

# Section IV

---

## MODELS FOR MULTILEVEL DATA



# Chapter 12

## AN INTRODUCTION TO GROWTH MODELING

DONALD HEDEKER

### 12.1. INTRODUCTION

Longitudinal studies are increasingly common in social sciences research. In these studies, subjects are measured repeatedly across time, and interest often focuses on characterizing their growth across time. Traditional analysis of variance methods for such growth curve analysis are described in Bock (1975). However, these traditional methods are of limited use because of restrictive assumptions concerning missing data across time and the variance-covariance structure of the repeated measures. The univariate “mixed-model” analysis of variance assumes that the variances and covariances of the dependent variable across time are equal (i.e., compound symmetry). Alternatively, the multivariate analysis of variance for repeated measures only includes subjects with complete data across time. Also, these procedures focus on estimation of group trends across time and provide little help in understanding about how specific individuals change across time. For these and other reasons, hierarchical linear models (HLMs) (Bryk & Raudenbush, 1992) have become the method of choice for growth modeling of longitudinal data.

Variants of HLMs have been developed under a variety of names: random-effects models (Laird &

Ware, 1982), variance component models (Dempster, Rubin, & Tsutakawa, 1981), multilevel models (Goldstein, 1995), two-stage models (Bock, 1989a), random-coefficient models (de Leeuw & Kreft, 1986), mixed models (Longford, 1987; Wolfinger, 1993), empirical Bayes models (Hui & Berger, 1983; Strenio, Weisberg, & Bryk, 1983), and random regression models (Bock, 1983a, 1983b; Gibbons, Hedeker, Waternaux, & Davis, 1988). A basic characteristic of these models is the inclusion of random subject effects into regression models to account for the influence of subjects on their repeated observations. These random subject effects thus describe each person’s growth across time and explain the correlational structure of the longitudinal data. In addition, they indicate the degree of subject variation that exists in the population of subjects.

Several features make HLMs especially useful in longitudinal research. First, subjects are not assumed to be measured on the same number of time points; thus, subjects with incomplete data across time are included in the analysis. The ability to include subjects with incomplete data across time is an important advantage relative to procedures that require complete data across time because (a) by including all data, the analysis has increased statistical power, and

---

AUTHOR’S NOTE: The author thanks David Kaplan and Michael Seltzer for helpful and constructive comments on an earlier version of this chapter. Preparation of this chapter was supported by National Institutes of Mental Health (NIMH) Grant MH44826.

(b) complete case analysis may suffer from biases to the extent that subjects with complete data are not representative of the larger population of subjects. Because time is treated as a continuous variable in HLMs, subjects do not have to be measured at the same time points. This is useful for analysis of longitudinal studies in which follow-up times are not uniform across all subjects. Both time-invariant and time-varying covariates can be included in the model. Thus, changes in the outcome variable may be due to both stable characteristics of the subject (e.g., their gender or race) as well as characteristics that change across time (e.g., life events). Finally, whereas traditional approaches estimate average change (across time) in a population, HLMs can also estimate change for each subject. These estimates of individual change across time can be particularly useful in longitudinal studies in which a proportion of subjects exhibit change across time that deviates from the average trend.

As these methods have developed, several textbooks describing HLMs for longitudinal data analysis, to various degrees, have been published (Brown & Prescott, 1999; Bryk & Raudenbush, 1992; Davis, 2002; Diggle, Liang, & Zeger, 1994; Goldstein, 1995; Hand & Crowder, 1996; Hox, 2002; Longford, 1993; Raudenbush & Bryk, 2002; Singer & Willett, 2003; Verbeke & Molenberghs, 2000). Similarly, several collected editions are available (Bock, 1989b; Collins & Sayer, 2001; Leyland & Goldstein, 2001; Moskowitz & Hershberger, 2002) containing a variety of HLM developments. Also, review, comparison, and/or tutorial articles on longitudinal data analysis treating HLMs have proliferated (Albert, 1999; Burchinal, Bailey, & Snyder, 1994; Cnaan, Laird, & Slasor, 1997; Delucchi & Bostrom, 1999; Everitt, 1998; Gibbons et al., 1993; Gibbons & Hedeker, 2000; Keselman, Algina, Kowalchuk, & Wolfinger, 1999; Lesaffre, Asefa, & Verbeke, 1999; Manor & Kark, 1996; Omar, Wright, Turner, & Thompson, 1999; Sullivan, Dukes, & Losina, 1999). Most of these articles concern continuous response variables, although ones dealing specifically with categorical outcomes have also appeared (Agresti & Natarajan, 2001; Fitzmaurice, Laird, & Rotnitzky, 1993; Gibbons & Hedeker, 1994; Hedeker & Mermelstein, 1996, 2000; Pendergast et al., 1996; Zeger & Liang, 1992).

Applications of growth modeling are steadily increasing and can be found in many different fields, including studies on alcohol (Curran, Stice, & Chassin, 1997), smoking (Niaura et al., 2002), HIV/AIDS (Gallagher, Cottler, Compton, & Spitznagel, 1997), drug abuse (Carroll et al., 1994; Halikas, Crosby, Pearson, & Graves, 1997), psychiatry (Elkin et al.,

1995; Serretti, Lattuada, Zanardi, Franchini, & Smeraldi, 2000), and child development (Campbell & Hedeker, 2001; Huttenlocher, Haight, Bryk, & Seltzer, 1991), to name a few. Not only do these articles illustrate the wide applicability of HLMs, but they also give a sense of how HLM results are typically reported in the various literatures. Thus, they can be very useful for investigators who are new to HLMs and their usage.

This chapter will focus on describing HLMs for continuous outcomes in a very practical way. We will first illustrate how HLMs can be seen as an extension of an ordinary linear regression model. Starting with a simple linear regression model, the model will slowly be extended and described to guide the reader going from familiar to less familiar territory. Following the descriptions of the statistical models, several HLM analyses will be presented using a longitudinal psychiatric data set. These analyses will illustrate many of the key features of HLMs for growth modeling. For further illustration, interested readers can download the data set and program files to replicate the analyses in this report from <http://www.uic.edu/~hedeker/long.html>.

## 12.2. HLMs FOR LONGITUDINAL DATA

To introduce HLMs, consider a simple linear regression model for the measurement  $y$  of individual  $i$  ( $i = 1, 2, \dots, N$  subjects) on occasion  $j$  ( $j = 1, 2, \dots, n_i$  occasions):

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + \varepsilon_{ij}. \quad (1)$$

Ignoring subscripts, this model represents the regression of the outcome variable  $y$  on the independent variable time (denoted  $t$ ). The subscripts keep track of the particulars of the data—namely, whose observation it is (subscript  $i$ ) and the relative order of the observation was made (the subscript  $j$ ). The independent variable  $t$  gives a value to the level of time and may represent time in weeks, months, and so forth. Because  $y$  and  $t$  carry both  $i$  and  $j$  subscripts, both the outcome variable and the time variable are allowed to vary by individuals and occasions.

In linear regression models, such as (1), the errors  $\varepsilon_{ij}$  are assumed to be normally and *independently* distributed in the population with zero mean and common variance  $\sigma^2$ . This independence assumption makes the model given in equation (1) an unreasonable one for longitudinal data. This is because the outcomes  $y$  are observed repeatedly from the same individuals, and so it is much more likely to assume that errors within an individual are correlated to some degree. Furthermore,

the above model posits that the growth, or change across time, is the same for all individuals because the model parameters describing growth ( $\beta_0$ , the intercept or initial level, and  $\beta_1$ , the linear change across time) do not vary by individuals. For both of these reasons, it is useful to add individual-specific effects into the model that will account for the data dependency and describe differential growth for different individuals. This is precisely what HLMs do. Thus, a simple HLM is given by

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + \nu_{0i} + \varepsilon_{ij}, \quad (2)$$

where  $\nu_{0i}$  represents the influence of individual  $i$  on his or her repeated observations.

To better reflect how this model characterizes an individual's influence on his or her observations, we can represent the model in a hierarchical or multilevel form. For this, it is partitioned into the within-subjects (or Level 1) model,

$$y_{ij} = b_{0i} + b_{1i} t_{ij} + \varepsilon_{ij}, \quad (3)$$

and the between-subjects (or Level 2) model,

$$\begin{aligned} b_{0i} &= \beta_0 + \nu_{0i}, \\ b_{1i} &= \beta_1. \end{aligned} \quad (4)$$

Here, the Level 1 model indicates that individual  $i$ 's response at time  $j$  is influenced by his or her initial level  $b_{0i}$  and time trend, or slope,  $b_{1i}$ . The Level 2 model indicates that individual  $i$ 's initial level is determined by the population initial level  $\beta_0$ , plus a unique contribution for that individual  $\nu_{0i}$ . Thus, each individual has his or her own distinct initial level. Conversely, the present model indicates that each individual's slope is the same; all are equal to the population slope  $\beta_1$ . Another way to think about it is that each person's trend line is parallel to the population trend determined by  $\beta_0$  and  $\beta_1$ . The difference between each individual's trend and the population trend is  $\nu_{0i}$ , which is constant across time.

The between-subjects, or Level 2, model is sometimes referred to as a "slopes as outcomes" model (Burstein, 1980). The hierarchical representation shows that just as within-subjects (Level 1) covariates can be included in the model to explain variation in Level 1 outcomes ( $y_{ij}$ ), between-subjects (Level 2) covariates can be included to explain variation in Level 2 outcomes (the subject's intercept  $b_{0i}$  and slope  $b_{1i}$ ). Note that combining the within- and between-subjects models (3) and (4) yields the previous single-equation model (2).

Because individuals in a sample are typically thought to be representative of a larger population

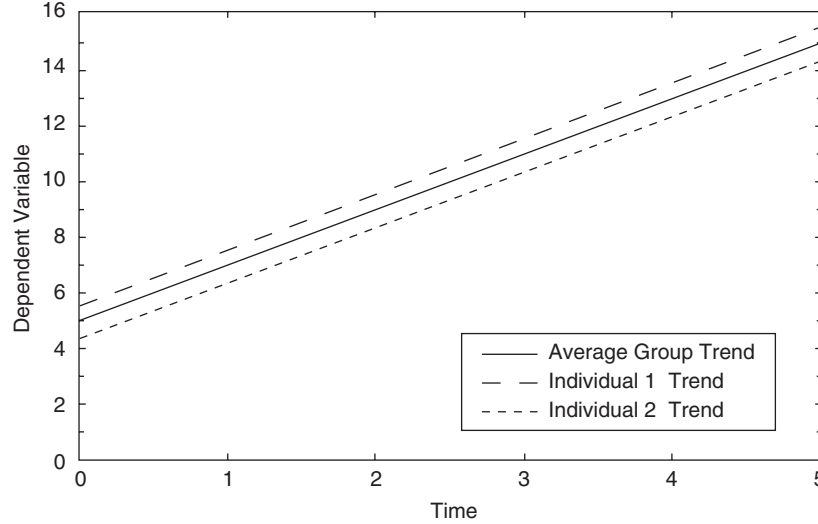
of individuals, the individual-specific effects  $\nu_{0i}$  are treated as random effects. That is,  $\nu_{0i}$  are considered to be representative of a distribution of individual effects in the population. The most common form for this population distribution is the normal distribution, with mean 0 and variance  $\sigma_v^2$ . In the model given by equation (2), the errors  $\varepsilon_{ij}$  are now assumed to be normally and *conditionally independently* distributed in the population with zero mean and common variance  $\sigma^2$ . *Conditional independence* here means conditional on the random individual-specific effects  $\nu_{0i}$ . Because the errors now have an influence due to individuals removed from them, this conditional independence assumption is much more reasonable than the ordinary independence assumption associated with (1). Because individuals deviate from the regression of  $y$  on  $t$  in a parallel manner (because there is only one subject effect  $\nu_{0i}$ ), this model is sometimes referred to as a random-intercepts model, with each  $\nu_{0i}$  indicating how individual  $i$  deviates from the model. Figure 12.1 represents this model graphically.

