Applying Mixed Regression Models to the Analysis of Repeated-Measures Data in Psychosomatic Medicine

EKIN BLACKWELL, MA, CARLOS F. MENDES DE LEON, PHD, AND GREGORY E. MILLER, PHD

Objective: Although repeated-measures designs are increasingly common in research on psychosomatic medicine, they are not well suited to the conventional statistical techniques that scientists often apply to them. The goal of this article is to introduce readers to mixed regression models, which provide a more flexible and accurate framework for managing repeated-measures data. **Methods and Results:** We begin with a summary of the advantages that mixed regression models have over conventional statistical techniques in the context of repeated-measures designs. Next, we outline the conceptual and mathematical underpinnings of mixed regression models for a nonstatistical audience. The article ends with two examples of how these models can be applied in psychosomatic research; one deals with a prospective investigation of depressive symptoms and change in body mass index in older adults and the other with a diary study of social interactions and cortisol secretion. **Conclusions:** Mixed regression models offer a flexible and powerful approach to analyzing repeated-measures data. They possess important advantages over more traditional strategies, and more widespread application of these models is likely to enhance the overall quality of psychosomatic research. **Key words:** mixed regression models, analysis of change, repeated measures, nested designs, random effects.

HLM = hierarchical linear model; **OLS** = ordinary least squares; **ANOVA** = analysis of variance; **BP** = blood pressure; **CES-D** = Center for Epidemiologic Studies–Depression; **BMI** = body mass index; **D** = dominance; **EPAQ** = Extended Version of the Personality Attributes Questionnaire.

INTRODUCTION

The purpose of this article is to introduce readers to mixed regression models for repeated-measures data. In epidemiologic studies, measurements typically occur over many months or years, because the interest is usually on changes in disease processes and outcomes. In diary studies, in which the interest is usually in short-term fluctuations in physiological or behavioral processes, measurement occasions span minutes, hours, or days. These measurements are often obtained at unequal intervals or at different time points for different individuals, and some measurements may be missing. This can pose problems for traditional methods like repeated-measures analysis of variance (ANOVA). Mixed regression models offer a flexible alternative for dealing with these unbalanced data sets.

Repeated Measurement Data Structures

Mixed regression models, a class of statistical models developed for the analysis of data structures with nested sources of variability, include hierarchical linear models (HLMs), growth curve models, and random coefficient models. They are rooted in time series analysis (1), mixed and variance components models (2), random effects models (3), and empiric Bayes models (4). Nesting occurs when units of observations at one level are clustered within higher-level observations. Examples of two-level nested structures include students within classrooms, children within families, families within neighborhoods, and spouses within couples. Nesting normally produces dependencies in the lower-level units. Sometimes this dependency is a nuisance that has to be controlled for, but it is often at the heart of the research question, specifically when the goal is to understand the source of the observed dependency. For example, in a data set in which families are nested within neighborhoods, the question may be how neighborhood-level characteristics (the higher-level units) such as social capital or median income relate to phenomena at the family level (the lower-level units) like children's educational achievement or delinquent behavior. Repeated-measures data represent a special kind of nesting in which units of measurement are nested within individuals. Dependency is the norm in repeated-measures data because observations obtained from the same individual tend to be correlated.

An important consideration in repeated-measures designs is whether to model time as a *fixed* or *random* factor. When the observed levels of a factor are the only ones of interest to the researcher, their effects on the outcome are fixed; they pertain to these levels only and cannot be generalized to levels that were not included in the study. This would occur, for example, in a study of immune functions at specific moments before, during, and after a discrete stressor; or in a study of well-being and health at specific intervals before, in the midst of, and after a woman goes through menopause. In other cases, the researcher is not interested in quantifying a measured outcome for any particular instance, but wants to infer its general pattern of change over time. The appropriate strategy in this situation is to model time as a random factor such that the moments of observation for each person represent a random sample of all possible observations during the study period. A longitudinal study tracking blood pressure (BP) levels throughout adulthood, or a diary study examining changes in BP during 10-minute intervals, are examples of this kind.

Traditional techniques like repeated-measures ANOVA can model time as either fixed or random, but violation of the sphericity assumption is a central concern. Sphericity dictates

From the Department of Psychology (E.B., G.E.M.), University of British Columbia, Vancouver, British Columbia, Canada; and the Rush Institute for Aging (C.F.M.d.L.), Rush-Presbyterian–St. Luke's Medical Center, Chicago, Illinois.

Address correspondence and reprint requests to Ekin Blackwell, MA, UBC Department of Psychology, 2136 West Mall, Vancouver, BC V6T 1Z4, Canada. E-mail: ekinb@interchange.ubc.ca

Received for publication April 1, 2005; revision received June 13, 2006.

This research was supported by grants from the National Institute of Environmental Health Sciences (ES 10902) and the National Institute on Aging (AG 11101), the Canadian Institutes of Health Research, the Michael Smith Foundation for Health Research, the National Alliance for Research on Schizophrenia and Depression, the National Heart, Lung, and Blood Institute, National Institute of Environmental Health Sciences, and the Social Sciences and Humanities Research Council of Canada.

DOI: 10.1097/01.psy.0000239144.91689.ca

0033-3174/06/6806-0870

that all pairwise difference scores between observations have equal variance, but this is rarely found in unbalanced data sets. Failure to meet the sphericity assumption results in a serious liberal bias, or type I error (5). Although techniques such as the Greenhouse-Geisser correction can compensate for sphericity violations, mixed regression models allow this assumption to be relaxed altogether. Researchers can specify a variety of covariance structures to account for specific patterns of correlations. For instance, a first-order autoregressive structure is appropriate for many diary studies, in which correlations between observations are assumed to be an exponentially decreasing function of time. The choice of structure depends on the design and assumptions of the study as well as the nature and timing of the measurements. For a more detailed discussion, see Petkova and Teresi, Nezlek, Nezlek, and Schwartz and Stone (6–9).

