CHAPTER EIGHT

Mixed-Effects Regression Models with Heterogeneous Variance: Analyzing Ecological Momentary Assessment (EMA) Data of Smoking

Donald Hedeker
Robin J. Mermelstein
University of Illinois at Chicago

Longitudinal studies are increasingly common in psychological and social sciences research. In these studies, subjects are measured repeatedly across time and interest often focuses on characterizing their growth or development across time. Mixed-effects regression models (MRMs) have become the method of choice for modeling of longitudinal data; variants of MRMs have been developed under a variety of names: Random-effects models. Laird and Ware (1982),variance component models (Dempster, Rubin, & Tsutakawa, 1981) , multilevel models (Goldstein, 1995), hierarchical linear models (Bryk & Raudenbush, 1992), two-stage models. Bock (1989), random coefficient models (Leeuw & Kreft, 1986), mixed models (Longford, 1987; Wolfinger, 1993), empirical Bayes models (Hui & Berger, 1983; Strenio, Weisberg, & Bryk, 1983), and random regression models (Bock, 1983b, 1983a; Gibbons, Hedeker, Waternaux, & Davis, 1988). A basic characteristic of these models is the inclusion of random subject effects into regression models in order to account for the influence of subjects on their repeated observations. These random effects reflect each person's growth or development across time, and explain the correlational structure of the longitudinal data. Additionally, they indicate the degree of subject variation that exists in the population of subjects.

There are several features that make MRMs especially useful in longitudinal research. First, subjects are not assumed to be measured on the same number of timepoints, 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 (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 MRMs, subjects do not have to be measured at the same timepoints. This is useful for analysis of longitudinal studies where 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, MRMs can also estimate change for each subject. These estimates of individual change across time can be particularly useful in longitudinal studies where a proportion of subjects exhibit change across time that deviates from the average trend.

In developmental research, MRMs have been used to describe and statistically compare growth or development across groups of subjects. For example, Huttenlocher, Haight, Bryk, and Seltzer (1991) used MRMs in studying gender differences in early vocabulary development. Additionally, these models have been used to examine the effects of contextual variables on growth or changes over time. In this regard, Neff and Karney (2004) examined (time-varying) negative stressors and their effect on marital satisfaction during the first four years of marriage. Other applications of MRMs can be found in many 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) to name a few.

Typically, statistical tests of the regression coefficients (i.e., the fixed effects) of the model are of primary interest. For example, the effect of gender on growth, or the time-varying effect of stress on satisfaction. Here, one tests whether or not the regression coefficients, which indicate the influence of the independent variables on the dependent variable, equal zero in the population (i.e., have zero slope). Usually, the error variance, which characterizes the within-subjects variance, and the variance parameters of the random effects, which characterize the between-subjects variance, are treated as being homogeneous across subject groups. However, longitudinal designs can allow relaxation of these homogeneity of variance assumptions, and indeed allow researchers the ability to model differences in variances, both between and within, across subject groups. The study of intraindividual variabiliity has received increasing attention in psychology(Fleeson, 2004; Hertzog & Nesselroade, 2003; Martin & Hofer, 2004; Nesselroade & Boker, 1994; Nesselroade & Schmidt McCollam, 2000; Nesselroade, 2001, 2004); these articles describe many

of the conceptual issues and some statistical approaches for examining such variation. MRMs can be used to broaden this study by assessing the determinants of both intraindividual (within-subjects) and interindividual (between-subjects) variation.

For example, in smoking research a common theme is that physical and subjective emotional reactions to smoking stabilize as one's experience with smoking increases. Indeed, one aspect of the concept of dependence is that responses to smoking become more internally stable or driven, and less dependent on external or situational contexts. To examine this issue, we present analyses of data from a longitudinal study of adolescent smoking. This study contains multilayered longitudinal data in that subjects are measured across three measurement waves, and at each wave data from 7 days are collected from each subject using hand-held computers ("ecological momentary assessments" or "real-time data capture"). This type of design follows the "bursts of measurement" approach described by Nesselroade (1991), and allows us to address several issues in terms of the stability of variance parameters, both at a given wave as well as across time. For a given measurement wave, we examine how the variances in these reactions to smoking vary across groups of subjects characterized by their smoking history. We also explore this issue longitudinally, across waves, to examine the degree to which variances change as adolescents progress in their smoking career. Standard software (e.g., SAS PROC MIXED) can be used to fit these models; several syntax examples are available from the first author on request. However, before we describe these heterogeneous variance models, which are the focus of this chapter, we begin with a basic introduction to MRMs for longitudinal data analysis. A more complete introduction can be found in Hedeker (2004).

MRMs FOR LONGITUDINAL DATA

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

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + \epsilon_{ij}. \tag{1}$$

This model represents the regression of the outcome variable y on the independent variable time (denoted t). The subscripts indicate whose observation it is (subscript i) and the relative timing of the observation (the subscript j). The actual timing is represented by the independent variable t which may represent time in weeks, months, etc. Both y and t carry the i and j subscripts, and so they are allowed to vary both by individuals and occasions. In a linear regression model, like Equation 1, the errors ϵ_{ij} are assumed to be normally and *independently* distributed in the population with zero mean and common variance σ^2_ϵ. This assumption of independence is generally unreasonable for longitudinal data. Instead, it is much more likely to assume that errors within an individual are correlated. Thus, individual-specific effects are added to the model to account for this

data dependency, as in

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + \upsilon_{0i} + \varepsilon_{ij} \ , \tag{2}$$

where the additional term υ_{0i} indicates the influence of individual i on his or her repeated observations. Specifically, β_0 is the overall population intercept, υ_{0i} is the intercept deviation for subject i, and β_1 is the overall population slope (i.e., the effect of time). Thus, individuals deviate from the regression of y on t in a parallel manner in this model.

Because individuals in a sample are usually thought to be representative of a larger population of individuals, the individual-specific effects υ_{0i} are treated as random effects. This population distribution is usually assumed to be a normal distribution with mean 0 and variance σ^2_υ. With the random effects υ_{0i} in Equation 2, the errors ϵ_{ij} are now assumed to be normally and *conditionally independently* distributed in the population with zero mean and common variance σ^2_ϵ. That is, the errors are independent conditional on the random individual-specific effects υ_{0i}. As 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).

