HIERARCHICAL LINEAR MODELS FOR THE ANALYSIS OF LONGITUDINAL DATA WITH APPLICATIONS FROM HIV/AIDS PROGRAM EVALUATION

Ann A. O'Connell
University of Connecticut
USA

In this paper, two examples of multilevel modeling as part of the analysis of data from HIV evaluation studies are presented. Strategies for teaching multilevel models for each type of data are discussed. The first, a panel study, uses multiple linear regression models to show how a hierarchical linear model can be developed. The second, a repeated cross-sectional design, uses simple analysis of variance models to show how a random coefficients model can be fit to the data. Complex multilevel models may be easier to understand and apply when broken down into these more familiar strategies. Analyses are presented using the HLM program and SAS.

Techniques for the analysis of data from longitudinal studies are one of the more challenging course content areas for multivariate methods classes. In most applied multivariate statistics courses, students are typically instructed in the analysis of repeated measures data through multivariate or univariate analysis of variance models. Analysis of variance approaches, however, can present severe limitations to the analysis of repeated measures data, due to the likelihood of unbalanced data and missing observations from longitudinal designs (e.g., Kenny, Bolger & Kashy, 2002), the treatment of time as a fixed rather than a random effect (e.g., van der Leeden, 1998), and assumptions about the nature of change across time (e.g., Wallace & Green, 2002). Advances through the past two decades or so have led to powerful and flexible models for describing data from longitudinal studies, particularly through the consideration of longitudinal data as multilevel in structure. However, availability of options for the statistical treatment of longitudinal data has not effectively translated to ease of application of these analytic strategies in practice (Mason, 1995). Particularly for future and current substantive researchers, who may be or have been successful at their required courses in statistics in graduate school but who would not classify themselves as professional statisticians, demystifying the technologically complex literature on multilevel models may improve the likelihood that analyses of data from longitudinal studies move beyond a rigid analysis of variance formulation. One way to approach this demystification is to break down these strategies into simpler and more familiar concepts and models such as multiple regression and analysis of variance. Two longitudinal examples are presented here to illustrate this approach. I focus on relatively simple models in both examples and represent rate of change as a linear trend.

DATA STRUCTURES

Longitudinal data can arise under scores of different research designs. To provide a context for some of these design characteristics from the field of AIDS prevention, consider the changes in attitudes towards condom use that are expected to occur as a result of participation in an HIV prevention program. When information is collected on the same individuals at two or more points in time, the result is repeated measures in the form of "panel data." The repeated observations of attitudes towards condom use can be seen as *nested* within each person in a study, yielding a hierarchical structure to the data. The first level of data is the *occasion* or *time* when the attitude information is collected. The second level of data is the actual person on which the data are collected, and there may be person characteristics such as gender, age, socio-economic status, or treatment group that could be used to help explain differences between people in their change across measurement occasions. This opportunity to *explain variability* in change relative to different levels of the data is an extremely useful characteristic of multilevel models. Additionally, within multilevel models it is entirely possible for individuals to vary regarding the number and spacing of time points. The ability to deal with missing occasions for some individuals and to work with designs where occasions may vary are two of the greatest benefits of the multilevel perspective (Bryk & Raudenbush, 1992; Hox, 2000; Snijders & Bosker, 1999).

Another type of data structure for which multilevel models are commonly employed arises from naturally clustered data. Evaluations of community-level interventions are one example of research studies where data are collected cross-sectionally across *communities or groups* over time, rather than following specific individuals. Persons clustered within the same community or group tend to be more similar than persons from a different community or group, a phenomenon known as the intraclass correlation (ICC) effect (Graubard & Korn, 1994; Kish, 1995; Longford, 1995). Traditional analysis of variance and multiple regression approaches ignore the impact of the ICC and grossly misrepresent the findings from clustered data sets (Bryk & Raudenbush, 1992; Hox, 2000; Kreft & deLeeuw, 1998; Murray, 1998). However, multilevel modeling allows the researcher to appropriately conduct simultaneous analysis of individuals and their community settings (Jones & Duncan, 1995).