It is also possible to analyze repeated-measures data in a multivariate framework that does not impose restrictions on the covariance matrix. However, this approach can lead to model overfitting and convergence problems, especially when the number of subjects is small relative to the number of repeated observations per subject. Unlike ANOVA techniques, mixed regression models do not require a "complete" data set, that is, the same number of assessments on all participants. Multiple imputation techniques provide another viable solution to the problem of incomplete data (10). Both strategies are superior alternatives to discarding data, at least when the data are assumed to be missing at random.

In the next sections, we consider another critical aspect of mixed regression models that sets them apart from ordinary least squares (OLS) techniques of estimation—their ability to *simultaneously* examine within- and between-person phenomena that contribute to change. The analysis of change can be thought of as being composed of two stages (11). In the first stage, we ask: What is the general shape or form of the process of change over time? Does it follow a linear or nonlinear pattern of increase or decrease or perhaps a cyclical pattern? We then use statistical modeling to derive a general form of change in the outcome variable from the individual-specific (within-person) patterns of change. In the second stage, we examine the heterogeneity in patterns of change between individuals and test the contribution of hypothesized factors to this heterogeneity. In the next section, we explain how these two steps fit into the general framework of the mixed regression model.

Basic Statistical Features of Mixed Regression Models for Repeated Measures

The mathematical underpinnings of the mixed regression model for repeated measures are presented in several excellent textbooks (12–15). Briefly, the most basic form of the two-level mixed regression model represents an outcome variable Y as a function of an intercept (β_0), a predictor variable A, and a random error term:

$$Y_{ij} = \beta_{0j} + \beta_{1j}A_{ij} + R_{ij} \quad (1)$$

The lower level (micro)units are indexed by i and the higher level (macro) units are indexed by j. The lower level units are measurement occasions and the higher-level units are individual subjects. In other words, Y_{ij} represents the outcome variable at time i for subject j. We can visualize the analysis as a set of j person-specific regression lines, one for each individual. Associated with each regression line are the coefficients β_{0j} and β_{1j}. These represent, respectively, the person-specific intercept and slope for individual j. These are presumed to vary among individuals.

Equation 1 is a general model for the first, or within-person, stage of analysis. In this phase, the analysis can take one of two forms. For example, in a longitudinal study examining changes in BP, the predictor variable A_{ij} represents time elapsed since the first measurement occasion. Replacing A_{ij} with the notation t_{ij} may be helpful in making this distinction. Thus, equation 1 is simply the regression of the outcome variable Y (BP) on time. This is commonly referred to as a growth curve model and is the conventional model for the analysis of change. The intercept β_{0j} is usually modeled such that it represents the person-specific value of Y at time "0," or baseline. Similarly, the slope β_{1j} represents the person-specific rate of change of Y over time.

This modeling approach differs from the diary example, in which the main interest was in how BP fluctuates with mood state. In such an analysis, BP at time i for subject j could be designated as the outcome variable Y_{ij} and mood state at the same time in the same subject as the predictor variable A_{ij}. In this case, the predictor A_{ij} is not time but a time-varying covariate. Analogous to the first example, intercept β_{0j} would represent the predicted BP value (Y) for subject j when mood (A) was rated as 0, and slope β_{1j} would represent the relationship between mood state and BP across sampling times for subject j.

In the second, or between-person, stage of analysis, the goal is to determine which person-level characteristics may explain differences in the within-person β coefficients. Each coefficient is further broken down into a group mean and a deviation from that mean as follows:

$$\beta_{0j} = \gamma_{00} + U_{0j} \quad (2)$$

$$\beta_{1j} = \gamma_{10} + U_{1j} \quad (3)$$

These equations are referred to as unconditional models, because they do not yet contain any level 2 predictors. γ_{00} and γ_{10} signify the mean intercept and slope across all individuals. These are considered to be fixed, or constant. U_{0j} and U_{1j} are random variables designating the specific amounts by which individual j deviates from these means. These quantities will vary randomly from study to study depending on which subjects have been selected from the population. The term random reflects the fact that the parameter estimates (intercepts and slopes) are free to vary across individuals; hence, they have a variance. Although β_{0j} is referred to as the *random intercept* and β_{1j} as the *random slope*, the term *mixed regression model* implies that these coefficients have both a fixed

and a random component. Just as the R_{ij} residuals are assumed normally distributed with variance σ^2, the U_{0j} and U_{1j} residuals are assumed normally distributed with variances τ_0^2 and τ_1^2. Because the random intercepts and random slopes are often correlated, U_{0j} and U_{1j} also have covariance τ_{01}.

So far we have only presented models with a single predictor at level 1 representing either time itself or a time-varying covariate. A level 1 predictor serves to reduce the variance of the R_{ij} residuals, signified by σ^2. If we want to explain even more of this unexplained variance, we can include additional level 1 predictors. Likewise, we can add any number of level 2 predictors to reduce the variances τ_0^2 and τ_1^2 associated with U_{0j} and U_{1j}. These predictors represent person-level characteristics such as sociodemographic information, personality features, and so on.

To illustrate, assume that Z denotes a single level 2 predictor called hostility. Equations 2 and 3 can be modified to reflect the contribution of hostility to the variability in the random intercept and random slope:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}Z_j + U_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}Z_j + U_{1j}$$

The equation for β_{0j} is known as an *intercepts as outcomes* model and the equation for β_{1j} as a *slopes as outcomes* model owing to the fact that the intercepts and slopes at level 1 (β_{0j} and β_{1j}) are treated as outcome variables at level 2, each having their own intercepts (γ_{00} and γ_{10}) and slopes (γ_{01} and γ_{11}). Like in the unconditional model (equations 2 and 3), γ_{00} and γ_{10} signify the mean intercept and slope, respectively, across all individuals for $Z = 0$. The parameter γ_{01} represents the slope of the regression equation predicting β_{0j} from Z_j, and the parameter γ_{11} represents the slope of the regression equation predicting β_{1j} from Z_j. In the longitudinal study tracking change in BP, β_{0j} would represent the association between hostility and BP at time 0 or baseline, and β_{1j} would represent the association between hostility and rate of change in BP over time (see Table 1 for a glossary of terms used in these models).