In this figure, the solid line represents the population average trend, which is based on  $\beta_0$  and  $\beta_1$ . Also depicted are two individual trends, one below and one above the population (average) trend. For a given sample, there are  $N$  such lines, one for each individual. The variance term  $\sigma_v^2$  represents the spread of these lines. If  $\sigma_v^2$  is near zero, then the individual lines would not deviate much from the population trend. In this case, individuals do not exhibit much heterogeneity in growth. Alternatively, as individuals differ from the population trend, the lines move away from the population trend line and  $\sigma_v^2$  increases. In this case, there is more individual heterogeneity in growth.

For longitudinal data, the above random-intercepts model is often too simplistic for a number of reasons. First, it is unlikely that the rate of growth, or trend across time, is the same for all individuals. It is more likely that individuals differ in their rates of growth across time. Not everyone changes at the same rate. Furthermore, the above model implies a compound symmetry assumption for the variances and covariances of the repeated measures. That is, both the variances and covariances across time are assumed to be the same, namely,

$$\begin{aligned} V(y_{ij}) &= \sigma_v^2 + \sigma^2 \\ C(y_{ij}, y_{ij'}) &= \sigma_v^2, \quad \text{where } j \neq j'. \end{aligned} \quad (5)$$

This assumption is usually untenable for most longitudinal data. In general, measurements at points

**Figure 12.1** Random-Intercepts HLM

close in time tend to be more highly correlated than measurements further separated in time. Also, in many studies, subjects are more similar at baseline and grow at different rates across time. Thus, it is natural to expect that variability will increase over time.

For these reasons, a more realistic HLM allows both the intercept and time trend to vary by individuals. For this, the Level 1 model is as before in (3), but the Level 2 model is augmented as

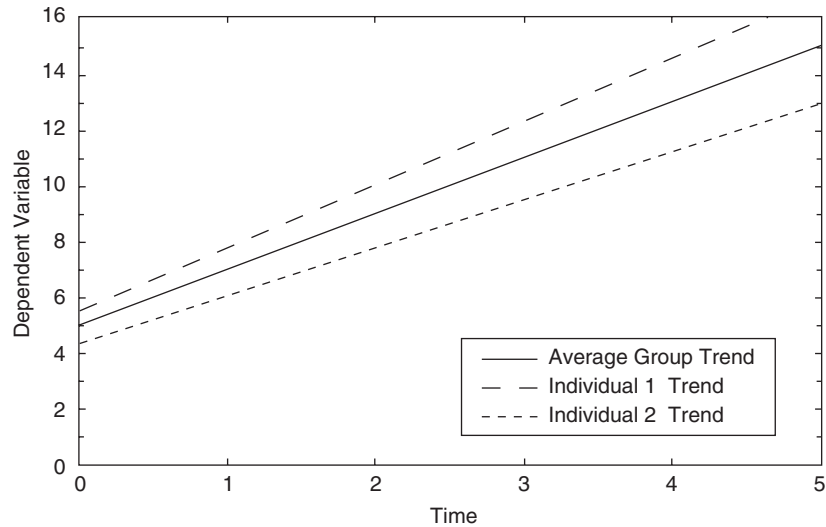
$$\begin{aligned} b_{0i} &= \beta_0 + \nu_{0i}, \\ b_{1i} &= \beta_1 + \nu_{1i}. \end{aligned} \quad (6)$$

In this model,  $\beta_0$  is the overall population intercept,  $\beta_1$  is the overall population slope,  $\nu_{0i}$  is the intercept deviation for subject  $i$ , and  $\nu_{1i}$  is the slope deviation for subject  $i$ . As before,  $\varepsilon_{ij}$  is an independent error term distributed normally with mean 0 and variance  $\sigma^2$ . The assumption regarding the independence of the errors is one of conditional independence; that is, they are independent conditional on  $\nu_{0i}$  and  $\nu_{1i}$ . With two random individual-specific effects, the population distribution of intercept and slope deviations is assumed to be bivariate normal  $N(0, \Sigma_\nu)$ , with the random-effects variance-covariance matrix given by

$$\Sigma_\nu = \begin{bmatrix} \sigma_{\nu_0}^2 & \sigma_{\nu_0\nu_1} \\ \sigma_{\nu_0\nu_1} & \sigma_{\nu_1}^2 \end{bmatrix}.$$

This model can be thought of as a personal trend or change model because it represents the measurements of  $y$  as a function of time, both at the individual ( $\nu_{0i}$  and  $\nu_{1i}$ ) and population ( $\beta_0$  and  $\beta_1$ ) levels. The intercept parameters indicate the starting point, and the slope parameters indicate the degree of change over time. The population intercept and slope parameters represent the overall (population) trend, whereas the individual parameters express how subjects deviate from the population trend. Figure 12.2 represents this model graphically.

Again, figure represents the population trend with the solid line and the trends from two individuals, who now deviate both in terms of the intercept and slope. Because the slope varies for individuals, this model allows the possibility that some individuals do not change across time, whereas others can exhibit dramatic change. The population trend is the average across the individuals, and the variance terms indicate how much heterogeneity there is in the population. Specifically, the variance term  $\sigma_{\nu_0}^2$  indicates how much spread there is around the population intercept, and  $\sigma_{\nu_1}^2$  represents the spread in slopes. To the degree that each individual's deviation from the population trend is only due to random error, these variance terms will approach zero. Alternatively, as each individual's deviation from the population trend is nonrandom but characterized by the individual trend parameters  $\nu_{0i}$  and  $\nu_{1i}$  as being nonzero, these variance terms will increase from zero. In addition, the covariance term,  $\sigma_{\nu_0\nu_1}$ , represents the

**Figure 12.2** Random-Intercept and Slopes HLM

degree to which the individual intercept and slope parameters covary. For example, a positive covariance term would suggest that individuals with higher initial values have greater positive slopes, whereas a negative covariance would suggest the opposite.

The coding of the time variable  $t$  has implications for the interpretation of the model parameters. For example, in growth models,  $t$  sometimes starts with the value zero for baseline and is incremented according to the measurement timeline (e.g., 1, 2, 3, 4 for, say, four monthly follow-ups). In this formulation, the intercept parameters ( $\beta_0$ ,  $\nu_{0i}$ , and  $\sigma_{\nu_0}^2$ ) then characterize aspects of the baseline time point. Alternatively,  $t$  can be expressed in centered form, where the average of time is subtracted from each time value (e.g.,  $-2, -1, 0, 1, 2$ ). In this case, the meaning of the intercept parameters changes to reflect aspects about the midpoint of time and not the baseline time point. As yet another coding choice, sometimes substantive interest focuses on the end of the measurement timeline. Here, time could be coded as  $-4, -3, -2, -1, \text{ and } 0$  (in this example with five time points), so that the intercept parameters reflect aspects of the final time point. The choice of which representation to use often depends on ease of interpretation and the hypotheses of interest.

The occasions range from  $j = 1$  to  $n_i$  in the model specification, with each person being measured on  $n_i$  time points. Because  $n$  carries the  $i$  subscript, each subject may vary in terms of the number of measured occasions. Furthermore, there are no restrictions on the

number of observations per individual; subjects who are missing at a given time point are not excluded from the analysis. Also, because the time variable  $t$  carries the  $i$  subscript, subjects can be measured on different occasions. The underlying assumption of the model is that the data that are available for a given individual are representative of how that individual deviates from the population trend across the timeframe of the study.

Regarding missing data, as Laird (1988) points out, HLMs for longitudinal data using maximum likelihood estimation provide valid statistical tests in the presence of ignorable nonresponse. By *ignorable nonresponse*, it is meant that the probability of nonresponse is dependent on observed covariates *and* previous values of the dependent variable from the subjects with missing data. The notion here is that if subject attrition is related to previous performance, in addition to other observable subject characteristics, then the model provides valid statistical inferences for the model parameters. Because many instances of missing data are related to previous performance or other subject characteristics, HLMs provide a powerful method for dealing with longitudinal data sets in the presence of missing data.

### 12.2.1. Matrix Formulation

A more compact representation of the model is afforded using matrices and vectors. This formulation is particularly useful in model programming and helps

to summarize statistical aspects of the model. For this, the HLM for the  $n_i \times 1$  response vector  $\mathbf{y}$  for individual  $i$  can be written as

$$\mathbf{y}_i(3) = (4)\mathbf{X}_i(5)\boldsymbol{\beta}(6) + (7)\mathbf{Z}_i(8)\mathbf{v}_i(9) + (10)\boldsymbol{\varepsilon}_i(11)$$

$$\begin{matrix} n_i \times 1 & n_i \times p & p \times 1 & n_i \times r & r \times 1 & n_i \times 1 \end{matrix} \quad (7)$$

with  $i = 1 \dots N$  individuals and  $j = 1 \dots n_i$  observations for individual  $i$ . Here,  $\mathbf{y}_i$  is the  $n_i \times 1$  dependent variable vector for individual  $i$ ,  $\mathbf{X}_i$  is the  $n_i \times p$  covariate matrix for individual  $i$ ,  $\boldsymbol{\beta}$  is the  $p \times 1$  vector of fixed regression parameters,  $\mathbf{Z}_i$  is the  $n_i \times r$  design matrix for the random effects,  $\mathbf{v}_i$  is the  $r \times 1$  vector of random individual effects, and  $\boldsymbol{\varepsilon}_i$  is the  $n_i \times 1$  residual vector.

For example, in the random intercepts and slopes HLM just considered, we would have

$$\mathbf{y}_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \dots \\ \dots \\ y_{in_i} \end{bmatrix} \quad \text{and} \quad \mathbf{X}_i = \mathbf{Z}_i = \begin{bmatrix} 1 & t_{i1} \\ 1 & t_{i2} \\ \dots & \dots \\ \dots & \dots \\ 1 & t_{in_i} \end{bmatrix}$$

for the data matrices and

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \quad \text{and} \quad \mathbf{v}_i = \begin{bmatrix} v_{0i} \\ v_{1i} \end{bmatrix}$$

for the population and individual trend parameter vectors, respectively. The distributional assumptions about the random effects and residuals are

$$\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_{n_i}),$$

$$\mathbf{v}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}_v).$$

As a result, it can be shown that the variance-covariance matrix of the repeated measures  $y$  is of the following form:

$$V(\mathbf{y}_i) = \mathbf{Z}_i \boldsymbol{\Sigma}_v \mathbf{Z}_i' + \sigma^2 \mathbf{I}_{n_i}. \quad (8)$$

For example, with  $r = 2$ ,  $n = 3$ , and

$$\mathbf{Z}_i = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix},$$

the variance-covariance matrix equals  $\sigma^2 \mathbf{I}_{n_i} +$

$$\begin{bmatrix} \sigma_{v_0}^2 & \sigma_{v_0}^2 + \sigma_{v_0 v_1} & \sigma_{v_0}^2 + 2\sigma_{v_0 v_1} \\ \sigma_{v_0}^2 + \sigma_{v_0 v_1} & \sigma_{v_0}^2 + 2\sigma_{v_0 v_1} + \sigma_{v_1}^2 & \sigma_{v_0}^2 + 3\sigma_{v_0 v_1} + 2\sigma_{v_1}^2 \\ \sigma_{v_0}^2 + 2\sigma_{v_0 v_1} & \sigma_{v_0}^2 + 3\sigma_{v_0 v_1} + 2\sigma_{v_1}^2 & \sigma_{v_0}^2 + 4\sigma_{v_0 v_1} + 4\sigma_{v_1}^2 \end{bmatrix},$$

which allows the variances and covariances to change across time. For example, if both  $\sigma_{v_0 v_1}$  and  $\sigma_{v_1}^2$  are

positive, then clearly the variance increases across time. Diminishing variance across time is also possible if, for example,  $-2\sigma_{v_0 v_1} > \sigma_{v_1}^2$ . Other patterns are possible depending on the values of these variance and covariance parameters.

