Rogosa, D. R. (1995). Myths and methods: "Myths about longitudinal research" plus supplemental questions. In J. Gottman (Ed.), *The Analysis of Change*, (pp. 2–65). Mahwah, NJ: Lawrence Erlbaum Associates.

Rogosa, D. R. & Willett, J. B. (1985). Understanding correlates of change by modeling individual differences in growth. *Psychometrika, 50*, 203–228.

Schulenberg, J., O'Malley, P. M., Bachman, J. G., Wadsworth, K. N., & Johnston, L. (1996). Getting drunk and growing up: Trajectories of frequent binge drinking during the transition to young adulthood. *Journal of Studies on Alcohol, 57*, 289–304.

Seltzer, M., & Raudenbush, S. (1997). The application of hierarchical linear models in developmental research. In P. J. Curran (Chair), *Comparing three modern approaches to longitudinal data analysis: An examination of a single developmental sample*. Symposium conducted at the 1997 meeting of the Society for Research on Child Development, Washington, DC.

Sher, K. J. (1991). *Children of alcoholics: A critical appraisal of theory and research*. Chicago, IL: University of Chicago Press.

Stoolmiller, M. (1994). Antisocial behavior, delinquent peer association, and unsupervised wandering for boys: Growth and change from childhood to early adolescence. *Multivariate Behavioral Research, 29*, 263–288.

Swaim, R. C., Oetting, E. R., Edwards, R. W., & Beauvais, F. (1989). Links from emotional distress to adolescent drug use: A path model. *Journal of Consulting and Clinical Psychology, 57*, 227–231.

Tucker, L. R. (1958). Determination of parameters of a functional relation by factor analysis. *Psychometrika, 23*, 19–23.

Willett, J. B. (1988). Measuring change: The difference score and beyond. In H. J. Walberg & G. D. Haertel (Eds.), *The international encyclopedia of educational evaluation*. Oxford, England: Pergamon Press.

Willett, J. B., & Sayer, A. G. (1994). Using covariance structure analysis to detect correlates and predictors of change. *Psychological Bulletin, 116*, 363–381.

# 2

# Longitudinal Data on Families: Growth Modeling Alternatives

Siek-Toon Khoo
*Arizona State University*

Bengt Muthén
*University of California, Los Angeles*

*Family data are often collected to study influences that occur within the context of home environments. Longitudinal family data are important for investigating and understanding adolescent development. The purpose of this chapter is to demonstrate a number of possible alternative methods for analyzing longitudinal family data with multiple siblings per household. Longitudinal family data is an example of hierarchical data with repeated observations over time nested within individuals and individuals nested within families. The repeated measurements offer opportunities for longitudinal modeling to study individual differences in the development. Two approaches to analyzing family data are discussed. The first is the multilevel approach that treats the siblings within families as statistically equivalent. The longitudinal family data are analyzed in a three-level model. The second approach is the multivariate approach that analyzes the longitudinal family data in a multivariate longitudinal model, taking the family as the unit of analysis and allowing multiple growth processes per family. Latent variable modeling is used for the second approach. It offers a flexible multivariate framework that is richer and presents more opportunities for the study of developmental growth and interrelationships of developments between family members. To illustrate the methodology, we use the subset of heavy drinking data spanning 15 years for 1,614 two-sibling families from the National Longitudinal Survey of Youth (NLSY). Individual and family background variables are included in the analyses as predictor variables.*

Longitudinal data are useful for studying individual differences in development. Family data with multiple siblings are often collected for genetic studies or for studies that investigate influences that occur within the context of home environments. Data obtained from ob-



#79

serving families over time are important for investigating and understanding adolescent development. Here, the interest is in modeling familial resemblance, interactions within families and influences of the family environment and family characteristics. The purpose of this chapter is to demonstrate a number of possible alternative methods for analyzing longitudinal family data with multiple siblings per household.

When individual data are collected from surveys that sample organizational units or naturally occurring clusters rather than individuals, the individuals are nested and observed within the larger units. This type of data has a hierarchical structure. For example, we may have students observed within classrooms, family members within households, or residents within neighborhoods. Observations and measurements may be made at each of the different levels of the hierarchy. In the case of sibling data, there are variables describing the individual siblings and there are variables describing the family. Since the individuals within the same larger unit share values on the higher level variables, observations on individuals within the larger units or clusters may be correlated. The nonindependence of these observations, if ignored, may give standard errors of parameter estimates that are underestimated if there is a substantial correlation between the observations. Aggregating and averaging the individual variables to the higher level and doing the analysis at the higher level would be throwing away valuable individual-level information. More importantly, ignoring the hierarchical structure may mean missing the opportunities to model the effects occurring at each level and to explicitly formulate and test hypotheses pertaining to each level.

In developmental research where repeated observations are made over time on a group of individuals, the multiple observations per individual can also be viewed as nested within persons. The repeated measurements offer opportunities for longitudinal modeling to study developmental trajectories and for describing individual differences in development over time. In longitudinal family data, we have repeated observations nested within individuals who, in turn, are nested within families. This structure constitutes three levels of hierarchical data.

In analyzing longitudinal family data with multiple siblings, there are several areas of interest. The first concerns the interest in describing individual development and change over time and the relationships between the developmental trajectories and the individual characteristics and experiences. The second concerns the study of how shared family environment and characteristics may influence the developmental trajectories. The third concerns the study of interrelationships of developments between siblings and sibling influences. For example, it is likely that the developmental process of an older sibling may affect that of a younger sibling. The fourth concerns the study of how family variables interact with individual characteristics in the developmental process. For example, ethnicity may interact with sibling order in its influence on a developmental process, with the older sibling being expected to take on more responsibilities in a certain culture. In order to answer these research questions efficiently, there is a need to take advantage of the newer methods and software now available.



Two approaches to analyzing family data are discussed in this chapter. The first is the multilevel approach that considers the siblings to be nested within families and treats the siblings within a family as statistically equivalent. This means that they behave the same statistically, sharing the same means and variances on the same set of variables. This approach models the longitudinal family data in a three-level hierarchical model with a separate model at each level. The second is the multivariate latent growth approach that takes the family as the unit of analysis. This approach considers the siblings to be different by sibling order and allows multiple growth processes per family in a multivariate context. Each sibling in the family has his or her own set of variables that can have different means and different variances. This approach takes advantage of the small cluster sizes of families. The within-family dependence is modeled within the family unit in the multivariate model.

## The Multilevel Approach

Recent advances in the analysis of hierarchical data draw on statistical techniques that have been given several different terms, for example, *random effects models*, *mixed effects models*, and *hierarchical linear models* (Bock, 1989; Bryk & Raudenbush, 1987, 1992; Goldstein, 1986, 1995; Laird & Ware, 1982; Strenio, Weisberg, & Bryk, 1983). Random effects modeling offers integrated and flexible approaches for the analysis of hierarchical data in hierarchical linear models or multilevel models. The analysis of longitudinal family data using multilevel modeling would call

for a three-level hierarchical model. These models can be estimated using readily available multilevel analysis software such as HLM (Bryk, Raudenbush, & Congdon, 1996), MLn (Woodhouse, Rabash, Goldstein, Yang, & Plewis, 1996) and SAS PROC MIXED (SAS Institute, Inc.,1995).

In the multilevel framework, a growth model is specified as a two-level model where the Level-1 units are the repeated measurements that are nested within the Level-2 units of individuals. The Level-1 model gives the within-subject model that represents individual growth trajectories. The Level-2 model gives the between-subject model that represents the variation of growth across individuals. When individuals are also nested in families, a Level-3 model can be formulated and tested to study variation across families and family-level influence. These multilevel models are discussed later.

## The Multivariate Approach

Family data can also be analyzed taking the family as the unit of analysis in multivariate models without throwing away individual information. The family data consist of data from each member of the family (allowing responses from each member to be represented by separate variables) and also shared family information. Cross-sectional family data have often been analyzed this way, especially in genetic analyses, in which researchers seek to model family resemblance. Neale, Walters, Eaves, Maes, and Kendler (1994) analyzed fears using multivariate models based on data from the twin pair and their parents. Pike, McGuire, Hetherington, Reiss, and Plomin (1996) studied family environment and adolescent depressive symptoms and antisocial behavior in a multivariate model using data from a sample of two-parent families with same-sex sibling pairs. These multivariate models are usually implemented in the latent variable framework using structural equation modeling software. Recent advances in latent growth curve modeling make it possible to extend this method to longitudinal family data.

Use of random effects models for growth modeling has been extended and applied in the context of latent variable modeling (McArdle & Epstein, 1987; Meredith & Tisak, 1984, 1990; Muthén, 1991, 1993, 1996; Muthén & Khoo, 1998; Willett & Sayer, 1994). This development lends powerful tools to the study of developmental



changes (see also the chapter by Curran in this volume). These latent growth curve models can be implemented using structural equation modeling software in a multivariate model instead of in a multilevel model. The combination of latent growth curve modeling and structural equation modeling offers great flexibility in dealing with multivariate outcomes, multiple processes, measurement models in the covariates, and modeling of mediational effects (Muthén & Curran, 1997).

Taking family as the unit of analysis, longitudinal family data can be analyzed using multivariate models with multiple growth trajectories, each corresponding to a sibling in the family. This type of analysis can be carried out in either the multilevel framework or the multivariate latent variable framework. In the multilevel framework, the multilevel context is only applied to the modeling of growth. At level 1 is the within-family measurement model with multiple sets of repeated measurements for each sibling (multivariate) and at level 2 is the between-family model in a two-level model. This technique is illustrated in Raudenbush, Brennan, and Barnett (1995) in a multivariate hierarchical model for studying psychological change within married couples. This was specified in a two-level model with two growth functions by using indicator variables for male and female. The data structure is seen as *repeated observations nested within families.* This technique combines growth modeling with cross-sectional methods for the study of matched pairs. A more natural and flexible framework for this type of analysis is the multivariate latent variable framework, mentioned in the last paragraph, that combines growth modeling with the flexibility of structural equation modeling in multivariate latent growth models.