Example Data Sets. The panel study was designed to investigate the effects of a behavioral intervention for seropositive gay males in the Northeastern United States. The intervention was based on the information-motivation-behavior theory of individual behavior change (Fisher & Fisher, 2000; Fisher, Kimble-Willcutts, Misovich, & Weinstein, 1998). Two versions of the intervention were used. Content remained the same, but the first group met multiple times over a period of eight weeks and the second group met over an entire weekend retreat. Using HLMv5 (Raudenbush, Bryk, Cheong, & Congdon, 2000) individual growth curves were fit to the data and intervention effectiveness was evaluated by change in slope of linear growth relative to individuals in a comparison condition. The repeated cross-sectional study evaluated an HIV prevention intervention for women in high-risk communities across the United States (Lauby, O'Connell, Stark, & Adams, 1999; Lauby, Smith, Stark, Person, & Adams, 2000). A cross-section of women within each community was sampled during each wave of data collection, and there were four annual waves of data collection in eight matched communities (four intervention and four comparison). Random coefficients models were fit to the data using SAS to assess the interaction of time by condition (the intervention effect).

Example 1: Longitudinal Panel Data. Self-confidence in the use of condoms is the outcome variable used for the first example. Change in individual growth was modeled as a linear trend, consistent with the availability of only three waves of data, with time measured as 0 (baseline), 3, or 6, representing number of months since baseline. The level 1 model includes time as the only predictor. In the example, no time-varying covariates are introduced at level one, but if necessary they could easily be incorporated into the model as additional independent variables. The level 1 model is: $Y_{ti} = \pi_{0i} + \pi_{1i}T_{ti} + e_{ti}$.

By convention, within person effects are indicated by the symbol π. Y_{ti} represents the self-confidence outcome for individual *i* measured at time *t*. T_{ti} represents the variable measurement occasion in terms of number of months from the baseline assessment for person *i* (0 is used as the baseline time point). The slope, π_{1i} , is the growth rate or growth parameter for the i^{th} person. This person-specific parameter represents the linear change or growth in self-confidence. The intercept, π_{0i}, represents the expected self-confidence rating of the person at baseline ($T_{ti} = 0$), also called initial status (Raudenbush & Bryk, 2002). The within-person residuals, e_{ti}, are assumed $N(0, \sigma^2)$.

At level 2, the goal is to investigate variations in the estimates of initial status for self-confidence (π_{0i}) and in the linear growth parameter (π_{1i}) attributable to intervention group (eightweek, weekend, or comparison group). If, for example, change occurs faster in an intervention group, the slopes for people in that group should be larger than in the comparison group. Similar to ordinary least squares (OLS) linear regression, dummy codes are created to represent group membership and are used as predictors of the level 1 effects. The between-subjects models are:

$$\pi_{0i} = \beta_{00} + \beta_{01}(D1)_i + \beta_{02}(D2)_i + r_{0i}$$
$$\pi_{1i} = \beta_{10} + \beta_{11}(D1)_i + \beta_{12}(D2)_i + r_{1i}$$

D1 and D2 represent the dummy codes used in the analysis with the comparison group as the referent. D1 takes on the value of 1 for the eightweek group; D2 takes on the value of 1 for the weekend group. Accordingly, β_{00} and β_{10} represent the expected baseline and slope, respectively, for the comparison group (when D1 and D2 are both 0). The coefficients for D1 and D2 indicate how much these expected values increase or decrease for each respective treatment group. There are two random effects at level 2, r_{0i} and r_{1i} , with variances τ_{00} and τ_{11}, respectively, and a covariance τ_{01}. The covariance reflects the relationship between initial status and rate of change. For a linear growth curve, these variances and the covariance can be used to find the correlation between initial status and rate of change. In general, a negative correlation suggests that low initial status would tend to be associated with larger slopes (stronger gains over time), and high initial status tends to be associated with less change. Intuitively this makes sense; individuals starting off as very confident in their use of condoms have less room to improve dramatically over time. There are some interesting conceptual similarities between the level 1 model and an ordinary least squares regression model of the outcome on time as the only predictor. Assuming there are 139 subjects each having three occasions of measurement, there could be 139 regression models based on three observations each. The intercepts from each of these 139 regressions represent the expected baseline response or starting point for that individual, and the slopes represent that individual's expected linear change in self-confidence for condom use as time increases by 1 unit (1 month). For example, the first person in the comparison group has an individual linear growth equation with an intercept of 3.939 and a slope of -.061. The first person in the eightweek group has an intercept of 3.648 and a slope of +.024.