Models with more than random intercepts and linear trends are also possible, as are models that allow autocorrelated errors; that is,  $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \sigma^2 \boldsymbol{\Omega}_i)$ . Here,  $\boldsymbol{\Omega}$  might, for example, represent an autoregressive (AR) or moving average (MA) process for the residuals. Autocorrelated error regression models are common in econometrics. Their application within an HLM formulation is treated by Chi and Reinsel (1989) and Hedeker (1989) and extensively described in Verbeke and Molenberghs (2000). By including both random effects and autocorrelated errors, a wide range of variance-covariance structures for the repeated measures is possible. This flexibility is in sharp contrast to the traditional ANOVA models, which assume either a compound symmetry structure (univariate ANOVA) or a totally general structure (MANOVA). Typically, compound symmetry is too restrictive, and a general structure is not parsimonious. HLMs, alternatively, provide these two and everything in between and so allow efficient modeling of the variance-covariance structure of the repeated measures.

### 12.3. HLM EXAMPLE

To illustrate an HLM application, we will consider data from a psychiatric study described in Reisby et al. (1977). This study focused on the longitudinal relationship between imipramine (IMI) and desipramine (DMI) plasma levels and clinical response in 66 depressed inpatients. Imipramine is the prototypic drug in the series of compounds known as tricyclic antidepressants and is commonly prescribed for the treatment of major depression (Seiden & Dykstra, 1977). Because imipramine biotransforms into the active metabolite desmethylimipramine (or desipramine), measurement of desipramine was also done in this study. Major depression is often classified in terms of two types. The first type, nonendogenous or reactive depression, is associated with some tragic life event such as the death of a close friend or family member, whereas the second type, endogenous depression, is not a result of any specific event and appears to occur spontaneously. It is sometimes held that antidepressant medications are more effective for endogenous depression (Willner, 1985). In this sample, 29 patients were classified as

nonendogenous, and the remaining 37 patients were deemed to be endogenous.

The study design was as follows. Following a placebo period of 1 week, patients received 225-mg/day doses of imipramine for 4 weeks. In this study, subjects were rated with the Hamilton depression (HD) rating scale (Hamilton, 1960) twice during the baseline placebo week (at the start and end of this week), as well as at the end of each of the 4 treatment weeks of the study. Plasma level measurements of both IMI and its metabolite DMI were made at the end of each treatment week. The sex and age of each patient were recorded, and a diagnosis of endogenous or nonendogenous depression was made for each patient. Although the total number of subjects in this study was 66, the number of subjects with all measures at each of the weeks fluctuated: 61 at Week 0 (start of placebo week), 63 at Week 1 (end of placebo week), 65 at Week 2 (end of first drug treatment week), 65 at Week 3 (end of second drug treatment week), 63 at Week 4 (end of third drug treatment week), and 58 at Week 5 (end of fourth drug treatment week). Of the 66 subjects, only 46 had complete data at all time points. Thus, complete case analysis under repeated-measures MANOVA, for example, would discard approximately one third of the data set. HLM, alternatively, uses the data that are available from all 66 subjects.

### 12.3.1. Heterogeneous Growth Model

The first model fit to these data corresponds to the within-subjects model (3) and the between-subjects model (6). Here, time is treated using incremental values from 0 to 5. The results are presented in Table 12.1.

Focusing first on the estimated regression parameters, this model indicates that patients start, on average, with an HD score of 23.58 and change by  $-2.38$  points each week. Lower scores on the HD reflect less depression, so patients are improving across time by about 2 points per week. The estimated HD score at Week 5 equals  $23.58 - (5 \times 2.38) = 11.68$ . In their report, Reisby et al. (1977) classified patients into three groups based on their final HD scores: Responders had scores below 8, partial responders were between 8 and 15, and nonresponders had final HD scores above 15. By this criterion, the average trend is in the partial response range at the final time point.

Both the intercept and slope are statistically significant ( $p < .0001$ ) by the so-called “Wald test” (Wald, 1943), which uses the ratio of the maximum likelihood parameter estimate to its standard error to determine

**Table 12.1** HLM Results for Level 1 Model (3) and Level 2 Model (6)

<i>Parameter</i>	<i>Estimate</i>	<i>SE</i>	<i>z</i>	<i>p &lt;</i>
$\beta_0$	23.58	0.55	43.22	.0001
$\beta_1$	-2.38	0.21	-11.39	.0001
$\sigma_{v0}^2$	12.63	3.47		
$\sigma_{v0v1}^2$	-1.42	1.03		
$\sigma_{v1}^2$	2.08	0.50		
$\sigma^2$	12.22	1.11		

NOTE:  $-2 \log L = 2219.04$ .

statistical significance. These test statistics (i.e.,  $z =$  ratio of the parameter estimate to its standard error) are compared to a standard normal frequency table to test the null hypothesis that the parameter equals 0. Alternatively, these  $z$ -statistics are sometimes squared, in which case the resulting test statistic is distributed as chi-square on 1 degree of freedom. In either case, the  $p$ -values are identical. The intercept being significant is not particularly meaningful; it just indicates that HD scores are different from zero at baseline. However, because the slope is significant, we can conclude that the rate of improvement is significantly different from zero in this study. On average, patients are improving across time.

For the variance and covariance terms, there are concerns with using the standard errors in constructing Wald test statistics, particularly when the population variance is thought to be near zero and the number of subjects is small (Bryk & Raudenbush, 1992). This is because variance parameters are bounded; they cannot be less than zero, and so using the standard normal for the sampling distribution is not reasonable. As a result, statistical significance is not indicated for the variance and covariance parameters in the tables. However, the magnitude of the estimates does reveal the degree of individual heterogeneity in both the intercepts and slopes. For example, although the average intercept in the population is estimated to be 23.58, the estimated population standard deviation for the intercept is  $3.55 (= \sqrt{12.63})$ . Similarly, the average population slope is  $-2.38$ , but the estimated population standard deviation for the slope equals 1.44, and so approximately 95% of subjects are expected to have slopes in the interval  $-2.38 \pm (1.96 \times 1.44) = -5.20$  to  $.44$ . That the interval includes positive slopes reflects the fact that not all subjects improve across time. Thus, there is considerable heterogeneity in terms of patients' initial level of depression and in their change across time. Finally, the covariance between the intercept and linear trend is negative; expressed as a correlation, it

equals  $-.28$ , which is moderate in size. This suggests that patients who are initially more depressed (i.e., greater intercepts) improve at a greater rate (i.e., more pronounced negative slopes). An alternative explanation, though, is that of a floor effect due to the HD rating scale. Simply put, patients with less depressed initial scores have a more limited range of lower scores than those with higher initial scores.

An interesting question, at this point, is whether the between-subjects model in equation (6) is necessary over that in equation (4). In other words, is the assumption of compound symmetry rejected or not? Fitting the more restrictive compound symmetry model (not shown) yields  $-2 \log L = 2285.14$ . Because these are nested models, they can be compared using a likelihood ratio test. For this, one compares the model deviance values (i.e.,  $-2 \log L$ ) to a chi-square distribution, where the degrees of freedom equals the number of parameters set equal to zero in the more restrictive model. In the present case,  $\chi_2^2 = 2285.14 - 2219.04 = 66.1$ ,  $p < .0001$ , for  $H_0 : \sigma_{v_0v_1} = \sigma_{v_1}^2 = 0$ . It should be noted that use of the likelihood ratio test for this purpose also suffers from the variance boundary problem mentioned above (Verbeke & Molenberghs, 2000). Based on simulation studies, it can be shown that the likelihood ratio test is too conservative (for testing null hypotheses about variance parameters)—namely, it does not reject the null hypothesis often enough. This would then lead to accepting a more restrictive variance-covariance structure than is correct. As noted by Berkhof and Snijders (2001), this bias can largely be corrected by dividing the  $p$ -value obtained from the likelihood ratio test (of variance terms) by 2. In the present case, it does not really matter, but this modification yields  $p < .0001/2 = .00005$ . Thus, there is clear evidence that the assumption of compound symmetry is rejected.

Using the estimated population intercept ( $\beta_0$ ) and slope ( $\beta_1$ ), we can estimate the average HD score across all time points. These are displayed in Table 12.2, along with the observed means and sample sizes at each time point.

As can be seen, there is close agreement between the observed and estimated means. Thus, the average change across time is very consistent with the posited linear change model. For a more quantitative assessment, the interested reader is referred to Kaplan and George (1998), who describe the use of econometric forecasting statistics to assess various forms of fit between observed and estimated means.

Similarly, we can address the fit of the observed variance-covariance matrix of the repeated measures,

**Table 12.2** Observed and Estimated Means

	Week					
	0	1	2	3	4	5
Observed	23.44	21.84	18.31	16.42	13.62	11.95
Estimated	23.58	21.21	18.82	16.45	14.07	11.69
Sample size	61	63	65	65	63	58

which is given below. These are calculated based on the pairwise data for the covariances and the available data for each of the variances.

$$V(\mathbf{y}) = \begin{bmatrix} 20.55 & & & & & \\ 10.50 & 22.07 & & & & \\ 10.20 & 12.74 & 30.09 & & & \\ 9.69 & 12.43 & 25.96 & 41.15 & & \\ 7.17 & 10.10 & 25.56 & 36.54 & 48.59 & \\ 6.02 & 7.39 & 18.25 & 26.31 & 32.93 & 52.12 \end{bmatrix}$$

Based on the model estimates, we get

$$\hat{V}(\mathbf{y}) = \mathbf{Z}\hat{\Sigma}_v\mathbf{Z}' + \hat{\sigma}^2\mathbf{I}$$

$$= \begin{bmatrix} 24.85 & & & & & \\ 11.21 & 24.08 & & & & \\ 9.79 & 12.52 & 27.48 & & & \\ 8.37 & 13.18 & 18.00 & 35.03 & & \\ 6.95 & 13.84 & 20.73 & 27.63 & 46.74 & \\ 5.53 & 14.50 & 23.47 & 32.44 & 41.41 & 62.60 \end{bmatrix},$$