In the multivariate latent variable framework, the growth model with multiple trajectories per family can be specified using the multiple sets of repeated multivariate outcomes. The trajectory of each sibling is modeled using his or her repeated observations over time. This is the measurement part of the model. The trajectories are then related to the individual and family predictor variables. This is the structural part of the model. The latent variable framework offers flexibility for studying developmental relationships in the structural part of the model, for example, in the modeling of mediational effects and the adding of measurement models in the predictor variables. Later, we describe the second approach in several multivariate latent growth models with

multiple processes to demonstrate possible ways of investigating developmental relationships between family members.

Besides the multilevel approach and the multivariate approach described, there are other approaches that can also be used for the analysis of longitudinal family data. For example, there is the multilevel latent growth approach (Muthén, 1997); which models the hierarchical nature of the data in a multilevel covariance framework. This model is very similar to the three-level multilevel approach. There are also the McArdle (1988) factor-of-curves and curve-of-factors methods that model the family curve. These are also multivariate latent growth models. Duncan and Duncan (1996) demonstrated these methods with substance use data.

The chapter by Muthén in this volume analyzes heavy drinking data from the National Longitudinal Survey of Youth (NLSY). The analyses on longitudinal family data in this chapter uses a subset of the same heavy drinking data from the NLSY. For illustrative purpose of this chapter, only the two-sibling families are included. The aim is to study the relationship in the development of heavy drinking between the older sibling and the younger sibling. This method can be generalized to families with more than two siblings. The subset of data used and the data structure are described in the next section.

## DATA AND GROWTH FUNCTIONS

The NLSY is a nationally representative household survey of young adults living in the United States in 1979. The sample includes every age-eligible household member. The two-sibling families from the NLSY data set are used to illustrate the models in this chapter. Table 2.1 shows the sibling ages in the 1,614 two-sibling families.

The ages shown are their ages in the year 1982. There are eight birth-year cohorts from 1957 to 1964. The age difference of most sibling pairs is about 2 years. Only about 16% of the families have siblings who are more than 3 years in age difference. There are longitudinal alcohol measures available covering ages 18 to 37. This chapter analyzes the frequency of heavy drinking variable covering ages 18 to 32. This variable is obtained from answers to the questionnaire item "How often have you had 6 or more drinks on one occasion during the last 30 days?" There are 7 response categories, from 0 to 6 (the categories are given in

TABLE 2.1
Sibling Ages in 1,614 Two-Sibling Families (Frequency Tables)

*Frequency of Siblings by Age*

| | *Older sibling* | | *Younger sibling* | | |
|---|---|---|---|---|---|
| *AGE1* | *Frequency* | *Percent* | *AGE2* | *Frequency* | *Percent* |
| 18 | 6 | 0.4 | 18 | 364 | 22.6 |
| 19 | 88 | 5.5 | 19 | 395 | 24.5 |
| 20 | 225 | 13.9 | 20 | 287 | 17.8 |
| 21 | 274 | 17.0 | 21 | 204 | 12.6 |
| 22 | 277 | 17.2 | 22 | 169 | 10.5 |
| 23 | 267 | 16.5 | 23 | 109 | 6.8 |
| 24 | 241 | 14.9 | 24 | 56 | 3.5 |
| 25 | 236 | 14.6 | 25 | 30 | 1.9 |

*Frequency of Sibling-pairs by Age*

| | *Age 2* | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *Age 1* | *18* | *19* | *20* | *21* | *22* | *23* | *24* | *25* | *Total* |
| 18 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 6 |
| 19 | 80 | 8 | 0 | 0 | 0 | 0 | 0 | 0 | 88 |
| 20 | 122 | 96 | 7 | 0 | 0 | 0 | 0 | 0 | 225 |
| 21 | 66 | 122 | 82 | 4 | 0 | 0 | 0 | 0 | 274 |
| 22 | 47 | 80 | 90 | 52 | 8 | 0 | 0 | 0 | 277 |
| 23 | 23 | 41 | 58 | 74 | 54 | 17 | 0 | 0 | 267 |
| 24 | 14 | 30 | 35 | 46 | 61 | 37 | 18 | 0 | 241 |
| 25 | 6 | 18 | 15 | 28 | 46 | 55 | 38 | 30 | 236 |
| Total | 364 | 395 | 287 | 204 | 169 | 109 | 56 | 30 | 1614 |

Table 2.2). For the illustrative purposes of this chapter, the outcome is treated as a continuous variable.

Some background variables describing individual characteristics and family characteristics are included in the analyses to demonstrate how these variables can be included in the models to study their influences on the outcome at individual level and at family level. The individual variables included are gender, early start (early onset of drinking), and dropping out of high school. The family variables included are ethnicity variables and family history of alcoholism variables. The variables are considered family-level variables if the siblings logically share the same value.

The early start variable, which represents early onset of drinking, is obtained from answers to the question "How old were you when you first started drinking, for example, having two or more drinks a week?" The early start variable is coded as 1 if the starting age is 14 or younger, and coded as 0 otherwise. The family history variables are obtained from answers to the question "Have any of your relatives listed on this card been alcoholics or problem drinkers at any time in their lives?" Three family history variables are constructed from the answers, based on whether alcoholism occurs only among first-order relatives (FH1), second or third-order relatives (FH23) or both first-order and second/third-order relatives (FH123).

The outcome variable and the background variables are summarized in Table 2.2.

TABLE 2.2

Descriptions of Variables

| | | |
|---|---|---|
| Heavy drinking | (HD, 0–6) | never (0), once (1), 2 or 3 times (2), 4 or 5 times (3), 6 or 7 times (4), 8 or 9 times (5), 10 or more times (6) |
| Gender | (Male,1/0) | male (1), female (0) |
| Early start | (ES,1/0) | 1 if onset on or before 14, 0 otherwise |
| High school dropout | (HSdrp,1/0) | 1 if not having completed high school by age 22 |
| Ethnicity | (Black,1/0) | Black (1), non-Black (0) |
| | (Hisp,1/0) | Hispanic (1), non-Hispanic (0) |
| Family history | (FH123,1/0) | 1 if family history among 1st and 2nd/3rd-degree relatives |
| | (FH1,1/0) | 1 if family history among 1st-degree relatives only |
| | (FH23,1/0) | 1 if family history among 2nd/3rd-degree relatives only |



The observed sample means of heavy drinking frequencies for both the older and the younger siblings over time are shown in Fig. 2.1.

It can be seen that the means follow a nonsymmetric trajectory that increases at a faster rate than it decreases. This means that a standard quadratic growth function will not do well in terms of model fit. The growth analyses of the same heavy drinking outcome in the Muthén chapter estimated some of the time steps to fit a very flexible functional form. In this chapter, a logarithmic transformation of the time scale using fixed time-steps is carried out to obtain the necessary growth function. The transformed time scale $\lambda_t$ is obtained by the following transformation:

$$\lambda_t = 2 * [\log_e(\text{Age}_t - 16) - \log_e(21 - 16)], \; t = 1,..,15 \qquad (2.1)$$

where $\text{Age}_t$ is the age of an individual at a particular time of measurement $t$. Preliminary growth analysis using this transformation had shown a good model fit to the estimated mean trajectories for both the older

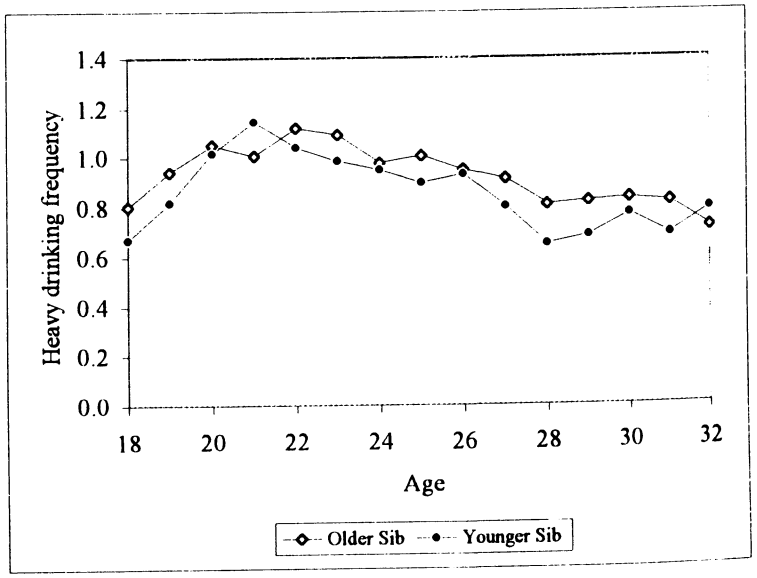


FIG. 2.1. Heavy drinking frequency.

and the younger siblings. The same logarithmic transformation of the time scale is used in both the multilevel modeling next, and the multivariate longitudinal modeling in later sections.

The growth function for the heavy drinking outcome variable in this chapter is specified as a quadratic function of $\lambda_t$. Since $\lambda_t = 0$ when $Age_t = 21$, the intercept of the growth function is defined at age 21. The growth function is explained further in the next sections.

Taking age as a basis for the time scale instead of using the measurement occasion means that there is a missing data structure due to the different age cohorts being interviewed at different ages. Issues concerning the missing data structure due to the multiple-cohort design are dealt with in the same way as the analyses in the Muthén chapter in this volume. The missing data are treated as missing completely at random, that is assuming that the different age cohorts are from the same population and there are no cohort effects. This same assumption is made for both the multilevel modeling and the multivariate longitudinal modeling.

## ANALYSIS USING THE MULTILEVEL MODELING APPROACH

The subset of longitudinal family data from NLSY consists of heavy drinking data spanning 15 years for 1,614 two-sibling families. One approach for analyzing this type of hierarchical longitudinal data is the multilevel approach. Bryk and Raudenbush (1987, 1992) among others have shown how a multilevel model involving growth can be formulated.