Before the development of mixed regression modeling techniques, these equations were often estimated in a two-step OLS estimation procedure. In the first step, β coefficients would be generated for each individual; these would then be treated as outcomes in a second estimation procedure. Mixed regression models provide more accurate standard errors compared with this approach because they differentiate between the sampling error associated with the estimated coefficients and the observation error associated with the coefficients themselves, and use shrunken estimates to correct for the effects of sampling variation (12).

When thinking about level 2 predictors, it is important to distinguish between time- varying (state-like) and time-invariant (trait-like) characteristics. In the example presented previously, hostility was conceptualized as a trait that remains stable across time but can vary across individuals. It is thus appropriately modeled as a level 2 predictor. One could also measure hostility in a way that captures within-person fluctuations over time, in which case it should be treated as a level 1 predictor. This is analogous to the way mood was modeled as a level 1 covariate of BP in the diary study example.

TABLE 1. Glossary of Symbols for Two-Level Mixed Regression Models in Repeated-Measures Designs

Symbol	Definition
Y_{ij}	Value of outcome measure at time *i* for individual *j*
A_{ij}	Value of level 1 predictor at time *i* for individual *j*
R_{ij}	Residual error in outcome measure at time *i* for individual *j*
β_{0j}	Person-specific intercept for individual *j*
β_{1j}	Person-specific slope for individual *j*
γ_{00}	Average intercept across all individuals
γ_{10}	Average slope across all individuals
Z_j	Value of level 2 predictor for individual *j*
γ_{01}	Slope of level 2 regression equation predicting β_{0j} from Z_j; denotes systematic deviations from γ_{00} resulting from Z_j
γ_{11}	Slope of level 2 regression equation predicting β_{1j} from Z_j; denotes systematic deviations from γ_{10} resulting from Z_j
U_{0j}	Level 2 residual denoting the unexplained deviation from γ_{00} for individual *j*
U_{1j}	Level 2 residual denoting the unexplained deviation from γ_{10} for individual *j*
σ^2	Variance of level 1 residuals R_{ij}
τ_0^2	Variance of level 2 residuals U_{0j}
τ_1^2	Variance of level 2 residuals U_{1j}
τ_{01}	Covariance of level 2 residuals U_{0j} and U_{1j}

Partitioning the within- and between-person sources of variability in outcome data produces more accurate estimates of associations between predictor and outcome variables relative to conventional regression and ANOVA approaches.

Centering

Centering involves shifting the zero point of a predictor's measurement scale to assign it a more meaningful reference. Because the intercept is defined with respect to this zero point, its interpretation will necessarily change. However, centering will not alter the interpretation of the slope in a linear model.

When a variable is uncentered, the intercept is the expected value of the outcome variable for a person whose score on the predictor variable equals 0. For example, in studies of infant development, it may be appropriate to structure the intercept so that it reflects values of the outcome variables at the time of birth ($t = 0$). However, imagine that a particular outcome variable such as number of spoken words was not assessed until 10 months of age ($t = 10$). The appropriate strategy in this case would be to shift the time scale so that the intercept reflects the value of this outcome variable at 10 months.

When the level 1 predictor is not time but rather a time-varying covariate of the outcome, centering can take the form of person-centering or grand-mean centering. When a level 1 predictor is person-centered, the intercept is the expected value of the outcome for a person whose score on the predictor variable equals the mean of his or her own observations. In a diary study, for instance, person-centering can be used to

MIXED REGRESSION MODELS

determine how each person's BP varies as a function of departures away from his or her average mood across the day. When a level 1 predictor is grand-mean centered, the intercept is the expected value of the outcome for a person whose score on the predictor variable equals the sample mean. So in a diary study predicting BP from mood, grand-mean centering can be used to relate each individual's BP readings to deviations away from the average mood of the group.

For level 2 (person-level) predictors, the relevant centering option is grand-mean centering. When centered in this manner, the intercept is the expected value of the outcome for a person whose score on the level 2 predictor variable equals the sample mean. This can be used to determine, for instance, how the relationship between BP and mood varies as a function of deviations away from the group mean on a trait-level characteristic such as hostility.

Note that we have limited our discussion of centering options to two-level nested structures, in which the highest level of nesting represents the person. When working with multiple groups represented by three-level data structures, group-mean centering is another option. Centering is discussed in greater detail elsewhere (6,7,9,13,15).

Applying Mixed Regression Models to Psychosomatic Research

Example 1: Depression and Change in Relative Weight in Older Adults

Obesity increases risk for a variety of adverse health outcomes, including mortality, cardiovascular disease, and diabetes (16–19). In older age, however, relative weight tends to decrease as a result of the loss of lean muscle mass that may result from declining physical health and activity levels (20–22). Because depression is associated with reduced activity levels and increased risk for chronic disease, we hypothesized that in older adults, higher levels of depressive symptoms would be associated with greater declines in relative weight over time.

The data are from a longitudinal study of 2812 older adults (23). Participants reported on their weight and height at baseline and during eight follow-up interviews, conducted at yearly intervals, for a total of nine waves of outcome data. The Center for Epidemiologic Studies–Depression (CES-D) scale was used to assess depressive symptoms (24). The within-person variable is time since baseline (in years). In addition to baseline CES-D, we included the between-person variables of age at baseline (in years) and sex as covariates. For the purpose of this analysis, we centered age at 75 and CES-D scores at the median value of 5.