Given these 139 regressions, with different intercepts and outcomes for individuals, it would be interesting to see whether or not these differences can be statistically accounted for by the intervention group; in essence, this would illustrate an intervention effect (threats to validity not withstanding). One approach to determining if the interventions made a difference relative to change in the comparison group would be to average the slope estimates for each group and test for differences across the groups. Conceptually, this is similar to subjecting the collection of level 1 slopes to a multiple linear regression, with dummy codes for group membership used to test for differences across the groups. Group differences at baseline could be investigated in a similar fashion. Two problems arise however. The first pertains to how the level 1 coefficients are estimated, and the second issue is how to use these level 1 estimates to assess intervention differences. The process of averaging the collection of slopes or intercepts assumes that they were estimated with equal reliability for all participants. This, however, is not the case, since some participants did not complete all survey administrations. While comparing "average" intercepts and slopes across the groups conceptually makes sense, better estimation and testing methods are available. Details on estimation principles may be found in more theoretical expositions of multilevel modeling (e.g. Bryk & Raudenbush, 1992; Raudenbush & Bryk, 2002; Snijders & Bosker, 1999), and I mention them only briefly here. First, rather than using OLS estimates of the level 1 intercepts and slopes in equation (1), which are free to vary among individuals, estimates of these coefficients are determined using empirical Bayes (EB). An empirical Bayes estimate for a participant, or unit, "utilizes not only data from that unit but also data from all other similar units" (Raudenbush & Bryk, 2002, p. 66). This estimation strategy allows for optimal estimates of individual intercepts and slopes given the presence of incomplete data. Second, once the EB estimates for the level 1 coefficients have been determined, we can return to our earlier question regarding differences in "average" intercepts or slopes between the intervention groups. Again, OLS does not offer the best estimates. Instead, generalized least squares (GLS) estimates are used, which are precision-weighted estimates taking into account the differential precision expected from contributions of different units. Finally, the residual variance at level 1, the variances of the random effects at level 2, and covariances between the level 2 random effects are determined through maximum likelihood (ML) techniques.

Example 1 Results: Panel Data. Table 1 provides the results. The first analysis is an unconditional model, where intervention effects are not assessed, only variability in intercepts and slopes is determined. The second analysis models the level 1 intercepts and slopes using two dummy codes to indicate group belongingness. Results show that when the intervention effect is not considered (unconditional model), the slopes seem to be flat on average ($\square_{10}$ = .014, *p*=.102), indicating close to zero slope and consequently no linear trend over time. However, considerable variability exists in the individual slopes across time ($\square_{11}$= .00089, *p*=.079). Initial status is high ($\square_{00}$ = 4.11, *p*=.000), but there is also considerable variation in initial status estimates of confidence for condom use ($\square_{00}$ = .17906, *p*=.000).

We now turn to the analysis designed to determine if treatment group can account for some of the variation in the slopes and initial status estimates. For the level two models conditioning the slope and intercept estimates on treatment group, two dummy coded variables are included as predictors (D1=eightweek group versus comparison, D2=weekend group versus comparison). The correlation between initial status and rate of linear growth was near zero (*r*=.046). The considerable variability in initial status could not be accounted for by group differences (τ_{00} = .17814, *p*=.000), with only .5% of the variability in initial status explained by treatment group ((.17906 -.17814)/.17906)). However, treatment group accounts for 34.83% of the variability in linear growth rates ((.00089 - .00058)/.00089). Relatively little variation in the time effect remains to be explained by other potential variables for this sample of individuals (τ_{11} = .00058, *p*=.127). The linear trend in the comparison group was positive but weak (β_{10} = .026, *p*=.120). The eightweek group showed slightly stronger improvement over the slope in the comparison group (for D1, β_{11} = .021, *p*=.438). Relative to the trend in the comparison group, the weekend group showed a decrease in rate of change (β_{12} = -.030, *p*=.130). The weekend group, however, had a high and non-variable self-confidence level on average across the three occasions (means for this group are: 4.14, 4.13, 4.12).