The NLSY data with repeated measurements nested within siblings who are in turn, nested within families can be modeled using a three-level model. Individual growth trajectories or the measurement model is specified at level 1. The observed outcome at each time point is specified as a function of a set of growth parameters over time with measurement error, assuming systematic growth trajectories. At level 2, the growth parameters that characterize individual growth trajectories are specified to vary across the population of siblings. Individual differences in growth can be studied by introducing predictor variables of individual characteristics at this between-individual level. At level 3, the model specified at the family level expresses variation across a population of families. Family-level predictor variables can be included in this



Level-3 model to describe variation across families. Two multilevel models are presented here: Model A and Model B. Model A is a three-level unconditional model without covariates. Model B includes individual covariates at level 2 and family covariates at level 3. A growth model with a linear growth rate and a quadratic growth rate is specified at level 1:

$$y_{tij} = \pi_{0ij} + \pi_{1ij} a_{tij} + \pi_{2ij} a_{tij}^{\ 2} + e_{tij},\ t = 1, 2, 3,... \qquad (2.2)$$

where $y_{tij}$ is the heavy drinking frequency for individual $i$ in family $j$ at time $t$. $a_{tij}$ is the time-related variable at time $t$ for individual $i$ in family $j$. This corresponds to the $\lambda_t$ described previously for individual $i$ in family $j$. The $\pi$'s are the growth parameters that vary across individuals. Note that $a_{tij}$ is 0 when $\lambda_t$ is 0 and this occurs at age 21. Therefore, $\pi_{0ij}$ is defined as the heavy drinking status at 21 years of age. The coefficient $\pi_{1ij}$ is the linear growth rate and $\pi_{2ij}$ is the quadratic growth rate, which gives the acceleration or the rate of increase of the linear growth rate. The term $e_{tij}$ is the random measurement error term for individual $i$ in family $j$ at time $t$, which is assumed normally distributed with a mean of 0 and common variance $\sigma^2$.

In Model A, each of the growth parameters from level 1 is specified as varying randomly across individuals from the mean values at level 2:

$$\begin{aligned} \pi_{0ij} &= \beta_{00j} + r_{0ij},\ r_{0ij} \sim N(0, \tau_{\pi 00}) \\ \pi_{1ij} &= \beta_{10j} + r_{1ij},\ r_{1ij} \sim N(0, \tau_{\pi 11}) \\ \pi_{2ij} &= \beta_{20j} + r_{2ij},\ r_{2ij} \sim N(0, \tau_{\pi 22}) \end{aligned} \qquad (2.3a)$$

where $\beta_{00j}$, $\beta_{10j}$, and $\beta_{20j}$ represent the growth parameter means of family $j$ and $r_{0ij}$, $r_{1ij}$, and $r_{2ij}$ are the random components of the growth parameters representing the deviations of individual $i$ from the family mean values. These random components are assumed normally distributed with variances $\tau_{\pi 00}$, $\tau_{\pi 11}$, and $\tau_{\pi 22}$, respectively and covariances $\tau_{\pi 01}$, $\tau_{\pi 02}$, and $\tau_{\pi 12}$.

In Model B, individual covariates are added to the Level-2 model:

$$\begin{aligned} \pi_{0ij} &= \beta_{00j} + \beta_{010}(Male)_{ij} + \beta_{020}(ES)_{ij} + \beta_{030}(HSdrp)_{ij} + r_{0ij} \\ \pi_{1ij} &= \beta_{10j} + \beta_{110}(Male)_{ij} + \beta_{120}(ES)_{ij} + \beta_{130}(HSdrp)_{ij} + r_{1ij} \\ \pi_{2ij} &= \beta_{20j} + \beta_{210}(Male)_{ij} + \beta_{220}(ES)_{ij} + \beta_{230}(HSdrp)_{ij} + r_{2ij} \end{aligned} \qquad (2.3b)$$

Here $\pi_{0ij}$, $\pi_{1ij}$, and $\pi_{2ij}$ are modeled as functions of individual covariates, Male, ES, and HSdrp. $\beta_{00j}$, $\beta_{10j}$, and $\beta_{20j}$ are the intercepts for family $j$ in the respective growth parameters and $\beta_{010}$, $\beta_{020}$,.., $\beta_{230}$ are the regression coefficients. These regression coefficients are modeled as fixed effects, that is, they are not allowed to vary across families. Here, the random components $r_{0ij}$, $r_{1ij}$, and $r_{2ij}$ have the same distributional assumptions and represent the deviations of individual $i$ from their model predicted values.

In Model A, the family growth parameters are specified to vary across a population of families at level 3:

$$\begin{aligned} \beta_{00j} &= \gamma_{000} + u_{00j},\ u_{00j} \sim N(0, \tau_{\beta 00}) \\ \beta_{10j} &= \gamma_{100} + u_{10j},\ u_{10j} \sim N(0, \tau_{\beta 11}) \\ \beta_{20j} &= \gamma_{200} + u_{20j},\ u_{20j} \sim N(0, \tau_{\beta 22}) \end{aligned} \qquad (2.4a)$$

where $\gamma_{000}$, $\gamma_{100}$, and $\gamma_{200}$ represent the grand means in this unconditional model and $u_{00j}$, $u_{10j}$, and $u_{20j}$ are the random parts representing the deviations of family $j$ from the overall means. $u_{00j}$, $u_{10j}$, and $u_{20j}$ are assumed normally distributed with variances $\tau_{\beta 00}$, $\tau_{\beta 11}$, and $\tau_{\beta 22}$, and their respective covariances.

In Model B, family-level covariates are introduced to the Level-3 model to describe variation across the population of families. A variable is considered to be a family-level variable if siblings within a family logically share the same value for the variable. In this analysis, ethnicity (Black and Hispanic) and family history variables (FH123, FH1, and FH23) are considered family variables. These predictor variable are added to the Level-3 model to study their influences on the family growth parameter intercepts:

$$\begin{aligned} \beta_{00j} &= \gamma_{000} + \gamma_{001}(Black)_j + \gamma_{002}(Hispanic)_j + \gamma_{003}(FH123)_j + \gamma_{004}(FH1)_j + \gamma_{005}(FH23)_j + u_{00j} \\ \beta_{10j} &= \gamma_{100} + \gamma_{101}(Black)_j + \gamma_{102}(Hispanic)_j + \gamma_{103}(FH123)_j + \gamma_{104}(FH1)_j + \gamma_{105}(FH23)_j + u_{10j} \\ \beta_{20j} &= \gamma_{200} + \gamma_{201}(Black)_j + \gamma_{202}(Hispanic)_j + \gamma_{203}(FH123)_j + \gamma_{204}(FH1)_j + \gamma_{205}(FH23)_j + u_{20j} \end{aligned} \qquad (2.4b)$$

In this Level-3 model $\beta_{00j}$, $\beta_{10j}$, and $\beta_{20j}$ are modeled as functions of family covariates, Black, Hispanic, FH123, FH1, and FH23. $\gamma_{000}$, $\gamma_{100}$, and $\gamma_{200}$ are the intercepts and $\gamma_{001}$, $\gamma_{002}$,.., $\gamma_{205}$ are the regression coefficients. Here, the random components, $u_{00j}$, $u_{10j}$, and $u_{20j}$ have the same distributional assumptions and represent the deviations of family $j$ from their model predicted values. Note that if predictor variables are included at level 2, then $\beta_{00j}$, $\beta_{10j}$, and $\beta_{20j}$ are intercept terms that represent conditional means after adjusted for the Level-2 predictor variables.



## Results of the Multilevel Modeling

Model A (characterized by Equations 2.2, 2.3a, 2.4a) and Model B (characterized by Equations 2.2, 2.3b, 2.4b) are estimated using the multilevel software HLM (Bryk, Raudenbush, & Congdon, 1996). Table 2.3 shows the results for Model A.

The fixed-effect estimates give the overall means for the three growth parameters. The variances of the growth parameters are partitioned among individuals into the within-family components and the between-family components. These are given by the Level-2 variance components and the Level-3 variance components in Table 2.3.

The estimated mean status of heavy drinking frequency at 21 years of age is 1.060 and has a significant variation across individuals (variance = 1.112, $\chi^2$= 5,134.2, df = 1,614, $p < 0.001$). The linear growth rate mean of -0.011 is not significantly different from zero (se = 0.018, $p = 0.526$), but there is significant variation (variance = 0.194, $\chi^2$ = 3,693.7, df = 1,614, $p < 0.001$) across individuals, showing that there are individuals whose growth rate is still increasing at 21, and there are also those whose rates are already on the decline at 21 with an average of about zero. The quadratic rate has a negative estimated mean of -0.127, which is significant (se = 0.014, $p < 0.001$). This shows that there is generally a deceleration in the growth rate of heavy drinking frequency. Heavy drinking frequency tends to increase, but with a slowing rate, and decreases after reaching a peak. The quadratic rate also has a significant variation (variance = 0.069, $\chi^2$ = 3,390.4, df = 1,614, $p < 0.001$) showing that the deceleration in the growth rate varies across individuals. In sum, these results show that there are a lot of individual differences in the drinking trajectories.

As shown by the $\chi^2$ statistics in Table 2.3, there are significant between-family variances for all three growth parameters that are given by the Level-3 variance components. These are considerable compared to the total variations (individuals' within-family components plus between-family components). The between-family variations are 25.9%, 31.7%, and 23.3% of the total variations for Status at age 21, linear rate, and quadratic rate, respectively.

Table 2.4 shows the results for Model B.