Analytic Strategy

We used mixed regression models to test the association between baseline CES-D scores and change in relative weight during the 8-year follow-up period in the 2209 subjects who had nonmissing CES-D data at baseline and ≥2 waves of nonmissing body mass index (BMI) outcome data. Their mean age at baseline was 73.8 and their mean BMI was 25.7 kg/m^2. The level 1 model was specified as follows:

$$\mathrm{BMI_{ij}} = \beta_{0j} + \beta_{1j}\mathrm{TIME_{ij}} + \mathrm{R_{ij}}$$

where $\mathrm{BMI_{ij}}$ represents the BMI value for person i at time j; β_{0j} represents the person-specific intercept, or baseline BMI value; β_{1j} represents the person-specific slope of change in BMI over time; and R_{ij} the residual error or deviation of the observed BMI values for each person i at each interview j. We then specified the level 2 model as follows:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}Z_j + U_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}Z_j + U_{1j}$$

where the person-specific β_{0j} intercept is modeled as a function of the overall intercept γ_{00} across all individuals, plus the systematic deviations from γ_{00} that are accounted for by fixed variables Z_j, plus the residual, unexplained deviation from β_{0j}, being U_{0j} with variance τ_0^2. Similarly, the person-specific β_{1j} slope is modeled as a function of the overall slope of change γ_{10} across all individuals, plus the systematic deviations from γ_{10} that are accounted for by fixed variables Z_j, plus the residual, unexplained deviations from β_{1j}, being U_{1j} with variance τ_1^2. The fixed Z_j predictor variables include age, sex, and CES-D scores, each of which may be associated with the intercept β_{0j} being the baseline value of BMI and with the slope β_{1j} being the rate of change in BMI over time. The main hypothesis is therefore tested by the term γ_{11} for the Z_j for CES-D, which represents the relationship between CES-D scores and the rate of change in BMI over time. This term represents the crosslevel interaction between the level 1 (within-person) variable, time, and the level 2 (between-person) variable, CES-D. We expected this term to be negative, because we hypothesized that higher CES-D scores would be associated with greater declines in relative weight over time (i.e., a more negative slope compared with the average slope). The analysis proceeded in two steps. At the first step (model A), we fit an unconditional level 2 model, that is, one without any level 2 predictor variables. At the next step (model B), we added the level 2 variables representing age, sex, and CES-D.

Depressive Symptoms and Rate of Change in Body Mass Index

The results of model A (see Table 2) indicate that the average BMI at baseline was 25.8. The unconditional estimate for $\beta_{1j,}$ represented by the γ_{10} for time (-0.206), indicates a decline in BMI averaging 0.206 kg/m^2 units per year. The variance components indicate significant random coefficients for the intercept ($U_{0j} = 18.557$, $p < .001$) and slope ($U_{1j} = 0.118$, $p < .001$). Thus, there is evidence for significant variability in the person-specific intercepts and the person-specific rates of change in BMI over time. This justifies our efforts to identify the determinants of the between-person variation in BMI and change in BMI. We can use these coefficients as a base with which to compare the results of

TABLE 2. Depressive Symptoms and Change in Body Mass Index During 8 Yr of Follow Up

	Model A		Model B	
	Coefficient (SE)	*p*	Coefficient (SE)	*p*
Fixed effects				
Intercept (γ_{00} for β_{0j})	25.789 (0.093)	<.001	25.754 (0.133)	<.001
Time (γ_{10} for β_{1j})	−0.206 (0.009)	<.001	−0.222 (0.013)	<.001
Age			−0.121 (0.014)	<.001
Age × time			−0.012 (0.002)	<.001
Male sex			−0.202 (0.180)	.26
Male sex * time			−0.002 (0.018)	>.50
CES-D score (γ_{01} for β_{0j})			−0.002 (0.012)	>.50
CES-D × time (γ_{11} for β_{1j})			−0.003 (0.001)	.03
	Variance (SD)	*p*	Variance (SD)	*p*
Random effects				
Level 1 residual (R_{ij})	1.994 (0.027)	<.001	1.994 (0.027)	<.001
Level 2 residuals[a]				
Intercept (U_{0j})	18.557 (0.581)	<.001	17.927 (0.562)	<.001
Linear slope (U_{1j})	0.118 (0.006)	<.001	0.113 (0.006)	<.001

[a] Covariance parameters of level 2 residuals are not included in the table but were included in the analysis.
CES-D = Center for Epidemiologic Studies–Depression; SE = standard error; SD = standard deviation.

subsequent models. The objective of entering additional predictors of BMI is to explain the variation in BMI initial status and rate of change that exists between persons. If we select relevant predictors of this process, we should expect a reduction in the size of these coefficients. The within-person residual variance (R_{ij}) reflects the difference between observed and predicted outcome scores resulting from the "random noise" in the instruments we use to measure outcomes. More reliable measurement instruments will normally reduce the magnitude of this variance component.

The results of model B indicate that CES-D scores were unrelated to baseline levels of BMI ($\gamma_{01} = -0.002$, $p > .50$). In contrast, CES-D scores were significantly associated with decline in BMI levels ($\gamma_{11} = -0.003$, $p = .03$). The coefficient may be interpreted as follows: the average decline in BMI per year, according to model B, is −0.222. It should be noted that this estimate only applies to subjects who have a 0 value on all the covariates in the model; in other words, a 75-year-old (centered value of age) woman with a CES-D score of 5 (centered value on the CES-D). Each one-point increase on the CES-D above a score of 5 is associated with a 0.003 greater rate of yearly decline in BMI, and each one-point decrease below a score of 5 is associated with a 0.003 smaller rate of yearly decline. Older age is associated with lower BMI at baseline ($\gamma_{01} = -0.121$, $p < .001$) and with faster decline in BMI over time ($\gamma_{11} = -0.012$, $p < .001$). Male sex was not significantly associated with BMI levels at baseline or with decline over time.

One way to illustrate the nature of the association between CES-D scores and change in BMI is to use the results of the last regression model (model B) to compute the predicted values of BMI over time as a function of different levels of CES-D (see Fig. 1). To that end, we selected CES-D scores in this cohort corresponding to the 10th percentile (CES-D = 0), the median or 50th percentile (CES-D = 5), and the 90th percentile (CES-D = 19). We then plotted the predicted values of BMI until the end of follow up after selecting specific values for age (75, the centered value for this variable) and sex (female), the other two variables in the model (see Fig. 1). Although there were only minimal differences in predicted BMI levels at baseline between the three different levels of CES-D, each higher CES-D score was associated with a greater decline in BMI over time.