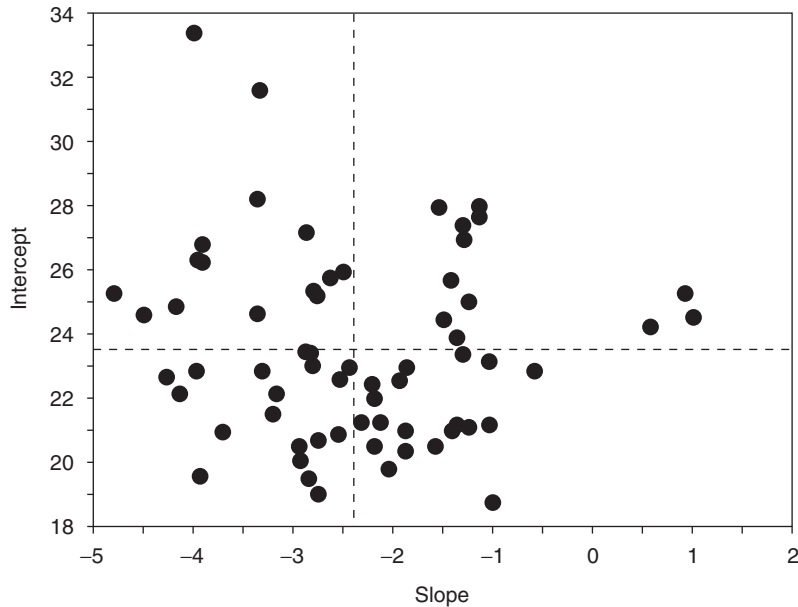
where the design matrix of the random effects and the estimates of the random-effects variance-covariance matrix are given by

$$\mathbf{Z}' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 & 5 \end{bmatrix},$$

$$\hat{\Sigma}_v = \begin{bmatrix} 12.63 & -1.42 \\ -1.42 & 2.08 \end{bmatrix},$$

and  $\hat{\sigma}^2 = 12.22$ . Given that this variance-covariance matrix of 21 elements is represented by four parameter estimates, the fit is reasonably good. The model is clearly picking up on the increasing variance across time and the diminishing covariance away from the diagonal.

Finally, estimates of the individual random effects,  $\hat{b}_{0i}$  and  $\hat{b}_{1i}$ , are often of interest. These are plotted in Figure 12.3. The dashed lines indicate the estimated population intercepts and slopes. Thus,  $\hat{v}_{0i}$  is represented by the horizontal distance between a point and the horizontal line, whereas  $\hat{v}_{1i}$  is represented by the vertical distance between a point and the vertical line.

**Figure 12.3** Reisby Data: Estimated Random Effects

This scatter plot reveals the wide range of observed intercepts and slopes in this sample. In particular, some patients are very depressed initially but improve to a great degree (upper left-hand corner). Similarly, some patients show little or no improvement over time (toward the right side).

It is worth noting that the estimates of the individual random effects, presented in Figure 12.3, are empirical Bayes (EB) estimates, which reflect a compromise between an estimate based solely on an individual's data and an estimate for the population of interest. Thus, they are not equivalent to ordinary least squares (OLS) estimates, which would only rely on an individual's data. An important advantage of EB estimates relative to OLS estimates is that they are not as prone to the undue influence of outliers. This is especially true when an individual has few measurements by which to base these estimates on. Because of this, the EB estimates are said to be *shrunk to the mean*, where the mean of the random effects equals zero in the population. The degree of shrinkage depends on the number of measurements an individual has. Thus, if a subject has few measurements, then the EB estimate will be smaller (in absolute value) than the corresponding OLS estimate. Alternatively, if the subject has many measurements across time, then the EB and OLS estimates would be very similar. These EB estimates are readily available from most HLM software programs.

### 12.3.2. Effect of Diagnosis on Growth

At this point, it may be interesting to examine whether we can explain some of the heterogeneity in intercepts and slopes, depicted in Figure 12.3, in terms of particular subject characteristics. For this, we will augment the Level 2 model to include a covariate  $DX$ , which equals 0 if the patient's diagnosis is nonendogenous (NE) and 1 if the patient is endogenous (E). This variable enters the Level 2 model rather than the Level 1 model because it varies only with subjects ( $i$ ) and not with time ( $j$ ).

$$\begin{aligned} b_{0i} &= \beta_0 + \beta_2 DX_i + \nu_{0i}, \\ b_{1i} &= \beta_1 + \beta_3 DX_i + \nu_{1i}. \end{aligned} \quad (9)$$

Now,  $\beta_0$  represents the average Week 0 HD level for NE patients, and  $\beta_1$  is the average HD weekly improvement for NE patients. Similarly,  $\beta_2$  represents the average Week 0 HD difference for E patients (relative to NE patients), and  $\beta_3$  is the average difference in HD weekly improvement rates for E patients (relative to NE patients). Thus,  $\beta_3$  represents the diagnosis-by-time interaction, indicating the degree to which the time trends vary by diagnostic group. In this augmented model,  $\nu_{0i}$  is the individual's deviation from his or her diagnostic group intercept, and  $\nu_{1i}$  is the individual's deviation from his or her diagnostic group slope. To the degree that the variable  $DX$  is

**Table 12.3** HLM Results for Level 1 Model (3) and Level 2 Model (9)

Parameter	Estimate	SE	<i>z</i>	<i>p</i> <
NE intercept $\beta_0$	22.48	0.79	28.30	.0001
NE slope $\beta_1$	-2.37	0.31	-7.59	.0001
E intercept	1.99	1.07	1.86	.063
difference $\beta_2$				
E slope $\beta_3$	-0.03	0.42	-0.06	.95
difference				
$\sigma_{v_0}^2$	11.64	3.53		
$\sigma_{v_0v_1}$	-1.40	1.00		
$\sigma_{v_1}^2$	2.08	0.50		
$\sigma^2$	12.22	1.11		

NOTE:  $-2 \log L = 2214.94$ 

useful in explaining intercept and slope variation, these individual deviations and their corresponding variances ( $\sigma_{v_0}^2$  and  $\sigma_{v_1}^2$ ) will be reduced. Results for this model are listed in Table 12.3.

A likelihood ratio test comparing this model to the previous one can be used to test the null hypothesis that the diagnosis-related effects (i.e.,  $\beta_2$  and  $\beta_3$ ) are zero. This yields  $\chi_2^2 = 2219.04 - 2214.94 = 4.1$ , which is not statistically significant. Inspection of the estimates in Table 12.3 reveals a marginally significant difference in terms of their initial scores, with endogenous patients about 2 points higher and absolutely no difference in their trends across time. This is also borne out if one compares the variance estimates from Tables 12.2 and 12.3. Notice that the intercept variance has diminished slightly from 12.63 to 11.64 as a result of the marginally significant intercept difference, whereas the slope variance is the same. Taken together, there is no real evidence that the two diagnostic groups differ in terms of their HD scores across time.

### 12.3.3. Curvilinear Growth Model

In many situations, it is too simplistic to assume that the change across time is linear. In the present example, for instance, it may be that the depression scores diminish across time in a curvilinear manner. A curvilinear trend would allow a leveling off of the improvement across time. This is clearly plausible for rating scale data, like the HD scores, where values below zero are impossible. Here, we will consider a curvilinear growth model by adding a quadratic term to the Level 1 model. More general polynomial growth models can also be obtained by adding

**Table 12.4** HLM Results for Level 1 Model (10) and Level 2 Model (11)

Parameter	Estimate	SE	<i>z</i>	<i>p</i> <
$\beta_0$	23.76	0.55	43.04	.0001
$\beta_1$	-2.63	0.48	-5.50	.0001
$\beta_2$	0.05	0.09	0.58	.56
$\sigma_{v_0}^2$	10.44	3.58		
$\sigma_{v_0v_1}$	-0.92	2.42		
$\sigma_{v_1}^2$	6.64	2.75		
$\sigma_{v_0v_2}$	-0.11	0.42		
$\sigma_{v_1v_2}$	-0.94	0.48		
$\sigma_{v_2}^2$	0.19	0.09		
$\sigma^2$	10.52	1.10		

NOTE:  $-2 \log L = 2207.64$ .

cubic terms, quartic terms, and so on to the Level 1 model.

$$y_{ij} = b_{0i} + b_{1i}t_{ij} + b_{2i}t_{ij}^2 + \varepsilon_{ij}. \quad (10)$$

Here,  $b_{0i}$  is the Week 0 HD level for patient  $i$ ,  $b_{1i}$  is the weekly linear change in HD for patient  $i$ , and  $b_{2i}$  is the weekly quadratic change in HD for patient  $i$ . This model can also be written as

$$y_{ij} = b_{0i} + (b_{1i} + b_{2i}t_{ij})t_{ij} + \varepsilon_{ij}$$

to point out that the overall effect of time is  $b_{1i} + b_{2i}t_{ij}$ —namely, it is not constant but changes across time. The Level 2 between-subjects model is now

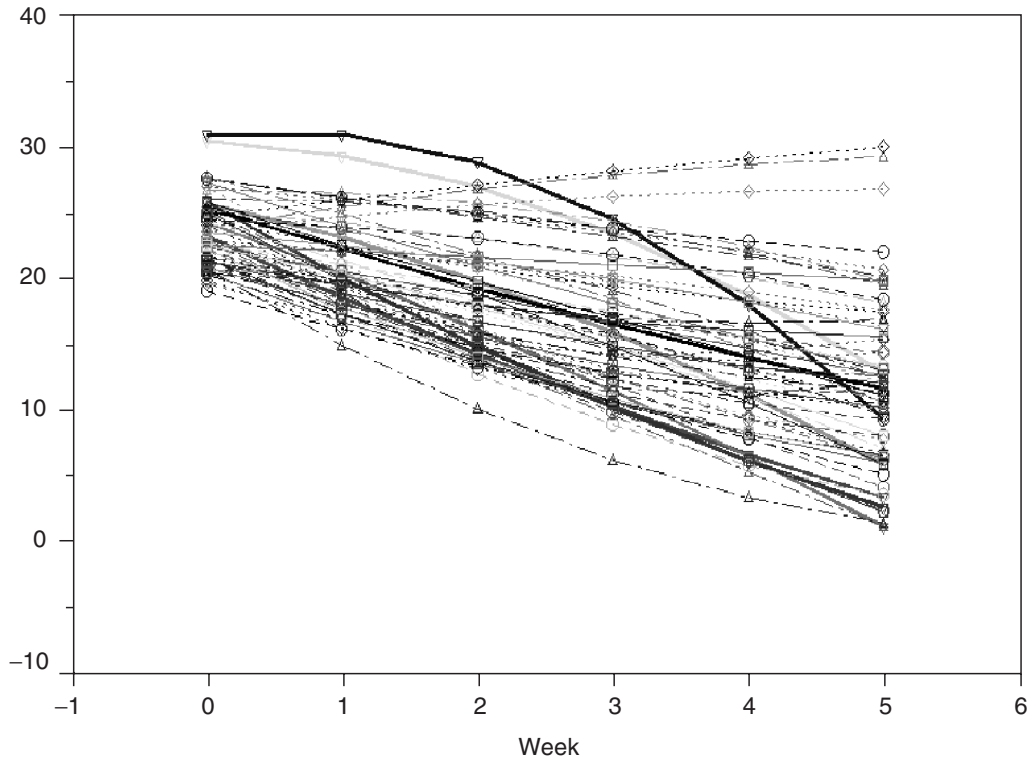
$$\begin{aligned} b_{0i} &= \beta_0 + v_{0i}, \\ b_{1i} &= \beta_1 + v_{1i}, \\ b_{2i} &= \beta_2 + v_{2i}, \end{aligned} \quad (11)$$

where  $\beta_0$  is the average Week 0 HD level,  $\beta_1$  is the average HD weekly linear change, and  $\beta_2$  is the average HD weekly quadratic change. Similarly,  $v_{0i}$  is the individual deviation from average intercept,  $v_{1i}$  is the individual deviation from average linear change, and  $v_{2i}$  is the individual deviation from average quadratic change. Thus, the model allows curvilinearity at both the population ( $\beta_2$ ) and individual ( $v_{2i}$ ) levels.