Looking at the fixed effects, FH123 have significant influence ($\gamma_{003}$ = 0.400, $p < 0.001$) on heavy drinking status at 21. This shows that individuals with family history in both first degree and the second/third degree relatives tend to drink more by age 21. Both Blacks (comparing

TABLE 2.3

Three-Level Hierarchical Model for Two-Sibling Families (Unconditional Model)

Number of two-sibling families = 1,614

| *Fixed Effect* | *Mean* | *SE* | *t-ratio* | *P-value* |
|---|---|---|---|---|
| Status at 21 years of age | 1.060 | 0.029 | 36.878 | 0.000 |
| Linear Growth Rate | -0.011 | 0.018 | -0.634 | 0.526 |
| Quadratic Growth Rate | -0.127 | 0.014 | -9.083 | 0.000 |
| *Random Effect* | *Variance Components* | *df* | $\chi^2$ | *p-value* |
| Level 1 | | | | |
| Measurement errors | 1.130 | | | |
| Level 2 | | | | |
| Individual Status at 21 | 1.112 | 1614 | 5134.2 | 0.000 |
| Individual Linear Growth Rate | 0.194 | 1614 | 3693.7 | 0.000 |
| Individual Quadratic Growth Rate | 0.069 | 1614 | 3390.4 | 0.000 |
| Level 3 | | | | |
| Family mean Status at 21 | 0.389 | 1613 | 2127.3 | 0.000 |
| Family mean Linear Growth Rate | 0.093 | 1613 | 1908.7 | 0.000 |
| Family mean Quadratic Growth Rate | 0.021 | 1613 | 1771.3 | 0.004 |
| Percentage of Variance Between Families | | | | |
| Status at 21 years of age | 25.9% | | | |
| Linear Growth Rate | 31.7% | | | |
| Quadratic Growth Rate | 23.3% | | | |
| Correlations Among the Random Effects | | | | |
| Level 2 | | | | |
| Status at 21 years of age | 1.000 | | | |
| Linear Growth Rate | -0.050 | 1.000 | | |
| Quadratic Growth Rate | -0.937 | 0.362 | 1.000 | |
| Level 3 | | | | |
| Status at 21 years of age | 1.000 | | | |
| Linear Growth Rate | -0.590 | 1.000 | | |
| Quadratic Growth Rate | -0.877 | 0.86 | 1.000 | |
| Deviance = 54,768.8 | | | | |
| Number of estimated parameters = 16 | | | | |



TABLE 2.4

Three-Level Hierarchical Model for Two-Sibling Families (With Covariates)

Number of two-sibling families = 1,614

| *Fixed Effect* | *Coefficients* | *SE* | *t-ratio* | *p-value* |
|---|---|---|---|---|
| Status at 21 years of age, $\pi_{0ij}$ | | | | |
| Intercept, $\gamma_{000}$ | 0.644 | 0.050 | 12.768 | 0.000 |
| Black, $\gamma_{001}$ | -0.575 | 0.064 | -9.041 | 0.000 |
| Hisp, $\gamma_{002}$ | -0.279 | 0.076 | -3.668 | 0.000 |
| FH123, $\gamma_{003}$ | 0.400 | 0.086 | 4.631 | 0.000 |
| FH1, $\gamma_{004}$ | 0.022 | 0.079 | 0.280 | 0.780 |
| FH23, $\gamma_{005}$ | 0.048 | 0.068 | 0.701 | 0.483 |
| Male, $\beta_{010}$ | 0.893 | 0.048 | 18.598 | 0.000 |
| ES, $\beta_{020}$ | 0.421 | 0.077 | 5.486 | 0.000 |
| HSdrp, $\beta_{030}$ | 0.179 | 0.063 | 2.846 | 0.005 |
| Linear Growth Rate, $\pi_{1ij}$ | | | | |
| Intercept, $\gamma_{100}$ | -0.052 | 0.036 | -1.566 | 0.117 |
| Black, $\gamma_{101}$ | 0.196 | 0.042 | 4.704 | 0.000 |
| Hisp, $\gamma_{102}$ | 0.124 | 0.049 | 2.548 | 0.001 |
| FH123, $\gamma_{103}$ | -0.138 | 0.056 | -2.447 | 0.015 |
| FH1, $\gamma_{104}$ | -0.021 | 0.052 | -0.410 | 0.682 |
| FH23, $\gamma_{105}$ | -0.110 | 0.044 | -2.480 | 0.013 |
| Male, $\beta_{110}$ | 0.031 | 0.033 | 0.938 | 0.349 |
| ES, $\beta_{120}$ | -0.114 | 0.051 | -2.219 | 0.026 |
| HSdrp, $\beta_{130}$ | 0.124 | 0.043 | 2.918 | 0.004 |
| Quadratic Growth Rate, $\pi_{2ij}$ | | | | |
| Intercept, $\gamma_{200}$ | -0.151 | 0.026 | -5.715 | 0.000 |
| Black, $\gamma_{201}$ | 0.171 | 0.033 | 5.156 | 0.000 |
| Hisp, $\gamma_{202}$ | 0.140 | 0.040 | 3.540 | 0.001 |
| FH123, $\gamma_{203}$ | -0.018 | 0.045 | -0.407 | 0.684 |
| FH1, $\gamma_{204}$ | 0.055 | 0.042 | 1.307 | 0.191 |
| FH23, $\gamma_{205}$ | 0.017 | 0.036 | 0.466 | 0.641 |
| Male, $\beta_{210}$ | -0.096 | 0.027 | -3.562 | 0.001 |
| ES, $\beta_{220}$ | 0.043 | 0.043 | 0.983 | 0.326 |
| HSdrp, $\beta_{230}$ | -0.017 | 0.035 | -0.505 | 0.613 |

*Table 2.4 continued*

*Table 2.4 continued*

| Random Effect | Variance Components | df | $\chi^2$ | p-value |
|---|---|---|---|---|
| Level 1 | | | | |
| Measurement errors, $e_{tij}$ | 1.135 | | | |
| Level 2 | | | | |
| Status at 21, $r_{0ij}$ | 0.829 | 1611 | 4706.2 | 0.000 |
| Linear Growth Rate, $r_{1ij}$ | 0.174 | 1611 | 3675.5 | 0.000 |
| Quadratic Growth Rate, $r_{2ij}$ | 0.064 | 1611 | 3358.2 | 0.000 |
| Level 3 | | | | |
| Status at 21, $u_{00j}$ | 0.334 | 1608 | 2146.9 | 0.000 |
| Linear Growth Rate, $u_{10j}$ | 0.718 | 1608 | 1883.9 | 0.000 |
| Quadratic Growth Rate, $u_{20j}$ | 0.017 | 1608 | 1758.3 | 0.005 |
| Deviance = 54,063.3 | | | | |
| Number of estimated parameters = 40 | | | | |

to non-Blacks) and Hispanics (comparing to non-Hispanics) have significantly lower drinking frequencies at 21 ($\gamma_{001}$ = -0.575, $p < 0.001$; $\gamma_{002}$ = -0.279, $p < 0.001$) after controlling for the other variables. All the individual covariates, Male, ES, and HSdrp, have positive influence on the drinking status at 21 ($\beta_{010}$ = 0.893, $p < 0.001$; $\beta_{020}$ = 0.421, $p < 0.001$; $\beta_{030}$ = 0.179, $p$ = 0.005). This indicates that males, early starters and high school dropouts tend to drink more at age 21.

For the regression of the linear rate at age 21, the intercept term is not significant ($\gamma_{100}$ = -0.052, $p$ = 0.117) and could have been set to 0 in the model. Both ethnicity groups, Blacks and Hispanics, have significantly higher linear rates at 21 ($\gamma_{101}$ = 0.196, $p$ = 0.000; $\gamma_{102}$ = 0.124, $p$ = 0.001). This may be because they started later and were still on the upward trend because these two group also had lower drinking frequencies at 21. Two of the family history variables (FH123, FH23) have significant negative influence on the linear rate ($\gamma_{103}$ = -0.138, $p$ = 0.015; $\gamma_{105}$ = -0.110, $p$ = 0.013). Early Start (ES) is also negatively related to the linear rate ($\beta_{120}$ = -0.114, $p$ = 0.026). This shows that individuals with family history and who started drinking early might have reached their peak and were probably slowing down by 21. HSdrp is positively related to the linear rate ($\beta_{130}$ = 0.124, $p$ = 0.004) showing that high school dropouts were still at an upward trend in their drinking pattern at age 21.



Only ethnicity variables and gender show significant effects on the quadratic growth rate. Both Blacks and Hispanics are positively related ($\gamma_{201}$ =0.171, $p < 0.001$; $\gamma_{202}$ = 0.140, $p$ = 0.001) to the quadratic growth rate, showing the linear rate is decreasing slower. Being male is negatively related to the quadratic growth rate, showing that growth rate is decreasing faster ($\beta_{210}$ = -0.096, p = 0.001). Taking the three growth parameters together, the results show that Blacks and Hispanics were not drinking as much at 21 but were on an upward trend that decreased slower, whereas males were drinking a lot at 21, but had reached their peak and were on a faster downward trend at 21.

Table 2.4 also shows that there are still considerable unexplained variances in the three growth parameters for both the within-family model at Level 2 and the between-families model at Level 3. There are probably more key covariates that can be included in the model to explain both the within-family variations across individuals and the between-family variations.

Multilevel models are very useful for partitioning the variations into the within-family variance components and the between-family variance components and for formulating and testing hypothesis at each level separately.

## ANALYSIS USING THE LATENT VARIABLE MODELING APPROACH

The growth model with a linear growth rate and a quadratic growth rate specified by Equation 2.2 can be reformulated to fit into the general latent variable framework (Jöreskog & Sorböm, 1979). In the latent growth formulation, the repeated measurements are strung out as a multivariate vector, and we have a regular single-level latent variable model where all the flexibility of structural equation modeling can be applied. Figure 2.2 shows such a model.

The outcome $Y_{ti}$ of individual i at time t is expressed as a quadratic function of the time-related variable $\lambda_t$,

$$Y_{ti} = 1.S_i + \lambda_t \, Lrate_i + \lambda_t^2 \, Qrate_i + \varepsilon_{ti}, \; t = 1,..,15 \qquad (2.5)$$

with the three latent factors $S_i$, $Lrate_i$, and $Qrate_i$ representing the three growth parameters, $S_i$ being the random intercept, $Lrate_i$ the linear

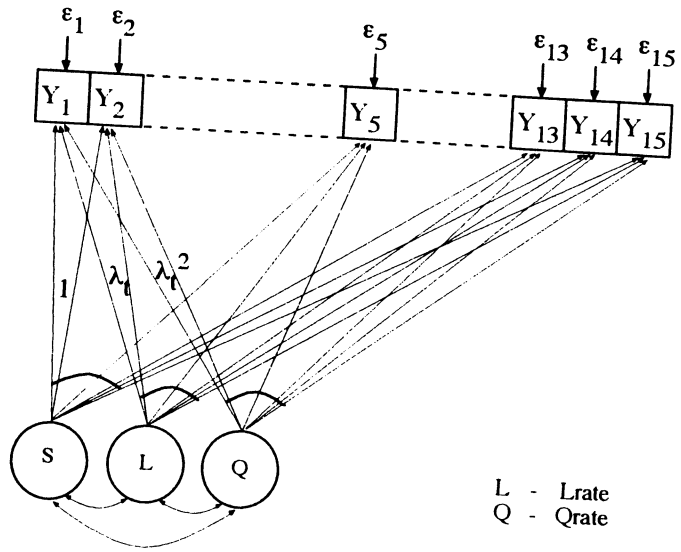


FIG. 2.2 A longitudinal growth model of heavy drinking.

growth rate, and $Qrate_i$ the quadratic growth rate or the acceleration. $\varepsilon_{ti}$ is the measurement error at time $t$ for individual $i$. In order to achieve the latent variable formulation, $\lambda_t$ is the fixed factor loading of $Y_{ti}$ on $L_i$. The factor loading of $Y_{ti}$ on $Q_i$ is $\lambda_t^2$.