Table 1 *HLM Results for Confidence in Behavioral Skills*

Level 1 parameters and their predictors	Unconditional Model	Conditional Model (treatment group as predictor)
Model for Intercepts (π_{0i})		
β_{00}	4.11** (*p*=.000)	4.19** (*p*=.000)
β_{01} (D1)	-----	-.210 (*p*=.118)
β_{02} (D2)	-----	-.058 (*p*=.575)
Variance (τ_{00})	.17906** (p=.000)	.17814** (*p*=.000)
Model for Slopes (π_{1i})		
β_{10}	.014 (*p*=.102)	.026 (*p*=.120)
β_{11} (D1)	-----	.021 (*p*=.438)
β_{12} (D2)	-----	-.030 (*p*=.130)
Variance (τ_{11})	.00089* (*p*=.079)	.00058 (*p*=.127)
Residual variance σ^2	.34633	.11889

** *p*<.01, * *p*<.10

Example 2: Repeated Cross-Sectional Data. Perceived advantages for consistent condom use ("pros") with one's main partner is the variable analyzed in this example. Recall that in the panel study, the data were treated as nested within the individual. In a repeated cross-sectional survey, different individuals are interviewed from within the same communities over time; the resulting data are nested within each community. This structure compels the researcher to consider an analysis where (1) intercepts (baseline characteristics) and slopes (time effect) may vary between communities; and (2) the positive intraclass correlation expected in the data due to ecological or geographical similarities of communities is accounted for in the determination of the intervention effect (time X condition interaction). In this research situation, the random coefficients model provides the most plausible representation of the data, since the data are naturally clustered within communities (Longford, 1995; Murray, 1998).

Understanding the model used to address these issues can be accomplished by first considering the situation for one city. Within one city, two areas are designated as either intervention or comparison communities: the condition effect. Four waves of data are collected from each of these groups or communities: the time effect. The simplest model to fit to the data represents time as a linear trend; such a model is reasonable when only one baseline survey is collected within each group or community (Murray, 1998). As a result, the problem is a straightforward analysis of variance design: the outcome Y=perceived advantages of condom use with one's main partner; Factor A is "condition" with two levels, factor B is "time" with 4 levels, and the interaction effect is also the intervention effect, "time by condition." If the intervention is successful at improving attitudes towards condom use, we would expect to see a departure from parallel slopes between conditions, and observe a greater linear slope estimate for the intervention condition. The model is familiar as the factorial analysis of variance model: $Y_{ijk} = \mu + \alpha_j + \beta_k + (\alpha\beta)_{jk} + \varepsilon_{ijk}$. We can improve on the representation of parameters in this model by using C_j to represent condition (0=comparison group, 1=treatment group), and by writing time as a fixed factor with a linear trend: $Y_{ijk} = \mu + C_j + T_{lin}t_k + (CT_{lin}t)_{jk} + \varepsilon_{ijk}$. In these equations, Y_{ijk} represents the ith person's attitude measure in the jth condition at the kth time point. Time is measured as 0, 1, 2, or 3, representing time in years from baseline. These models can be easily fit and compared for each of the four cities in the data set.