This variance σ^2_υ represents the between-subjects variance and indicates the degree of heterogeneity in the population of subjects. In contrast, the residual variance σ^2_ϵ is the within-subjects variance. It is often of interest to express the between-subjects variance in terms of an intraclass correlation (ICC), namely,

$$ICC = \frac{\sigma^2_\upsilon}{\sigma^2_\upsilon + \sigma^2_\epsilon}. \tag{3}$$

This ratio of the between-subjects variance σ^2_υ to the total variance $\sigma^2_\upsilon + \sigma^2_\epsilon$ represents the degree of association of the longitudinal data within subjects. Specifically, it indicates the proportion of variance in the data, conditional on the model covariates, that is attributable to individuals. As the heterogeneity in the population of subjects increases, so does the ICC. Conversely, as subjects are more similar to each other, the ICC diminishes.

An extension of the random-intercepts model that is popular for longitudinal data, is the random trend model. In this model, subjects deviate in terms of the intercept and the trend across time:

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + \upsilon_{0i} + \upsilon_{1i} t_{ij} + \epsilon_{ij} \tag{4}$$

where, β_0 is the overall population intercept, β_1 is the overall population slope, υ_{0i} is the intercept deviation for subject i, υ_{1i} is the slope deviation for subject i, and ϵ_{ij} is an independent error term distributed normally with mean 0 and variance σ^2_ϵ. The errors are independent conditional on both υ_{0i} and υ_{1i}. With two random individual-specific effects, the population distribution of intercept and slope deviations is assumed to be a bivariate normal $\mathcal{N}(0, \mathbf{\Sigma}_\upsilon)$, where $\mathbf{\Sigma}_\upsilon$ is a

2×2 variance–covariance matrix. This model can be thought of as a personal trend or development model because it represents the measurements of y as a function of time, both at the individual (υ_{0i} and υ_{1i}) and population (β_0 and β_1) levels. The intercept parameters indicate the starting point, and the slope parameters indicate the degree of change over timepoints. The population intercept and slope parameters represent the trend for the population, whereas the individual parameters express how the individual deviates from the population trend.

Whereas the random-intercept model posits that the between-subjects variance σ^2_υ is constant across time, the random trend model allows this variance to change across time, since for a particular timepoint t the between-subjects variance equals:

$$\sigma^2_{\upsilon_0} + 2t\sigma_{\upsilon_0\upsilon_1} + t^2\sigma^2_{\upsilon_1}. \tag{5}$$

Notice that if both $\sigma_{\upsilon_0\upsilon_1}$ and $\sigma^2_{\upsilon_1}$ are positive, then the between-subjects variance increases across time. Diminishing variance across time is also possible if, for example, $-2\sigma_{\upsilon_0\upsilon_1} > \sigma^2_{\upsilon_1}$.

Often a researcher is interested in assessing the influence of covariates, such as treatment group, on the responses across time. For this, covariates that either do not change over time (time invariant) or that vary across measured occasions (time varying) can be added to the model:

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 x_i + \beta_3 x_{ij} + \upsilon_{0i} + \upsilon_{1i} t_{ij} + \epsilon_{ij} \ . \tag{6}$$

Here, β_2 is the coefficient for the time invariant covariate x_i, and β_3 is the coefficient for the time varying covariate x_{ij}. Interactions between the covariates can be included in the same way as interactions are included into an ordinary multiple regression model. For example, x_i might represent the treatment group that a subject is assigned to (for the course of the study), and x_{ij} might be the treatment by time interaction that is obtained as the product of x_i by t_{ij}. Similarly, contextual variables that are time invariant or time varying are easily added to the model.

Thus far, we have only allowed for a linear time effect, but one could also have higher order polynomials in the model (e.g., quadratic and cubic trends) to model nonlinear changes across time. Additionally, these trends can be treated as random effects to further allow nonlinearity in terms of both the population and individual trends. The complexity of the model for time clearly depends on the number of timepoints. With few timepoints (e.g, three or four) a linear effect for time might be reasonable, whereas if there are many timepoints (e.g., five or more), this would be less plausible. More general models of time are also possible (Wu & Zhang, 2002), as are models that allow for random effects for other time-varying predictors; an example of this type of random coefficients model for psychological data is presented in Hedeker, Flay, and Petraitis (1996). In the longitudinal example presented in this chapter there are only three timepoints, and so only a linear time effect is considered.

HETEROGENEOUS VARIANCE MODELS

Most applications of MRMs focus on estimation and testing of the regression coefficients. In this section, we focus on extensions of the basic random intercepts model that allows estimation and testing of variance parameters. Specifically, we show how the model can be extended to allow heterogeneous within- and between-subjects variance. We first consider the situation of two groups of subjects with heterogeneous variances, and then extend to situations of multiple groups. Additionally, these same models can be used to examine whether the variation in the dependent variable changes or is constant across groups of observations (e.g., different contexts).

Varying ICCs by Groups

Suppose that we are considering a random-intercept model, but there is interest in allowing the ICC to vary by groups of subjects, e.g., by gender. For this, let M_i and F_i represent indicator variables (i.e., variables coded either 0 or 1) for males and females, respectively. Then to allow the between-subjects variance to vary across gender groups, these two indicators would be treated as random effects. Additionally, because males and females are distinct groups of subjects, these two random effects need to be specified as independent, that is,

$$\mathbf{\Sigma}_\upsilon = \begin{bmatrix} \sigma^2_{\upsilon_M} & 0 \\ 0 & \sigma^2_{\upsilon_F} \end{bmatrix} . \tag{7}$$

The two between-subjects variance parameters $\sigma^2_{\upsilon_M}$ and $\sigma^2_{\upsilon_F}$ would indicate the degree of heterogeneity in the population of males and females, respectively. Similarly, one can also specify gender-varying within-subject variance using these same dummy-codes,

$$\sigma^2_\epsilon = M_i\sigma^2_{\epsilon_M} + F_i\sigma^2_{\epsilon_F} . \tag{8}$$

Now, there are two ICCs — one for each gender group:

$$\begin{aligned} \text{ICC} &= \sigma^2_{\upsilon_M}/(\sigma^2_{\upsilon_M} + \sigma^2_{\epsilon_M}) \quad \text{for Males} \\ &= \sigma^2_{\upsilon_F}/(\sigma^2_{\upsilon_F} + \sigma^2_{\epsilon_F}) \quad \text{for Females.} \end{aligned}$$

Comparing such a model that allows heterogeneous variances to a model that does not, allows one to test whether the within- and between-subjects variance varies by group or not. In other words, are the within- and between-subjects variances a function of group, or are they constant across groups. Additionally, this kind of heterogeneous ICC model can easily be generalized to more than two groups. In this situation, with k groups, there would be k between-subjects variance parameters and k within-subjects parameters.