Longitudinal sibling data can be analyzed using multivariate models with multiple growth trajectories, one for each sibling. This is possible because of the small number of siblings per family. The unit of analysis is the family, and the growth factors of the siblings can be interrelated by correlating the factors or by adding structural paths as in any standard structural equation model. In this way, the nonindependence between siblings within family is modeled within the structural model. These multivariate growth models can be estimated using any structural equation modeling (SEM) software that handles mean structures. That covers practically all the commercially available SEM software, for example, EQS (Bentler, 1985), LISCOMP(Muthén, 1987), LISREL (Jöreskog & Sörbom, 1989), and MPlus (Muthén & Muthén, 1998).

This multivariate approach is illustrated through Model C and Model D. The main aim of these examples is not in the theory-building but in pointing out possibilities. Both models consist of two growth processes, one for the older sibling and one for the younger sibling. These models



can be generalized to any reasonable number of siblings in a family. The analysis is carried out without covariates in Model C. Individual variables, gender (Male), early start (ES), and high school dropout (HSdrp) and family variables, ethnicity (Black, Hispanics), and family drinking history (FH123, FH1, FH23) are included in Model D.

The growth trajectories are characterized by quadratic growth functions with three growth factors each (*S21*, *LRate*, *QRate*), where *S21* is the individual status factor at 21 years of age, *LRate* is the individual linear growth rate factor, and *QRate* is the individual quadratic growth rate factor. The factors are subscripted with *O* for the older sibling and with *Y* for the younger sibling. The model of longitudinal heavy drinking for family $j$ is given by two simultaneous growth functions:

$$O_{jt} = S21_{oj} + \lambda_t\, Lrate_{oj} + \lambda_t^2\, Qrate_{oj} + \varepsilon_{jot} \qquad (2.6)$$

for the older sibling, and

$$Y_{jt} = S21_{Yj} + \lambda_t\, Lrate_{Yj} + \lambda_t^2\, Qrate_{Yj} + \varepsilon_{jYt} \qquad (2.7)$$

for the younger sibling,

where $O_{jt}$ and $Y_{jt}$ are heavy drinking frequency at time $t$ = (18,19,..,32) for the older sibling and the younger sibling, respectively. The $\lambda_t$s are the time-related variables; $\varepsilon_{jOt}$ and $\varepsilon_{jYt}$ are the measurement errors. If the equivalent parameters are equated across siblings, and if we assume that the measurement error variances are common across time and across siblings, then the pair of growth functions above would reduce to Equation (2.2) for the HLM model at level 1.

Model C (Fig. 2.3) is the basic family growth model with the two quadratic growth trajectories, one for each sibling.

All the measurement error variances are allowed to be different across sibling groups and across time. The factor variances are also allowed to be different across sibling groups. In this general preliminary model, the growth factor means are allowed to be different across siblings. The equality of each parameter across siblings can be tested by reanalyzing the same model but with the parameter in question equated across siblings. The two models are said to be nested. The equality hypothesis can then be tested using the chi-square difference test with 1 degree of freedom of the two nested models. The growth factors are

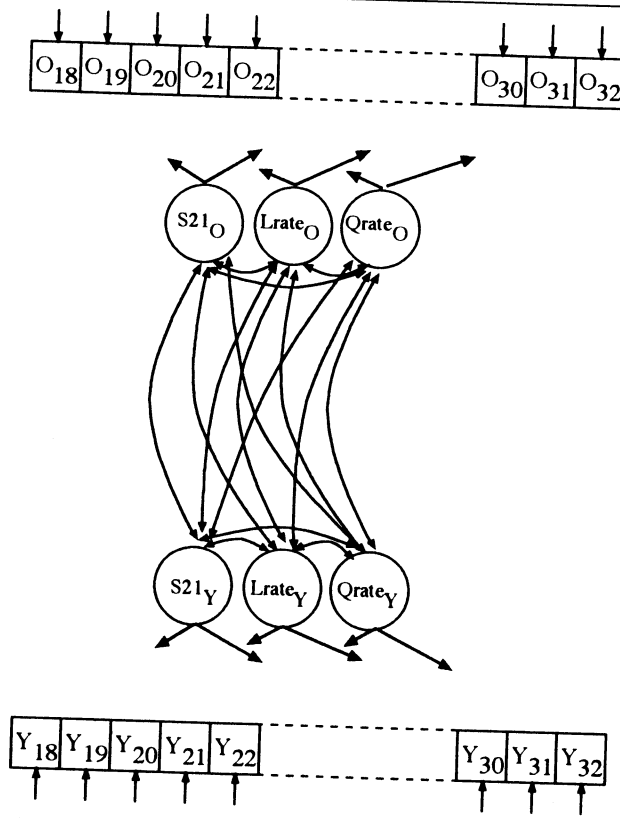


FIG. 2.3. A Longitudinal Growth Model of Heavy Drinking for Two-Sibling Families - Model C.

correlated among one another to investigate the strength of associations among the two sets of growth factors.

In Model D (see Fig. 2.4), the growth model remains the same as in Model C but with covariates introduced into the model.

Each sibling's individual covariates are allowed to influence his or her own growth trajectory by having paths from the covariates to the three growth parameters. The family-level covariates are allowed to influence both trajectories by including paths from these covariates to each of the six growth parameters. Because it is not possible to include all the predictors that will explain the associations among the growth parameters, the residual unexplained variances are allowed to correlate freely. In this model, the coefficients of family covariates and individual covariates are allowed to be different across siblings. The equality of each of these coefficients across the two sibling groups can be tested by



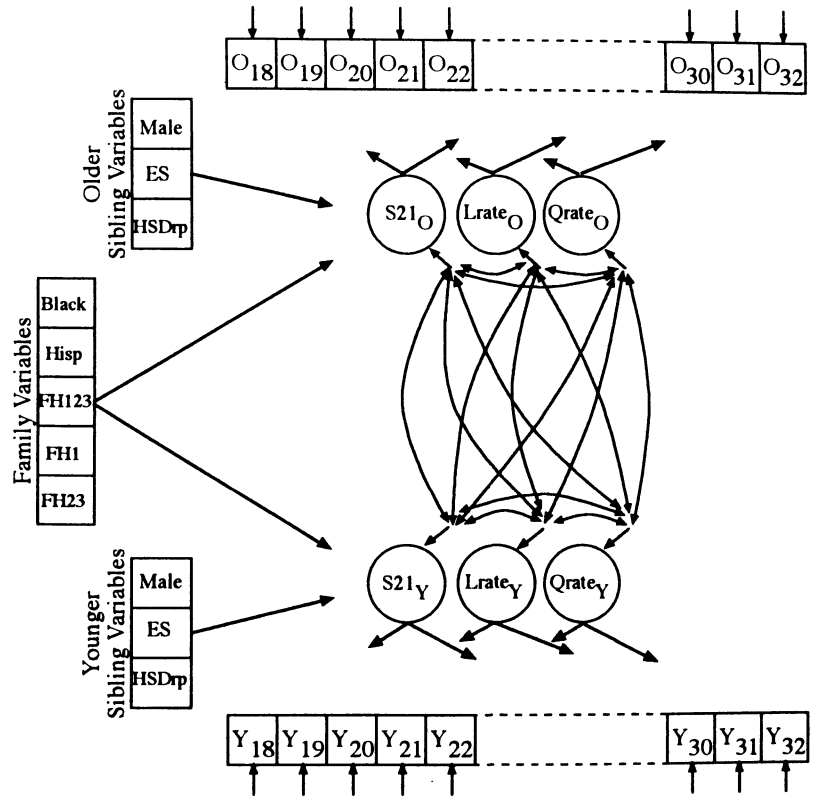


FIG. 2.4. A Longitudinal Growth Model of Heavy Drinking for Two-Sibling Families - Model D.

using a chi-square difference test of two nested models as before. It would be interesting to investigate whether a family variable has the same impact on the drinking patterns of the older siblings and the younger siblings. The influence of a family variable may interact with sibling order. For example, a family environment with alcoholic parents may affect an older child and a younger child differently.

The growth factors are allowed to be freely correlated in Model D as in Model C. Model D can be modified to include direct paths among the growth factors. If we hypothesize that the trajectories of the older siblings influence those of the younger siblings, directional paths can be introduced from $S21_O$ to $S21_Y$, $LRate_Y$ and $QRate_Y$, from $LRate_O$ to

*LRate*$_Y$ and *QRate*$_Y$ and from *QRate*$_O$ to *QRate*$_Y$. Such a possible revised model is shown in Figure 2.4R.

The two multivariate models are estimated using Mplus, a newly developed modeling software (Muthén & Muthén, 1998) that handles the multiple-cohort data structure in the NLSY data.

## Results of the Multivariate Modeling

Table 2.5 shows the results of Model C.

