Published in final edited form as:

Psychosom Med. 2012 November ; 74(9): 925–936. doi:10.1097/PSY.0b013e3182736971.

Multilevel Modeling in Psychosomatic Medicine Research

Nicholas D. Myers, Ph.D.,

Department of Educational and Psychological Studies, University of Miami

Ahnalee M. Brincks, M.A.,

Department of Epidemiology and Public Health, University of Miami

Allison J. Ames, M.S.,

Department of Educational and Psychological Studies, University of Miami

Guillermo J. Prado, Ph.D.,

Department of Epidemiology and Public Health, University of Miami

Frank J. Penedo, Ph.D., and

Department of Psychology, University of Miami

Catherine Benedict, M.A.

Department of Psychology, University of Miami

Abstract

The primary purpose of this manuscript is to provide an overview of multilevel modeling for *Psychosomatic Medicine* readers and contributors. The manuscript begins with a general introduction to multilevel modeling. Multilevel regression modeling at two-levels is emphasized because of its prevalence in psychosomatic medicine research. Simulated datasets based on some core ideas from the Familias Unidas effectiveness study are used to illustrate key concepts including: communication of model specification, parameter interpretation, sample size and power, and missing data. Input and key output files from *Mplus* and SAS are provided. A cluster randomized trial with repeated measures (i.e., three-level regression model) is then briefly presented with simulated data based on some core ideas from a cognitive behavioral stress management intervention in prostate cancer.

Keywords

mixed regression models; random coefficient models; hierarchical linear models; centering; power; missing data

Multilevel data, sometimes referred to as hierarchical data, are common in psychosomatic research. A common hierarchical structure has repeated observations nested within participants (1,2). Time-varying data from the repeated observations constitute Level-1 variables, while time-invariant data from the individual participants constitute the Level-2 variables. A multilevel model (MLM) can be defined as a statistical model that has two or more distinct hierarchical levels, with variables at each of these levels, and a theoretical interest in relationships at one or more of, or across, these levels (3).

Cross-sectional data can also be viewed as multilevel. For example, data may be collected from patients who are receiving treatment for the same illness within different hospitals. Or, with respect to behavioral interventions, patients may receive a psychological treatment from different therapists. Suppose that the relevant population of hospitals (or therapists) was viewed as large and that some of the randomly sampled hospitals were assigned to an experimental treatment condition while other hospitals were assigned to the control condition (e.g., treatment as usual). From a conceptual perspective, variability in patient outcomes may be attributable to both patient-level characteristics (e.g., severity of condition) as well as hospital-level characteristics (e.g., experimental treatment versus control condition). A MLM enables researchers to partition variability in the outcome(s) by level (e.g., patient- and hospital-level) and allows for modeling within and across these levels simultaneously.

From a statistical perspective, a key concern is to what degree the assumption of independence of observations is violated. If the outcome measure for a patient who receives treatment at Hospital A is more similar to the outcome measures for other patients at Hospital A than to the outcomes for a patient at Hospital B, the independence assumption is violated. When this non-independence is non-trivial and is attributable to at least one cluster variable (e.g., the hospital; the therapist), imposing a fixed effects model that ignores the non-independence is often problematic from both an empirical perspective (e.g., increased Type I error) and a theoretical perspective (e.g., inability to model relationships at, and across, different levels).

A review of *Psychosomatic Medicine* from 2006–2010 revealed that 33 studies utilized a MLM (see Supplemental Digital Content 1). There is considerable variability within this set of studies, and across applications of MLMs in other disciplines (4), with regard to some core methodological issues: communication of model specification, parameter interpretation, sample size and power, and missing data. Modern approaches to address these core methodological issues are illustrated with example datasets in this manuscript (i.e., Example 1: Familias Unidas; Example 2: Stress Management). Prior to these illustrations we provide a brief general introduction to MLMs.

MLMs: A Brief Introduction

Researchers may recognize MLMs from multiple labels including mixed models, random coefficient models, covariance components models, and hierarchical linear and non-linear models. We adopt the nomenclature of Hox and Roberts (1) and classify MLMs into two groups: multilevel regression models (MLRMs) and multilevel structural equation models (MLSMs). MLRMs have an observed outcome or outcomes at Level-1, regression coefficients as outcomes at Level-2 and above, and generally observed independent variables. MLSMs have latent or observed outcomes at each level, latent or observed independent variables, and greater flexibility than MLRMs with regard to the path models that may be imposed. There are types of MLMs, however, that do not fit neatly into either category (e.g., multilevel survival analysis).

Multilevel substantive theories predated the ability to estimate what are currently considered MLMs (5–7). MLMs were conceptualized only a few decades ago and initially based on MLRMs (8–9). Initial algorithms that were implemented for the estimation of these models followed (10–13). In the late 1980s and early 1990s, several MLM texts were presented (14–16). In each of these texts, a primary focus was given to cross-sectional and/or longitudinal two-level linear regression models. Application of this type of MLM is most common within *Psychosomatic Medicine* (see Supplemental Digital Content 1). Therefore, this manuscript primarily focuses on this particular implementation of a MLRM unless otherwise noted.

It is worthwhile to provide a conceptual summary of an early forerunner of the MLRM that was alluded to in the previous paragraph: Regression coefficients as outcomes. Under such an approach a linear regression model was fitted to Level-1 (e.g., patient) data separately within each Level-2 unit (e.g., hospital). The regression coefficients (i.e., intercept and slope(s)) then were saved in a “Level-2 dataset” where Level-2 predictors were then specified (i.e., addressing a “unit of analysis” problem). Problems with such an approach included: sub-optimal estimation of regression coefficients within individual Level-2 units; forcing variability in the regression coefficients into a restrictive and often unrealistic covariance structure; and, the approach was inefficient (conceptually as well as empirically) and generally imprecise (particularly in regard to accurate estimation of standard errors). The ideas, however, that some Level-1 predictive relationships should be free to randomly vary across Level-2 units, and that some of this variance should be explained by Level-2 characteristics, both remain central to the present day MLRM.

Deciding how to specify Level-1 coefficients at Level-2 is a key consideration in MLRMs.¹ Specifying a Level-1 coefficient as random at Level-2 implies that the magnitude of the coefficient truly varies (beyond sampling error and that which can be accounted for in the structural model) across the relevant population of Level-2 units (e.g., mean risky sexual behavior of adolescents truly varies across a specified population of schools).² Specifying a Level-1 coefficient as non-randomly varying at Level-2 implies that the magnitude of the coefficient has a finite number of true values that can be adequately handled in the structural part of the model, thereby precluding the need for an accompanying random effect for that coefficient (e.g., the magnitude of the true within-school effect of adolescent functioning in school on risky sexual behavior of adolescents depends on mean family functioning in each school). Specifying a Level-1 coefficient as fixed at Level-2 implies that the magnitude of the coefficient has only one true value. Treating all Level-1 coefficients initially as random and then working backwards based on only statistical information from a particular non-large sample, while conceptually alluring, has known problems with regard to accurately recovering a true model (19). Therefore deciding how to specify Level-1 coefficients at Level-2 is a key decision that should be guided by theoretical (e.g., primary versus secondary predictors) and empirical information. The relevant decisions made within the subsequent examples of the current manuscript were made only to illustrate modeling possibilities.

MLM Notation and Software

As with any quantitative methodology with more than one set of contributors, notation, terminology, and software for MLM vary. Notation and terminology used in this manuscript are most consistent with Raudenbush and Bryk (19), which is consistent with the studies listed in Supplemental Digital Content 1. The software used for demonstration is *Mplus* (20) and SAS (21) but many other software options are available (22).

Consider the notation for independent variables in MLMs. Independent variables at Level-1 are denoted X_{qij} where q identifies the variable (e.g., X_{1ij} , X_{2ij} , ..., X_{kij}). The subscript i denotes a particular observation at Level-1. The subscript j denotes a particular cluster at Level-2. In the repeated measures example, X_{123} identifies variable X_1 measured at time 2 for person 3. In a cross-sectional example, X_{123} identifies variable X_1 measured for person 2 within hospital 3. Independent variables at Level-2 are often denoted in one of two ways. When a variable is measured at Level-2, it is denoted W_{qs} where q communicates which

¹It is possible to specify all of the effects as fixed effects and then use bootstrapping for clustered data to get accurate standard error estimates. Estimates of the variance components of the random effects could also be found with bootstrapping (17).

²Another view of random effects could be framed within the Neyman-Scott problem (18), where the unit effects may be construed as a nuisance to be addressed via the inclusion of random effects in order to increase the likelihood of observing consistent fixed effects.