Fitting this model yields the results given in Table 12.4.

Comparing this model to that of Table 12.1 (i.e., a model with  $\beta_2 = \sigma_{v_2}^2 = \sigma_{v_0v_2} = \sigma_{v_1v_2} = 0$ ) yields a deviance of 11.4, which is statistically significant on 4 degrees of freedom. This is interesting given that the Wald test for  $\beta_2$  is clearly nonsignificant. In fact, comparing the above model to one with  $\sigma_{v_2}^2 = \sigma_{v_0v_2} = \sigma_{v_1v_2} = 0$  (not shown) yields a deviance

**Figure 12.4** Reisby Data: Estimated Curvilinear Trends



of 11.0. Nearly all of the improvement in model fit is through the inclusion of the quadratic term as a random effect and not as a fixed effect. This suggests that although the trend across time is essentially linear at the population level, it is curvilinear at the individual level.

Figure 12.4 contains a plot of the individual trend estimates from this model. These are obtained by calculating  $\hat{y}_{ij} = \hat{b}_{0i} + \hat{b}_{1i}t_{ij} + \hat{b}_{2i}t_{ij}^2$ , for  $t = 0, 1, \dots, 5$ , and then connecting the time point estimates for each individual.

The plot makes apparent the wide heterogeneity in trends across time, as well as the increasing variance in HD scores across time. Some individuals have accelerating downward trends, suggesting a delay in the drug effect. Alternatively, others have decelerating downward trends, which are consistent with a leveling off of the drug effect. Some individuals even have positive trends, indicating a worsening of their depressive symptoms across time. This is not too surprising given that antidepressants, such as imipramine, are known to be ineffective for some patients. The

figure is also interesting in showing that many of the individual trend lines are approximately linear. Thus, the improvement that the curvilinear model provides in describing change across time is perhaps modest.

Finally, the fit of the observed variance-covariance matrix of the repeated measures is provided as follows:

$$\hat{V}(\mathbf{y}) = \mathbf{Z}\hat{\Sigma}_v\mathbf{Z}' + \hat{\sigma}^2\mathbf{I}$$

$$= \begin{bmatrix} 20.96 & & & & & \\ 9.41 & 23.86 & & & & \\ 8.16 & 15.57 & 31.07 & & & \\ 6.68 & 16.08 & 23.11 & 38.31 & & \\ 4.98 & 14.88 & 23.26 & 30.12 & 45.98 & \\ 3.06 & 11.97 & 20.98 & 30.09 & 39.29 & 59.11 \end{bmatrix},$$

where

$$\mathbf{Z}' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 & 5 \\ 0 & 1 & 4 & 9 & 16 & 25 \end{bmatrix}$$

$$\hat{\Sigma}_v = \begin{bmatrix} 10.44 & -0.92 & -0.11 \\ -0.92 & 6.64 & -0.94 \\ -0.11 & -0.94 & 0.19 \end{bmatrix}.$$

By comparing this matrix with the observed variance-covariance matrix, presented earlier, we see that the estimated variances are close to the observed, and the model is clearly picking up the pattern of diminishing covariance away from the diagonal and at the earlier time points. Comparing this model to one with a totally general variance-covariance structure (not shown) yields a likelihood ratio  $\chi^2_{14} = 14.9$ , which is not statistically significant. Thus, this curvilinear model with seven variance-covariance parameters ( $\sigma^2$  and six unique parameters in  $\Sigma_v$ ) provides a parsimonious fit of the variance-covariance matrix  $V(\mathbf{y})$ , which, being of dimension  $6 \times 6$ , has 21 unique elements. More details on methods for assessing and comparing model fit of the variance-covariance structure are described by Wolfinger (1993) and Grady and Helms (1995).

12.3.4. Orthogonal Polynomials

For trend models, it is often beneficial to represent the polynomials in orthogonal form (Bock, 1975). Mathematically, this avoids collinearity problems that can result from using multiples of  $t$  ( $t^2, t^3$ , etc.) as regressors. To see this, consider a curvilinear trend model with three time points. Then,  $t = 0, 1$ , and  $2$ , whereas  $t^2 = 0, 1$ , and  $4$ ; these two variables are nearly perfectly correlated. To counter this, time is sometimes expressed in centered form—for example,  $(t - \bar{t}) = -1, 0$ , and  $1$  and  $(t - \bar{t})^2 = 1, 0$ , and  $1$ . If there is the same number of observations at the three time points, this centering removes the correlation between the linear and quadratic trend components entirely. In the more usual situation of nonequal numbers of observations across time, this greatly diminishes the correlation between the polynomials. Another aspect of centering time is that the meaning of the model intercept changes. In the previous raw form of time, the intercept represented differences at the first time point (i.e., when time = 0). Alternatively, in centered form, the model intercept represents differences at the midpoint of time. For this reason, the intercept is often referred to as the constant or grand mean term in models using centered regressors.

An additional advantage of using orthogonal polynomials, over simply centering time, is that the polynomials are put on the same scale. Thus, their estimated coefficients can be compared in terms of their magnitude in the same way as standardized beta coefficients in ordinary regression analysis. For equal time intervals, tables of orthogonal polynomials can be found in Pearson and Hartley (1976), and Bock (1975) also

**Table 12.5** HLM Results for Orthogonal Polynomial Version of Level 1 Model (10) and Level 2 Model (11)

Parameter	Estimate	SE	z	P <
$\beta_0$	43.24	1.37	31.61	.0001
$\beta_1$	-9.94	0.86	-11.50	.0001
$\beta_2$	0.31	0.54	0.58	.56
$\sigma^2_{v_0}$	111.91	21.60		
$\sigma_{v_0v_1}$	37.99	10.92		
$\sigma^2_{v_1}$	37.04	8.90		
$\sigma_{v_0v_2}$	-10.14	6.19		
$\sigma_{v_1v_2}$	-0.82	3.50		
$\sigma^2_{v_2}$	7.23	3.50		
$\sigma^2$	10.52	1.10		

NOTE:  $-2 \log L = 2207.64$ .

describes how orthogonal polynomials can be obtained for unequal time intervals. For the current situation with six equally spaced time points, these are given as

$$\mathbf{X}' = \mathbf{Z}' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ -5 & -3 & -1 & 1 & 3 & 5 \\ 5 & -1 & -4 & -4 & -1 & 5 \end{bmatrix} \begin{matrix} / \sqrt{6} \\ / \sqrt{70} \\ / \sqrt{84} \end{matrix}$$

Notice that these row vectors are independent of each other. Also, by dividing the values by the square root of the quantities on the right, which are simply the sum of squared values in a row, these polynomials have the same scale. Thus, these terms are simultaneously made independent of each other and standardized to the same (unit) scale. This holds exactly when the number of observations at each time point is equal and approximately so when they are unequal.

Fitting this orthogonal polynomial trend model yields the results given in Table 12.5.

Comparing the regression coefficients, as before, we see that only the constant and linear terms are significant. These terms also dominate in terms of magnitude; not only is the quadratic term nonsignificant, but it is also negligible. Thus, at the population average level, the trend is unquestionably linear. Turning to the variance estimates, we see that the estimated constant variance ( $\hat{\sigma}^2_{v_0}$ ) is much larger than the estimated linear trend component ( $\hat{\sigma}^2_{v_1}$ ), which is much larger than the estimated quadratic trend component ( $\hat{\sigma}^2_{v_2}$ ). In terms of relative percentages, these three represent 71.7, 23.7, and 4.6, respectively, of the sum of the estimated individual variance terms. Thus, at the individual level, there is heterogeneity in terms of all three components but with diminishing return as the order of the polynomial increases. This analysis then quantifies what Figure 12.4 depicts.

Inspection of the covariance terms reveals a strong positive association between the constant and linear terms ( $\hat{\sigma}_{v_1 v_0}^2 = 37.99$ , expressed as a correlation = .59). This seems to be in contrast with the results for this term from the previous analysis in Table 12.4, in which there was a slight negative association between the intercept and linear terms ( $\hat{\sigma}_{v_1 v_0}^2 = -.92$ , expressed as a correlation =  $-.11$ ). The reason for this apparent discrepancy is that in Table 12.4, the intercept represents the first time point, whereas the constant term in Table 12.5 represents the midpoint in time. Thus, an individual's linear trend is both negatively associated with his or her baseline depression level and positively associated with his or her mid-study depression level. Subjects with higher initial depression levels have slightly more negative linear slopes and, as a result, lower values at mid-study.

Finally, notice that the log-likelihood value is identical in Tables 12.5 and 12.6. Thus, the two solutions are equivalent; one is simply a reexpressed version of the other. Because of this, one can derive the results from Table 12.5 based on those from Table 12.6 and vice versa. Because the orthogonal polynomial representation greatly reduces any collinearity and scale differences in the regressors, it is computationally easier to obtain. For this reason, in cases where numerical difficulties are occurring with analyses using raw time values, investigators might consider using orthogonal polynomials instead.

### 12.3.5. Growth Model With Time-Varying Covariates

In this section, we examine the effects of the time-varying drug plasma levels IMI and DMI. Because an inspection of the data indicated that the magnitude of these measurements varied greatly between individuals (from 4 to 312 mg/L for IMI and from 0 to 740 mg/L for DMI), a log transformation is used for these covariates. This helps to ensure that the estimated regression coefficients are not unduly influenced by extreme values on these covariates. Also, these variables,  $\ln$  IMI and  $\ln$  DMI, are expressed in grand-mean centered form so that the model intercept represents HD scores for patients with average drug levels. To obtain the grand-mean centered versions of these variables, we subtract the variable's sample mean from each observation. For notational simplicity in the model equations,  $I_{ij}$  and  $D_{ij}$  will represent the grand-mean centered versions of  $\ln$  IMI and  $\ln$  DMI, respectively, in what follows. Also, whereas the previous models considered HD outcomes from

Weeks 0 to 5, the models of this section only include HD outcome data from Weeks 2 to 5. This is because the drug plasma levels are not available at the first two time points of the study (i.e., Week 0, or baseline, and Week 1, or the end of the drug washout period). Although HLM does allow incomplete data across time, data must be complete within a given time point (in terms of both the dependent variable and covariates) for that time point to be included in the analysis. Thus, the analyses that follow are for the 4-week period following the drug washout period, with  $t_{ij}$  coded as 0, 1, 2, and 3 for these four respective time points. As a result, the intercept represents HD scores for Week 2 of the study (i.e., when  $t_{ij} = 0$ ).