This model does not have any predictor variables. Here, we are interested in examining the goodness of fit for the growth model and

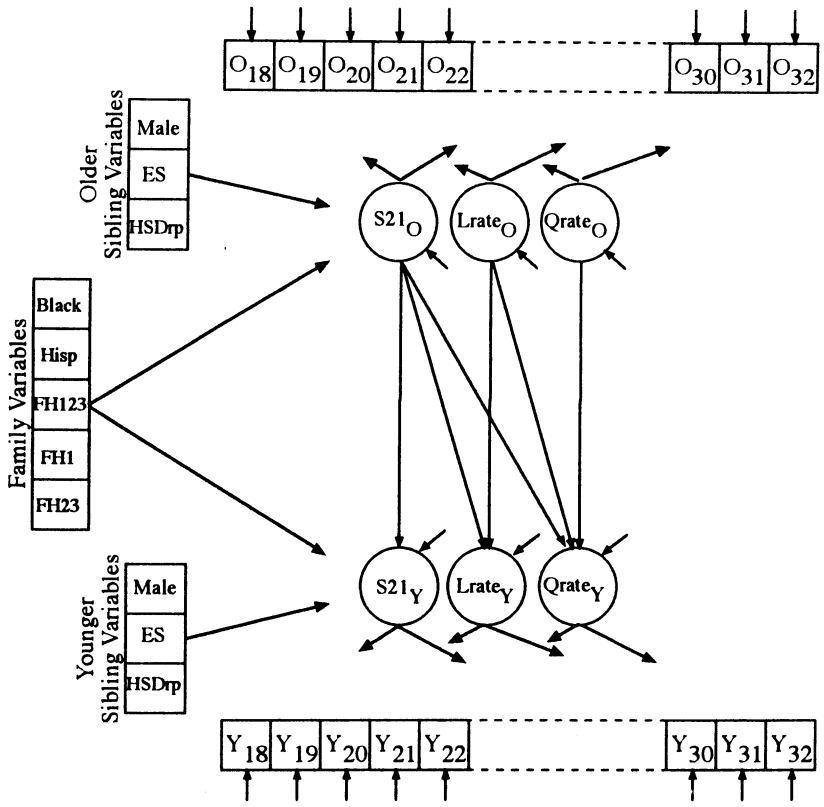


FIG. 2.4R. A Longitudinal Growth Model of Heavy Drinking for Two-Sibling Families - Revised Model D.

TABLE 2.5

Multivariate Longitudinal Model for Two-Sibling Families (Without Covariates)

Number of two-sibling families = 1,614

$\chi^2$ = 762.434 (df = 438, $p$ = 0.000); RMSEA = 0.0214, *p(RMSEA < 0.05) = 1*

| | *Older Sibling* | | | *Younger Sibling* | | |
|---|---|---|---|---|---|---|
| | *Estimate* | *SE* | *t-ratio* | *Estimate* | *SE* | *t-ratio* |
| Means | | | | | | |
| Status at 21 years of age | 1.069 | 0.038 | 27.811 | 1.018 | 0.034 | 30.049 |
| Linear Growth Rate | 0.006 | 0.039 | 0.153 | 0.029 | 0.016 | 1.812 |
| Quadratic Growth Rate | -0.066 | 0.018 | -3.774 | -0.088 | 0.010 | -8.780 |
| Variances | | | | | | |
| Status at 21 years of age | 1.345 | 0.081 | 16.629 | 1.400 | 0.066 | 21.274 |
| Linear Growth Rate | 0.075 | 0.045 | 1.665 | 0.101 | 0.013 | 7.544 |
| Quadratic Growth Rate | 0.058 | 0.15 | 3.943 | 0.035 | 0.007 | 4.998 |

| *Covariances & Correlations* | *Older Sibling* | | | *Younger Sibling* | | |
|---|---|---|---|---|---|---|
| Older Sibling | *OS21* | *OL* | *OQ* | *YS21* | *YL* | *YQ* |
| Status at 21 years of age OS21 | 1.345 | | | | | |
| | *(1.000)* | | | | | |
| Linear Growth Rate OL | 0.065 | 0.075 | | | | |
| | *(0.205)* | *(1.000)* | | | | |
| Quadratic Growth Rate OQ | -0.193 | -0.002 | 0.058 | | | |
| | *(-0.691)* | *(-0.030)* | *(1.000)* | | | |

Table 2.5 continued

*Table 2.5 continued*

| Covariances & Correlations | Older Sibling | | | Younger Sibling | |
|---|---|---|---|---|---|
| Younger Sibling | | | | | |
| Status at 21 years of age YS21 | 0.376 | -0.173 | 0.025 | 1.400 | |
| | *(0.274)* | *(-0.534)* | *(0.088)* | *(1.000)* | |
| Linear Growth Rate YL | -0.043 | 0.040 | -0.002 | 0.290 | 0.101 |
| | *(-0.117)* | *(0.460)* | *(-0.026)* | *(0.771)* | *(1.000)* |
| Quadratic Growth Rate YQ | -0.031 | 0.019 | 0.001 | -0.15 | -0.003 | 0.035 |
| | *(-0.143)* | *(0.371)* | *(0.022)* | *(-0.678)* | *(-0.050)* | *(1.000)* |



to compare in general the two sets of growth parameters in terms of their means and variances. We will also look at the strength of associations among the growth parameters.

Figure 2.5 shows the mean trajectories of the siblings over time as estimated in Model C.

This model has a chi-square value of 762.43 with 438 degrees of freedom ($p = 0.000$, $N = 1{,}614$). A test of close fit using the root mean squared error of approximation (RMSEA; Browne & Cudeck, 1993) shows an acceptable model fit (RMSEA = 0.0214, p(RMSEA < 0.05) = 1). Looking at the drinking status at age 21, the older siblings average slightly higher than the younger siblings (1.069, 1.018). The linear-rate means appear to be not significantly different from 0 (at 0.05 level) for both siblings showing that on average, the heavy drinking patterns tend to peak around 21 years of age for both siblings. The linear growth rate variance at age 21 is significant for the younger siblings but not for the older siblings showing that although most of the older siblings peaked at 21, there are considerable variations for the younger siblings. It is interesting to compare this result to the findings for Model A where there is only one set of growth parameters for both siblings.




FIG. 2.5. Estimated Mean Trajectories

The difference in variation in the linear rate between the two sibling groups would not have been detected in Model A. The quadratic-rate means are negative and significant for both siblings (-0.066, -0.088). The variances of all the growth factors are significant except for the linear growth rate of the older siblings, showing considerable individual differences in status at age 21 and growth over time. The variance–covariances of the two sets of growth factors are also shown in Table 2.5. The corresponding correlations are given in parentheses underneath the covariances. For the within-siblings factor covariances, there is a significant negative association between the status at 21 and the quadratic rate that is consistent across both siblings (r = -0.691, -0.678). This shows that individuals who peak higher would tend to decrease at a faster rate in terms of heavy drinking. Looking at factor covariances across siblings, there are significant positive associations between the status at age 21 (r = 0.274), and a significant positive association between the linear rates at age 21 (r = 0.460), showing that there are some similarities in the drinking patterns of the siblings. There is also a significant negative association between the older siblings' linear growth rate at 21 and the younger siblings' drinking status at 21 (r = -0.534) and a significant negative association between the older siblings' status at 21 and the younger siblings' quadratic growth rate. Family-level predictor variables are introduced in Model D in an attempt to model these associations.

An individual's trajectory or growth pattern is established by taking all three growth factors together. For example, a high status at 21 with near-zero linear rate and a negative quadratic rate characterize a trajectory that is peaking at age 21 (near zero rate) and on a downward trend (zero rate decreasing to become negative rate). A low status at 21 with a positive linear rate and a negative quadratic rate characterize a trajectory low at 21 that is still on the up trend but is slowing down.

Table 2.6 shows the results of Model D.

With individual-level and family-level predictor variables in the model, we are interested in the influences of these predictors on the trajectories of the two siblings. Model D has a chi-square value of 1,077.85 with 720 degrees of freedom ($p$ = 0.000, $N$ = 1,614), a test of close fit shows a RMSEA (e) of 0.00175 and p(e < = 0.05) = 1, which also gives acceptable model fit. Table 2.6 mainly shows the results from the added structural part of the model. The estimates are the regression coefficients in regressing the growth factors/parameters on the individ-

TABLE 2.6

Multivariate Longitudinal Model for Two-Sibling Families (With Covariates) $N$ = 1,614

$\chi^2$ = 1,077.85( df = 720, $p$ = 0.000); RMSEA = 0.0175, p(RMSEA < 0.05) = 1

| | *Older Sibling* | | | *Younger Sibling* | | |
|---|---|---|---|---|---|---|
| | *Estimate* | *SE* | *t-ratio* | *Estimate* | *SE* | *t-ratio* |
| Status 21 Regression | | | | | | |
| Intercept | 0.687 | 0.071 | 9.703 | 0.612 | 0.060 | 10.181 |
| Individual variables | | | | | | |
| Gender (Male=1, Female=0) | 0.777 | 0.072 | 10.733 | 0.908 | 0.061 | 14.882 |
| Early Start | 0.550 | 0.130 | 4.232 | 0.367 | 0.092 | 3.979 |
| High School-Dropout | 0.065 | 0.091 | 0.713 | 0.198 | 0.078 | 2.540 |
| Family variables | | | | | | |
| Black | -0.532 | 0.088 | -6.067 | -0.621 | 0.075 | -8.314 |
| Hispanic | -0.239 | 0.106 | -2.267 | -0.253 | 0.090 | -2.826 |
| FH123 | 0.375 | 0.120 | 3.127 | 0.260 | 0.102 | 2.544 |
| FH1 | 0.163 | 0.105 | 1.550 | 0.110 | 0.090 | 1.226 |
| FH23 | 0.020 | 0.098 | 0.203 | 0.116 | 0.083 | 1.389 |
| Linear Growth Rate Regression | | | | | | |
| Intercept | -0.274 | 0.081 | -3.363 | -0.019 | 0.030 | -0.628 |
| Individual variables | | | | | | |
| Gender (Male=1, Female=0) | 0.181 | 0.084 | 2.142 | 0.050 | 0.031 | 1.609 |

ual covariates and the family covariates. In this preliminary model, the coefficients of family covariates are allowed to be different across siblings to investigate differential impact. Equality of the coefficients across siblings can be tested by comparing nested models but is not carried out in this illustrative example.