Inspection of the variance components reveals that addition of the fixed effects in the level 2 model accounts for relatively small portions of the between-person variability in initial status in BMI (intercept) and rate of change in BMI (slope). We can compare the change in each of the random coefficients from model A to model B to determine the proportion of explained variance. For the intercept, the proportion of additional variance may be computed as ([18.557 − 17.927]/18.557) × 100%, or approximately 3.4%. Similarly, the reduction in the random slope is ([0.118 − 0.113]/0.118) × 100% or 3.1%. In other words, the fixed effects account for approximately 3.4% of the variation in baseline BMI values that exists between persons in this sample and 3.1% of the variation in the linear rate of change in BMI.

Example 2: Communal Orientation Moderates the Association Between Diurnal Cortisol Rhythm and Abrasive Social Interactions

Dominance has been implicated as a risk factor for cardiovascular disease in prospective studies (e.g., (25,26)). Although reactivity to social stress is thought to be a mechanism for this link, laboratory studies of acute social strain have yielded mixed findings (27–30). We reasoned that this inconsistency might be resolved by considering person–environ-

MIXED REGRESSION MODELS

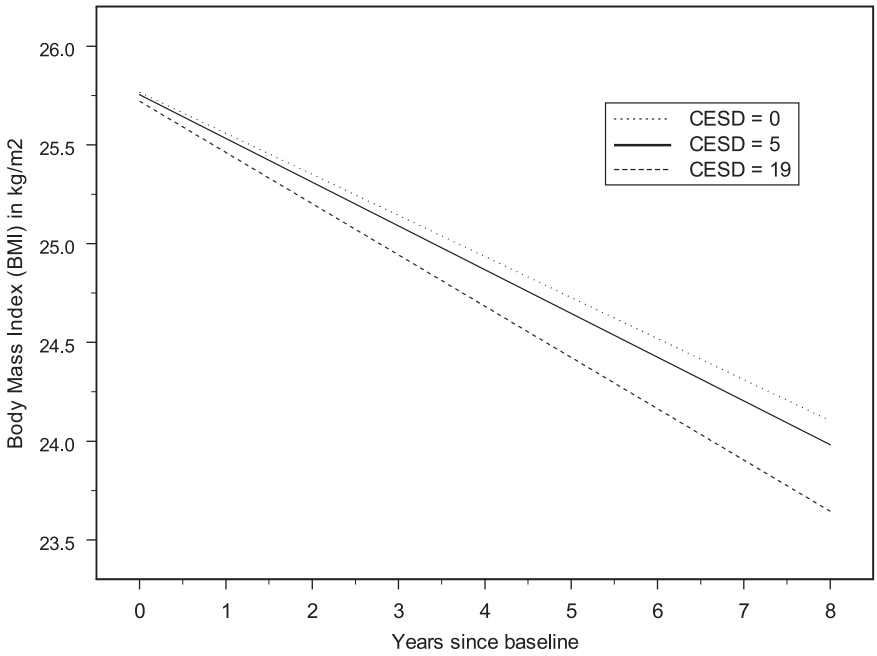

Figure 1. Change in body mass index as a function of depressive symptoms.

ment fit. People are more prone to physiological distress when their preferred behavioral style clashes with situational demands (e.g., (31)). Because social conflict should allow dominant individuals to display their preferred style, which is to exert control, they should evidence few signs of physiological stress. In contrast, individuals who are low in dominance may exhibit greater signs of physiological stress during discordant interactions, because these situations may challenge their normally passive stance.

We recently investigated whether dominance moderates the relationship between abrasive social interactions and diurnal cortisol secretion in daily life. Cortisol secretion usually peaks in the early morning and then declines throughout the day (32). Exposure to chronic stressors may result in a flattened diurnal rhythm (33,34). Disruptions in the regular pattern of cortisol release could have ramifications for a variety of biologic processes in the immune, circulatory, metabolic, and nervous systems (33,35) and may predict physical health problems (33,36,37).

Data were obtained from 87 healthy volunteers (82% female, mean age 29 years), 52 of whom were instructed to collect data across 4 nonconsecutive days and 35 of whom were instructed to collect data across 3 nonconsecutive days. A handheld computer sounded a daily alarm at 1, 4, 9, and 11 hours after the participant's planned wakeup time. Thus, ambulatory data collection occurred four times a day over a period of either 3 or 4 days. Our outcome measure was diurnal cortisol rhythm, which was computed by log-transforming the raw values and then regressing the transformed values on sampling times for each day separately. Note that we did not use individual cortisol assessments as our outcome measure, but rather the estimated *cortisol rhythm* we computed for each day per participant. Thus, we have a two-level model with cortisol rhythm across days nested within persons. Although this data set in fact represents three levels of data collection (cortisol output at each time-point nested with days in turn nested within persons), we reduced it to a two-level structure for the purpose of this example. In our own work, we have analyzed these data with both three- and two-level models; the results were similar.

Abrasive interactions were assessed using three questions from the Diary of Ambulatory Behavioral States (38). Each item used a 5-point scale ranging from 0 (none at all) to 4 (extremely). We collapsed the responses to these three questions to create an average measure of abrasive interactions at each time point. Because our outcome measure (cortisol rhythm) represents a day-level variable, we further collapsed these averages across the four time points on each day to obtain a day-level measure for abrasive interactions. However, the day-level responses were highly skewed with few ratings in the 1 to 4 range. We therefore recoded these scores to a dichotomous variable, with 0 indicating no abrasive interactions on that day and 1 indicating one or more abrasive interactions. This indicator created two groups of roughly equal frequencies representing low and high levels of abrasive interactions. This variable was used as the level 1 predictor in the ensuing analyses. Dominance (D) was measured using the Unmitigated Agency subscale of the Extended Version of the Personality Attributes Questionnaire (EPAQ) (39).