Trends Across Groups

When there are more than two groups, it may be of interest and/or parsimonious to estimate a trend in the variances across the k groups. This makes sense if the grouping variable reflects some kind of ordering of subjects. For instance, suppose that a grouping variable g_i is ordered as 0 = low, 1 = med, and 2 = high in terms of some attribute. In our example, we consider three ordered groups based on their previous smoking history. To allow the between-subjects variance to increase across these three groups, one could specify an intercept and the variable g_i as two independent random effects. Here,

$$\mathbf{\Sigma}_v = \begin{bmatrix} \sigma^2_{v_0} & 0 \\ 0 & \sigma^2_{v_g} \end{bmatrix} \tag{9}$$

where $\sigma^2_{v_0}$ represents the intercept variance and $\sigma^2_{v_g}$ reflects how the between-subjects variance varies across groups. Note that the between-subjects (BS) variance is equal to a function of these two parameters,

$$\text{BS variance} \;=\; \sigma^2_{v_0} + g_i^2 \sigma^2_{v_g} \tag{10}$$

which shows why the coding of g_i is important. For example, with three groups

$$\begin{aligned} \text{BS variance} &= \sigma^2_{v_0} \\ &= \sigma^2_{v_0} + \sigma^2_{v_g} \\ &= \sigma^2_{v_0} + 4\sigma^2_{v_g} \end{aligned}$$

for the three groups coded as $g_i = 0, 1, 2$ respectively. This coding of g_i assumes that the variance increases across groups, because variances can never be negative. Reverse coding (i.e., 2, 1, 0) can be used if the variance decreases across groups. It is also possible to use non-integer coding for g, e.g., $\sqrt{g_i} = 0, 1, \sqrt{2}, \sqrt{3}, \ldots$, to permit a strict linear increase of the BS variance in variance units, namely, BS variance $= \sigma^2_{v_0} + g_i \sigma^2_{v_g}$.

Incorporating varying between-subjects variance is easily accomplished within the mixed model. The primary feature is that the random effects (i.e., either the group dummy codes, or the intercept and the group variable g_i) need to be treated as independent, since subjects belong to only one of the subject groups. Most popular software programs for mixed model analysis (e.g., SAS PROC MIXED, SPSS MIXED, MLwiN, HLM) allow these specifications.

For estimating a trend in the within-subjects (WS) variance across subject groups, a log-linear representation is often used in ordinary multiple regression (Aitkin, 1987; Harvey, 1976), and implemented within a mixed model in SAS PROC MIXED. Here, in terms of the grouping variable g_i, this representation

would be

$$\text{WS variance} \;=\; \sigma_\epsilon^2 \exp(g_i \tau) \; . \tag{11}$$

Notice that the WS variance would equal σ_ϵ^2 for the first group ($g_i = 0$), $\sigma_\epsilon^2 \exp \tau$ for $g_i = 1$, and $\sigma_\epsilon^2 \exp 2\tau$ for $g_i = 2$. This representation thus allows for both increasing and decreasing variance across groups. For example, if $\tau > 0$ then variance increases over groups, while if $\tau < 0$ then it decreases over the groups. Including trends in both the BS and WS variance across groups, the ICC for a particular group, coded g_i, would equal:

$$\text{ICC} = \frac{\sigma_{v_0}^2 + g_i^2 \sigma_{v_g}^2}{\sigma_{v_0}^2 + g_i^2 \sigma_{v_g}^2 + \sigma_\epsilon^2 \exp(g_i \tau)} \; . \tag{12}$$

The log-linear modeling of the WS variance allows it to depend on more than one variable, namely,

$$\text{WS variance} \;=\; \sigma_\epsilon^2 \exp(\boldsymbol{w}_{ij}' \boldsymbol{\tau}) \; . \tag{13}$$

where $\boldsymbol{w}_{ij}$ is a $s \times 1$ vector of variables influencing the within-subjects variance. This regression-like structure allows estimation of separate WS variance by group, as well as more complicated multiple regression-like forms for the WS variance. In this way, one can examine whether contextual variables are related to the WS variance. For example, it may be the case that the variation in responses to smoking is very different for adolescents depending on whether or not they are smoking alone or with friends.

By combining modeling of the WS variance with the inclusion of random effects in the mixed model (to model the BS variance), a wide variety of heterogeneous variance models can be estimated and compared. The illustrations that follow give a sense of some of these possibilities and their usefulness in psychological research.

ILLUSTRATION: ADOLESCENT SMOKING STUDY

Data for the analyses reported here come from a longitudinal, natural history study of adolescent smoking (Mermelstein, Hedeker, Flay, & Shiffman, 2002). Students included in the longitudinal study were either in grade 8 or 10 at baseline, and self-reported on a screening questionnaire 6 to 8 weeks prior to baseline that they either had never smoked, but indicated a probability of future smoking, or had smoked in the past 90 days, but had not smoked more than 100 cigarettes in their lifetime. Written parental consent and student assent were required for participation. A total of 562 students completed the baseline measurement wave. The longitudinal study utilized a multimethod approach to assess adolescents at three time points: Baseline, 6 months, and 12 months. The data collection modalities included self-report questionnaires, a week-long time/event sampling method via palmtop computers (Ecological Momentary Assessments), and in-depth interviews.

Data for the analyses presented here came from the ecological momentary assessments. Adolescents carried the hand held computers with them at all times during the 7 consecutive day data collection period at each wave and were trained to both respond to random prompts from the computers and to event record (initiate a data collection interview) smoking episodes. Immediately after smoking a cigarette, participants completed a series of questions on the hand held computers. Questions included ones about place, activity, companionship, mood, and other subjective items. The hand held computers date and time-stamped each entry. For inclusion in the analyses reported here, adolescents must have smoked at least two cigarettes during the 7-day baseline data collection period; 100 adolescents met this inclusion criterion. We used a cutpoint of two smoking reports because of our interest in modeling variation.