Looking at the coefficients in Table 2.6, it is interesting to see that there are many coefficients that appear similar across siblings but also several coefficients that appear to be quite different across siblings. For the regressions on the drinking status at 21, the coefficients appear to be similar across siblings except for the coefficients of HSdrp (0.065,0.198), which is significant for the younger siblings and not for the older siblings. This result shows that a younger sibling who is a high-school dropout has the tendency to be drinking more at age 21 compared to a non-high school dropout, but not so for an older sibling. Both being male and starting drinking early are related to a significantly higher drinking status at 21, this is the same for both siblings. With the family variables, Blacks (compared to non-Blacks) and Hispanics (compared to non-Hispanics) have a significantly lower drinking status at 21 for both siblings. In terms of family history, both siblings drink more at age 21 if they have family drinking history among the first-degree relatives as well as the second- or third-degree relatives.

For the linear growth rate regression, there appears to be differential impact for early starters (ES). This is negatively significant for the younger sibling only, which shows that younger siblings who started drinking early tend to have lower linear rate at 21. This may mean that they peak earlier, and by 21, they are either at the peak or already on a downward path. It is interesting to note that only 8.4% of the older siblings are early starters, whereas 12.6% of the younger siblings are early starters. There also appears to be greater gender differences for the older siblings than for the younger siblings (0.181, 0.05). This coefficient (Male effect) is positive and significant for the older siblings and not for the younger siblings. For the family history variables, only FH1 has significant influence and only for the older siblings, showing that their heavy drinking frequency is at a higher growth rate. This seems to indicate that the older sibling is affected more by family environment with alcoholic parents or other first-degree relatives. High school dropouts also have a higher growth rate at 21 and for both siblings. The Black effect has a positive influence on the linear growth rate at 21 for both siblings.

*Table 2.6 continued*

| | Older Sibling | | | Younger Sibling | | |
|---|---|---|---|---|---|---|
| | *Estimate* | *SE* | *Iratio* | *Estimate* | *SE* | *Iratio* |
| Early Start | -0.007 | 0.151 | -0.045 | -0.133 | 0.047 | -2.836 |
| High School-Dropout | 0.306 | 0.106 | 2.900 | 0.096 | 0.040 | 2.418 |
| Family variables | | | | | | |
| Black | 0.312 | 0.100 | 3.104 | 0.080 | 0.037 | 2.140 |
| Hispanic | -0.016 | 0.121 | -0.135 | 0.071 | 0.045 | 1.574 |
| FH123 | 0.187 | 0.137 | 1.362 | 0.019 | 0.051 | 0.367 |
| FH1 | 0.307 | 0.120 | 2.555 | -0.036 | 0.045 | -0.811 |
| FH23 | -0.077 | 0.112 | -0.689 | 0.009 | 0.042 | 0.215 |
| Quadratic Growth Rate Regression | | | | | | |
| Intercept | 0.036 | 0.037 | 0.995 | -0.095 | 0.019 | -4.981 |
| Individual variables | | | | | | |
| Gender (Male=1, Female=0) | -0.106 | 0.038 | -2.793 | -0.071 | 0.020 | -3.591 |
| Early Start | -0.057 | 0.068 | -0.843 | 0.035 | 0.030 | 1.181 |
| High School-Dropout | -0.115 | 0.048 | -2.420 | -0.005 | 0.025 | -0.196 |
| Family variables | | | | | | |
| Black | -0.017 | 0.045 | -0.383 | 0.103 | 0.024 | 4.346 |
| Hispanic | 0.081 | 0.054 | 1.492 | 0.069 | 0.028 | 2.419 |
| FH123 | -0.130 | 0.062 | -2.117 | -0.023 | 0.032 | -0.721 |
| FH1 | -0.124 | 0.054 | -2.306 | 0.024 | 0.028 | 0.860 |
| FH23 | 0.024 | 0.050 | 0.477 | -0.003 | 0.050 | -0.130 |

The Male effect has significant negative coefficients on the quadratic growth rate for both siblings showing decelerations. Taking gender effects on all three growth parameters together, it shows that older male siblings tend to drink more than female older siblings at age 21; they also tend to be on a steeper trajectory that is decelerating faster compared to the older female siblings. The younger male siblings also tend to drink more than the females at age 21. There is no significant difference in the rate of increase in the drinking frequency, but the males tend to decelerate faster from a higher drinking status at 21 compared to younger female siblings.

Both Blacks and Hispanics have significant positive coefficients on the quadratic growth rate, probably showing a slower decline for the younger siblings only. Both high school dropouts and having family history involving first-degree relatives seem to result in negative coefficients on the quadratic growth rate, showing a slowing ascent that is leveling out faster or a faster decline but only for the older siblings. For older siblings with family history, this may be a result of having a higher growth rate at 21 than those without family history to start with.

To illustrate some of these differential effects of the predictor variables on the heavy drinking patterns of the older siblings and the younger siblings, we contrast the effects by showing some estimated mean trajectories in Fig. 2.6a to Fig. 2.6d. In Fig. 2.6a, we show the estimated average heavy drinking trajectories (from age 20 to age 26) of the older siblings and the younger siblings who are males from Black families with family history of alcoholism among the first-degree relatives (FH1 or FH123). A combined trajectory for both siblings is also shown; this is obtained from the multilevel analysis (Model B).

Figure 2.6b shows the same set of trajectories for Black males without family history.

Here, the trajectories approximately coincide showing practically no difference between the older sibling group and the younger sibling group. In contrast, Fig. 2.6a shows that family history of alcoholism has impact on both the older siblings and the younger siblings, but the older siblings are more severely affected. The mean trajectory of the younger siblings shows a flat trajectory that is elevated by about 0.2 points, yet the mean trajectory of the older siblings shows an upward trend that results in an increase of about 0.4 points when peaking at age 24, when it starts to decline slowly.

Figures 2.6c and 2.6d show the same set of trajectories for White males.



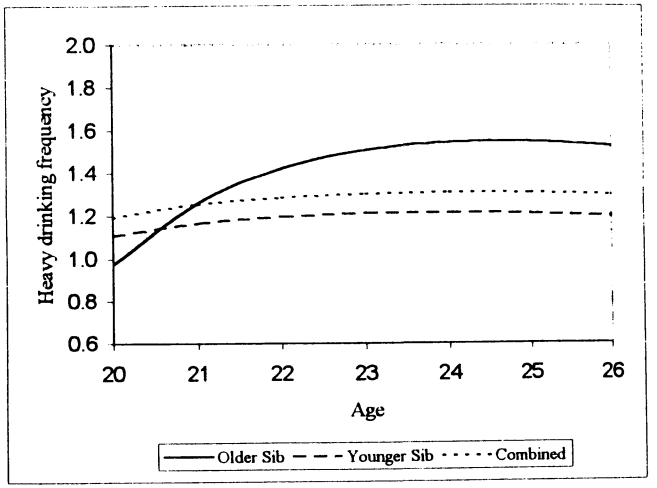


FIG. 2.6a. Black Male with Family History (estimated mean trajectories)

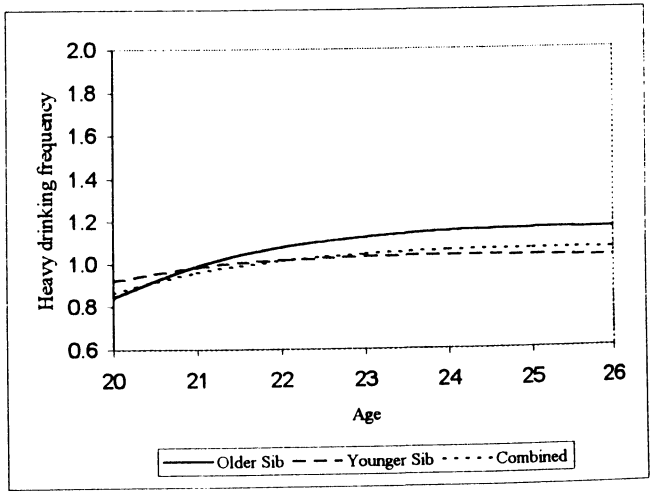


FIG. 2.6b. Black Male without Family History (estimated mean trajectories)

The White males have, in general, a higher heavy drinking frequency level than the Black males. Figure 2.6d shows the trajectories for White males without family history. Here again, the trajectories approximately coincide showing no difference between the older sibling group and the

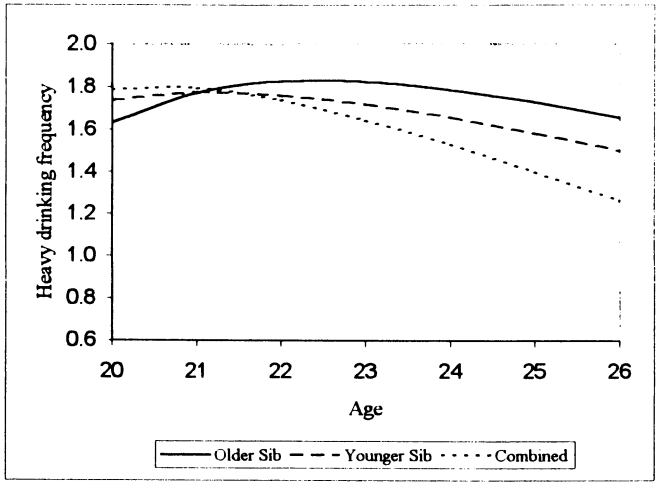


FIG. 2.6c. White Male with Family History (estimated mean trajectories)

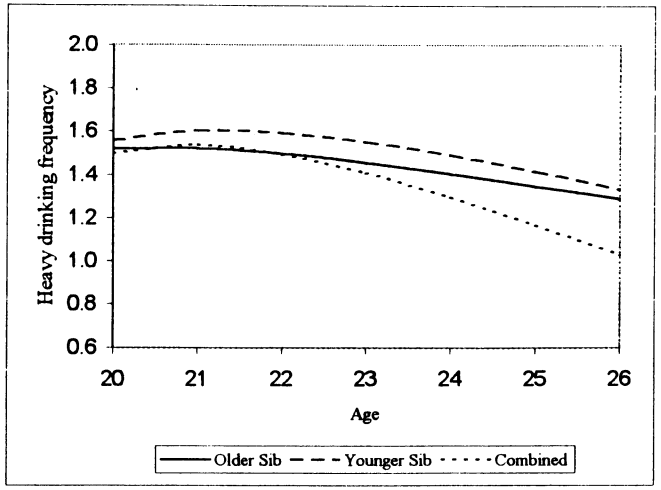


FIG. 2.6d. White Male without Family History (estimated mean trajectories)

younger sibling group. Figure 2.6c shows a pattern that is quite consistent with the results shown in Figure 2.6a for Black males with family history of alcoholism. The older siblings from the White families are also affected more by family history of alcoholism compared to the



younger siblings. The mean trajectory of the younger siblings shows a trajectory that is elevated by about 0.2 points and the mean trajectory of the older siblings shows an upward trend that results in an increase of about 0.4 points when peaking at age 22 and then starts to decline slowly.

