of MLM parameters

Population fixed effects	Variance of unique effects	Correlation of unique effects	Population error variance
β_{00}	σ_{r0}^2	τ_{01}	σ_e^2
β_{01}	σ_{r01}^2		
β_{10}			
β_{11}			

tions. In the case of our example, this involves substituting the level 2 model equations (6 and 7) into the level 1 model (5). (In the case of a 3 level model, we would substitute the level 3 equations into the level 2 equations, then substitute the level 2 equations into the level 1 model, etc.). The resulting full model would look as follows:

$$y_{it} = [\beta_{00} + \beta_{01}(Gender_i) + r_{0i}] + [\beta_{10} + \beta_{11}(Gender_i) \\ + r_{1i}](Time_{it}) + e_{it} \tag{13}$$

Maximum likelihood estimation (full or restricted) is generally used to estimate the model's parameter estimates, which have the advantage of dealing with missing data in arguably the most optimal way. Estimation of the model generally assumes that errors are normally distributed and independent.

Parameter Interpretation
Once the model has been specified and fit parameter estimates, their corresponding confidence intervals, and significance levels need to be interpreted (table 1).

Statistical Significance

In general, the statistical significance of a test evaluating a model parameter can be interpreted in much the same way as in other statistical tests. Significance tests of parameters test if the parameter is significantly different from zero (or some other specified value). These parameters are the fixed effects of the full model. When interpreting the meaning of significant parameters, it is important to remember that we are estimating intercepts and slopes, not scores. Thus when we want to interpret β_{00}, it is helpful to go back to our original equations (5, 6 and 7). Looking at these equations, we remind ourselves that β_{00} is the intercept of the equation predicting π_{0i}, the intercept of our level 1 equation. Thus in this model, β_{00} provides information about initial status on our variable of interest. Similarly, β_{01} provides information about gender differences in the intercept.

Error Variance Interpretation

Significance of error terms does not have exactly the same meaning as significance of parameters. In fact, the meaning of significant error variance changes depending on the level at which the error term is located.

Random Effects. The error variance from the level 1 model corresponds to true error, or amount of variance not accounted for by the full model. This error term is implicitly involved in the significance tests of the model parameters because its variance is important for the computation of the standard errors of the fixed effects.

Unique Effects. Error terms at higher levels, however, have to do with individual and group differences. *In general*, the unique effects at level 2 (sometimes referred to as random effects) refer to individual differences, and are not literally error in the traditional sense, as they represent the deviation from the individuals' estimate (or parameter) and the estimated (or population) fixed effect. In our model (equations 6 and 7) r_{0i} and r_{1i} thus refer to the deviation between the fixed effect (either intercept or slope) for the appropriate condition and the individuals' estimated change coefficient. Significant unique effects at level 2 indicate whether or not the parameter of interest differs significantly *among individuals* (i.e. if there is variability in the unique effects). For example, if r_{0i} is the random effect in the prediction of π_{0i}, the intercept, if r_{0i} is statistically significant, it would indicate that the intercept for each individual included in our model is best modeled with their own unique intercept instead of just using a single intercept to model all individuals. In other words, there are significant differences among individuals on the intercept. Interpretation of the random effect for the slope, r_{1i}, is conducted in a similar manner. At higher levels, the unique effects often describe differences among groups instead of differences among individuals. For example, this type of situation might arise when participants are nested within different research sites or have different initial values of BP after dialysis.

As can be seen by table 1, the actual error terms are not reported in output, but rather the variance of the error terms. Many times another parameter, τ (tau), is reported which corresponds to the correlation between the random effect for the intercept and the random effect for the slope (or other correlations for the unique effects when more level 1 variables are included). Thus, τ can be interpreted as the relationship between initial status and rate of change.