These 100 adolescents began the study with varying amounts of cigarette smoking experience. Adolescents were divided into three groups based on their lifetime smoking levels: Those who had smoked less than 6 cigarettes in their lifetimes (n = 18), representing very novice smokers; Those who had smoked between 6 and 99 cigarettes in their lifetimes ($n = 48$), representing a group of irregular or experimental smokers; and those who had smoked 100 or more cigarettes during their lifetimes ($n = 34$), representing more regular smokers. This (trichotomous) ordered smoking history variable is the grouping variable used in the first set of analyses. We chose not to treat smoking history as a continuous variable because the effect of this variable is not thought to be the same across its levels. That is, the effect of lifetime smoking history in adolescents is substantively not the same contrasting, say, individuals who have smoked 5 and 6 cigarettes versus those who have smoked 100 and 101 cigarettes. Of course, higher order polynomials could be used to model such a nonlinear effect, but the trichotomous treatment of this variable is simpler to interpret and is based on useful, though somewhat arbitrary, cutpoints of this smoking history variable for adolescents.

The dependent variable concerns the responses to questions asking about subjective physiological sensations immediately after and prior to smoking a cigarette. Specifically, the subjects rated their subjective physiological sensations in terms of two items, "*Sick*" and "*Buzzed*," on an analog ladder-type scale, by moving a stylus to the appropriate point on the ladder scale. Immediately after smoking the cigarette, subjects turned on their hand-held computer to complete a variety of questions. They were first asked about their moods and feelings "right now" (after smoking) and later about how they felt just before smoking. The definition of these items varied slightly for the before and after assessments. Specifically, for the after assessment, they were prompted as:

- Sick: Think about how you feel right now: Do you feel sick?
- Buzzed: Think about how you feel right now: Do you feel buzzed?

whereas, for the before assessment the prompts were:

- Sick: Now think about the time just before you smoked: I felt sick.
- Buzzed: Now think about the time just before you smoked: I felt buzzed.

For all four questions, subjects rated their feelings on the 1 (*not at all*) to 10 (*very*) ladder. Correlations of these two items were modest (.24 for sick and buzzed after responses, and .36 for sick and buzzed before responses), however separate analyses of these two items did not reveal any major substantive differences, and so here we only present analyses considering these two items together. Specifically, for each subject, an average of the "*Sick*" and "*Buzzed*" after responses and an average of the before responses were obtained, and a change score (after – before) was calculated. These represent the changes in the level of subjective physiological sensations attributable to smoking a cigarette.

Because participants could smoke more than two cigarettes during the assessment week, there were multiple observations per subject. In all, there was a total of 517 observations clustered within these 100 subjects. A question of interest is whether the change (after – before) in physiological sensations varies between smoking groups defined by level of smoking experience. We would expect, for example, that both means and variances diminish as smoking level increases. In particular, we address this in terms of differences in means and variances across these groups. Table 8.1 presents some descriptive statistics broken down by smoking history group.

The BS and WS variance estimates were obtained separately for each group using a random-intercepts MRM for each. These results suggest that there is not a great deal of difference in means. However, both BS and WS variances decrease with increasing smoking history. This supports notions from smoking research that physical reactions to smoking stabilize as one's experience with smoking increases.

We address these observations more formally using models with varying BS and WS variance parameters. Specifically, we consider models of separate variance across smoking groups, and also trend in variance across the smoking groups. These variance parameterizations are summarized in Table 8.2. Note that the separate variance parameterizations require three parameters, whereas the trend versions only require two parameters. Also, because the variance decreased across the groups, the specification of the group codes was reversed for the BS variance, as mentioned earlier. Results are presented in Table 8.3.

Nine models are estimated by considering three representations for both the BS and WS variance: Common variance, group trend in variance, and separate group variance. Model selection can be based on the model deviance, and two penalized versions of the deviance: The Akaike Information Criterion (AIC; Akaike (1973)) and the Bayesian Information Criterion (BIC; Schwarz (1978)). The penalties are for model complexity and the BIC penalty is greater than the AIC penalty. Thus, BIC tends to suggest simpler models relative to AIC. In the present case, AIC would select Model IIc as best, whereas BIC selects Model Ib. Because they are not nested, these two models cannot be compared to each other using a likelihood-ratio

TABLE 8.1
Changes in Physiological Sensations by Smoking Group: Numbers of Subjects, Observations, Means, Variances, and ICCs

Smoking Group	*Subjects* N	*Observations* $\sum n_i$	*Mean* $\bar{y}_{ij}$	*BS Variance* $\hat{\sigma}^2_{\upsilon}$	*WS Variance* $\hat{\sigma}^2_{\epsilon}$	*ICC*
LO: < 6 cigs	18	50	.54	2.69	4.23	.39
MID: 6-99 cigs	48	183	.80	.89	3.55	.20
HI: 100+ cigs	34	284	.50	.66	1.54	.30

ICC = intraclass correlation

TABLE 8.2
Variance Parameterizations

Group	*Separate Variance*		*Trend in Variance*			
	WS Variance	*BS Variance*	*Code*	*WS Variance*	*Code*	*BS Variance*
< 6 cigs	$\sigma^2_{\epsilon_{LO}}$	$\sigma^2_{\upsilon_{LO}}$	0	σ^2_{ϵ}	2	$\sigma^2_{\upsilon_0} + 4\sigma^2_{\upsilon_1}$
6-99 cigs	$\sigma^2_{\epsilon_{MID}}$	$\sigma^2_{\upsilon_{MID}}$	1	$\exp(\tau)\,\sigma^2_{\epsilon}$	1	$\sigma^2_{\upsilon_0} + \sigma^2_{\upsilon_1}$
100+ cigs	$\sigma^2_{\epsilon_{HI}}$	$\sigma^2_{\upsilon_{HI}}$	2	$\exp(2\tau)\,\sigma^2_{\epsilon}$	0	$\sigma^2_{\upsilon_0}$