Level-1 coefficient, β_{qj} ; a given variable was specified to predict and s uniquely identifies the variable (e.g., W_{q1j} , W_{q2j} ; etc.). In a cross-sectional example, W_{03j} (or W_{21j} as another example) identifies β_{0j} (or β_{2j}) as the relevant outcome, the third predictor of β_{0j} (or the first predictor of β_{2j}), for the j th hospital. Note that W_{03j} and W_{21j} could actually denote the same hospital-level predictor variable. When a variable is measured at Level-1 and then aggregated at Level-2, it is denoted as $\bar{X}_{qs,j}$ to make clear that observations of the variable were aggregated across i individuals within each j group (or across i time points within each j person). The potential importance of including the means of Level-1 predictors at Level-2 within relevant medical statistical models (e.g., modeling of ecological momentary assessment data) has been discussed for some time (23). Fixed effects at Level-2, γ_{qs} , adopt the previously defined subscript notation. Subscripts are dropped in this manuscript when their function is redundant.

Centering Predictors in MLMs

Centering predictors is an important aspect of MLMs that has long been misunderstood (24). Manifestations of such misunderstanding include omission of centering decisions in published papers (rendering regression coefficients impossible to interpret) and basing centering decisions on only empirical information (e.g., trying various centering options in search of statistical significance). Centering Level-1 predictors should be informed by substantive multilevel theory, well-defined research questions, and empirical information. While a full review of centering Level-1 predictors is beyond the scope of this manuscript, thorough reviews are available elsewhere (25,26).

Parallel in time to the publication of initial texts on MLMs, a literature emerged on the consequences of centering Level-1 predictors within MLMs (27–31). Primary options debated were the use of raw scores (RAS; no centering), grand mean centering (GMC; subtracting the mean across all Level-1 observations from each observation), and centering within cluster (CWC; subtracting the group mean within each cluster from each observation within the relevant cluster). Within this manuscript, GMC will be communicated as follows

$X_{qij} - \bar{X}_{q\cdot}$ at Level-1 and $\bar{X}_{qs,j} - \bar{\bar{X}}_{qs}$ at Level-2, where a double bar indicates the mean of the group means. CWC will be denoted $X_{qij} - \bar{X}_{qj}$; CWC does not apply at Level-2 because there are no upper-level units for these observations to be nested within (i.e., Level-3).

Kreft, de Leeuw, and Aiken (32) mathematically defined conditions under which imposing different centering options (RAS, GMC, and CWC) for Level-1 predictors in two common specifications of MLMs resulted in theoretically equivalent models. The two models of primary focus were (a) random coefficients regression and (b) intercepts and slopes as outcomes – both of these models are introduced in the subsequent section. Within both models, RAS and GMC are equivalent to each other and nonequivalent to CWC. That two parameterizations of a MLM, which differ only by centering decision(s), are statistically equivalent does not necessarily mean that both models do an equally good job of representing specific research questions and/or theory-driven hypotheses – a point that will be made clear in the Familias Unidas example. Prior to proceeding to the Familias Unidas example it is worth noting that centering (as described above) within a MLM can also be conceptualized within a broader framework as one way to adjust the estimated within-cluster effect in the presence of confounding by cluster (33–35).³

³The cited studies (33–35) provide a thorough account of when the estimated within-cluster effect is likely to be accurately adjusted in the presence of confounding by cluster under various models (e.g., conditional linear/logistic regression, general(ized) linear mixed models) and link functions (e.g., identity, logit, log-log). The approach taken in the current manuscript (e.g., Equation 6) has been shown to produce consistent and efficient estimates of the within-cluster effect when the fitted model is correctly specified (33,34). Although the approach taken in the current manuscript is both practical and useful (particularly for the type of examples reviewed in this manuscript) it also has been shown to be a special case of a more general approach (35).

Example 1: Familias Unidas

The HIV epidemic constitutes a major public health concern in the U.S. and around the globe. Although HIV has affected persons of all ages, from 2005 to 2008, the rates of HIV infection for those 15–19 years of age have increased (36). The prevalence of unprotected sexual intercourse among adolescents is one of the factors responsible for the high incidence of new HIV cases. Results from The CDC's Youth Risk Behavior Surveillance Survey indicate that 46% of high school students reported having had sexual intercourse, and 13.8% reported four or more lifetime sexual partners (37). Although these national statistics are alarming, what may be even more disconcerting is that Hispanics, the largest and fastest growing minority group in the United States (37), appear to be at heightened risk for engaging in HIV risk behaviors, including unprotected sexual behavior, and consequently at disproportionate risk of HIV infection (38).

Ecodevelopmental theory is a conceptual framework that has been put forth to explain risk and protective processes linked to HIV risk behaviors in Hispanic youth (39–41). Risky sexual behavior (RSB) is a primary health outcome within an ecodevelopmental model predicting HIV risk behaviors. Theory-based predictors of RSB include perceived peer abstinence (PA), adolescent functioning in school (SF), parental and adolescent Americanism gap (GAP), and family functioning (FAM).

Ecodevelopmental theory data are often collected within middle schools and/or high schools, where only one adolescent per family is sampled. Sampling through schools provides researchers with access to a large number of adolescents but also introduces the possibility that the data collected may be non-independent. From a statistical perspective, risky sexual behavior is measured at Level-1 (i.e., adolescent) but these data may be non-independent based on the school (i.e., Level-2) that an adolescent is nested within. From a conceptual perspective, risky sexual behavior may be partially explained by independent variables at Level-1 (e.g., family functioning), Level-2 (e.g., average gap between parental and adolescents Americanism at the school-level), and cross-level interactions (e.g., family functioning* average gap between parental and adolescents Americanism at the school-level).

The first example dataset was simulated based on some core ideas from an ongoing randomized trial evaluating the relative effectiveness of Familias Unidas (42), where RSB ($M = 0.04$, $SD = 1.06$, minimum = -4.16 , maximum = 5.13) was a composite variable that was continuous and normally distributed. Familias Unidas is a Hispanic-specific, family-based preventive intervention designed to reduce risks and increase protection for substance use and unprotected sexual behavior in Hispanic adolescents. Familias Unidas is guided by ecodevelopmental theory and aims to prevent substance use and unprotected sexual behavior by increasing positive parenting, family support of the adolescent, parental involvement, as well as by improving general parent-adolescent communication and parent-adolescent communication specific to substance use, unsafe sexual behavior, and HIV. In the Familias Unidas effectiveness study, a total of 744 Hispanic youth and their parents are being recruited from 24 middle schools in Miami-Dade County. Participants are assessed at baseline, randomized to either Familias Unidas or a Community Practice control condition, and reassessed at 6, 18, and 30-months post baseline. In a preliminary baseline dataset, adolescents ($N = 460$) were nested within middle schools ($J = 16$) and within-cluster size was unbalanced ($\bar{n}_j = 28.75$) with the smallest school having 22 adolescents and the largest school having 32 adolescents.

A larger dataset using some characteristics from the Familias Unidas dataset was created because multilevel modeling generally is a large sample technique (43); a point that will be

elaborated upon in the section on sample size and power. In the simulated dataset (i.e., dataset_1), adolescents ($N = 2869$) were nested within schools ($J = 100$) and within-cluster size was unbalanced ($\bar{n}_j = 28.69$) with the smallest cluster having 22 observations and the largest cluster having 32 observations. Because dataset_1 through dataset_3 were simulated, the results that follow should be viewed as only a means for providing a relevant context for understanding multilevel modeling.

MLRM and Familias Unidas

The Familias Unidas simulated dataset is used to illustrate key concepts in a MLRM including: communication of model specification, parameter interpretation, sample size and power, and missing data. Model specification follows the general strategy outlined in Raudenbush and Bryk (19) where MLRMs are introduced by way of six “sub-models.” In practice, a single MLM may be fitted when sufficient a priori theory exists, but proceeding in this incremental way is consistent with the purpose of this manuscript. For spatial reasons, however, most of these sub-models (i.e., Model 2: Means as Outcomes Regression Model; Model 3: One-Way ANCOVA with Random Effects; Model 4: Non-Randomly Varying Slopes; Model 5: Random Coefficients Regression) are elaborated upon in Supplemental Digital Content 2 only (along with Equation 2, Equation 3, Equation 4, and Equation 5). The first four models are random intercept models. The final two models are random intercepts and slopes models. The alpha for all hypothesis tests was set equal to .05.

Communication of model specification is provided in equation form and also via a path diagram. Each equation is written in uncombined form with variable acronyms (e.g., RSB_{ij}) and then in combined form with general notation (e.g., Y_{ij}) in Supplemental Digital Content 2. The uncombined equation is a way to emphasize the meaning of the parameters by writing equations at each level. The combined equation is an equivalent, compact expression that collapses the uncombined equation across levels. An interpretation of only key terms (to reduce redundancy) follows each equation. The causal language used (e.g., effect) communicates the direction of the relationship assumed by the model specification. In practice, any causal inference would also need to account for design considerations and substantive theory in addition to the model specification. A path diagram of each model specification (44) is also provided in Supplemental Digital Content 2 (see Figures 1–6). Providing path diagrams in addition to equations is done to provide future researchers with options for how to succinctly and clearly communicate model specification.

Model 1: One-Way ANOVA with Random Effects⁴

A primary purpose of Model 1 is to partition the unconditional variance in the outcome(s) by level to assess the degree of non-independence of observations. Degree of non-independence is typically assessed by the intraclass correlation coefficient: $ICC = \tau_{00} / (\tau_{00} + \sigma^2)$. Simulation research has shown potential consequences (e.g., inflated Type I) of ignoring the non-independence (e.g., imposing a fixed effects model) when the ICC is as small as .05 (45).