However, one of the primary goals of this evaluation study was to determine if the intervention was effective across all the cities included in this study, with the desire to be able to generalize to other communities similar to the ones included in the intervention. Therefore, a model was developed to include group (or city) as a random effect ($\mathbf{G}_{k:l}$) in an expanded prediction model. Structurally, the data collected is treated as nested within each city. Using SAS Proc Mixed, the following (general) linear random coefficients model was fit to the data: $Y_{i:jk:l} = \mu + C_l + T_{lin}t_j + T_{lin}t_jC_l + \mathbf{G}_{k:l} + T_{lin}t_j\,\mathbf{G}_{k:l} + \Sigma\beta_0(X_{0i:jk:l} - \overline{X}_{0\ldots}) + \varepsilon_{i:jk:l}$. Random effects are indicated in bold print. The outcome represents the perceived advantages of condom use for the ith person nested within the jth time period in the kth group within the lth community. The random coefficients model allows the intercepts and slopes to vary between level 2 units, the communities, but no attempt is made to explain this variability (Raudenbush & Bryk, 2002). In this model, heterogeneity among the group specific slopes and intercepts are represented by the random effects: $\mathbf{G}_{k:l}$ and $T_{lin}t_j\,\mathbf{G}_{k:l}$. With the clustering effect inherent in the data, failure to include these random effects in the model (as if all individuals in the same condition, regardless of city, were combined and a factorial analysis run to compare time trend across the two conditions), would seriously inflate the type I error and misrepresent the findings. A propensity score adjustment was used in these analyses. This is a composite covariate used to adjust for demographic differences related to selection effects. Variance components are determined for those sources of variation in the model that involve a random factor. These include the variability among the group:condition specific intercepts ($\sigma^2_{g:c}$), the variation among the group:condition specific slopes ($\sigma^2_{t(lin)g:c}$), and variation among members (error variance) within time by group:condition ($\sigma^2_{m:jk:l}$). The random coefficients analysis also provides an estimate of the covariance (if any) between the slopes and the intercepts.

Example 2 Results: Repeated Cross Sectional Data. The random coefficients model allows each community to have its own slope and intercept. Results of the analysis are shown in Table 2.

Table 2 *Linear Random Coefficients Model for Perceived Advantages of Condom Use with Main Partner*

RANDOM EFFECTS				
Variance Covariance Matrix				
	Intercepts	Slopes		
Intercepts	.0053	-.0032		
Slopes	-.0032	.0036		
Individual level random effects: $\sigma^2_{m:jk:l} = \sigma^2_e = .4729$				
FIXED EFFECTS				
Source	NDF	DDF	Type III F	p
Cond	1	6	0.97	.3616
Year	1	6	3.05	.1313
CxYear	1	6	0.89	.4056
Segment	4	6405	5.16	.0004

The effect of interest is the intervention effect and is represented by the Time by Condition term in the model. For the pros of condom use with a main partner, the findings were not as strong as expected. Encouraging women to use condoms with their main partner, even when that partner may be putting them at risk for HIV, presents an enormous challenge to intervention researchers, so sensitive tests of intervention effects are very important. Although repeated cross-sectional designs are becoming more common in the research community, particularly in understanding health promotion efforts (Koepsell, et al., 1992; Murray, 1998; von Korff, Koepsell, Curry & Diehr, 1992;), our data were fairly limited due to the short time series and the small number of communities included in the study. Increasing the number of communities, or lengthening the series in which data are collected, can enhance the ability to detect intervention effects across varying groups.

CONCLUSIONS

Although multilevel analyses can be quite complex, I have tried here through simplified analyses to focus on their relationships to familiar and well-understood analytic tools, such as multiple regression analysis and analysis of variance, which was the primary goal of this paper. Instructors of multilevel techniques can begin to demystify some of the complexity by constantly referring back to these less cumbersome analyses. While no analysis can fully compensate for all of the design issues affecting the two studies presented here, models that truly reflect the structure of the data are clearly going to offer the most useful information as we continue to seek best practice in health promotion efforts and in the analysis of change. While there is much more to the world of multilevel modeling than I could possibly cover here, I hope that the approach laid out provides a helpful starting point for those teaching others or working with multilevel models themselves. In terms of learning how best to improve health or decrease risk from disease for individuals or communities, the challenges of becoming comfortable with multilevel modeling strategies may at first appear daunting, but the benefits of their use may surely be even greater.

ACKNOWLEDGEMENT

The research presented here was supported in part by AIDS Project Hartford, Hartford, Connecticut, USA. The community-level intervention was funded through the Centers for Disease Control and Prevention, Behavioral Interventions Research Branch, Atlanta, GA.