Notation

Before an example is provided, it is important to discuss the customization of models. Looking at the equations written above, we see the subscript *i* on each parameter, which as mentioned before, stands for *individual*. When this subscript is placed on the parameter, the model will estimate parameters for each separate individual. However, in some cases there may be reason to believe that all individuals will change in the same way, thus individual parameters for each person would not be necessary, which is the way general linear model approaches deal with such issues for change parameters other than the intercept. This should be an empirical question, not an a priori limitation of the model. For example, research might suggest that individuals start at different places (i.e. have different intercepts) but all grow at the same rate (have the same slope, which is the repeated measures approach under the sphericity assumption). So a model could be created:

$$y_{it} = \pi_{0i} + \pi_1(Time) + e_{it}, \tag{14}$$

corresponding to the sphericity assumption discussed previously. Or if it is believed that men and women change differently, but within groups, all women show the same trajectory and all men show the same trajectory, we reduce the model as shown in equations 15, 16 and 17. Notice here that all the *i* subscripts have been removed indicating that the parameters no longer vary for each individual, but rather the group parameters are assumed representative for all individuals in the group:

$$y_{it} = \pi_0 + \pi_1(Time_{it}) + e_{it} \tag{15}$$

$$\pi_0 = \beta_{00} + \beta_{01}(Gender) \tag{16}$$

$$\pi_1 = \beta_{10} + \beta_{11}(Gender) \tag{17}$$

Model Comparisons

As demonstrated by the previous discussion, many different models of the same data can be postulated. To determine which model provides the best fit to the data, we can use model comparison techniques. When models are nested within one another (i.e. when one model can be written as a special case of another model) and use identical data, model comparisons can be executed through use of the *deviance statistic*. The deviance statistic is a statistic which most multilevel model programs will include in the output or that can be obtained indirectly from the output. The deviance statistic quantifies the fit of a model compared to the saturated model (i.e. a model that fits perfectly). Provided sample size is not small and normality of the errors hold, the difference in deviance statistics between nested models is distributed asymptotically as a χ^2 statistic with the degrees of freedom equal to the difference in the number of model parameters estimated [7].

If models are not nested, however, the deviance statistic is not an appropriate method for comparison. In these cases, the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC) statistics can be used. The AIC and BIC can be used to compare any two models, regardless of whether they are nested, as long as the data used is identical. The smaller the AIC/BIC, the better fit the model provides. The AIC and BIC simultaneously consider error and parsimony, so a model that has a smaller error term might not be considered 'better' because additional parameters were necessary to achieve that level of fit. We add a cautionary note that the use of the AIC or BIC statistics for model com-
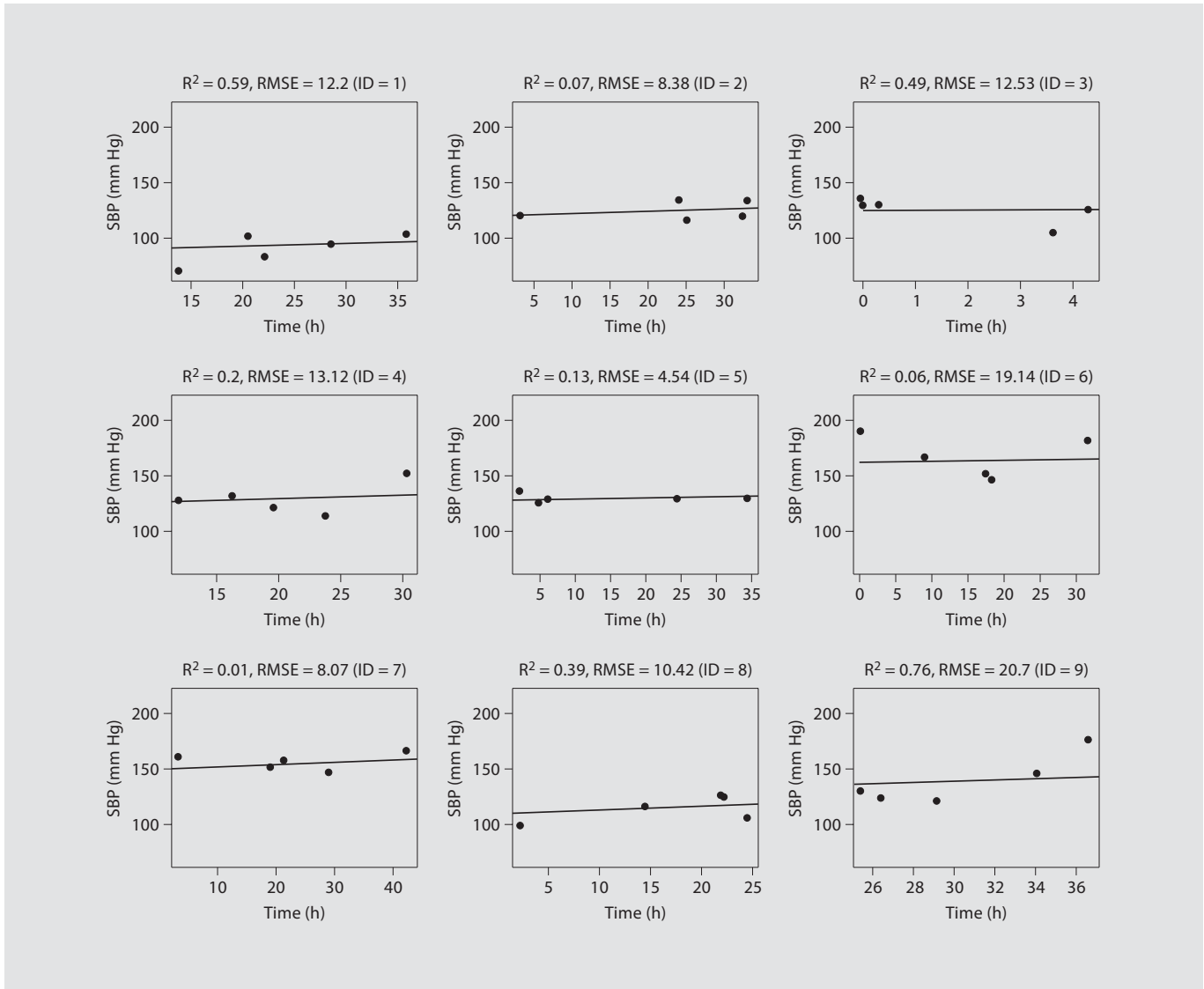
Fig. 2. Individual trajectory plots of SBP as a function of time after dialysis for the first 9 patients for five randomly selected awake measurement occasions.