The first Level 1 model is given by

$$y_{ij} = b_{0i} + b_{1i}t_{ij} + b_{2i}I_{ij} + b_{3i}D_{ij} + \varepsilon_{ij}, \quad (12)$$

where  $b_{0i}$  is the Week 2 HD level for patient  $i$  under average levels of both  $\ln$  IMI and  $\ln$  DMI,  $b_{1i}$  is the weekly change in HD for patient  $i$ ,  $b_{2i}$  is the patient's change in HD due to  $\ln$  IMI, and  $b_{3i}$  is the change in HD due to  $\ln$  DMI. The between-subjects model is given as

$$\begin{aligned} b_{0i} &= \beta_0 + \nu_{0i}, \\ b_{1i} &= \beta_1 + \nu_{1i}, \\ b_{2i} &= \beta_2, \\ b_{3i} &= \beta_3, \end{aligned} \quad (13)$$

where  $\beta_0$  is the average Week 2 HD level for patients with average  $\ln$  IMI and  $\ln$  DMI values,  $\beta_1$  is the average HD weekly change,  $\beta_2$  is the average HD difference for a unit change in  $\ln$  IMI, and  $\beta_3$  is the average HD difference for a unit change in  $\ln$  DMI. Also,  $\nu_{0i}$  is the individual intercept deviation, and  $\nu_{1i}$  is the individual slope deviation. Notice that the Level 2 model indicates that the drug effects could also be treated as random. This would be accomplished by adding  $\nu_{2i}$  and  $\nu_{3i}$  to the model and would allow individual variation in terms of the drug-level effect on HD scores. Given that antidepressants such as IMI and DMI are not effective for all individuals, it is plausible that the drug levels are more strongly related to changes in depression for some individuals, whereas for others they are less so. Similarly, one could add individual-level covariates (e.g., endogenous/nonendogenous group) into the models for  $b_{2i}$  and  $b_{3i}$  to examine whether the drug effects vary with individual-level covariates. Again, it is feasible that the drug effects on outcome are stronger for endogenous than nonendogenous patients. Although these possibilities will not be considered

**Table 12.6** HLM results for Level 1 model (12) and Level 2 Model (13)

Parameter	Estimate	SE	z	p <
Intercept $\beta_0$	18.17	0.71	25.70	.0001
Time slope $\beta_1$	-2.03	0.28	-7.15	.0001
In IMI $\beta_2$	0.60	0.85	0.71	.48
In DMI $\beta_3$	-1.20	0.63	-1.90	.06
$\sigma_{v_0}^2$	24.83	5.79		
$\sigma_{v_0v_1}^2$	-0.72	1.74		
$\sigma_{v_1}^2$	2.73	0.95		
$\sigma^2$	10.46	1.37		

NOTE:  $-2 \log L = 1502.5$ .

**Table 12.7** HLM Results for Level 1 Model (14) and Level 2 Model (13)

Parameter	Estimate	SE	z	p <
Intercept $\beta_0$	-5.18	0.66	-7.87	.0001
Slope $\beta_1$	-1.97	0.29	-6.90	.0001
In IMI $\beta_2$	0.63	0.82	0.77	ns
In DMI $\beta_3$	-1.97	0.60	-3.26	.0014
$\sigma_{v_0}^2$	20.50			
$\sigma_{v_0v_1}$	0.84			
$\sigma_{v_1}^2$	2.78			
$\sigma^2$	10.53			

NOTE:  $-2 \log L = 1498.8$ .

here, an example of an HLM allowing such individual variation in relationships is described by Hedeker, Flay, and Petraitis (1996).

Fitting the present model yields the results given in Table 12.6.

It is interesting to note that neither of the drug levels seems to be significantly related to the depression scores across time. However, note that the model given in (12) specifies that a person's drug level is related to his or her depression score at that same time point. It might be more plausible to instead posit that a person's drug level is related to his or her *change* in depression score, or improvement, at that same time point. For this, the following alternative Level 1 model is considered:

$$(y_{ij} - y_{i0}) = b_{0i} + b_{1i}t_{ij} + b_{2i}I_{ij} + b_{3i}D_{ij} + \varepsilon_{ij}, \quad (14)$$

where  $y_{i0}$  is the individual's HD score at baseline (or at Week 1 for those few subjects with a missing baseline score). This yields the results presented in Table 12.7.

Interestingly, now the effect of DMI, the metabolite of IMI, is highly significant and negative. Thus, greater

**Table 12.8** Correlation Between HD Scores and Plasma Levels (Natural Log Units)

Drug	Week 2	Week 3	Week 4	Week 5
HD total score				
IMI	-0.034	-0.038	-0.003	-0.189
DMI	-0.177	-0.075	-0.246	-0.293*
HD change from baseline				
IMI	-0.049	-0.106	-0.046	-0.240
DMI	-0.366*	-0.281*	-0.363*	-0.361*

NOTE: \* $p < .05$ .

DMI values are associated with greater improvement (i.e., more negative HD change scores). However, the parent drug IMI is not significantly related to HD change scores; in fact, its coefficient is positive. It is important to remember that the model estimates the IMI effect, controlling for the DMI effect, and vice versa. These two drug levels are moderately correlated with each other ( $r = .18, .23, .22,$  and  $.18$  for the four respective time points), and so the results above are not necessarily indicative of the marginal relationships of each drug with depression scores. Correlations of the drug plasma levels with the HD scores, both raw and expressed as change scores, are given in Table 12.8. These bear out the fact that the drug levels are much more associated with the HD change scores than the actual scores.

These correlations also show the greater association between HD change scores and DMI, rather than IMI, drug levels.

*12.3.5.1. Within- and Between-Subjects Effects for Time-Varying Covariates*

When time-varying covariates are included in an HLM, as in the manner of the last analysis, an assumption is made that the between- and within-subjects effects of these variables are equal. To see this, express the time-varying covariates  $I_{ij}$  and  $D_{ij}$  as

$$I_{ij} = \bar{I}_i + (I_{ij} - \bar{I}_i),$$

$$D_{ij} = \bar{D}_i + (D_{ij} - \bar{D}_i),$$

where  $\bar{I}_i$  and  $\bar{D}_i$  are the means of these two time-varying covariates computed for each individual. Thus, the first term following the equality represents the individual's mean on the time-varying covariate (i.e., a between-subjects variable), and the second term represents the individual's deviation around his or her

mean (i.e., a within-subjects variable). Including both of these terms into the HLM yields

$$(y_{ij} - y_{i0}) = b_{0i} + b_{1i}t_{ij} + b_{2i}(I_{ij} - \bar{I}_i) + b_{3i}(D_{ij} - \bar{D}_i) + \varepsilon_{ij}, \quad (15)$$

and

$$\begin{aligned} b_{0i} &= \beta_0 + \beta_4\bar{I}_i + \beta_5\bar{D}_i + \nu_{0i}, \\ b_{1i} &= \beta_1 + \nu_{1i}, \\ b_{2i} &= \beta_2, \\ b_{3i} &= \beta_3, \end{aligned} \quad (16)$$

for the Level 1 and Level 2 models. Thus, the total effect of IMI, for example,

$$\beta_2(I_{ij} - \bar{I}_i) + \beta_4\bar{I}_i,$$

is partitioned into its within- and between-subjects effects (i.e.,  $\beta_2$  and  $\beta_4$ , respectively). The between-subjects part indicates the degree to which the individual's average drug level is related to his or her average depression level, averaging across time. In other words, it may be that subjects with consistently high drug levels have consistently low depression scores. Alternatively, the within-subjects component represents the degree to which variation in an individual's drug level is associated with a change in his or her depression scores (i.e., a within-subject change). Thus, it may be that a higher relative drug level for an individual is associated with a lower relative depression score for that individual at a particular time point. If these two are equal ( $\beta_2 = \beta_4$ ), then the IMI effect is

$$\beta_2(I_{ij} - \bar{I}_i) + \beta_2\bar{I}_i = \beta_2 I_{ij},$$

which is exactly what was used in the last analysis. Thus, we implicitly assumed that the within- and between-subjects effects of these two drug levels were the same in the previous analysis. This assumption can be tested by comparing the model specified by (14) and (13) with the more general model of (15) and (16). Table 12.9 includes the results of this latter analysis.

Comparing the two models yields a likelihood-ratio statistic of  $\chi^2_2 = 3.0$ , which is not statistically significant. Thus, the assumption of homogeneity of the between- and within-subjects regressions cannot be rejected for these data. Inspecting the estimated coefficients for DMI supports this:  $-1.8$  and  $-2.4$  for the within- and between-subjects effects, respectively. Conversely, the estimates for IMI are very different and even of the opposite sign. However, neither is statistically significant, and the standard

**Table 12.9** HLM Results for Level 1 Model (15) and Level 2 Model (16)

Parameter	Estimate	SE	z	p <
Intercept $\beta_0$	-5.09	0.66	-7.71	.0001
Slope $\beta_1$	-2.02	0.29	-6.94	.0001
Within In IMI $\beta_2$	2.44	1.46	1.68	.10
Within In DMI $\beta_3$	-1.80	1.00	-1.80	.075
Between In IMI $\beta_4$	-0.31	1.00	-0.31	ns
Between In DMI $\beta_5$	-2.37	0.80	-2.97	.004
$\sigma^2_{\nu_0}$	20.32			
$\sigma_{\nu_0\nu_1}$	0.50			
$\sigma^2_{\nu_1}$	2.83			
$\sigma^2$	10.38			

NOTE:  $-2 \log L = 1495.8$ .

errors for these two IMI estimates are quite large. In conclusion, for these data, there is not sufficient evidence to reject the assumption of equality in the within- and between-subjects effects for these two drug levels.

### 12.3.5.2. Time Interactions With Time-Varying Covariates

In some cases, it can be of substantive interest to examine whether there are interactions between a time-varying covariate and time. For example, one might posit that the relationship between the time-varying covariate and the outcome either increases or decreases across time. This is clearly plausible in the present example because the effectiveness of antidepressants is not thought to be immediate but instead to develop over time (Reisby et al., 1977). Thus, it is of interest to examine the degree to which the effects of the time-varying drug plasma levels on the change in depression scores vary across time. To explore this possibility, we can augment the Level 1 model to include the time interactions, namely,

$$\begin{aligned} (y_{ij} - y_{i0}) &= b_{0i} + b_{1i}t_{ij} + b_{2i}I_{ij} \\ &+ b_{3i}D_{ij} + b_{4i}(I_{ij} \times t_{ij}) \\ &+ b_{5i}(D_{ij} \times t_{ij}) + \varepsilon_{ij}, \end{aligned} \quad (17)$$

with the accompanying Level 2 model,

$$\begin{aligned} b_{0i} &= \beta_0 + \nu_{0i}, \\ b_{1i} &= \beta_1 + \nu_{1i}, \\ b_{2i} &= \beta_2, \\ b_{3i} &= \beta_3, \\ b_{4i} &= \beta_4, \\ b_{5i} &= \beta_5. \end{aligned} \quad (18)$$

To correctly interpret the model parameters, one should remember that the drug levels have been grand-mean centered, that the week variable equals 0 for the second week of the study, and that interpretation of the “main effects” is altered when interactions are present (i.e., they represent the effect of the variable when the interacting variable equals 0). Thus, in this model,  $\beta_0$  represents the average Week 2 HD change score for patients with average drug levels,  $\beta_1$  is the average weekly change in HD change scores for patients with average drug levels,  $\beta_2$  is the HD change score difference for a unit change of  $\ln$  IMI at Week 2, and  $\beta_3$  represents the HD change score difference per unit change of  $\ln$  DMI at Week 2. One can think of  $\beta_2$  as the regression slope corresponding to the plot of HD change scores versus  $\ln$  IMI levels considering Week 2 data only (with the caveat that this regression slope is really a partial regression slope adjusting for the other drug level). Similar comments apply for interpreting  $\beta_3$  in terms of  $\ln$  DMI. Turning to the interactions  $\beta_4$  and  $\beta_5$ , these indicate the per week change in the drug effects on the HD change scores. In terms of the plot analogy, these interactions correspond to the change in (partial) regression slopes associated with separate weekly plots of HD change scores versus drug levels as one goes across the weeks—in other words, how the slope for a given drug varies across time. Finally,  $v_{0i}$  represents the individual intercept deviation, and  $v_{1i}$  is the individual time-slope deviation. Table 12.10 lists the results of this analysis.