This model is fully unconditional in that there are no predictors at either level.⁵ This model can be written:

⁴This model also can be referred to as a Nested 1-way ANOVA with the Level-s units treated as a random factor.

⁵Non-multilevel modeling options for such data include treating the cluster variable as a fixed factor and thereby limiting inferences to the given sample of cluster units; and, modeling only at Level-1 while applying a statistical correction for the non-independence by cluster (46–48).

$$\begin{aligned} RSB_{ij} &= \beta_{0j} + r_{ij} \\ \beta_{0j} &= \gamma_{00} + u_{0j} \end{aligned} \quad (1)$$

RSB_{ij} = observed risky sexual behavior for the i^{th} adolescent in the j^{th} school

β_{0j} = mean risky sexual behavior in the j^{th} school

γ_{00} = grand mean risky sexual behavior across schools

r_{ij} = unique effect of the i^{th} adolescent on risky sexual behavior in the j^{th} school

u_{0j} = unique effect of the j^{th} school on risky sexual behavior

$\sigma^2 = \text{Var}(r_{ij})$ = unconditional adolescent-level variance in risky sexual behavior

$\tau_{00}^2 = \text{Var}(u_{0j})$ = unconditional school-level variance in risky sexual behavior

Table 1 provides parameter estimates, standard errors, and model-data fit information for this model and all subsequent models. Supplemental Digital Content 3 provides input and key output for this model and all subsequent models.⁶ The ICC was equal to .064, which implied that 6.4% of the variance in risky sexual behavior was attributable to school-level differences. The estimated grand mean risky sexual behavior across schools, $\hat{\gamma}_{00}$, was equal to 0.040. The unconditional school-level variance estimate, $\hat{\tau}_{00}$, around this average was equal to 0.072 and was statistically significant, $p < .001$.⁷ Table 2 provides descriptive statistics at Level-1 (i.e., “within” or adolescent-level) and Level-2 (i.e., “between” or school-level) due to evidence of non-independence by school. This information can be used to describe the bivariate associations that exist at each level.

Model 6: Intercepts and Slopes as Outcomes

Suppose that the interest is in modeling adolescent- and school-level effects on risky sexual behavior while specifying random intercepts and a mix of fixed, non-randomly varying, and random slopes. Level-1 predictors are specified to address research questions based on ecodevelopmental theory (e.g., what is the effect of perceived peer abstinence on risky sexual behavior at the adolescent-level?). Level-2 predictors of β_{0j} and β_{2j} are specified to address research questions based on ecodevelopmental theory (e.g., what is the effect of mean perceived peer abstinence on mean risky sexual behavior at the school-level?). For a fuller explanation of the wider array of research questions see Model 2 – Model 5 in Supplemental Digital Content 2. Suppose further that there is reason to believe (1) that the magnitude of the within-school effect of family functioning on risky sexual behavior depends, in part, on the level of mean gap between parental and adolescents Americanism (i.e., a cross-level statistical interaction), and (2) that there will be substantial residual variance in the set of β_{4j} unaccounted for. This model can be written:

⁶Results reported in the text of this manuscript are from *Mplus* output. Regardless of statistical program used to fit a particular MLM, individual Level-1 coefficients are not model parameters per se. Estimating reasonably accurate Level-1 coefficients can be a complex task (often accomplished with a Bayesian framework) and is an area of research (51).

⁷Testing the statistical significance of the variance of a Level-2 random effect, u_{qj} , is a somewhat controversial issue in MLM. The problem is $H_0: \tau_{qq} = 0$ occurs at the boundary of the parameter space because the true value of a variance cannot be negative. Corrections to instances where such a hypothesis test is imposed have been offered (e.g., 52).

$$\begin{aligned}
 RSB_{ij} = & \beta_{0j} + \beta_{1j} (PA_{ij} - \overline{PA}_{.j}) + \beta_{2j} (SF_{ij} - \overline{SF}_{.j}) + \beta_{3j} (GAP_{ij} - \overline{GAP}_{.j}) + \beta_{4j} (FAM_{ij} - \overline{FAM}_{.j}) + r_{ij} \\
 \beta_{0j} = & \gamma_{00} + \gamma_{01} (\overline{PA}_{.j} - \overline{PA}_{..}) + \gamma_{02} (\overline{SF}_{.j} - \overline{SF}_{..}) + \gamma_{03} (\overline{GAP}_{.j} - \overline{GAP}_{..}) + \gamma_{04} (\overline{FAM}_{.j} - \overline{FAM}_{..}) + u_{0j} \\
 & \beta_{1j} = \gamma_{10} \\
 & \beta_{2j} = \gamma_{20} + \gamma_{24} (\overline{FAM}_{.j} - \overline{FAM}_{..}) \\
 & \beta_{3j} = \gamma_{30} \\
 & \beta_{4j} = \gamma_{40} + \gamma_{43} (\overline{GAP}_{.j} - \overline{GAP}_{..}) + u_{4j}
 \end{aligned} \tag{6}$$

u_{4j} = residual within-school effect of risky sexual behavior regressed on family functioning (after controlling for the effect of peer abstinence, adolescent functioning in school, and gap between parental and adolescents Americanism at Level-1) after controlling for the effect of mean gap between parental and adolescents Americanism⁸

τ_{44} = variance of the u_{4j}

τ_{04} = school-level covariance between the u_{0j} and the u_{4j}

As can be readily viewed in the combined form of Equation 6 (see Figure 6 in Supplemental

Digital Content 2), γ_{24} represents a cross-level interaction: $\gamma_{24} \left((X_{2ij} - \overline{X}_{2..}) * (\overline{X}_{4.j} - \overline{X}_{4..}) \right)$. One way to interpret this interaction is that the magnitude of the effect of adolescent functioning in school on risky sexual behavior depends on the level of mean family functioning. The statistical significance of this cross-level interaction, $\hat{\gamma}_{24} = -0.009$, $p < .001$ (see Table 1), supported rejection of the null hypothesis that the magnitude of the effect of adolescent functioning in school on risky sexual behavior did not depend on level of mean family functioning: $H_0: \gamma_{24} = 0$.

A question unanswered by testing $H_0: \gamma_{24} = 0$ was: at what values of mean family functioning ($M = 0.00$, $SD = 1.13$, $Min = -2.13$, $Max = 3.45$) was the effect of adolescent functioning in school ($Min = -42.19$, $Max = 38.11$) on risky sexual behavior statistically significant? More formally, what were the regions of significance of mean family functioning for the simple slope of risky sexual behavior on the effect of adolescent functioning in school (53,54)? Figure 1 plots simple slopes of risky sexual behavior on the effect of adolescent functioning in school on a few values of mean family functioning: $-SD$, M , $+SD$. The slope was positive when mean family functioning was low, $-SD$, but the slope became increasingly negative as mean family functioning increased.

Figure 2 identifies the regions of significance of mean family functioning where the effect of adolescent functioning in school on risky sexual behavior was statistically significant and non-significant. When mean family functioning was “low” (i.e., -1.37) the effect of adolescent functioning in school on risky sexual behavior was statistically significant and positive. When mean family functioning was “low-average” (i.e., $-1.37 < \text{mean family functioning} < -0.17$) the effect of adolescent functioning in school on risky sexual behavior was not statistically significant. When mean family functioning was “average or greater” (i.e., -0.17) the effect of adolescent functioning in school on risky sexual behavior was statistically significant and negative.

⁸An alternative interpretation, offered by an anonymous reviewer, is that u_{4j} is that piece of the effect of the within-school effect of family functioning on risky sexual behavior (after controlling for the effects of peer abstinence, adolescent functioning in school, and gap between parental and adolescents Americanism at Level-1), β_{4j} , that is not accounted for by the mean gap between parental and adolescents Americanism.

The interpretation of the cross-level interaction, γ_{43} , followed the same form as that of γ_{24} . As can be viewed in Table 1, mean gap between parental and adolescents Americanism was a statistically significant predictor of the within-school effect of family functioning on risky sexual behavior: $\hat{\gamma}_{43} = 0.017$, $p < .001$. The statistical significance of $\hat{\tau}_{44} = 0.015$, $p < .001$ suggested that there was substantial residual variance in the β_{4j} slopes unaccounted for by mean gap between parental and adolescents Americanism.

Determining explained variance in multilevel models, particularly when at least one Level-1 slope coefficient is not fixed, is a complex task and an active area of research. As can be observed in Supplemental Digital Content 3, in the current model, information regarding variance accounted for was not provided. When a slope is not fixed (i.e., TYPE = TWOLEVEL RANDOM) *Mplus* does not provide variance accounted for in the output. An ad hoc procedure for determining proportion of variance accounted for was detailed in Raudenbush and Bryk (19).⁹

At Level-1 the relevant equation can be written:

$$R_1^2 = \frac{\sigma_{\text{baseline}}^2 - \sigma_{\text{full}}^2}{\sigma_{\text{baseline}}^2} \quad (7)$$

A key consideration is which model is designated the baseline model and which model is designated the full model. Using Model 1 as the baseline model and Model 6 as the full model, 32.8% of the Level-1 variance was accounted for by the set of Level-1 predictors.