These preliminary results have generated some hypotheses to be tested. For example, which of the coefficients are equal across siblings and which are not equal? There may also be some challenging new research questions for the substantive researchers. For example, for those coefficients that are not equal and show differential effects across siblings, would there be plausible explanations in the current developmental theories as to why the effects are different?

## CONCLUSIONS

Two approaches to analyzing family data were discussed. The first is the multilevel approach that takes the hierarchical data structure into account and deals efficiently with the three-level hierarchical data. The multilevel modeling allows the introduction of individual covariates and family covariates into the appropriate levels, thus encouraging clear conceptualization of relationships at each level. In the modeling of individual development and change over time, the hierarchical linear modeling framework is very flexible about the time structure because observations are viewed as nested within individuals and not as a fixed set of measurements across individuals. The multilevel modeling also allows for the partitioning of variances into the across-siblings, within-family components and the between-family components.

In this approach, we can formulate and test hypotheses explicitly at the individual level and at the family level separately. Another advantage of this approach is that if the families are again nested in larger organizational units, for instance, in neighborhoods, then the modeling can further be extended to another higher level to study neighborhood effects on families. Models can be built to answer questions effectively in the first and second areas of interest in modeling longitudinal family data mentioned in the introduction. The three-level model will not model sibling relationships and sibling influences mentioned as the third area of interest. However, interaction effects between family variables and individual characteristics including sibling order mentioned

in the fourth area of interest can be studied in some limited ways by introducing interaction terms at Level 2 as in traditional linear regression models.

The second approach is the multivariate approach that analyzes the longitudinal family data in a multivariate latent growth model. This approach takes the family as the unit of analysis and allows the siblings to have different growth processes. Having multiple growth processes in the model, one for each sibling, allows separate modeling of the developmental growth of each of the siblings. It takes into account the nonindependence of the siblings within family by modeling the dependence within the covariance structure but this approach does not explicitly partition the variance into the within-family component and the between-family component. Models can be set up to answer questions from all four areas of interests in analyzing longitudinal family data mentioned earlier. It is useful when studying small numbers of siblings per family, for example, studying first-born and second-born; oldest child, middle child, and youngest child; or twins. This method allows for the testing of differential effects of family variables on the development of siblings due to sibling order by comparing the goodness of fit of explicitly specified nested models.

The multivariate approach also has the advantage of the full utility of the structural equation modeling framework. We can have multiple indicators for the outcome variables as well as for the predictor variables. The model can include direct effects as well as indirect effects through mediators. Combined with multiple-group analysis, this approach presents more opportunities for the study of similarities and differences in the developmental growth of the siblings. One example is to have the two-sibling families divided into a same-gender siblings group and an opposite-gender siblings group. Multiple-group analyses can then be carried out to see if the developmental relationships between siblings and influences of family variables are the same across the two groups. It would also be interesting to investigate if families with family history of alcoholic problems are the same as those without in terms of siblings' developmental relationships and effects of other family predictor variables.

The multilevel approach and the multivariate approach described are useful in different research settings and objectives and with different data structures. The multivariate latent growth approach is rich and full of potential to make it possible to answer many research questions in the developmental area. The intent of this chapter is to stimulate analysis



using this framework. The illustrative examples serve as an introduction to this approach.

## ACKNOWLEDGMENT

This research was supported by Grant 1 R21 AA10948-01A1 from NIAAA.

## REFERENCES

Bentler, P. M. (1985). Theory and implementation of EQS: A structural equations program. Los Angeles: BMDP Statistical Software.

Bock, R. D. (1989). *Multilevel analysis of educational data*. San Diego, CA: Academic Press.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.

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

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

Bryk, A. S., Raudenbush, S. W., & Congdon, R. T. (1996). *HLM: Hierarchical Linear and Nonlinear Modeling with the HLM/2L and HLM/3L Programs*. Chicago, IL: Scientific Software International, Inc.

Duncan, S. C., & Duncan, T. E. (1996). A multivariate latent growth curve analysis of adolescent substance use. *Structural Equation Modeling, 3*, 323–347.

Goldstein, H. I. (1986). Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrics, 73*, 43–56.

Goldstein, H. I. (1995). *Multilevel Statistical Models*. London: Edward Arnold.

Jöreskog, K. G., & Sörbom, D. (1979). *Advances in factor analysis and structural equation models*. Cambridge, MA: Abt Books.

Jöreskog, K. G., & Sörbom, D. (1989). LISREL 7 user's reference guide. Mooresville, IN: Scientific Software.

Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics, 38*, 963–974.

McArdle, J. J., & Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Development, 58*, 110–133.

McArdle, J. J. (1988). Dynamic but structural equation modeling of repeated measures data. In R. B. Cattel & J. Nesselroade (Eds.), *Handbook of Multivariate Experimental Psychology* (2nd ed.; pp. 561–614). New York: Plenum.

Meredith, W., & Tisak, J. (1984). *"Tuckerizing" curves*. Paper presented at the annual meeting of the Psychometric Society, Santa Barbara, California.

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

Muthén, B. O. (1987). LISCOMP: Analysis of linear structural equations with a comprehensive measurement model. Mooresville, IN: Scientific Software.

Muthén, B. O. (1991). Analysis of longitudinal data using latent variable models with varying parameters. In L. Collins & J. Horn (Eds.), *Best methods for the analysis of change: Recent advances, unanswered questions, future directions* (pp. 1–17). Washington DC: American Psychological Association.

Muthén, B. O. (1993). Latent variable modeling of growth with missing data and multilevel data. In C. R. Rao & C. M. Cuadras (Eds.), *Multivariate analysis: Future directions 2* (pp. 199–210). Amsterdam: North-Holland.

luthén, B. O. (1996). *Longitudinal studies of achievement growth using latent variable modeling.* Technical report # 412, National Center for Research on Evaluation, Standards, and Student Testing (CRESST), University of California, Los Angeles.
luthén, B. O. (1997). Latent variable modeling of longitudinal data and multilevel data. In A. E. Raftery (Ed.), *Sociological methodology* (pp. 453–480). Washington, DC: Blackwell.
luthén, B.O., & Curran, P. J. (1997). General longitudinal modeling of individual differences in experimental designs: A latent variable framework for analysis and power estimation. *Psychological Methods, 2,* 371–402.
luthén, B.O., & Khoo, S. T. (1998). Longitudinal studies of achievement growth using latent variable modeling. *Learning and Individual Differences* V10 N2, 73–101.
luthén, L. K., & Muthén, B. O. (1998). Mplus User's Guide: The Comprehensive Modeling Program for Applied Researchers. Los Angeles: Muthén & Muthén.
'eale, M. C., Walters, E. E., Eaves, L. J., Maes, H. H., & Kendler, K. S. (1994). Multivariate genetic analysis of twin-family data on fears: Mx models. *Behavior Genetics, 24,* 119–139.
ike, A., McGuire, S., Hetherington, E. M., Reiss, D., & Plomin, R. (1996). Family environment and adolescent depressive symptoms and antisocial behavior: A multivariate genetic analysis. *Developmental Psychology, 32,* 590–603.
audenbush, S. W., Brennan, R. T., & Barnett, R. C. (1995). A multivariate hierarchical model for studying psychological change within married couples. *Journal of Family Psychology, 9,* 161–174.
AS Institute (1995). SAS/STAT Software:Changes and Enhancements, Release 6.11, Cary, NC: Author.
renio, J. L. F., Weisberg, H. I., & Bryk, A. S. (1983). Empirical Bayes estimation of individual growth curve parameters and their relationship to covariates. *Biometrics, 39,* 71–86
Villett, J. B., & Sayer, A. G. (1994). Using covariance structure analysis to detect correlates and predictors of individual change over time. *Psychological Bulletin, 116,* 363–381.
Voodhouse, G., Rabash, J., Goldstein, H., Yang, M., & Plewis, I. (1996). MLn, version 1.0a. London: Multilevel Models Project, Institute of Education.

# 3

# The Natural History of Smoking: A Pattern–Mixture Random-Effects Regression Model

Donald Hedeker, PhD
*University of Illinois at Chicago*

Jennifer S. Rose, PhD
*Indiana University*

*This chapter describes and illustrates use of random-effects regression models (RRM) to examine the natural history of smoking from adolescence to adulthood. For longitudinal data analysis, RRMs are useful because they allow for the presence of missing data, time-varying or invariant covariates, and subjects measured at different time points. Thus, a key advantage of RRM is that it can accommodate unbalanced longitudinal data, where a sample of subjects are not all measured at each and every time point. Also, variants of RRM have been developed to model dichotomous and ordinal outcomes, that are common in substance use research. A pattern–mixture approach (Little, 1995) can also be accommodated within RRM to further handle and describe the influence of missing data in longitudinal studies. For this approach, subjects are first divided into groups depending on their missing-data pattern, and then variables based on these groups are used as model covariates. Researchers are then able to examine the effect of missing-data patterns on the outcome(s) of interest. In this chapter, we illustrate these methods using an example from a study examining smoking status from early adolescence to young adulthood.*

Longitudinal studies play a prominent role in investigations of substance use. In these studies, the same individuals are measured repeatedly on a number of variables over a series of time points. For example,