Analytic Strategy

The nested design of our study (either 3 or 4 days of data collection nested within each participant) allowed us to use mixed regression modeling to test whether the within-person relationship between abrasive social interactions and diurnal

patterns of cortisol secretion varies as a function of the between-person characteristic D. A two-level model was constructed with random slopes and intercepts at the first level modeled as outcomes at the next level up. At level 1 (day-level), the between-day variability in cortisol rhythm for each individual was modeled as a function of abrasive social interactions (AI):

$$\text{Cortisol rhythm}_{ij} = \beta_{0j} + \beta_{1j}(AI) + R_{ij} \qquad (4)$$

In this equation, the abrasive interactions covariate has been person-centered. In general, level 1 centering will reduce problems in estimation because it reduces the covariance between the intercept and the slope (7). Centering also enables us to ask how each individual's cortisol rhythm changes in response to fluctuations from his or her own usual pattern of interactions. That is, the questions becomes: How does a person's cortisol rhythm differ on days when he or she has high compared with low levels of abrasive interactions? As evident from this question, the analysis is completely within-person; centering eliminates the influence of between-person differences in abrasive interactions on the outcome.

The regression coefficients β_{0j} and β_{1j} can be understood as follows: For each individual j, β_{0j} represents the expected value of diurnal cortisol rhythm on average interaction days (i.e., when abrasive exchanges are at person j's mean daily level), and β_{1j} represents how diurnal cortisol rhythm varies in response to deviations from person j's average levels of abrasive interactions. The residual error associated with each observation is denoted by R_{ij}.

At level 2 (person-level), the between-person variability in β_{0j} and β_{1j} were modeled as a function of dominance:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(D_j) + U_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}(D_j) + U_{1j}$$

Note that the dominance variable was grand-mean centered so that D_j reflects the amount by which each person deviates from the sample mean. In other words, $D_j = (D_j - D_{mean})$. This centering strategy allows for a more meaningful interpretation of the intercept, because a score of 0 falls outside of the range of possible dominance scores. With the dominance score centered in this manner, the intercept parameters γ_{00} and γ_{10} denote the expected values of β_{0j} (diurnal cortisol rhythm on typical interaction days) and β_{1j} (the relationship between diurnal cortisol rhythm and abrasive interactions) for a person whose dominance score equals the sample mean. The slope parameter γ_{01} estimates the relationship between β_{0j} and D, or the main effect of D. It signifies whether diurnal cortisol rhythm on typical interaction days varies according to between-person differences in dominance. However, our main interest is in the slope parameter γ_{11}, which estimates the relationship between β_{1j} and D, or the D $\times$ AI interaction effect. Importantly, this parameter enables us to test our hypothesis that the relationship between diurnal cortisol rhythm and abrasive interactions is moderated by this trait. The terms U_{0j} and U_{1j} capture unexplained variability in β_{0j} and β_{1j} after including this predictor.

A standard assumption of mixed regression models is that within-person residuals are independent. However, as mentioned earlier, this is not always the case in diary studies in which observations are so closely spaced together within the same person. Models with different variance–covariance structures can be compared using a likelihood ratio test, which compares the deviances of models with different restrictions. Deviance is a measure of lack of fit between the model and the data. Using this option, we compared the fit of a standard homogenous level 1 variance model (which also assumes uncorrelated level 1 residuals) to the fit of an autoregressive model, which assumes that level 1 residuals are correlated in a time-decreasing fashion. The χ^2 statistic indicated that the two models did not differ significantly for our data set ($\chi^2 = 10.92$, $p = .141$). We therefore assumed an uncorrelated residual structure in the ensuing analysis. This lack of autocorrelation in our data set may be the result of the wide time intervals between measurements. In studies in which measurement intervals are shorter, an autoregressive structure often provides a better fit.

The analysis followed the same sequence as in the previous example. At the first step (model A), we specified an unconditional model without the level 2 predictor D. At the next step (model B), we added the level 2 predictor D. The coefficient for the D $\times$ AI interaction represents the test of the hypothesis. This two-step procedure allows us to compare the random coefficients of model A and model B and calculate the proportion of variance that is explained by including the level 2 predictor.

Abrasive Interactions, Cortisol Rhythm, and Dominance

The data were analyzed using the HLM software package (40). However, the same model may also be fitted in other software packages such as SAS (procedure MIXED) (41) and SPSS (command VARCOMP) (42). Results of the analysis are presented in Table 3. For the unconditional level 2 model (model A), the expected (mean) cortisol rhythm for the overall sample on average interaction days is $\gamma_{00} = -0.053$, which is significantly different from 0. The estimated value of γ_{10} is 0.003, which is nonsignificant. This suggests a very weak association between cortisol rhythm and abrasive interactions *in the sample as a whole.*

We now turn to the results of the level 2 model (model B). The values for γ_{00} and γ_{10} now have slightly different interpretations. These parameters estimate, respectively, the expected values of cortisol rhythm on average interaction days and the relationship between cortisol rhythm and abrasive interactions for a person whose D score equals the sample mean ($D_{mean} = 2.1$). The estimates were not appreciably different from those obtained in the unconditional level 2 model (model A).

Importantly, and confirming our hypothesis, the level 2 model reveals a significant D $\times$ AI interaction effect ($\gamma_{11} =$

MIXED REGRESSION MODELS

TABLE 3 Cortisol Rhythm and Abrasive Social Interactions as a Function of Dominance

	Model A		Model B	
	Coefficient (SE)	*p*	Coefficient (SE)	*p*
Fixed Effects				
Intercept (γ_{00} for β_{0j})	−0.053 (.003)	<.001	−0.053 (.003)	<.001
AI slope (γ_{10} for β_{1j})	0.003 (.005)	>.50	0.003 (.005)	>.50
D (γ_{01} for β_{0j})			0.001 (.004)	>.50
D × AI (γ_{11} for β_{1j}) 0.021			−0.012 (.005)	.021
	Variance (SD)	*p*	Variance (SD)	*p*
Random effects				
Level 1 residual (R_{ij})	0.00080 (0.028)	NA[a]	0.00079 (0.028)	NA[a]
Level 2 residuals				
Intercept (U_{0j})	0.00039 (0.020)	<.001	0.00039 (0.020)	<.001
Linear slope (U_{1j})	0.00016 (0.012)	>.25	0.00015 (0.012)	>.25

[a] The *p* value has not been reported here because HLM software does not provide a significance test for level 1 residuals.
SE = standard error; SD = standard deviation; NA = not available; HLM = hierarchical linear model.