parison tends to be very subjective. An added complication arises when the AIC and BIC statistics provide conflicting evidence. Thus, the AIC and BIC should generally only be used when models are not nested, given that exact procedures are available in such situations [7].

Another statistic that may be reported is the Pseudo R^2. When fitting models, it is the hope that the addition of additional explanatory variables will add to the explanation of the outcome variable. As a measure of this, a Pseudo R^2 can be calculated which represents *the proportional reduction in residual variance between two nested models*. This Pseudo R^2 statistic should not be confused with the Pseudo R^2 statistic used most often in the GLM (especially multiple regression), which has well known properties and provides a measure of the total variability explained.

The most simple model tested (in our case, this will be model 1 of our example, the no slope means only model) provides the baseline variance for comparison. Pseudo R^2 can then be calculated by

$$\text{Pseudo } R^2 = \frac{\sigma_{\varepsilon 0}^2 - \sigma_{\varepsilon M}^2}{\sigma_{\varepsilon 0}^2}, \tag{18}$$

where $\sigma_{\varepsilon 0}^2$ is the population error variance for the baseline model (e.g. intercept only) and $\sigma_{\varepsilon M}^2$ is the population error variance for the more rich (e.g. straight line change) model [6]. It should be noted that in some circumstances, it is possible for Pseudo R^2 statistic to be negative because it is possible the inclusion of an additional predictor can increase the magnitude of the variance component [6, 7]. When a negative Pseudo R^2 statistic occurs, the