REFERENCES

Bryk, A.S., & Raudenbush, S.W. (1992). *Hierarchical linear models: Applications and data analysis methods*. CA: Sage.

Fisher, J.D., & Fisher, W.A. (2000). Theoretical approaches to individual-level change in HIV risk behavior. In Peterson and DiClemente (Eds.), *Handbook of HIV prevention* (pp. 3-55). New York: Plenum.

Fisher, J.D., Kimble-Willcutts, D.L., Misovich, S.J., & Weinstein, B. (1998). Dynamics of sexual risk behavior in HIV-infected men who have sex with men. *AIDS and Behavior, 2*(2), 101-113.

Graubaud, B.I. & Korn, E.L. (1994). Regression analysis with clustered data. *Statistics in Medicine, 13*, 509-522.

Hox, J.J. (2000). Multilevel analyses of grouped and longitudinal data. In T.D Little, K.U. Schnabel, and J.Baumert (Eds.), *Modeling longitudinal and multilevel data* (p. 15-32). Mahwah, NJ: Lawrence Erlbaum.

Jones, K., & Duncan, C. (1995). Individuals and their ecologies: Analysing the geography of chronic illness within a multilevel modelling framework. *Health and Place, 1*(1), 27-40.

Kenny, D.A., Bolger, N., & Kashy, D.A. (2002). Traditional methods for estimating multilevel models. In D.S. Moskowitz & S.L. Hershberger (Eds.), *Modeling intraindividual variability with repeated measures data: methods and applications* (pp. 1-24). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Kish, L. (1995, 1965). *Survey Sampling*. New York, NY: John Wiley and Sons.

Koepsell, T.D., et al., (1992). Selected methodological issues in evaluating community-based health promotion and disease prevention programs. *Annual Reviews in Public Health, 13*, 31-57.

Kreft, I. & de Leeuw, J. (1998). *Introducing Multilevel Modeling*. London: Sage.

Korn, E.L. & Graubard, B. I. (1999). *Analysis of health surveys*. New York: John Wiley.

Lauby, J., O'Connell, A. A., Stark, M., and Adams, J. (1999). *Analysis of cross-sectional surveys to evaluate a community-level HIV prevention intervention for women*. Presented at the National HIV Prevention Conference. Atlanta.

Lauby, J.L, et al., (2000). A community-level HIV prevention intervention for inner city women: Results of the women and infants demonstration trial. *American Journal of Public Health, 90*(2), 216-222.

Longford, N.T. (1995). Random coefficient models. In G. Arminger et al. (Eds.), *Handbook of statistical modeling for the social and behavioral sciences* (pp. 519-577). New York: Plenum Press.

Mason, W.M. (1995). Comment. *Journal of Educational and Behavioral Statistics, 20*, 221-227.

Murray, D.M. (1998). *Design and analysis of group-randomized trials*. New York: Oxford University Press.

Raudenbush, S.W., & Bryk, A.S. (2002). *Hierarchical linear models: applications and data analysis methods*. CA: Sage.

Raudenbush, S., et al. (2000). *HLM 5: hierarchical linear and nonlinear modeling*. Lincolnwood: SSI.

Shadish, W. R., Cook, T.D., & Campbell, D.T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin Company.

Snijders, T. & Bosker, R. (1999). Multilevel Analysis. London: Sage.

Van der Leeden, R. (1998). Multilevel analysis of longitudinal data. In C. Bijleveld and L. Van der Kamp (Eds.), *Longitudinal data analysis: Designs, models, and methods* (pp. 269-317). London: Sage.

Von Korff, M., Koepsell, T., Curry, S. & Diehr, P. (1992). Multi-level analysis in epidemiological research on health behaviors and outcomes. *American Journal of Epidemiology, 135*(10), 1077-1082.

Wallace, D., & Green, S.B. (2002). Analysis of repeated measures designs with linear mixed models. In D.S. Moskowitz & S.L. Hershberger (Eds.), *Modeling intraindividual variability with repeated measures data: methods and applications* (pp. 103-134). Mahwah, New Jersey: Lawrence Erlbaum Associates.