−0.012, $p = .02$). This indicates that the strength of the within-person association between abrasive interactions and cortisol rhythm is moderated by trait dominance; the relationship between cortisol rhythm and abrasive interactions becomes increasingly negative for each D score above the sample mean and increasingly positive for each D score below the sample mean. So for individuals with high trait dominance, cortisol rhythms tend to be steeper on days when abrasive interactions exceed usual levels but flatter on days when abrasive interactions fall below usual levels. The finding of flatter (i.e., more dysregulated) rhythms on low conflict days was surprising, because we had expected cortisol rhythms to be similar for dominant persons regardless of frequency of conflict. It may be that for persons with a strong need for interpersonal control, neutral social interactions are perceived as potential threats rather than opportunities for affiliation. The reverse pattern is observed for individuals with low trait dominance whose cortisol rhythms tend to be flatter on days when abrasive interactions exceed usual levels and steeper on days when abrasive interactions fall below usual levels. Thus, the biologic consequences of conflict appear to be higher for persons who are low in dominance, possibly because they view conflict as more challenging or anxiety-provoking. However, in the absence of conflict, they may have a biologic advantage compared with their high-dominance counterparts. This interaction effect is illustrated in Figure 2 for D scores corresponding to the 25th, 50th, and 75th percentiles of the distribution.

Before leaving this example, we draw attention to the variance components statistics in Table 3, which tell us how much of the variability in the outcome measure remains unaccounted for at each level of the model. By comparing the random coefficients from model A and model B, we can determine the proportion of between-person variance that is explained by inclusion of the level 2 predictor variable D. For the random intercept, there was no detectable change in variance between the two models. That is, the fixed effects did not contribute in any appreciable way to the between-person variability in cortisol rhythms on average interaction days. The proportion of change in the slope variance may be computed as ([0.00016 − 0.00015]/0.00016) × 100%, which is approximately 6%. This indicates that the fixed effects account for approximately 6% of the between-person variability in the relationship between cortisol rhythm and daily abrasive interactions.

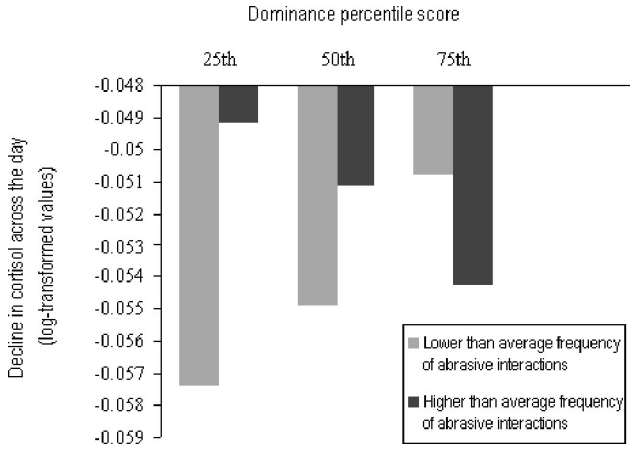

Figure 2. Diurnal cortisol rhythm as a function of dominance on days when the frequency of abrasive social interactions is higher or lower than usual.

CONCLUSION

Mixed regression models provide powerful tools for the analysis of change in repeated-measures studies. They can be applied in a variety of settings, ranging from relatively small-scale laboratory studies in which data are collected over the course of minutes, hours, or days, to large-scale epidemiologic investigations with follow-up periods lasting years. Many psychosomatic studies could benefit from the use of these models, because they often involve repeated measurements of outcomes within individuals or other types of nested data.

These models will not necessarily produce *different* results compared with more traditional methods. In general, however, they tend to yield more *accurate* results, thereby increasing the likelihood that the findings will be replicable.