Table 2. Equations of the four blood pressure models tested

Model 1	
Level 1	$y_{it} = \pi_{0i} + e_{it}$
Level 2	$\pi_{0i} = \beta_{00} + r_{0i}$
Model 2	
Level 1	$y_{it} = \pi_{0i} + \pi_{1i}(Hours_{it}) + e_{it}$
Level 2	$\pi_{0i} = \beta_{00} + r_{0i}$
	$\pi_{1i} = \beta_{10} + r_{1i}$
Model 3	
Level 1	$y_{it} = \pi_{0i} + \pi_{1i}(Hours_{it}) + e_{it}$
Level 2	$\pi_{0i} = \beta_{00} + \beta_{01}(BPMed_i) + r_{0i}$
	$\pi_{1i} = \beta_{10} + r_{1i}$
Model 4	
Level 1	$y_{it} = \pi_{0i} + \pi_{1i}(Hours_{it}) + e_{it}$
Level 2	$\pi_{0i} = \beta_{00} + \beta_{01}(BPMed_i) + r_{0i}$
	$\pi_{1i} = \beta_{10} + \beta_{11}(BPMed_i) + r_{1i}$

Pseudo R^2 is largely uninterpretable and should not generally be used [7, 8]. Similar logic can be applied in each of the level 2 (or beyond) equations, which we do not discuss here.

Empirical Example

Kelley et al. [11] fit a series of complicated change models to ambulatory BP data. Ambulatory BP monitoring was performed after the mid-week hemodialysis session for 44 h. Ambulatory BPs were recorded every 20 min during the day (6 a.m. to 10 p.m.) and every 30 min during the night (10 p.m. to 6 a.m.). For illustrative purposes, we randomly sampled five measurement occasions from each of the individuals when they were awake. An appendix has been provided that includes the SAS syntax necessary to reproduce the results given momentarily (R code is available directly from the authors). The SAS (and R) code as well as data will be available on the Internet at the following location: www.karger.com/doi/10.1159/000131102.

Figure 2 shows the SBP values for the first 9 patients as a function of time after dialysis. Close examination of the plots within figure 2 suggests that there may be a slight positive linear trend (the full data reported in Kelley et al. [11] clearly show a positive linear and oscillating trajectory). We decided to further examine SBP changes over time by comparing the results of four different MLMs. The formal equations for these models are displayed in table 2. Model 1 is the baseline intercept only model, which does not include explanatory variables (i.e. it is unconditional). Model 2 is the basic longitudinal model that includes an intercept and linear slope of hours after dialysis as an explanatory variable in the level 1 model. Thus model 2 explicitly considers time after dialysis not including any explanatory variables of the level 1 parameters. Models 3 and 4 each include an explanatory variable (whether or not BP medicine was taken) in the level 2 model (thus making it a conditional model). Model 3 uses the presence of BP medicine to explain the intercept alone, while in model 4 it is used to explain both the intercept and the slope.

Table 3 compares the results of the 4 different models. Analysis of the fit statistics, namely the AIC, BIC and χ^2 test of deviances, are all in agreement that model 3 provides the best fit to the data. Thus, the addition of time after dialysis and BP medicine as an explanatory variable for the intercept significantly added to the model. Use of BP medicine to explain the slope, however, did not yield a significant effect. The average BP for individuals not using BP medication immediately following dialysis was 110.7 (β_{00}, p < 0.001). For individuals using BP medication, average BP immediately following dialysis was 128.1 (i.e. 110.66 + 17.47; β_{01}, p < 0.001). The effect of time following dialysis was significant (β_{10}, p < 0.001) indicating for every 1 h change in time, BP increases by 0.24 units. There is a relatively small negative correlation estimated between initial status and rate of change ($\tau = -0.17$), indicating that individuals who begin with a higher BP tend to experience less change over time due to BP medication than do individuals who begin with a lower BP.

Discussion

Some researchers have already realized the benefit of MLM in nephrological research and have addressed research questions with success [12, 13]. Perhaps a more tangible reason why MLMs should be considered is because funding agencies want to ensure researchers are using appropriate research methods. All other things being equal, we believe a researcher proposing an MLM approach to analyzing change would be much more likely to obtain funding than a researcher proposing a simple change model that may well violate the models' assumptions (e.g. independence of errors and/or sphericity) and not adequately address the research questions of interest. As demonstrated, not only do MLMs provide a flexible framework which can be molded to the situation at hand,