Comparing this model to the one without the drug-by-time interaction (i.e., from Table 12.7) yields a likelihood ratio statistic of  $\chi^2_2 = 6.8$ , which is statistically significant at the .05 level. Thus, there is evidence that the drug effects on depression do vary across time. Inspecting the estimates and their test statistics in Table 12.10 reveals that it is DMI, not IMI, that is interacting significantly with time. Specifically, DMI has an initial Week 2 effect that is significant ( $p < .017$ ), indicating that higher levels of DMI are associated with greater improvement on the HD scale at this time point, and this beneficial effect of DMI gets more pronounced across time ( $p < .01$ ). Concretely, the benefit of a one-unit change in  $\ln$  DMI at Week 2 is a 1.5-point reduction on the HD change score, whereas by the last time point, it is a 4.5-point reduction ( $3 \times 1.5$ ).

At first glance, it might seem a bit unusual that the DMI-by-time interaction is so highly significant given the reported correlations in Table 12.8. To better understand this, consider the simple linear regression slopes that are obtained from regressing HD change scores on  $\ln$  DMI values at each of the four time points

**Table 12.10** HLM Results for Level 1 Model (17) and Level 2 Model (18)

Parameter	Estimate	SE	z	p <
Intercept $\beta_0$	-5.12	0.65	-7.82	.0001
Time slope $\beta_1$	-1.94	0.28	-7.04	.0001
$\ln$ IMI $\beta_2$	0.40	0.87	0.46	ns
$\ln$ DMI $\beta_3$	-1.51	0.62	-2.43	.017
$\ln$ IMI by time $\beta_4$	0.16	0.41	0.39	ns
$\ln$ DMI by time $\beta_5$	-0.90	0.34	-2.65	.01
$\sigma^2_{v_0}$	20.24			
$\sigma_{v_0v_1}$	0.99			
$\sigma^2_{v_1}$	2.50			
$\sigma^2$	10.35			

NOTE:  $-2 \log L = 1492.0$ .

separately: These are  $-2.081$ ,  $-2.195$ ,  $-3.370$ , and  $-3.3765$ , respectively. These regression slopes provide clearer evidence of the DMI-by-time interaction, as they increase (in absolute value) more dramatically across time than the analogous correlations in Table 12.8. Why do these two sets of descriptive statistics suggest different conclusions? Remembering that the correlation is essentially a scale-free representation of the slope (i.e.,  $r = \hat{\beta} s_x/s_y$ ), it is clear that the scales of the dependent and independent variables play a role here. Interestingly, the scale of these two go in opposite directions across time; the standard deviations of the HD change scores increase (5.38, 6.51, 7.35, and 7.88 across the four time points), whereas the standard deviations of the  $\ln$  DMI values decrease (.95, .84, .79, and .76 across these same four time points). Thus, the metric for the slopes across time is very different (i.e., the ratio of standard deviations  $s_x/s_y$  equals .18, .13, .11, and .10, respectively), which explains why the simple slopes and correlations are not in such close agreement and why the significant DMI-by-time interaction of the HLM is a bit at odds with the apparent consistent pattern of the correlations across time. As this final HLM and the descriptive statistics make clear, it is the scale-dependent slope of DMI (i.e., how much change in depression is associated with a unit change in this blood level) that is increasing across time, not the scale-free association.

## 12.4. DISCUSSION

As demonstrated, HLM provide a useful way of analyzing longitudinal data. Specifically, HLM allows for the presence of missing data, irregularly spaced measurements across time, time-varying and invariant covariates, accommodation of individual-specific

deviations from the average time trend, and estimation of the population variance associated with these individual effects. In addition, methods and software exist for the analysis of continuous and categorical outcomes. Perhaps the most popular feature of HLM is its treatment of missing data. As has been illustrated, subjects are not assumed to be measured at the same number of time points. Because there are no restrictions on the number of observations per individual, subjects who are missing at a given interview wave are not excluded from the analysis. The assumption of the model is that the data that are available for a given subject are representative of that subject's deviation from the average trends across time (which are estimated based on the whole sample).

A slightly more sophisticated approach for handling missing data is to group subjects based on their available data pattern across time. For example, subjects might be classified as complete-data subjects or incomplete-data subjects. This between-subjects classification variable can then be included in the analysis to examine the degree to which these two types of subjects differ in terms of the outcome variable. Interactions can also be included to see if the treatment group-related effects vary by missing data pattern. This approach has been called *pattern-mixture modeling* by Little (1993, 1994, 1995). Hedeker and Gibbons (1997) illustrate the use of this approach as applied to psychiatric clinical trials data. Verbeke and Molenberghs (2000) further describe the pattern-mixture approach in much greater statistical detail, including how these models can be used to assess the sensitivity of the results to different assumptions about the missing data. Although not applied in this chapter, the pattern-mixture approach provides a further way of dealing with missing data in longitudinal studies.

Statistical software to perform HLM analysis has proliferated, especially for continuous outcomes: HLM 5 (Raudenbush, Bryk, Cheong, & Congdon, 2000), SAS PROC MIXED, MLwiN (Goldstein et al., 1998), and MIXREG (Hedeker & Gibbons, 1996b), to mention a few programs. For categorical data, software has become available for dichotomous (EGRET [CYTEL, 1999]) and ordinal or nominal outcomes (SAS PROC NLMIXED, HLM 5, MLwiN, and GLLAMM [Rabe-Hesketh, Pickles, & Skrondal, 2001]; MIXOR [Hedeker & Gibbons, 1996a]; MIXNO [Hedeker, 1999]). Of course, software for nominal and ordinal outcomes can be used to fit models for dichotomous outcomes. Review articles comparing some of these software programs include van der Leeden, Vrijburg, and de Leeuw (1996) and de Leeuw and Kreft (2001).

This chapter has focused on the modeling aspects of HLM without discussion of parameter estimation. In nearly all of the software programs for continuous outcomes, a combination of two complementary methods has generally been used: empirical Bayes (EB) methods for estimation of the individual effects (e.g.,  $v_{0i}$ ) and maximum likelihood (ML) methods for estimation of variance and covariance parameters (e.g.,  $\sigma^2$ ,  $\sigma_{v_0}^2$ ,  $\sigma_{v_1}^2$ , and  $\sigma_{v_0v_1}$ ) and covariate effects ( $\beta$ ). Iterative solutions to estimate these two sets of parameters have been described using the EM algorithm (Bryk & Raudenbush, 1992; Laird & Ware, 1982) and the Fisher scoring algorithm (Bock, 1989a; Longford, 1987). Because these models are more complex than ordinary fixed-effects regression models, it is sometimes the case that the iterative procedure does not converge to a solution. If this occurs, it is often because the model is overly complex relative to the data being used to estimate it, and so model simplification is necessary. Although it is not always apparent why a particular model does not converge, building models in a sequential piecemeal manner can help to isolate where troubles occur.

In the example, repeated observations were observed nested within individuals. In the terminology of multi-level analysis (Goldstein, 1995) and hierarchical linear models (Raudenbush & Bryk, 2002), this is termed a *two-level data structure*, with individuals representing Level 2 and the nested repeated observations Level 1. The models that we have presented are thus referred to as two-level models. Individuals themselves, though, are often observed clustered within some higher level unit, for example, a classroom, clinic, or worksite. Cross-sectional clustered data can also be considered as two-level data, with the clusters representing Level 2 and the clustered subjects Level 1. Analysis of cross-sectional clustered data using HLM is discussed by Hedeker, Gibbons, and Flay (1994) and Hedeker, McMahon, Jason, and Salina (1994). In some studies, subjects are clustered and also repeatedly measured, resulting in three levels of data: the cluster (Level 3), individual (Level 2), and repeated observation (Level 1). Analysis of three-level data is described in Goldstein (1995), Raudenbush and Bryk (2002), Longford (1993), and Gibbons and Hedeker (1997).

Because longitudinal designs are increasingly used in the social sciences, it is important that statistical methods are developed and used to extract the most out of these longitudinal data sets. HLM provides an attractive approach for addressing some key questions that emerge from longitudinal designs. It is hoped that this chapter has helped in increasing the understanding of

these methods and their potential for use in analyzing longitudinal outcomes.