At Level-2 relevant equations can be written:

$$R_{\beta_{0j}}^2 = \frac{\tau_{00\text{baseline}} - \tau_{00\text{full}}}{\tau_{00\text{baseline}}} \quad (8)$$

$$R_{\beta_{4j}}^2 = \frac{\tau_{44\text{baseline}} - \tau_{44\text{full}}}{\tau_{44\text{baseline}}} \quad (9)$$

Given that Level-2 predictors were specified for both β_{0j} and β_{4j} , both Equation 8 and Equation 9 were used to determine variance accounted for at Level-2. Model 5 (see Supplemental Digital Content 2) is used as the baseline model and Model 6 as the full model.¹⁰ Mean peer abstinence, mean adolescent functioning in school, mean gap between parental and adolescents Americanism, and mean family functioning jointly accounted for 12% of the variance in the unadjusted risky sexual behavior school means (i.e., β_{0j}). Mean gap between parental and adolescents Americanism accounted for 3.75% of the variance in the within-school slopes of risky sexual behavior regressed on family functioning (i.e., β_{4j}).

Sample Size and Power in MLRM

Three practical questions often arise when designing and/or analyzing data from an experiment: “what N is needed in order to use a particular statistical method?”, “how much power (π) do I have with a fixed N ?”, and “what N is needed in order to achieve a fixed

⁹There are limitations (e.g., possibility of negative variance accounted for) to this approach to variance explained (52). Other approaches, including a single index of explained variance for the entire model, are beginning to emerge (55).

¹⁰Both Equation 8 and Equation 9 assume that the specification of the Level-1 model is constant in the baseline model and the full model.

level of π ?”. A definitive answer to each of these somewhat under-defined questions is not possible across the myriad of model-data conditions observed in practice. Each of these questions is generally more difficult to answer under a MLRM as compared to a fixed effects linear regression model because there are multiple sample sizes to consider (i.e., n_j , J , and N) and the random part of the model is more complex (e.g., $u_{0j} + r_{ij}$ in Equation 1 versus r_i in Figure 2 of Supplemental Digital Content 2).

Sample Sizes in General

Our response to the question, “what N is needed in order to use a MLRM?”, is informed by summarizing the relevant simulation research (56). MLRM generally is a large sample technique that often uses maximum likelihood (ML) estimation. Results from full ML are summarized below. Estimation of the fixed effects, γ , is generally unbiased. Estimation of the standard errors for γ have been shown to be negatively biased, increasing Type I error rate, when $J < 50$; although this bias generally is small when $J \approx 30$ (57–59). Accurate estimation of the random component at Level-1, σ^2 , tends to be robust. Accurate estimation of the random component(s) at Level-2, $\mathbf{T}$ (where $\mathbf{T}$ is the Level-2 variance, $\tau_{qq'}$ -covariance, $\tau_{qq'}$, matrix), tends to be the most difficult, particularly when $J < 100$; although estimates may be adequate (if somewhat negatively biased) when $J \approx 50$ (57,60). Using restricted ML may markedly improve the estimation of $\mathbf{T}$ when J is not large (57,60).

Increasing J tends to be more important than increasing n_j for the purpose of maximizing the possibility of producing precise and powerful estimates. This trend along with the more specific findings summarized in the previous paragraph has informed a set of general guidelines (56). The 30/30 (i.e., $(J = 30)/(n_j = 30)$) guideline is offered for when the research questions are focused on γ without cross-level interactions. When the research questions are focused on cross-level interactions within γ the guideline is modified: 50/20. When the research questions are focused on $\mathbf{T}$ the guideline is 100/10. Like all sample size related guidelines, a limitation of the above guidelines is that they do not take fully into account study-specific characteristics.

Sample Sizes and Power in a Specific Study

Our response to the questions, “how much power (π) do I have with fixed values of n_j and J ?” and “what n_j and/or J is/are needed in order achieve a fixed level of π ?”, is demonstrated by using a flexible two-step Monte Carlo approach in *Mplus* (see Supplemental Digital Content 3). A full accounting of this approach, input and output, is available elsewhere (20,61). The first step is referred to as a real data analysis. The second step is referred to as a Monte Carlo simulation. This approach fits within the more general framework of using Monte Carlo methods in data analysis (62). The contribution of using Monte Carlo methods (e.g., simulating data) for a clinical trial with real data is toward improved application of a particular statistical model (e.g., MLRM) for a particular clinical trial, and hence resultant conclusions, and not to statistical theory per se (63).

Returning to the Familias Unidas study, another dataset (i.e., dataset_2) was generated following the same protocol used to generate dataset_1 with the exception that $N = 633$, $J = 25$, and $\bar{n}_j = 25.32$, which probably more closely resembles sample sizes in many trials due to budgetary constraints. Suppose that Model 6 is of interest (see Equation 6) and that answering the question “how much π do I have with fixed values of n_j and J ?” for each parameter of interest is desired. Parameters of interest, θ_i , were all statistically significant parameter estimates, except the main effects of the variables involved in the cross-level interactions, from Model 6 in Table 1. In Step 1, Model 6 was fitted to dataset_2 and parameter estimates were saved. In Step 2 these parameter estimates were then treated as population values for a Monte Carlo simulation where 10,000 datasets, with $\bar{n}_j = 25.32$ and J

$= 25$, were simultaneously generated and analyzed. Observed power for θ_i , $\hat{\pi}_i$ was defined as the percentage of replications that rejected $H_0: \theta_i = 0$. A commonly used desired $\hat{\pi}_i$ of 80% was adopted (64).

As displayed in Table 1, $\hat{\pi}_i$ for $\hat{\gamma}$ ranged from 33.4% to 97.5%. Two fixed effects, γ_{03} and γ_{43} , had low power and in both cases the simulation suggested that there was less than 50/50 chance that the relevant null hypothesis would be rejected. Three fixed effects, γ_{10} , γ_{24} and γ_{30} , were sufficiently powered and in all cases the simulation suggested that there was a very good chance that the relevant null hypothesis would be rejected. A similar range of $\hat{\pi}_i$ was observed for the variance components: from 53.6% to 100%. Clearly, in this example and most examples in practice, an answer to the question “how much π do I have with fixed values of n_j and J ” will be complex (e.g., sufficiently powered for primary research questions while insufficiently powered for non-primary research questions). Providing an answer to this question, however, may provide a useful context for evaluating results from a trial (e.g., chance of committing a Type II error).

Given the results summarized in the previous paragraph, and assuming that n_j is fixed, suppose the following question is asked: “what J is needed in order to achieve $\hat{\pi}_i = 80\%$ for each θ_i ?”. Through relatively minor manipulation of input for Step 2 it is determined that the answer to the question is $J \approx 115$. Because sampling more Level-2 units often requires more resources than sampling more Level-1 units, however, derivation of an optimal design in practice frequently considers trade-offs (in addition to $\hat{\pi}_i$) due to budgetary constraints and other study-specific characteristics (65–67). Software has been developed for this purpose (e.g., 68).

Missing Data in MLRM

A key to understanding the missing data literature is to comprehend Rubin’s taxonomy of missing data (69). Missing completely at random (MCAR) is when the propensity for missing data on a variable, G , is unrelated to values of other variables in the model and is also unrelated to values of G itself. Missing at random (MAR) is when the propensity for missing data on G is related to values of one or more other variables in the model but unrelated to values of G itself. Missing not at random (MNAR) is when the propensity for missing data on G is related to values of G itself. Which mechanism the missing data actually follow is typically impossible to identify with relative certainty because MCAR is the only missing data mechanism for which a statistical test exists (70). Modern approaches which address missing data generally assume MAR (71) although approaches that assume MNAR are beginning to be used with more frequency (72–74).

Incomplete data has long been a difficult issue in MLRM when a non-trivial proportion (e.g., 5%; 71) of data are missing for both the outcome(s) and at least one predictor (75,76).¹¹ Common default approaches in many statistical packages are to delete Level-1 observations with any incomplete Level-1 data (especially predictors), and, to delete all observations within a cluster if any Level-2 data (especially predictors) are missing within that cluster. These defaults may inefficiently use the observed data, assume MCAR, and may yield untrustworthy results (77,78). The use of a few simple strategies that directly provide a single set of observations for missing data, such as last observation carried forward and class mean imputation, has generally been discouraged (79–81).

¹¹The problem generally is considerably easier to address if incomplete data are confined to the outcome(s) and the missing data can reasonably be viewed as MAR.

Likelihood-based methods for handling missing data in MLRM address the missing data and fit the MLRM in a single step. Likelihood-based methods are preferred when the missing data are in the outcome(s) only, the specification of the MLRM is correct, and MAR can reasonably be assumed (76). Missing data in the predictors generally cannot be handled in likelihood-based methods (with typical model commands) although some programs (e.g., *Mplus*) do allow missing data on auxiliary variables (82). Likelihood-based methods can, however, handle missing data in the predictors if the predictors are brought into the model by specifying additional model commands that make distributional assumptions regarding the predictors (see Supplemental Digital Content 3). A current and thorough review of likelihood-based methods for handling missing data in medical research is available elsewhere (81).