(LR) test. However, each of these models can be compared to the homogeneous BS and WS variance model (Model Ia). For Model Ib, the LR chi-square statistic equals 2037.4 - 1992.8 = 44.6, which is highly significant on 1 degree of freedom. Similarly for Model IIc, the LR χ^2 equals 48.7 on 3 degrees of freedom, which again is highly significant. Both models reject the notion of homogeneous WS variance, and Model IIC additionally rejects homogeneity of BS variance. Table 8.4 lists the parameter estimates for three of the models: Homogeneous WS and BS variance, Model Ib (best BIC), and Model IIc (best AIC). The first model is for comparison only since, as shown earlier, this model is rejected in favor of either Model Ib or IIc. Statistical tests for specific parameters can be obtained by dividing the parameter estimate by its standard error (i.e., Wald statistic) and comparing to a standard normal distribution (at $\alpha = .05$, the two-tailed critical value is 1.96 or approximately 2). Although Wald tests are routinely used for fixed-effects parameters, for variance parameters their use is dubious (see Verbeke & Molenberghs, 2000, pages 64-65). In terms of the fixed effects there are no statistically significant results. Thus, the three smoking history groups are similar in terms of the average level of change in subjective physiological sensations.

TABLE 8.3
Change in Subjective Physical Sensation Before vs After Smoking: Model Comparisons.

Model	*Fixed*	*p*	*BS Variance*	*r*	*WS Variance*	*s*	*Deviance*	*AIC*	*BIC*
Ia	I + G	3	I	1	I	1	2037.4	2041.4	2046.6
Ib	I + G	3	I	1	I + T	2	1992.8	1998.8	2006.6
Ic	I + G	3	I	1	I + G	3	1991.0	1999.0	2009.4
IIa	I + G	3	I + T	2	I	1	2027.0	2033.0	2040.8
IIb	I + G	3	I + T	2	I + T	2	1991.6	1999.6	2010.1
IIc	I + G	3	I + T	2	I + G	3	1988.7	1998.7	2011.7
IIIa	I + G	3	I + G	3	I	1	2027.0	2035.0	2045.4
IIIb	I + G	3	I + G	3	I + T	2	1991.6	2001.6	2014.6
IIIc	I + G	3	I + G	3	I + G	3	1988.5	2000.5	2016.2

Turning to the variance estimates, based on both Models Ib and IIc, it is clear that the WS variance decreases across smoking history groups. The latter model estimates separate variances across these groups, while the former estimates a decreasing trend across the groups. Based on this trend estimate of Model Ib, the WS variance is estimated as 6.62, $6.62 \times \exp(-.72) = 3.22$, and $6.62 \times \exp(2 \times -.72) = 1.57$ for groups LO, MID, and HI, respectively. Thus, the WS estimates for Models Ib and IIc are in close agreement, except for the LO group which has the fewest numbers of subjects and observations. Additionally, Model IIc posits that the BS variance diminishes across smoking history groups, from .63 for HI, to $.63 + .42 = 1.05$ for MID, and $.63 + 2^2 \times .42 = 2.31$ for LO.

These estimates can be used to generate ICCs based on each model. For the naive Model Ia, we obtain:

$$\text{ICC} = \frac{1.23}{1.23 + 2.41} = .34$$

as the common ICC for all groups. For Model Ib, we get:

$$\begin{aligned} \text{LO ICC} &= \frac{.80}{.80 + 6.62} = .11 \\ \text{MID ICC} &= \frac{.80}{.80 + [\exp(-.72) \times 6.62]} = .20 \\ \text{HI ICC} &= \frac{.80}{.80 + [\exp(2 \times -.72) \times 6.62]} = .34 \end{aligned}$$

for the three groups. Similarly, the ICCs based on model IIc are:

$$\text{LO ICC} = \frac{.63 + 4 \times .42}{(.63 + 4 \times .42) + 4.33} = .35$$

TABLE 8.4
Change in Subjective Physical Sensation Before vs After Smoking: REML Estimates & Standard Errors (se) for Select Models.

	Model Ia		*Model Ib*		*Model IIc*	
Term	*Estimate*	*se*	*Estimate*	*se*	*Estimate*	*se*
Fixed Effects						
Intercept	.71	.35	.63	.43	.71	.47
MID cigs	.12	.40	.18	.47	.11	.52
HI cigs	-.24	.41	-.16	.46	-.24	.50
BS Variance						
Intercept variance	1.23	.31	.80	.24	.63	.22
Group trend (LO=2, MID=1, HI=0)					.42	.32
WS Variance						
Intercept variance	2.41	.17	6.62	1.18		
Group trend (LO=0, MID=1, HI=2)			-.72	.11		
LO cigs					4.33	1.10
MID cigs					3.50	.42
HI cigs					1.54	.14
Deviance $(-2 \log L)$	2037.4		1992.8		1988.7	

$$\text{MID ICC} = \frac{.63 + .42}{(.63 + .42) + 3.50} = .23$$

$$\text{HI ICC} = \frac{.63}{.63 + 1.54} = .29.$$

Based on these estimates, Model IIc is the best in terms of matching the stratified ICCs presented in Table 8.1. Also, Model Ib does badly in terms of the LO group, suggesting that Model IIc can be chosen as the "best" model both in terms of AIC and also in matching the stratified results. Thus, there is evidence of decreasing variation in physical subjective sensation change across smoking history groups in terms of both BS and WS variance.

Variation in Variation Across Time: Baseline to 12 months

The foregoing results suggest clear differences in physical subjective sensation variation across smoking history groups, and it is tempting to attribute these differences to smoking history. Although this conclusion is certainly logical, it could be that another variable, and not smoking history, is responsible for the group differences in physical subjective sensation variation. To address this possibility, we examined a subset of subjects with longitudinal data across time. Specifically, we are interested in seeing whether the cross-sectional variation patterns observed earlier are also evident longitudinally as subjects progress in their smoking careers.

Because of our focus on variance modeling (both between- and within-subjects), we included subjects who provided two or more smoking reports at two or more timpoints (baseline, 6-, and 12-month follow-ups). In all, there were 46 subjects meeting these criteria: 37 provided data at baseline (mean number of reports = 8.05), 40 at 6-months (mean number of reports = 7.93), and 37 at 12-months (mean number of reports = 11.92). In terms of the smoking history grouping variable used in the foregoing analyses , 2, 8, and 27 of these 46 subjects were in the LO, MID, and HI groups, respectively. Additionally, 9 of these 46 subjects were not included in the earlier analyses because they provided no smoking reports at baseline.