Table 3. Results of multilevel model

	Parameter	Model			
		1	2	3	4
Fixed effects					
π_{0i}	β_{00}	129.96*** (2.02)	125.20*** (2.23)	110.66*** (4.82)	109.96*** (5.32)
	β_{01}			17.47*** (5.17)	18.31*** (5.82)
π_{1i}	β_{10}		0.24*** (0.07)	0.24*** (0.07)	0.29 (0.16)
	β_{11}				−0.06 (0.18)
Random effects					
	σ_e^2	229.25	191.60	191.60	191.55
	σ_{r0}^2	467.34	463.97	464.14	419.35
	σ_{r1}^2		0.19	0.19	0.19
	τ_{01}		−0.17	−0.17	−0.19
Fit statistics					
	deviance	5,516.12	5,479.37	5,468.39	5,468.23
	AIC	5,522.12	5,491.38	5,482.39	5,484.29
	BIC	5,535.44	5,518.05	5,513.51	5,519.85
Model comparisons			$\chi^2_{12} = 36.75$**	$\chi^2_{13} = 47.73$***	$\chi^2_{14} = 47.89$***
				$\chi^2_{23} = 10.98$***	$\chi^2_{24} = 11.14$***
					$\chi^2_{34} = 0.16$
Pseudo R^2 (level 1)			$R^2_{12} = 0.16$	$R^2_{13} = 0.16$	$R_{14} = 0.16$

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, standard error appears in parentheses.
The subscripts on the χ^2 and Pseudo R^2 statistics denote the two models being compared.

they also provide more information about change which simpler models cannot provide.

MLMing is becoming more and more common in the research community and several statistical software packages are available for this purpose. Widely used packages include HLM, Mplus, Mx, MLwiN, SAS (specifically the MIXED and NLMIXED procedures), STATA (specifically the xt set of commands), SPSS (the MIXED command), and R and Splus (specifically the lme4 and nlme packages). We have provided SAS code and the example data as a supplement to this article (www.karger.com/doi/10.1159/000131102).

In summary, although there are statistical models of change such as repeated measures ANOVA that are easier to implement and simpler to interpret, they may not be valid in complex real-world situations. Multilevel models may be needed to understand the true nature of change, whether it be interindividual, intraindividual or both.

References

1 Nesselroade JR: Interindividual differences in intraindividual change; in Collins LM, Horn JL (eds): Best Methods for the Analysis of Change: Recent Advances, Unanswered Questions, Future Directions. Washington, American Psychological Association, 1982, pp 92–105.
2 Laird NM, Ware JH: Random-effects models for longitudinal growth. Biometrics 1982;38: 963–974.
3 Myers JL, Well AD: Research Design and Statistical Analysis. Mahwah, Lawrence Erlbaum Associates, 2003.
4 Schafer JL: Analysis of Incomplete Multivariate Data. London, Chapman & Hall/CRC, 1997.

5   Schafer JL: Multiple imputation with PAN; in Collins LM, Sayer AG (eds): New Methods for the Analysis of Change. Washington, American Psychological Association, 2001, pp 357–387.

6   Hedeker D, Gibbons RD: Longitudinal Data Analysis. Hoboken, Wiley, 2006.

7   Singer JD, Willett JB: Applied Longitudinal Data Analysis. New York, Oxford University Press, 2003.

8   Raudenbush SW, Bryk AS: Hierarchical Linear Models Applications and Data Analysis Methods. Thousand Oaks, Sage Publications, 2002.

9   Pinheiro JC, Bates DM: Mixed Effects Models in S and S-Plus. New York, Springer, 2000.

10 Davidian M, Giltinan DM: Nonlinear Models for Repeated Measures Data. Boca Raton, CRC Press, 1995.

11 Kelley K, Light RP, Agarwal R: Trended cosinor change model for analyzing hemodynamic rhythm pattern in hemodialysis patients. Hypertension 2007;50:143–150.

12 Fink JC, Hsu VD, Zhan M, et al: Center effects in anemia management of dialysis patients. J Am Soc Nephrol 2007;18:646–653.

13 Wright JT Jr, Bakris G, Greene T, et al: African American Study of Kidney Disease and Hypertension Study Group. Effect of blood pressure lowering and antihypertensive drug class on progression of hypertensive kidney disease: results from the AASK trial. JAMA. 2002;288:2421–2431; erratum in: JAMA 2006;295:2726.