Multiple imputation methods for handling missing data in MLRM address the missing data, and then fits the MLRM to complete datasets, in separate steps. The first step addresses the missing data problem by imputing multiple datasets while acknowledging the multilevel structure of the data.¹² The second step fits the complete datasets to the MLRM and combines the results to provide a single set of parameter estimates and standard errors (83,84). Multiple imputation of multilevel data is generally the preferred approach in MLRM (under the MAR assumption) because of its greater flexibility than likelihood-based approaches (76). It should be noted, however, that much methodological research remains to be done with regard to exploring the full range of possible missing data scenarios (and the relative performance of likelihood-based versus multiple imputation approaches) in MLRM.

Multiple imputation of multilevel data can be performed in *Mplus* under Bayesian estimation (85). Both unrestricted and restricted models can be used for imputation but for multilevel data specifying a restricted model may increase the likelihood of convergence under model-data conditions frequently observed in practice (e.g., a large number of variables, continuous and categorical variables, and a non-large number of Level-2 units). A review of the unrestricted model options available in *Mplus* is available elsewhere (85).

Returning to the Familias Unidas study, dataset_2 was altered (i.e., dataset_2_md) such that approximately 10% of cases were missing data on both RSB (i.e., the outcome) and FAM (i.e., a predictor). Using the default setting of most statistical packages, 323 cases would have been ignored when fitting Model 6 to dataset_2_md. Instead, prior to fitting Model 6 to the incomplete dataset, five imputed datasets were generated in *Mplus* (i.e., Step 1) under a restricted approach (i.e., Model 6 - see Supplemental Digital Content 3). Generating five imputed datasets was consistent with the small amount of missing data (71). In Step 2, Model 6 was fitted to the five imputed datasets in *Mplus* (see Supplemental Digital Content 3) and the results were consistent with those previously listed in Table 1.

Example 2: Stress Management

Three-level MLRMs occasionally appear in the literature (see Supplemental Digital Content 1). When three-level MLRMs are used the data structure often is repeated measures nested within people who are nested within Level-3 units (e.g., treatment provider, dyads, etc.). Thus, for example 2, we briefly present a cluster randomized trial with repeated measures based on some core ideas from a real intervention: cognitive behavioral stress management in prostate cancer (86). Fuller treatments of three-level MLRMs are available elsewhere (19).¹³

¹²Addressing the missing data problem by imputing multiple datasets, while ignoring the multilevel structure of the data, can be described as multiple imputation-flat file. There is evidence that multiple imputation of multilevel data outperforms multiple imputation-flat file for handling missing data that follow a multilevel structure (76).

Prostate cancer is the most common cancer among men in the United States, excluding skin cancer. There are over 2.2 million current survivors of prostate cancer, with over 240,000 new diagnoses estimated each year (87). Radical prostatectomy and radiation therapy are the standard treatment options for localized disease. Although both treatments are highly effective in disease management and extending survival, men often experience chronic and debilitating treatment-related side effects, such as urinary and/or bowel incontinence and erectile dysfunction, that challenge quality of life (88,89). Sexual dysfunction, for example, is associated with increased distress, depression and anxiety, and lowered self-esteem, quality of marital/partner relationships and general QOL (90–93). Psychosocial interventions tailored specifically for post-treatment prostate cancer survivors have been shown to have positive effects on survivors' ability to adjust to and cope with disease- and treatment-related changes.

The cognitive-behavioral stress management (CBSM) intervention is a 10-week, group-based intervention designed to improve quality of life through improvements in cognitive-behavioral stress management skills. It was modified from the original protocol designed by Antoni et al. (94) in order to address the specific needs and concerns of men coping with localized prostate cancer. A review of the CBSM intervention has been reported elsewhere (95). Briefly, intervention groups met once per week (over the course of 10 weeks) for approximately 2 hours, which consisted of a 90-minute didactic component and 30-minute relaxation training. Participants were instructed in a number of stress management skills, including identifying distorted or irrational thoughts, effective coping strategies, anger management, assertiveness training, and eliciting social support. The control condition consisted of a group-based, one-day seminar that lasted for approximately 4 hours during which participants were instructed in the same stress management skills that were taught in the CBSM intervention sessions.

Suppose that the CBSM intervention example summarizes a relatively controlled and large trial. Treatment is randomly assigned at the cluster-level, where 50 therapists are to administer the control condition and 50 therapists are to administer the experimental condition ($J = 100$). Participants are randomly assigned to a therapist, with 10 participants assigned to each therapist ($I = 1000$). A quality of life measure is taken repeatedly from each participant at four equally-spaced measurement occasions ($M = 4$) that are common across participants: immediately prior to the trial (i.e., baseline); 1-week into the trial; 2-weeks into the trial; and three weeks into the trial. The nesting structure can be conceptualized as repeated quality of life measures (i.e., Level-1) within participants (i.e., Level-2) within therapists (i.e., Level-3). The hypothesis is that the average linear change in quality of life will be greater for patients in the treatment condition compared to those in the control condition, *after accounting for the cluster ("therapist") effect*. This model can be written:

$$\begin{aligned}
 \text{quality of life}_{mij} &= \pi_{0ij} + \pi_{1ij}A_{mij} + e_{mij} \\
 \pi_{0ij} &= \beta_{00j} + r_{0ij} \\
 \pi_{1ij} &= \beta_{10j} + r_{1ij} \\
 \beta_{00j} &= \gamma_{000} + \gamma_{001}W_j + u_{00j} \\
 \beta_{10j} &= \gamma_{100} + \gamma_{101}W_j + u_{10j}
 \end{aligned} \tag{10}$$

where, A represents time (and is coded 0 at baseline, 1 at week-1, 2 at week-2, and 3 at week-3) and W is a dummy variable for treatment condition (0 = one-day educational intervention, 1 = CBSM intervention). A dataset was generated (i.e., dataset_3) consistent with this example. Equation 10 was fitted to dataset_3 (see Supplemental Digital Content 3

¹³The Monte Carlo approach in *Mplus* for sample size and power determination described previously in the Familias Unidas example could be extended to this example (20,61).

for input and key output).¹⁴ Interpretation of a few key results follows but a fuller set of interpretations and the relevant path model are both available in Supplemental Digital Content 2.

The fixed effects are typically emphasized in three-level MLRMs. In this example, the estimate of γ_{001} was not statically significant, $\hat{\gamma}_{001} = 1.437$, $p = .207$. This result provided evidence that the difference in expected quality of life at baseline for participants in the experimental condition as compared to the control condition was not greater than what we would expect by chance, which was consistent with the random assignment of participants to therapists. The estimate of γ_{100} was statically significant, $\hat{\gamma}_{100} = 1.206$, $p < .001$. This result provided evidence that the expected rate of weekly linear change in quality of life for participants in the control condition was greater than what we would expect by chance, which was consistent with the one-day educational seminar that this group received (i.e., the control group received a standard treatment). Perhaps most importantly, the estimate of γ_{101} was statically significant, $\hat{\gamma}_{101} = 0.843$, $p = .004$. This result provided evidence that the difference in the expected rate of weekly linear change in quality of life for participants in the experimental condition as compared to the control condition was greater than what we would expect by chance, which was consistent with the CBSM intervention that this group received (i.e., in addition to the standard treatment received by the control group).¹⁵

The random part of three-level MLRMs also is important. In this example, the variance of the level-1 error term, e_{mij} was statically significant, $\hat{\sigma}^2 = 46.823$, $p < .001$. This result provided evidence for residual within-participant heterogeneity and thus the possibility of adding time-varying covariates (e.g., urinary incontinence, erectile dysfunction, pain) to the model if theoretically defensible. The variance of each Level-2 error term, r_{0ij} and r_{1ij} was statically significant, $\tau_{\pi 00} = 125.026$, $p < .001$, and, $\tau_{\pi 11} = 4.960$, $p < .001$. This result provided support for allowing each participant to have her/his own change trajectory relative to other participants within the same therapist. The observed level of between-participant heterogeneity would support adding participant-level covariates (e.g., age, comorbid medical conditions, pre-treatment physical functioning) to the model if theoretically defensible. The variance of each Level-3 error term, u_{00j} and u_{10j} was statically significant, $\tau_{\beta 00} = 19.168$, $p < .001$ and $\tau_{\beta 11} = 0.531$, $p = .043$. This result provided support for allowing each therapist to have her/his own average participant trajectory relative to the other therapists – even after controlling for the effect of the treatment condition. The observed level of between-therapist heterogeneity would support adding therapist-level covariates to the model (e.g., gender, years of experience) if theoretically defensible.