REFERENCES

1. Elston RC, Grizzle JE. Estimation of time-response curves and their confidence bands. Biometrics 1962;18:148–59.
2. Cochran WG, Cox GM. Experimental Designs, 2nd ed. Oxford, UK: John Wiley & Sons, 1957.
3. Laird NM, Ware JH. Random-effects models for longitudinal data. Biometrics 1982;38:963–74.
4. Lindley DV, Smith AFM. Bayes estimates for the linear model. J Roy Stat Soc Ser B 1972;34:1–41.
5. Jennings JR. Editorial policy on analyses of variance with repeated measures. Psychophysiology 1987;24:474–5.
6. Petkova E, Teresi J. Some statistical issues in the analyses of data from longitudinal studies of elderly chronic care populations. Psychosom Med 2002;64:531–47.
7. Nezlek JB. Multilevel random coefficient analyses of event- and interval-contingent data in social and personality psychology research. Pers Soc Psychol Bull 2001;27:771–85.
8. Nezlek JB. Using multilevel random coefficient modeling to analyze social interaction diary data. J Soc Pers Relat 2003;20:437–69.
9. Schwartz JE, Stone AA. Strategies for analyzing ecological momentary assessment data. Health Psychol 1998;17:6–16.
10. Sinharay S, Stern HS, Russell D. The use of multiple imputation for the analysis of missing data. Psychol Methods 2001;6:317–29.
11. Singer JD, Willett JB. Applied Longitudinal Analysis. Modeling Change and Event Occurrence. New York: Oxford University Press, 2003.
12. Snijders T, Bosker R. Multilevel Analysis. London: Sage, 1999.
13. Raudenbush SW, Bryk AS. Hierarchical Linear Models: Applications and Data Analysis Methods, 2nd ed. Thousand Oaks, CA: Sage, 2002.
14. Willett JB. Questions and answers in the measurement of change. In: Rothkopf EZ, ed. Review of Research in Education. American Educational Research Association, 1988:345–422.
15. Kreft IGG, de Leeuw J. Introducing Multilevel Modeling. Thousand Oaks, CA: Sage, 1998.
16. Durazo-Arvizu RA, McGee DL, Cooper RS, Liao Y, Luke A. Mortality and optimal body mass index in a sample of the US population. Am J Epidemiol 1998;147:739–49.
17. Rexrode KM, Carey VJ, Hennekens CH, Walters EE, Colditz GA, Stampfer MJ, Willett WC, Manson JE. Abdominal adiposity and coronary heart disease in women. JAMA 1998;280:1843–8.
18. Seeman T, Mendes de Leon C, Berkman L, Ostfeld A. Risk factors for coronary heart disease among older men and women: a prospective study of community-dwelling elderly. Am J Epidemiol 1993;138:1037–49.
19. Weinstein AR, Sesso HD, Lee IM, Cook NR, Manson JE, Buring JE, Gaziano JM. Relationship of physical activity vs body mass index with type 2 diabetes in women. JAMA 2004;292:1188–94.
20. Dey DK, Rothenberg E, Sundh V, Bosaeus I, Steen B. Height and body weight in the elderly: a 25-year longitudinal study of a population aged 70 to 95 years. Eur J Clin Nutr 1999;53:905–14.
21. Dziura J, Mendes de Leon C, Kasl S, DiPietro L. Can physical activity attenuate aging-related weight loss in older people? The Yale Health and Aging Study, 1982–1994. Am J Epidemiol 2004;159:759–67.
22. Flynn MA, Nolph GB, Baker AS, Martin WM, Krause G. Total body potassium in aging humans: a longitudinal study. Am J Clin Nutr 1989; 50:713–7.
23. Cornoni-Huntley J, Ostfeld AM, Taylor JO, Wallace RB, Blazer D, Berkman LF, Evans DA, Kohout FJ, Lemke JH, Scherr PA. Established populations for epidemiologic studies of the elderly: study design and methodology. Aging Clin Exp Res 1993;5:27–37.
24. Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. Applied Psychological Measurement 1977;1: 384–401.
25. Houston BK, Chesney MA, Black GW, Raglan DR. Social dominance and 22-year all-cause mortality in men. Psychosom Med 1997;59:5–12.
26. Siegman AW, Kubzansky LD, Kawachi I, Boyle S, Vokonas PS, Sparrow D. A prospective study of dominance and coronary heart disease in the normative aging study. Am J Cardiol 2000;86:145–9.
27. Rejeski WJ, Gagne M, Parker PE, Koritnik DR. Acute stress reactivity from contested dominance in dominant and submissive males. Behav Med 1989;15:118–24.
28. Rejeski WJ, Parker PE, Gagne M, Koritnik DR. Cardiovascular and testosterone responses to contested dominance in women. Health Psychol 1990;9:35–47.
29. Newton TL, Bane CM, Flores A, Greenfield J. Dominance, gender, and cardiovascular reactivity during social interaction. Psychophysiology 1999;36:245–52.
30. Sgoifo A, Braglia F, Costoli T, Musso E, Meerlo P, Ceresini G, Troisi A. Cardiac autonomic reactivity and salivary cortisol in men and women exposed to social stressors: relationship with individual ethological profile. Neurosci Biobehav Rev 2003;27:179–88.
31. Engebretson TO, Matthews KA, Scheier MF. Relations between anger expression and cardiovascular reactivity: reconciling inconsistent findings through a matching hypothesis. J Pers Soc Psychol 1989;57:513–21.
32. Stone AA, Schwartz JE, Smyth J, Kirschbaum C, Cohen S, Hellhammer D, Grossman S. Individual differences in the diurnal cycle of salivary free cortisol: a replication of flattened cycles for some individuals. Psychoneuroendocrinology 2001;26:295–306.
33. Heim C, Ehlert U, Hellhammer DH. The potential role of hypocortisolism in the pathophysiology of stress-related bodily disorders. Psychoneuroendocrinology 2000;25:1–35.
34. Miller GE, Cohen S, Ritchey AK. Chronic psychological stress and the regulation of pro-inflammatory cytokines: a glucocorticoid-resistance model. Health Psychol 2002;21:531–41.
35. Weiner H. Perturbing the Organism: The Biology of Stressful Experience. Chicago: University of Chicago Press, 1992.
36. Sephton SE, Sapolsky RM, Kraemer HC, Spiegel D. Diurnal cortisol rhythm as a predictor of breast cancer survival. J Natl Cancer Inst 2000;92:994–1000.
37. Smyth JM, Ockenfels MC, Gorin AA, Catley D, Porter LS, Kirschbaum C, Hellhammer DH, Stone AA. Individual differences in the diurnal cycle of cortisol. Psychoneuroendocrinology 1997;22:89–105.
38. Kamarck TW, Shiffman SM, Smithline L, Goodie JL, Paty JA, Gyns M, Jong JY. Effects of task strain, social conflict, and emotional activation on ambulatory cardiovascular activity: daily life consequences of recurring stress in a multiethnic adult sample. Health Psychol 1998;17:17–29.
39. Spence JT, Helmreich RL, Holahan CK. Negative and positive components of psychological masculinity and femininity and their relationships to self-reports of neurotic and acting out behaviors. J Pers Soc Psychol 1979;37:1673–82.
40. Bryk AS, Raudenbush SW, Congdon RJ. HLM: Hierarchical Linear and Nonlinear Modeling With the HLM/2L and HLM/3L Programs. Chicago: Scientific Software International, 1996.
41. Little RC, Milliken GA, Stroup WW, Wolfinger RD. SAS System for Mixed Models. Cary, NC: SAS Institute, 1996.
42. Norusis MJ. SPSS 11.0 Guide to Data Analysis. Upper Saddle River, NJ: Prentice Hall, 2002.