These 46 subjects were classified into two groups based on a separate latent growth analysis, which is briefly summarized here. This analysis was based on data obtained from a time-line follow-back interview at four questionnaire waves: Baseline, 6-, 12-, and 18-month follow-ups. At each wave interviewers guided adolescents through a structured calendar-based recall of their smoking, noting the amount (even a puff) smoked on specific days over the past months. These data represent counts of numbers of cigarettes per day, over time, anchored to specific real-time calendars. For each subject, and at each wave, we calculated the following summary variable: The number of cigarettes smoked per day for each of the two 90-day intervals within the six-month interval. Since this outcome was highly non-normal it was categorized into an ordinal variable as follows: 0 = 0 cigarettes (none); 1 = greater than 0 cigarettes but less than 1 cigarette per month (quarterly); 2 = greater than or equal to 1 cigarette per month but less than 1 cigarette per week (monthly); 3 = greater than or equal to 1 cigarette per week but less than 1 cigarette per day (weekly); 4 = greater than or equal to 1 cigarette per day (daily). Subjects who gave at least one non-zero ordinal response across time were then analyzed in the latent growth analysis. This analysis identified five latent smoking groups, which we labeled as triers, escalators, quitters, rapid escalators and regular smokers.

The analysis in this chapter uses only a fraction of the subjects that were used in the latent growth analysis (because of our criteria that a subject must have two or more smoking reports at two or more timepoints). Of the 46 subjects included here, 18 were classified as smoking escalators or rapid escalators and 28 as regular smokers in the latent growth analysis. We treat escalators and rapid escalators as a single group because of the small numbers. Figure 8.1 plots the means by these two groups for this ordinal cigarette rate variable across time.

As can be seen, the groups have very different average trajectories across time. The means that the regular smokers are at or near daily smoking at all timepoints, whereas escalators are a bit under monthly smoking initially, but increase to between weekly and daily smoking by the end. In terms of the timing, timepoints 0 and 1 in the figure correspond to the two 90-day periods before baseline, timepoints 2 and 3 to the two 90-day periods before the 6-month follow-up, timepoints 4 and 5 to the two 90-day periods before the 12-month follow-up, and timepoints 6 and

7 to the two 90-day periods before the 18-month follow-up. Since the analysis in this chapter focuses on smoking reports at baseline, 6-, and 12- month follow-ups, the figure means from timepoints 0 to 5 are most useful in getting a sense of how the two groups here compare.

FIGURE 8.1
Smoking group means across time:

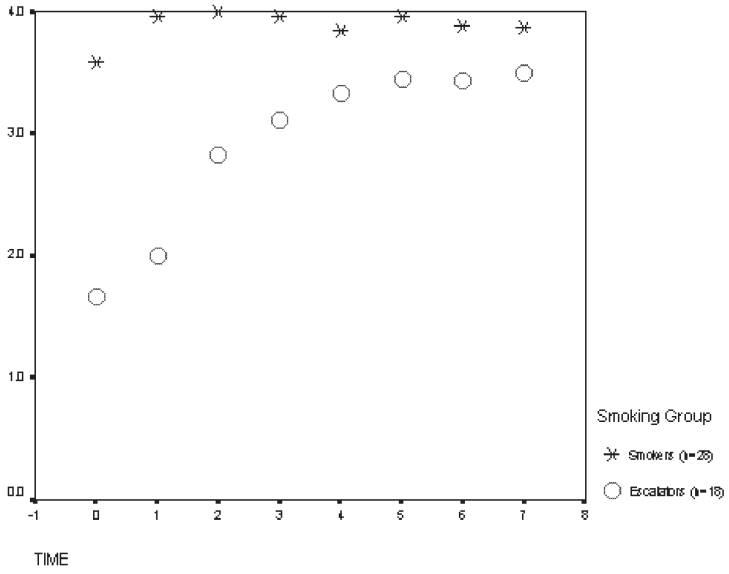

Table 8.5 lists descriptive information regarding the smoking reports from these two groups. It appears that the means diminish across time, but that there are no large group differences in terms of the means. The variances also diminish across time, and whereas the variances for the groups are very different at baseline, they are quite similar at 12 months.

To these data, we fit five models of WS variance: Common WS variance, group varying, time varying, group and time varying, and group by time varying. The latter model is the one of most interest because it would suggest that the pattern of WS variance across time was different for the two groups. Based on smoking theory and our earlier cross-sectional results , we hypothesized that escalators would have a much greater decrease in variation in their physical sensations across time than smokers.

For BS variance, we fit four models: common variance, group varying, time varying, and group by time varying. The common variance model is simply a random-intercepts model, whereas the group-varying model includes separate random intercepts for the two groups. The time varying model includes a random subject intercept and a random time trend, which are correlated. This yields three parameters in the 2×2 variance–covariance matrix: Intercept variance, time

variance, and the covariance between the intercept and slope. As noted earlier, for the random trend model, the BS variance is then a function of time. The final BS variance model specifies separate 2×2 variance–covariance matrices for each of

TABLE 8.5
Changes in Physiological Sensations by Time and Smoking Group: Numbers of Subjects, Observations, Means, Variances, and ICCs.

Time	*Group*	*N*	$\sum n_i$	$\bar{y}$	*BS Variance*	*WS Variance*	*ICC*
Baseline	Escalators*	12	43	.71	1.53	2.94	.34
	Smokers	25	255	.58	.48	1.69	.22
6 months	Escalators*	15	77	.49	.09	1.82	.05
	Smokers	25	239	.44	.30	1.38	.18
12 months	Escalators*	15	139	.17	.20	1.41	.13
	Smokers	22	302	.21	.21	1.45	.13

* escalators and rapid escalators combined

the two smoking groups. This corresponds to a group by time modeling of the BS variance, because the BS variance is a function of time for each group separately. Table 8.6 lists results for these models.