It may have been ideal if evidence had been observed to support treating β_{00j} as fixed (i.e., dropping W_j and u_{00j}) and β_{10j} as non-randomly varying based on treatment condition (i.e., dropping u_{10j}). Treating β_{00j} as fixed, $\beta_{00j} = \gamma_{000}$, would have suggested that mean participant quality of life at baseline was relatively constant across therapists. Conceptually, such a result could have been interpreted as the result of successful random assignment of participants to therapists – giving each therapist a similar patient profile, with regard to mean quality of life, at baseline. Treating β_{10j} as non-randomly varying, $\beta_{10j} = \gamma_{100} + \gamma_{101} W_j$, would have suggested that mean participant rate of weekly linear change in quality of life was relatively constant across therapists after controlling for the effect of treatment condition. Conceptually, such a result could have been interpreted as the average rate of linear change in quality of life being independent of which therapist a participant was

¹⁴Mplus takes a multivariate (not univariate as in some dedicated MLRM software) approach to change modeling. Thus, the program considers a three-level MLRM to be a two-level model.

¹⁵Note that γ_{101} is a cross-level interaction. The regions of significance approach described in the Familias Unidas example could be extended to this example (53,54), including in a way that could identify at which time points the difference between groups is statistically significant.

assigned to, after adjusting for the effect of treatment condition. What may be theoretically ideal may not be observed in practice and the complex error structure of three-level MLRMs allow for flexible modeling of complex phenomenon often observed in cluster randomized trials with repeated measures.

Summary

Multilevel data, sometimes referred to as hierarchical data, are common in psychosomatic research. A common hierarchical structure has repeated observations nested within participants (e.g., Example 2: Stress Management) but cross-sectional data can also be viewed as multilevel (e.g., Example 1: Familias Unidas). When non-independence in the data is non-trivial and is attributable to at least one cluster variable (e.g., adolescents nested within schools; repeated measures nested within participants who are nested within therapists), imposing a model that ignores the non-independence is often problematic from both an empirical perspective (e.g., increased Type I error) and a theoretical perspective (e.g., inability to model relationships at, and across, different levels).

MLMs, particularly MLRMs, are beginning to appear with some frequency in psychosomatic research (see Supplemental Digital Content 1). Some core methodological issues within applications of MLMs that require careful attention include: communication of model specification, parameter interpretation, sample size and power, and missing data. Examples of how to communicate the model specified and how to interpret the model parameters were provided in Example 1, Example 2, and Supplemental Digital Content 1. Special sections were dedicated to both sample size and power in MLRM, and, to missing data in MLRM. Supplemental Digital Content 3 provides annotated input and key output from *Mplus* and SAS. Readers are referred to extant resources for excellent didactic treatments of longitudinal data from a mixed model (2) and/or latent growth curve model (96) perspective. The current manuscript, in combination with these existing statistical corner manuscripts (2,96), combine to provide readers and contributors to *Psychosomatic Medicine* with a coherent and reasonably broad set of resources for potentially interesting possibilities with regard to modeling non-independence of observations.

We note that up to this point we have focused much of our discussion on MLRMs with a continuous and normally distributed outcome (i.e., what is sometimes referred to as the hierarchical linear model) and where the Level-1 link function is the identity. MLRMs, as well as relevant software programs (e.g., PROC GLIMMIX), can model other types of outcomes (e.g., binary, ordinal, count) as well (i.e., what are sometimes referred to as hierarchical generalized linear models). For a full discussion of how MLRMs accommodate (e.g., Level-1 Sampling Model, Level-1 link function) these other types of outcomes readers are referred to some relevant resources (97–99). Finally, while MLRMs were emphasized in this manuscript, we note that MLMs have also been implemented within a more general latent variable framework that may be used with considerable frequency in future research (e.g., MLSMs; 100–102).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The Cognitive Behavioral Stress Management in Prostate Cancer study was supported by NCI grant # 1P50CA84944 awarded to Mike Antoni.

The Familias Unidas effectiveness study is supported by NIDA grant # 1R01DA025192-01A1 and NIAAA grant #3R01DA025192-01A1S1 awarded to Guillermo Prado.

Glossary

MLM(s)	multilevel model(s)
MLRM(s)	multilevel regression model(s)
MLSMs	multilevel structural equation model(s)
RAS	raw score
GMC	grand mean centering
CWC	centering within cluster
RSB	risky sexual behavior
PA	perceived peer abstinence
SF	adolescent functioning in school
GAP	gap between parental and adolescents Americanism
FAM	family functioning
ICC	intraclass correlation coefficient
ML	maximum likelihood
CBSM	cognitive-behavioral stress management
A	time
W	treatment condition

References

1. Sullivan LM, Dukes KA, Losina E. Tutorial in biostatistics: An introduction to hierarchical linear modelling. *Stat Med*. 1999; 18:855–88. [PubMed: 10327531]
2. Blackwell E, Mendes de Leon CF, Miller GE. Applying mixed regression models to the analysis of repeated-measures data in psychosomatic medicine. *Psychosom Med*. 2006; 68:870–78. [PubMed: 17079709]
3. Hox, JJ.; Roberts, JK. Multilevel analysis: Where we were and where we are. In: Hox, JJ.; Roberts, JK., editors. *Handbook of advanced multilevel analysis*. New York: Routledge; 2010. p. 3–11.
4. McCoach, BD. Hierarchical linear modeling. In: Hancock, GR.; Mueller, RO., editors. *Reviewer's guide to quantitative methods in the social sciences*. New York: Routledge; 2010. p. 123–40.
5. Robinson WS. Ecological correlations and the behavior of individuals. *Sociol Rev*. 1950; 15:351–57.
6. Lazarsfeld, PF.; Menzel, H. On the relation between individual and collective properties. In: Etzioni, A., editor. *A sociological reader on complex organizations*. New York: Holt, Rinehart & Winston; 1969. p. 499–516.
7. Cronbach LJ, Webb N. Between- and within-class effects in a reported aptitude-by-treatment interaction: Reanalysis of a study by G. L. Anderson *J of Educ Psychol*. 1975; 67:717–24.
8. Lindley DV, Smith AFM. Bayes estimates for the linear model. *J R Stat Soc Series B Stat Methodol*. 1972; 34:1–41.
9. Smith AFM. A general Bayesian linear model. *J R Stat Soc Series B Stat Methodol*. 1973; 35:61–75.
10. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc Series B Stat Methodol*. 1977; 39:1–8.

11. Dempster AP, Rubin DB, Tsutakawa RK. Estimation in covariance components models. *J Am Stat Assoc.* 1981; 76:341–53.
12. Goldstein H. Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika.* 1986; 73:43–56.
13. Longford N. A fast scoring algorithm for maximum likelihood estimation in unbalanced models with nested random effects. *Biometrika.* 1987; 74:817–27.
14. Goldstein, H. Multilevel models in education and social research. London: Griffin; 1987.
15. Bock, RD. Multilevel analysis of educational data. New York: Academic Press; 1989.
16. Bryk, AS.; Raudenbush, SW. Hierarchical linear models. Newbury Park, CA: Sage; 1992.
17. Goldstein, H. Bootstrapping in multilevel models. In: Hox, JJ.; Roberts, JK., editors. *Handbook of advanced multilevel analysis.* New York: Routledge; 2010. p. 163-172.
18. Neyman J, Scott EL. Consistent estimates based on partially consistent observations. *Econometrica.* 1948; 16:1–32.
19. Raudenbush, SW.; Bryk, AS. Hierarchical linear models. 2. Newbury Park, CA: Sage; 2002.
20. Muthén, LK.; Muthén, BO. *Mplus User's Guide.* 6. Los Angeles, CA: Muthén & Muthén;; 1998–2010.
21. SAS Institute Inc. SAS 9.2 (English) Help and Documentation. Cary, NC: SAS Institute Inc; 2002–2008.
22. Roberts, JK.; McLeod, P. Software options for multilevel models. In: O'Connell, AA.; McCoach, BD., editors. *Multilevel modeling of educational data.* Charlotte, NC: Information Age Publishing; 2008. p. 427-67.
23. Schwartz JE, Stone SA. Strategies for analyzing ecological momentary assessment data. *Health Psychol.* 1998; 17:6–16. [PubMed: 9459065]
24. Enders CK, Tofighi D. Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychol Methods.* 2007; 12:121–38. [PubMed: 17563168]
25. Algina, J.; Swaminathan, H. Centering in two-level nested designs. In: Hox, JJ.; Roberts, JK., editors. *Handbook of advanced multilevel analysis.* New York: Routledge; 2010. p. 285-312.
26. Myers ND, Brincks AM, Beauchamp MR. A tutorial on centering in cross-sectional two-level models. *Meas Phys Educ Exerc Sci.* 2010; 14:275–94.
27. Longford N. Contextual effects and group means. *Multilevel Modelling Newsletter.* 1989; 1:5, 11.
28. Longford N. To center or not to center. *Multilevel Modelling Newsletter.* 1989; 1:7, 11.
29. Plewis I. Comment on “centering” predictors in multilevel analysis. *Multilevel Modelling Newsletter.* 1989; 1:6, 11.
30. Raudenbush SW. “Centering” predictors in multilevel analysis: Choices and consequences. *Multilevel Modelling Newsletter.* 1989; 1:10–12.
31. Raudenbush SW. A response to Longford and Plewis. *Multilevel Modelling Newsletter.* 1989; 1:8–11.
32. Kreft IGG, de Leeuw J, Aiken LS. The effect of different forms of centering in hierarchical linear models. *Multivariate Behav Res.* 1995; 30:1–21.
33. Brumback BA, Dailey AB, Brumback LC, Livingston MD, He Z. Adjusting for confounding by cluster using generalized linear mixed models. *Stat Probab Lett.* 2010; 80:1650–1654.
34. Goetgeluk S, Vansteelandt S. Conditional generalized estimating equations for the analysis of clustered and longitudinal data. *Biometrics.* 2008; 64:772–780. [PubMed: 18047524]
35. Neuhaus JM, McCulloch CE. Separating between- and within-cluster covariate effects by using conditional and partitioning methods. *J R Stat Soc Series B Stat Methodol.* 2006; 68:859–872.
36. Centers for Disease Control and Prevention. HIV Surveillance Report 2008. 2010. p. 20<http://www.cdc.gov/hiv/topics/surveillance/resources/reports/>
37. Ramirez, RR.; de la Cruz, GP. The Hispanic population in the United States: March 2002 (Current Population Report P20-545). Washington, DC: U.S. Census Bureau; 2003.
38. Prejean J, Song R, Hernandez A, Ziebell R, Green T, Walker F, Lin LS, An Q, Mermin J, Lansky A, Hall I. Estimated HIV Incidence in the United States, 2006–2009. *PLoS ONE.* 2011; 6:e17502.10.1371/journal.pone.0017502 [PubMed: 21826193]