## REFERENCES

- Agresti, A., & Natarajan, R. (2001). Modeling clustered ordered categorical data: A survey. *International Statistical Review*, *69*, 345–371.
- Albert, P. S. (1999). Longitudinal data analysis (repeated measures) in clinical trials. *Statistics in Medicine*, *18*, 1707–1732.
- Berkhof, J., & Snijders, T. A. B. (2001). Variance component testing in multilevel models. *Journal of Educational and Behavioral Statistics*, *26*, 133–152.
- Bock, R. D. (1975). *Multivariate statistical methods in behavioral research*. New York: McGraw-Hill.
- Bock, R. D. (1983a). The discrete Bayesian. In H. Wainer & S. Messick (Eds.), *Modern advances in psychometric research* (pp. 103–115). Hillsdale, NJ: Lawrence Erlbaum.
- Bock, R. D. (1983b). Within-subject experimentation in psychiatric research. In R. D. Gibbons & M. W. Dysken (Eds.), *Statistical and methodological advances in psychiatric research* (pp. 59–90). New York: Spectrum.
- Bock, R. D. (1989a). Measurement of human variation: A two stage model. In R. D. Bock (Ed.), *Multilevel analysis of educational data* (pp. 319–342). New York: Academic Press.
- Bock, R. D. (Ed.). (1989b). *Multilevel analysis of educational data*. New York: Academic Press.
- Brown, H., & Prescott, R. (1999). *Applied mixed models in medicine*. New York: John Wiley.
- Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.
- Burchinal, M. R., Bailey, D. B., & Snyder, P. (1994). Using growth curve analysis to evaluate child change in longitudinal investigations. *Journal of Early Intervention*, *18*, 403–423.
- Burstein, L. (1980). The analysis of multilevel data in educational research and evaluation. In D. Berliner (Ed.), *Review of research in education* (Vol. 8, pp. 158–233). Washington, DC: American Educational Research Association.
- Campbell, S. K., & Hedeker, D. (2001). Validity of the test of infant motor performance for discriminating among infants with varying risks for poor motor outcome. *Journal of Pediatrics*, *139*, 546–551.
- Carroll, K. M., Rounsaville, B. J., Nich, C., Gordon, L. T., Wirtz, P. W., & Gawin, F. (1994). One-year follow-up of psychotherapy and pharmacotherapy for cocaine dependence. *Archives of General Psychiatry*, *51*, 989–997.
- Chi, E. M., & Reinsel, G. C. (1989). Models for longitudinal data with random effects and AR(1) errors. *Journal of the American Statistical Society*, *84*, 452–459.
- Cnaan, A., Laird, N. M., & Slasor, P. (1997). Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. *Statistics in Medicine*, *16*, 2349–2380.
- Collins, L. M., & Sayer, A. G. (Eds.). (2001). *New methods for the analysis of change*. Washington, DC: American Psychological Association.
- Curran, P. J., Stice, E., & Chassin, L. (1997). The relation between adolescent and peer alcohol use: A longitudinal random coefficients model. *Journal of Consulting and Clinical Psychology*, *65*, 130–140.
- CYTEL. (1999). *Egret for Windows*. Cambridge, MA: Author.
- Davis, C. S. (2002). *Statistical methods for the analysis of repeated measurements*. New York: Springer.
- de Leeuw, J., & Kreft, I. (1986). Random coefficient models for multilevel analysis. *Journal of Educational Statistics*, *11*, 57–85.
- de Leeuw, J., & Kreft, I. (2001). Software for multilevel analysis. In A. H. Leyland & H. Goldstein (Eds.), *Multilevel modelling of health statistics* (pp. 187–204). New York: John Wiley.
- Delucchi, K., & Bostrom, A. (1999). Small sample longitudinal clinical trials with missing data: A comparison of methods. *Psychological Methods*, *4*, 158–172.
- Dempster, A. P., Rubin, D. B., & Tsutakawa, R. K. (1981). Estimation in covariance component models. *Journal of the American Statistical Society*, *76*, 341–353.
- Diggle, P., Liang, K.-Y., & Zeger, S. L. (1994). *Analysis of longitudinal data*. New York: Oxford University Press.
- Elkin, I., Gibbons, R. D., Shea, M. T., Sotsky, S. M., Watkins, J. T., Pilkonis, P. A., & Hedeker, D. (1995). Initial severity and differential treatment outcome in the NIMH treatment of depression collaborative research program. *Journal of Consulting and Clinical Psychology*, *63*, 841–847.
- Everitt, B. S. (1998). Analysis of longitudinal data: Beyond MANOVA. *British Journal of Psychiatry*, *172*, 7–10.
- Fitzmaurice, G. M., Laird, N. M., & Rotnitzky, A. G. (1993). Regression models for discrete longitudinal responses. *Statistical Science*, *8*, 284–309.
- Gallagher, T. J., Cottler, L. B., Compton, W. M., & Spitznagel, E. (1997). Changes in HIV/AIDS risk behaviors in drug users in St. Louis: Applications of random regression models. *Journal of Drug Issues*, *27*, 399–416.
- Gibbons, R. D., & Hedeker, D. (1994). Application of random-effects probit regression models. *Journal of Consulting and Clinical Psychology*, *62*, 285–296.
- Gibbons, R. D., & Hedeker, D. (1997). Random effects probit and logistic regression models for three-level data. *Biometrics*, *53*, 1527–1537.
- Gibbons, R. D., & Hedeker, D. (2000). Application of mixed-effects models in biostatistics. *Sankhya, Series B*, *62*, 70–103.
- Gibbons, R. D., Hedeker, D., Elkin, I., Watermaux, C. M., Kraemer, H. C., Greenhouse, J. B., et al. (1993). Some conceptual and statistical issues in analysis of longitudinal psychiatric data. *Archives of General Psychiatry*, *50*, 739–750.
- Gibbons, R. D., Hedeker, D., Watermaux, C. M., & Davis, J. M. (1988). Random regression models: A comprehensive approach to the analysis of longitudinal psychiatric data. *Psychopharmacology Bulletin*, *24*, 438–443.
- Goldstein, H. (1995). *Multilevel statistical models* (2nd ed.). New York: Halstead.
- Goldstein, H., Rasbash, J., Plewis, I., Draper, D., Browne, W., Yang, M., et al. (1998). *A user's guide to MLwiN*. London: Institute of Education, University of London.
- Grady, J. J., & Helms, R. W. (1995). Model selection techniques for the covariance matrix for incomplete longitudinal data. *Statistics in Medicine*, *14*, 1397–1416, 1995.
- Halikas, J. A., Crosby, R. D., Pearson, V. L., & Graves, N. M. (1997). A randomized double-blind study of carbamazepine

- in the treatment of cocaine abuse. *Clinical Pharmacology and Therapeutics*, 62, 89–105.
- Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology and Neurosurgical Psychiatry*, 23, 56–62.
- Hand, D., & Crowder, M. (1996). *Practical longitudinal data analysis*. New York: Chapman & Hall.
- Hedeker, D. (1989). *Random regression models with autocorrelated errors*. Unpublished doctoral dissertation, University of Chicago, Department of Psychology.
- Hedeker, D. (1999). MIXNO: A computer program for mixed-effects nominal logistic regression. *Journal of Statistical Software*, 4(5), 1–92.
- Hedeker, D., Flay, B. R., & Petraitis, J. (1996). Estimating individual differences of behavioral intentions: An application of random-effects modeling to the theory of reasoned action. *Journal of Consulting and Clinical Psychology*, 64, 109–120.
- Hedeker, D., & Gibbons, R. D. (1996a). MIXOR: A computer program for mixed-effects ordinal probit and logistic regression analysis. *Computer Methods and Programs in Biomedicine*, 49, 157–176.
- Hedeker, D., & Gibbons, R. D. (1996b). MIXREG: A computer program for mixed-effects regression analysis with autocorrelated errors. *Computer Methods and Programs in Biomedicine*, 49, 229–252.
- Hedeker, D., & Gibbons, R. D. (1997). Application of random-effects pattern-mixture models for missing data in longitudinal studies. *Psychological Methods*, 2, 64–78.
- Hedeker, D., Gibbons, R. D., & Flay, B. R. (1994). Random-effects regression models for clustered data: With an example from smoking prevention research. *Journal of Consulting and Clinical Psychology*, 62, 757–765.
- Hedeker, D., McMahon, S. D., Jason, L. A., & Salina, D. (1994). Analysis of clustered data in community psychology: With an example from a worksite smoking cessation project. *American Journal of Community Psychology*, 22, 595–615.
- Hedeker, D., & Mermelstein, R. J. (1996). Application of random-effects regression models in relapse research. *Addiction*, 91(Suppl.), S211–S229.
- Hedeker, D., & Mermelstein, R. J. (2000). Analysis of longitudinal substance use outcomes using random-effects regression models. *Addiction*, 95(Suppl. 3), S381–S394.
- Hox, J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Lawrence Erlbaum.
- Hui, S. L., & Berger, J. O. (1983). Empirical Bayes estimation of rates in longitudinal studies. *Journal of the American Statistical Association*, 78, 753–759.
- Huttenlocher, J. E., Haight, W., Bryk, A. S., & Seltzer, M. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology*, 27, 236–248.
- Kaplan, D., & George, R. (1998). Evaluating latent growth models through ex post simulation. *Journal of Educational and Behavioral Statistics*, 23, 216–235.
- Keselman, H. J., Algina, J., Kowalchuk, R. K., & Wolfinger, R. D. (1999). A comparison of recent approaches to the analysis of repeated measurements. *British Journal of Mathematical and Statistical Psychology*, 52, 63–78.
- Laird, N. M. (1988). Missing data in longitudinal studies. *Statistics in Medicine*, 7, 305–315.
- Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, 38, 963–974.
- Lesaffre, E., Asefa, M., & Verbeke, G. (1999). Assessing the goodness-of-fit of the Laird and Ware model—an example: The Jimma infant survival differential longitudinal study. *Statistics in Medicine*, 18, 835–854.
- Leyland, A. H., & Goldstein, H. (Eds.). (2001). *Multilevel modelling of health statistics*. New York: John Wiley.
- Little, R. J. A. (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association*, 88, 125–133.
- Little, R. J. A. (1994). A class of pattern-mixture models for normal incomplete data. *Biometrika*, 81, 471–483.
- Little, R. J. A. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association*, 90, 1112–1121.
- Longford, N. T. (1987). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. *Biometrika*, 74, 817–827.
- Longford, N. T. (1993). *Random coefficient models*. New York: Oxford University Press.
- Manor, O., & Kark, J. D. (1996). A comparative study of four methods for analysing repeated measures data. *Statistics in Medicine*, 15, 1143–1159.
- Moskowitz, D. S., & Hershberger, S. L. (Eds.). (2002). *Modeling intraindividual variability with repeated measures data*. Mahwah, NJ: Lawrence Erlbaum.
- Niaura, R., Spring, B., Borrelli, B., Hedeker, D., Goldstein, M. G., Keuthen, N., et al. (2002). Multicenter trial of fluoxetine as an adjunct to behavioral smoking cessation treatment. *Journal of Consulting and Clinical Psychology*, 70, 887–896.
- Omar, R. Z., Wright, E. M., Turner, R. M., & Thompson, S. G. (1999). Analysing repeated measures data: A practical comparison of methods. *Statistics in Medicine*, 18, 1587–1603.
- Pearson, E. S., & Hartley, H. O. (1976). *Biometrika tables for statisticians* (Vol. 1). London: Biometrika Trust.
- Pendergast, J. F., Gange, S. J., Newton, M. A., Lindstrom, M. J., Palta, M., & Fisher, M. R. (1996). A survey of methods for analyzing clustered binary response data. *International Statistical Review*, 64, 89–118.
- Rabe-Hesketh, S., Pickles, A., & Skrondal, A. (2001). GLLAMM: A class of models and a Stata program. *Multilevel Modelling Newsletter*, 13, 17–23.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models* (2nd ed.). Thousand Oaks, CA: Sage.
- Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., & Congdon, R. (2000). *HLM 5: Hierarchical linear and nonlinear modeling*. Chicago: Scientific Software International.
- Reisby, N., Gram, L. F., Bech, P., Nagy, A., Petersen, G. O., Ortmann, J., et al. (1977). Imipramine: Clinical effects and pharmacokinetic variability. *Psychopharmacology*, 54, 263–272.
- Seiden, L. S., & Dykstra, L. A. (1977). *Psychopharmacology: A biochemical and behavioral approach*. New York: Van Nostrand Reinhold.
- Serretti, A., Lattuada, E., Zanardi, R., Franchini, L., & Smeraldi, E. (2000). Patterns of symptom improvement during antidepressant treatment of delusional depression. *Psychiatry Research*, 94, 185–190.
- Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis*. New York: Oxford University Press.

- Strenio, J. 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.
- Sullivan, L. M., Dukes, K. A., & Losina, E. (1999). An introduction to hierarchical linear modelling. *Statistics in Medicine*, *18*, 855–888.
- van der Leeden, R., Vrijburg, K., & de Leeuw, J. (1996). A review of two different approaches for the analysis of growth data using longitudinal mixed linear models. *Computational Statistics and Data Analysis*, *21*, 583–605.
- Verbeke, G., & Molenberghs, G. (2000). *Linear mixed models for longitudinal data*. New York: Springer.
- Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society*, *54*, 426–482.
- Willner, P. (1985). *Depression: A psychobiological synthesis*. New York: John Wiley.
- Wolfinger, R. D. (1993). Covariance structure selection in general mixed models. *Communications in Statistics, Simulation and Computation*, *22*, 1079–1106.
- Zeger, S. L., & Liang, K.-Y. (1992). An overview of methods for the analysis of longitudinal data. *Statistics in Medicine*, *11*, 1825–1839.