Based on BIC values, the model with group by time WS variance and time BS variance is "best." Though not presented, the AIC criterion would also select this model. Likelihood ratio tests also support selection of this model, which we dub the WS(GT) and BS(T) model. For example, comparing the selected model to the WS(G+T) and BS(T) model (i.e., the Model 1 row directly above our selected model) yields $\chi^2_1 = 3568.0 - 3562.5 = 5.5, p < .05$, supporting the hypothesis that the WS variance varies across time differently for the two groups. In terms of the BS variance, the similar hypothesis of group by time influence is not supported. Comparing the selected model relative to the WS(GT) and BS(GT) model (i.e., the model 1 column directly to the right of our selected model) yields $\chi^2_3 = 3562.5 - 3562.1 = 0.4$, which is clearly non-significant. Thus, while the BS variance does vary across time, there is no evidence that it additionally varies across time by group. It should be noted, however, that the numbers of subjects in these two groups are relatively small and so our ability to detect group by time differences in BS variance is limited.

Table 8.7 lists the parameter estimates of our selected model, as well as estimates from the simplest model including only one WS variance and one BS variance parameter, the WS(I) and BS(I) model. Inspecting the fixed effects yields similar conclusions from both models. The group by time interaction is not significant, but the time effect approaches significance and is negative (i.e., $z = -.31/.16 = 1.94, p < .063$). Removing the non-significant interaction (not shown) yields a

TABLE 8.6
Change in Subjective Physical Sensation Before vs After Smoking: Model Deviance and *BIC*.

		BS Variance			
WS Variance	*Intercept*	*Group*	*Time*	*Group × Time*	*s*
Intercept	3599.9	3598.8	3577.7	3575.9	1
	3607.6	*3610.2*	*3593.0*	*3602.7*	
Group	3597.8	3597.1	3574.8	3573.4	2
	3609.3	*3612.4*	*3593.9*	*3604.0*	
Time	3590.0	3589.4	3571.8	3570.3	2
	3601.5	*3604.7*	*3590.9*	*3601.0*	
Group + Time	3587.0	3586.7	3568.0	3566.9	3
	3602.3	*3605.9*	*3591.0*	*3601.4*	
Group × Time	3581.6	3581.6	3562.5	3562.1	4
	3600.8	*3604.6*	*3589.3*	*3600.4*	
r	1	2	3	6	

Note. s = number of WS variance parameters and r = number of BS variance parameters. For all models: fixed effects = Intercept + Group + Time + Group × Time.

TABLE 8.7
Change in Subjective Physical Sensation Before vs After Smoking: REML Estimates & Standard Errors (se) for Two Models.

	Model BS(I) WS(I)		*Model BS(T) WS(GT)*	
Term	*Estimate*	*se*	*Estimate*	*se*
Fixed Effects				
Intercept	.77	.22	.81	.28
Group (smokers = 1)	-.13	.25	-.15	.32
Time (0, 1, 2)	-.24	.12	-.31	.16
Group by time	.06	.13	.10	.19
BS Variance				
Intercept variance	.29	.09	.54	.20
Int, time covariance			-.22	.10
Time variance			.14	.06
WS Variance				
Intercept variance	1.65	.07	3.04	.60
Group			-.62	.21
Time			-.40	.12
Group by time			.32	.14
$-2 \log L$	3599.9		3562.5	

highly significant main effect of time (i.e., $z = -.235/.082 = -2.88, p < .006$) Thus, the (change in) physical sensation means of both groups of subjects are decreasing across time.

Comparing the two models in Table 8.7 via a likelihood ratio test supports the BS(T) and WS(GT) model; $\chi^2_5 = 3599.9 - 3562.5 = 37.4, p < .001$. Examining the WS variance terms, all estimates exceed their standard errors by 1.96 and so are significant based on the Wald test. The estimates indicate that initially the smokers have significantly less WS variance (group $z = -.62/.21 = -2.89, p < .004$), that the WS variance decreases significantly across time for the escalators (time $z = -.40/.12 = 3.21, p < .002$), and that the time trend for smokers is significantly less than for escalators (group by time $z = .32/.14 = 2.32, p < .02$). Turning to the BS variance terms, it's clear that there is significant subject heterogeneity in their intercepts and slopes, and that these two are modestly negatively associated.

Table 8.8 presents estimated BS and WS variances across time for both groups based on the WS(GT) and BS(T) model. Comparing these to the stratified estimates obtained in Table 8.5 gives a sense of the degree of model fit. As can be seen, the model estimates agree well with the stratified results of the WS variances for both groups across time. These estimates clearly support our earlier interpretation that the two groups are quite dissimilar initially but converge across time. As our selected model did not include any group-related terms in terms of the BS variance, the estimated BS variances are the same for the two groups, but do vary across time. As Table 8.8 indicates, the estimated BS variance diminishes quite a bit following baseline, though it is relatively consistent at the two follow-ups. This pattern agrees reasonably well with the stratified results in Table 8.5 for the smokers, but not for the escalators. Again, though, this latter group only has a total of 18 subjects, and so precise estimation of BS variances is limited for this relatively small sample.

TABLE 8.8
Change in subjective physical sensation before vs after smoking: Estimated variances based on the BS(T) and WS(GT) model.

Time	*Group*	*BS variance*	*WS variance*
Baseline	Escalators*	.54	3.04
	Smokers	.54	1.64
6 months	Escalators*	.24	2.04
	Smokers	.24	1.52
12 months	Escalators*	.22	1.37
	Smokers	.22	1.40

* escalators and rapid escalators combined.

CONCLUSION

This chapter has illustrated how mixed models for longitudinal data can be used to model differences in variances, and not just means, across subject groups and time. As such, these models can help to identify predictors of both within-subjects and between-subjects variation, and to test psychological hypotheses about these variances.

As an example, we used these models to examine important aspects of the development of dependency to cigarette smoking among adolescent smokers. One of the key concepts in dependence is the development of tolerance, or the diminishing of effects of a substance with its continued use. A common experience, reported by adolescents, during early trials of cigarette smoking is feeling "sick" or "buzzed" after smoking a cigarette, and the equally common notion is that these subjective feelings diminish over time as one's experience with smoking increases. However, to date, researchers have been able to examine changes in these subjective experiences primarily through paper and pencil, retrospective questionnaire reports. Thus, it has been difficult to document adequately exactly how symptoms of dependence develop or with what level of smoking experience. Our analyses indicated an increased consistency of subjective physiological responses as experience with smoking increased, both cross-sectionally and longitudinally. Our data thus provide one of the few ecologically valid examinations of the development of tolerance. Adolescents' self-reports, in real time, of their subjective responses to smoking changed over time as a function of experience with smoking. Importantly, too, these changes were relatively dramatic over a 1-year period. Thus, these analytic models provide a way to examine changes in subjective intrapersonal experiences that have not been readily feasible before.