39. Prado GJ, Schwartz SJ, Maldonado-Molina M, Huang S, Pantin HM, Lopez B, Szapocznik J. Ecodevelopmental $\times$ intrapersonal risk: Substance use and sexual behavior in Hispanic adolescents. *Health Educ Behav*. 2009; 36:45–51. [PubMed: 18326053]
40. Prado GJ, Huang S, Maldonado-Molina M, Bandiera F, Schwartz SJ, de la Vega P, Brown CG, Pantin H. An empirical test of Ecodevelopmental theory in predicting HIV risk behaviors among Hispanic youth. *Health Educ Behav*. 2010; 37:97–114. [PubMed: 20130302]
41. Pantin HM, Schwartz SJ, Sullivan S, Prado GJ, Szapocznik J. Ecodevelopmental HIV prevention programs for Hispanic adolescents. *Am J Orthopsychiatry*. 2004; 74:545–58. [PubMed: 15554814]
42. Prado G, Pantin HM. Reducing substance use and HIV health disparities among Hispanic youth in the U.S.A: The familias unidas program of research. *Psychosocial Interventions*. 2011; 20:63–73.
43. Moerbeek, M.; Teerenstra, S. Optimal design in multilevel experiments. In: Hox, JJ.; Roberts, JK., editors. *Handbook of advanced multilevel analysis*. New York: Routledge; 2010. p. 257-81.
44. Curran PJ, Bauer DJ. Building path diagrams for multilevel models. *Psychol Methods*. 2007; 12:283–97. [PubMed: 17784795]
45. Julian MW. The consequences of ignoring multilevel data structures in nonhierarchical covariance modeling. *Struct Equ Modeling*. 2001; 8:325–52.
46. Muthén BO. Multilevel covariance structure analysis. *Sociol Methods Res*. 1994; 22:376–98.
47. Satorra, A.; Bentler, PM. Corrections to test statistics and standard errors on covariance structure analysis. In: von Eye, A.; Clogg, CC., editors. *Latent variable analysis*. Thousand Oaks, CA: Sage; 1994. p. 399-419.
48. Stapleton, LM. Using multilevel structural equation modeling techniques with complex sample data. In: Hancock, GR.; Mueller, RO., editors. *Structural equation modeling: A second course*. Greenwich, CT: Information Age Publishing; 2006. p. 345-83.
49. Lüdtke O, Marsh HW, Robitzsch A, Trautwein U, Asparouhov T, Muthén B. The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies. *Psychol Methods*. 2008; 13:203–29. [PubMed: 18778152]
50. Marsh HW, Lüdtke O, Robitzsch A, Trautwein U, Asparouhov T, Muthén B, Nagengast B. Doubly-latent models of school contextual effects: Integrating multilevel and structural equation approaches to control measurement and sampling errors. *Multivariate Behav Res*. 2009; 44:764–802.
51. Raudenbush SW, Willms J. The estimation of school effects. *J Educ Behav Stat*. 1995; 20:307–335.
52. Snijders, T.; Bosker, R. *Multilevel analysis*. Thousand Oaks, CA: Sage; 1999.
53. Bauer DJ, Curran PJ. Probing Interactions in Fixed and Multilevel Regression: Inferential and Graphical Techniques. *Multivariate Behav Res*. 2005; 40:373–400.
54. Preacher KJ, Curran PJ, Bauer DJ. Computational tools for probing interaction effects in multiple linear regression, multilevel modeling, and latent curve analysis. *J Educ Behav Stat*. 2006; 31:437–48.
55. Roberts, JK.; Monaco, JP.; Stovall, H.; Foster, V. Explained variance in multilevel models. In: Hox, JJ.; Roberts, JK., editors. *Handbook of advanced multilevel analysis*. New York: Routledge; 2010. p. 219-30.
56. Hox, J. *Multilevel analysis: Techniques and applications*. 2. Mahwah, NJ: Erlbaum; 2010.
57. Maas CJM, Hox JJ. Robustness issues in multilevel regression analysis. *Stat Neerl*. 2004; 58:127–37.
58. Maas CJM, Hox JJ. The influence of violation of assumptions on multilevel parameter estimates and their standard errors. *Comput Stat Data Anal*. 2004; 46:427–40.
59. Van der Leeden, R.; Meijer, E.; Busing, F. Resampling multilevel models. In: de Leuw, J.; Meijer, E., editors. *Handbook of multilevel analysis*. New York: Springer; 2008. p. 401-33.
60. Browne WJ, Draper D. Implementation and performance issues in the Bayesian and likelihood fitting of multilevel models. *Comput Stat*. 2000; 15:391–420.
61. Muthén LK, Muthén BO. How to use a Monte Carlo study to decide on sample size and determine power. *Struct Equ Modeling*. 2002; 9:599–620.
62. Gentle, JE. *Random number generation and Monte Carlo methods*. 2. New York: Springer; 2003.

63. Myers ND, Ahn S, Ying J. Sample size and power estimates for a confirmatory factor analytic model in exercise and sport: A Monte Carlo approach. *Res Q Exer Sport*. 2011; 82:412–23.
64. Cohen, J. *Statistical power for the behavioral sciences*. 2. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
65. Snijders TAB, Bosker R. Standard errors and sample sizes for two-level research. *J Educ Behav Stat*. 1993; 18:237–59.
66. Raudenbush SW, Liu X. Statistical power and optimal design for multisite randomized trials. *Psychol Methods*. 2000; 5:199–213. [PubMed: 10937329]
67. Moerbeek M, van Breukelen GJP, Berger M. Design issues for experiments in multilevel populations. *J Educ Behav Stat*. 2000; 25:271–84.
68. Optimal Design. Version 2.01. [Computer Software]. New York, NY: William T. Grant Foundation; 2011.
69. Rubin DB. Inference and missing data. *Biometrika*. 1976; 63:581–90.
70. Little RJA. A test of missing completely at random for multivariate data with missing values. *J Am Stat Assoc*. 1988; 83:1198–1202.
71. Schafer JL, Graham JW. Missing data: Our view of the state of the art. *Psychol Methods*. 2002; 7:147–177. [PubMed: 12090408]
72. Enders CK. Missing data at random models for latent growth curve analyses. *Psychol Methods*. 2011; 16:1–16. [PubMed: 21381816]
73. Muthén BO, Asparouhov T, Hunter AM, Leuchter AF. Growth modeling with nonignorable dropout: Alternative analyses of the STAR*D antidepressant trial. *Psychol Methods*. 2011; 16:17–33. [PubMed: 21381817]
74. Enders CK. A primer on the use of modern missing-data methods in psychosomatic medicine research. *Psychosom Med*. 2006; 68(3):427–436. [PubMed: 16738075]
75. Longford, NT. Missing Data. In: de Leuw, J.; Meijer, E., editors. *Handbook of multilevel analysis*. New York: Springer; 2008. p. 377–99.
76. Van Buuren, S. Multiple imputation of multilevel data. In: Hox, JJ.; Roberts, JK., editors. *Handbook of advanced multilevel analysis*. New York: Routledge; 2010. p. 173–96.
77. Little RJA. Regression with missing X's: A review. *J Am Stat Assoc*. 1992; 87:1227–37.
78. Little, RJA.; Rubin, DB. *Statistical analysis with missing data*. 2. New York: Wiley; 2002.
79. Little RJA, Yau L. Intent-to-treat analysis for longitudinal studies with drop-outs. *Biometrics*. 1996; 52:1324–33. [PubMed: 8962456]
80. Schafer JL, Schenker N. Inference with imputed conditional means. *J Am Stat Assoc*. 2000; 95:144–154.
81. Beunckens C, Molenberghs G, Thijs H, Verbeke G. Incomplete hierarchical data. *Stat Methods Med Res*. 2007; 16:457–92. [PubMed: 17656453]
82. Collins LM, Schafer JL, Kam CH. A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychol Methods*. 2001; 6:330–51. [PubMed: 11778676]
83. Rubin, DB. *Multiple imputation for nonresponse in surveys*. New York: Wiley; 1987.
84. Schafer, JL. *Analysis of incomplete multivariate data*. London: Chapman & Hall; 1997.
85. Asparouhov, T.; Muthén, B. *Multiple imputation with Mplus (Version 2)*. 2010. Retrieved from <http://www.statmodel.com/download/Imputations7.pdf>
86. Penedo FJ, Molton I, Dahn JR, Shen BJ, Kinsinger D, Traeger L, Siegel S, Schneiderman N, Antoni M. A randomized clinical trial of group-based cognitive-behavioral stress management in localized prostate cancer: development of stress management skills improves quality of life and benefit finding. *Ann Behav Med*. 2006; 31:261–70. [PubMed: 16700640]
87. American Cancer Society. *Cancer facts & figures*. Atlanta, GA: American Cancer Society; 2011.
88. Hoffman RM, Gilliland FD, Penson DF, Stone SN, Hunt WC, Potosky AL. Cross-sectional and longitudinal comparisons of health-related quality of life between patients with prostate carcinoma and matched controls. *Cancer*. 2004; 101:2011–19. [PubMed: 15452835]
89. Miller DC, Sanda MG, Dunn RL, Montie JE, Pimentel H, Sandler HM, McLaughlin WP, Wei JT. Long-term outcomes among localized prostate cancer survivors: Health-related quality-of-life