More applications of this class of models clearly exist in psychology. For example, many questions of both normal development and the development of psychopathology address the issue of variability or stability in emotional responses to various situations and/or contexts. Often, a concern is with the range of responses an individual gives to a variety of stimuli or situations, and not just with the overall mean level of responsivity. These models also allow us to examine hypotheses about cross-situational consistency of responses as well.

In order to reliably estimate variances, one needs a fair amount of both WS and BS data. Modern data collection procedures, such as ecological momentary assessments (EMA) and/or real-time data captures, provide this opportunity. Such designs are in keeping with the "bursts of measurement" approach described by Nesselroade (1991), who called for such an approach in order to assess intraindividual variability. As noted by Nesselroade, such bursts of measurement increase the research burden in several ways, however they are necessary for studying intraindividual variation.

This chapter has focused on models for continuous normally-distributed outcomes, and has illustrated how such models can be estimated with standard soft-

ware. Because ordinal data are often obtained in many research areas as well, we are currently extending these procedures for ordinal data. For example, many variables in psychology are measured using Likert scales or other similar types of ordinal categories. Admittedly, there is typically more information in continuous than ordinal responses, so the ability to model variances in ordinal data may not be as general as what is possible using the methods presented in this chapter. Thus, we hope to examine the degree to which these models of variation can be applied to ordinal outcomes.

ACKNOWLEDGMENTS

Thanks are due to Siu Chi Wong for statistical analysis, and to Todd Little for organizing the conference and for his helpful comments on a previous draft of this chapter. This work was supported by National Institutes of Mental Health grant MH56146, National Cancer Institute grant CA80266, and by a grant from the Tobacco Etiology Research Network, funded by the Robert Wood Johnson Foundation. Correspondence to Donald Hedeker, Division of Epidemiology & Biostatistics (M/C 923), School of Public Health, University of Illinois at Chicago, 1603 West Taylor Street, Room 955, Chicago, IL, 60612-4336. e-mail: hedeker@uic.edu

REFERENCES

Aitkin, M. (1987). Modelling variance heterogeneity in normal regression using GLIM. *Applied Statistics*, *36*, 332-339.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle . In B. N. Petrov & F. Csaki (Eds.), *Second international symposium on information theory* (pp. 267–281). Budapest: Academiai Kiado.

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 Associates.

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. (1989). 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.

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

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. *The Journal of Pediatrics*, *139*, 546-551.

Carroll, K. M., Rounsaville, B. J., Nich, C., Gordon, L., Wirtz, P., & Gawin, F. (1994). One-year follow-up of psychotherapy and pharmacotherapy for cocaine dependence. *Archives of General Psychiatry*, *51*, 989-997.

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.

Dempster, A. P., Rubin, D. B., & Tsutakawa, R. K. (1981). Estimation in covariance component models. *Journal of the American Statistical Society*, *76*, 341-353.

Elkin, I., Gibbons, R. D., Shea, M. T., Sotsky, S. M., Watkins, J. T., Pilkonis, P. A., et al. (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.

Fleeson, W. (2004). Moving personality beyond the person-situation debate. *Current Directions in Psychological Science*, *13*, 83-87.

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., Waternaux, 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 edition.* New York: Halstead Press.

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.

Harvey, A. C. (1976). Estimating regression models with multiplicative heteroscedasticity. *Econometrica*, *44*, 461-465.

Hedeker, D. (2004). An introduction to growth modeling. In D. Kaplan (Ed.), *The SAGE handbook of quantitative methodology for the social sciences* (pp. 215–234). Thousand Oaks, CA: Sage.

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.

Hertzog, C., & Nesselroade, J. R. (2003). Assessing psychological change in adulthood: An overview of methodological issues. *Psychology and Aging*, *18*, 639-657.

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.

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

Leeuw, J. de, & Kreft, I. (1986). Random coefficient models for multilevel analysis. *Journal of Educational Statistics*, *11*, 57-85.

Longford, N. T. (1987). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. *Biometrika*, *74*, 817-827.

Martin, M., & Hofer, S. M. (2004). Intraindividual variability, change, and aging: Conceptual and analytical issues. *Gerontology*, *50*, 7-11.

Mermelstein, R., Hedeker, D., Flay, B., & Shiffman, S. (2002). *Situational versus intra-individual contributions to adolescents' subjective mood experience of smoking.* Savannah, GA: Annual Meeting for the Society for Research on Nicotine and Tobacco.

Neff, L. A., & Karney, B. R. (2004). How does context affect intimate relationships? linking external stress and cognitive processes within marriage. *Personality and Social Psychology Bulletin*, *30*, 134-148.

Nesselroade, J. R. (1991). The warp and woof of the developmental fabric. In R. Downs, L. Liben, & D. Palarmo (Eds.), *Visions of development, the environment, and aesthetics: The legacy of Joachim F. Wohlwill.* (pp. 213–240). Hillside, NJ: Lawrence Earlbaum Associates.

Nesselroade, J. R. (2001). Intraindividual variability in development within and

between individuals. *Eur Psychol*, *6*, 187-193.

Nesselroade, J. R. (2004). Intraindividual variability and short-term change. *Gerontology*, *50*, 44-47.

Nesselroade, J. R., & Boker, S. M. (1994). Assessing constancy and change. In T. Heatherton & J. Weinberger (Eds.), *Can personality change?* (pp. 503-541). Washington, DC: American Psychological Association.

Nesselroade, J. R., & Schmidt McCollam, K. M. (2000). Putting the process in developmental processes. *International Journal of Behavioral Development*, *24*, 295-300.

Niaura, R., Spring, B., Borrelli, B., Hedeker, D., Goldstein, M., 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.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461-464.

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.

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.

Verbeke, G., & Molenberghs, G. (2000). *Linear mixed models for longitudinal data*. New York: Springer.

Wolfinger, R. D. (1993). Covariance structure selection in general mixed models. *Communications in Statistics, Simulation and Computation*, *22*, 1079-1106.

Wu, H., & Zhang, J.-T. (2002). Local polynomial mixed-effects models for longitudinal data. *Journal of the American Statistical Association*, *97*, 883-889.