- changes after radical prostatectomy, external radiation, and brachytherapy. *J Clin Oncol*. 2005; 23:2772–80. [PubMed: 15837992]
90. Beck AM, Robinson JW, Carlson LE. Sexual intimacy in heterosexual couples after prostate cancer treatment: What we know and what we need to learn. *Urol Oncol*. 2009; 27:137–43. [PubMed: 18367118]
 91. Chambers SK, Schover L, Halford K, Cluttons S, Ferguson M, Gordon L, Gardiner RA, Occhipinti S, Dunn J. ProsCan for Couples: randomised controlled trial of a couples-based sexuality intervention for men with localised prostate cancer who receive radical prostatectomy. *BMC Cancer*. 2008; 8:226–32. [PubMed: 18687149]
 92. Helgason AR, Adolfsson J, Dickman P, Fredrikson M, Arver S, Steineck G. Waning sexual function the most important disease-specific distress for patients with prostate cancer. *Br J Cancer*. 1996; 73:1417–21. [PubMed: 8645589]
 93. Penson DF, Feng Z, Kuniyuki A, McClerran D, Albertsen PC, Deapen D, et al. General quality of life 2 years following treatment for prostate cancer: what influences outcomes? Results from the prostate cancer outcomes study. *J Clin Oncol*. 2003; 21:1147–54. [PubMed: 12637483]
 94. Antoni M, Baggett L, Ironson G, LaPerriere A, August S, Klimas N, Schneiderman N, Fletcher MA. Cognitive-behavioral stress management buffers distress responses and immunologic changes following notification of HIV-1 seropositivity. *J Consult Clin Psychol*. 1991; 59:906–15. [PubMed: 1774375]
 95. Penedo, FJ.; Antoni, MH.; Schneiderman, N. Cognitive-behavioral stress management for prostate cancer recovery. Oxford University Press, Inc; New York: 2008.
 96. Llabre MM, Spitzer S, Siegel S, Saab PG, Schneiderman N. Applying Latent Growth Curve Modeling to the Investigation of Individual Differences in Cardiovascular Recovery From Stress. *Psychosom Med*. 2004; 66:29–41. [PubMed: 14747635]
 97. Agresti, A. An introduction to categorical data analysis. John Wiley; New York: 1996.
 98. Hedeker D, Gibbons R. MIXOR: A computer program for mixed-effects ordinal probit and logistic regression analysis. *Comput Methods Programs Biomed*. 1996; 49:157–176. [PubMed: 8735023]
 99. Schall R. Estimation in generalized linear models with random effects. *Biometrika*. 1991; 40:719–727.
 100. Muthén B. Latent variable modeling in heterogeneous populations. *Psychometrika*. 1989; 54:557–85.
 101. Muthén B. Multilevel covariance structure analysis. *Sociol Methods Res*. 1994; 22:376–98.
 102. Muthén, B.; Asparouhov, T. Beyond multilevel regression modeling: Multilevel analysis in a general latent variable framework. In: Hox, JJ.; Roberts, JK., editors. *Handbook of advanced multilevel analysis*. New York: Routledge; 2010. p. 15-40.

Figure 1.
Simple slopes showing the effect of adolescent school functioning on risky sexual behavior for a few values of mean family functioning.

Figure 2. Regions of significance of mean family functioning (mean family functioning -1.37 , or, mean family functioning -0.17) where the effect of adolescent functioning in school on risky sexual behavior was statistically significant.

Table 1
Familias Unidas: Parameter Estimates, (Standard Errors), Observed Power, and Model-Data Fit

Parameter	Sub-models fitted to dataset 1					Power	
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	J = 25 J = 100
Fixed Effects							
γ_{00}	0.040 (0.033)	0.039 (0.031)	0.039 (0.032)	0.030 (0.036)	0.040 (0.033)	0.039 (0.031)	-----
γ_{01}		0.020 (0.031)	0.023 (0.032)	0.016 (0.034)		0.025 (0.031)	-----
γ_{02}		0.030 (0.042)	0.037 (0.044)	0.025 (0.046)		0.037 (0.041)	-----
γ_{03}		0.036* (0.014)	0.035* (0.014)	0.036* (0.015)		0.037* (0.014)	33.4% 79.7%
γ_{04}		-0.007 (0.047)	-0.010 (0.047)	-0.005 (0.048)		-0.009 (0.048)	-----
γ_{10}			-0.017*** (0.003)	-0.016*** (0.003)	-0.015*** (0.003)	-0.015*** (0.003)	80.8% 100.0%
γ_{20}			-0.009** (0.003)	-0.008* (0.003)	-0.008* (0.003)	-0.007* (0.003)	-----
γ_{24}				-0.011*** (0.002)		-0.009*** (0.002)	97.5% 100.0%
γ_{30}			0.011*** (0.002)	0.010*** (0.002)	0.010*** (0.002)	0.010*** (0.002)	95.1% 100.0%
γ_{40}			-0.034* (0.016)	-0.033* (0.016)	-0.026 (0.014)	-0.027* (0.013)	-----
γ_{43}						0.017*** (0.004)	40.6% 90.4%
Variance Components							
σ^2	1.058*** (0.053)	1.058*** (0.053)	1.012*** (0.048)	0.961*** (0.047)	0.747*** (0.019)	0.711*** (0.019)	100.0% 100.0%
τ_{00}	0.072*** (0.013)	0.061*** (0.011)	0.065*** (0.012)	0.070*** (0.013)	0.083*** (0.013)	0.073*** (0.011)	53.6% 100.0%
τ_{44}					0.018*** (0.003)	0.015*** (0.002)	96.9% 100.0%
τ_{04}					0.001 (0.005)	-0.005 (0.004)	-----
Model-Data Fit							
Deviance	8412	8401	8280	8185	7679	7567	
# par	3	7	11	12	9	15	
AIC	8418	8415	8302	8209	7697	7597	
BIC	8436	8457	8368	8281	7751	7686	

Note.

\$watermark-text

\$watermark-text

\$watermark-text

* $p < .05$;
** $p < .01$;
*** $p < .001$;
par = number of parameters estimated.

Table 2

Descriptive Statistics at Level-1 and Level-2

Level-1	
	Covariance/Correlation Matrix
Risky sexual behavior	1.06 -.06 .02 .15 -.12
Peer abstinence	-.070 135.26 -.83 .06 -.31
Adolescent functioning in school	0.20 -.121.68 159.28 -.05 .24
Parental and adolescent Americanism gap	1.74 8.03 -6.92 131.77 -.26
Family functioning	-.045 -13.00 10.99 -10.68 13.31
Level-2	
	Covariance/Correlation Matrix
Risky sexual behavior	0.07 -.05 .02 .37 -.16
Peer abstinence	-.007 21.23 -.95 -.20 -.63
Adolescent functioning in school	0.02 -13.84 9.91 .05 .67
Parental and adolescent Americanism gap	0.32 -2.87 0.44 9.77 -.33
Family functioning	-.005 -3.29 2.37 -1.17 1.27
Means	0.04 0.04 -0.02 0.04 -0.01

Note. Correlation values are above the diagonal, variance values are on the diagonal, and covariance values are below the diagonal within each Covariance/Correlation Matrix table.