Modeling Achievement Trajectories When Attrition Is Informative

Betsy J. Feldman and Sophia Rabe-Hesketh

JOURNAL OF EDUCATIONAL AND BEHAVIORAL STATISTICS published online 12 October 2012
DOI: 10.3102/1076998612458701

JOURNAL OF EDUCATIONAL AND BEHAVIORAL STATISTICS OnlineFirst, published on October 12, 2012 as doi:10.3102/1076998612458701

Journal of Educational and Behavioral Statistics
2012, Vol. 00, No. 0, pp. 1–34
DOI: 10.3102/1076998612458701

Modeling Achievement Trajectories When Attrition Is Informative

Betsy J. Feldman
University of Washington

Sophia Rabe-Hesketh
University of California, Berkeley, and Institute of Education, University of London

In longitudinal education studies, assuming that dropout and missing data occur completely at random is often unrealistic. When the probability of dropout depends on covariates and observed responses (called missing at random *[MAR]), or on values of responses that are missing (called* informative *or* not missing at random *[NMAR]), inappropriate analysis can cause biased estimates. NMAR requires explicit modeling of the missingness process together with the response variable. In this article, we review assumptions needed for consistent estimation of hierarchical linear growth models using common missing-data approaches. We also suggest a joint model for the longitudinal data and missingness process to handle the situation where data are NMAR. The different approaches are applied to the NELS:88 study, as well as simulated data. Results from the NELS:88 analyses were similar between the MAR and NMAR models. However, use of listwise deletion and mean imputation resulted in significant bias, both for the NELS:88 study and simulated data. Simulation results showed that incorrectly assuming MAR leads to greater bias for the growth-factor variance–covariance matrix than for the growth factor means, the former being severe with as little as 10% missing data and the latter with 40% missing data when departure from MAR is strong.*

Keywords: *longitudinal, NELS:88, not missing at random, nonignorable missing data, simulation*

Educators and education researchers are increasingly using longitudinal data to assess the impact of interventions and to track children's progress through school. Rivkin (2007) argues that, to investigate students' progress, it is necessary to look at growth and change at the student level, and, when possible, associate that with teacher performance, in addition to school and district characteristics. The U.S. government programs, No Child Left Behind and Race to the Top (U.S. Department of Education, 2009), are also encouraging states to collect and use longitudinal data to investigate teacher and school performance by looking at individual student growth in achievement instead of considering only school- or district-level mean test scores (Ladd & Lauen, 2009), and the Institute of Education Sciences (U.S. Department of Education) offers grants to help states set up such longitudinal data collection systems. At this time, 42 states have received grants to set up student-level longitudinal data collection systems for their schools (http://nces.ed.gov/Programs/SLDS/summary.asp). Moreover, the American Recovery and Reinvestment Act also required that states institute educational longitudinal data collection systems to receive their full funding (McNeil, 2009).

When longitudinal data are used, missed assessments and attrition are inevitable, and inappropriate handling of missing data can lead to biased results (Enders, 2001; Peugh & Enders, 2004). Our goal, in this article, is to clarify the assumptions invoked by the most commonly used methods for dealing with missing data in longitudinal research, as well as the risks incurred when those assumptions are incorrect. Specifically, assumptions must be made regarding the reasons for data being missing, or the *missingness mechanism*. The missingness mechanism is formally expressed in terms of how the probability of a response being missing depends on other observed and unobserved variables, as well as on the response itself. We describe different types of missingness mechanisms and describe the assumptions invoked when using traditional approaches, such as listwise deletion (also referred to as complete case analysis), mean imputation (or mean substitution), and standard maximum likelihood estimation (MLE) with all available data.

Although listwise deletion and mean substitution have been shown to be problematic (e.g., Schafer & Graham, 2002), such methods are still pervasive in education research. Peugh and Enders (2004) reviewed methods used for handling missing data in 23 education journals in 1999 and 2003. They found that only 34% and 74% of studies identified as having missing data acknowledged the problem in 1999 and 2003, respectively, and that the vast majority of those used listwise or pairwise deletion. Jeličić, Phelps, and Lerner (2009) found similar results in their review of 100 longitudinal studies published between 2000 and 2006 in *Child Development*, *Developmental Psychology*, or *Journal of Research on Adolescence*. Recent examples of the use of listwise deletion in education journals include Marcotte, Bailey, Borkoski, and Kienzl (2005), Cai, Wang, Moyer, Wang, and Nie (2011), and Kim, Petscher, Schatschneider, and Foorman (2010). A similar approach to listwise deletion is dropping individuals with fewer than two or three time points (for an example, see Bouffard, Vezeau, Roy, & Lengelé, 2011). Mean imputation appears to be used less often (e.g., see Corriveau et al., 2009; McNiece & Jolliffe, 1998; Metzger, Dawes, Mermelstein, & Wakschlag, 2011). Some authors impute conditional means, taking into account a small number of covariates (e.g., Borman et al., 2007; D'Agostino, 2000; McNiece, Bidgood, & Soan, 2004). Standard MLE with all available data is one of the recommended modern approaches for handling missing data (for examples using MLE, see Grimm, Steele, Mashburn, Burchinal, & Pianta, 2010; Kieffer, 2011; Montague, Enders, & Cavendish, 2011).

In addition to these traditional approaches, we describe, in detail, the random-coefficient-dependent model for nonignorable missingness—often the most appropriate model for achievement trajectories with missing data due to dropout—and we provide code for its estimation in Mplus (for more types of nonignorable missing-data models in Mplus, see Enders, 2011; Muthén, Asparouhov, Hunter, & Leuchter, 2011). To demonstrate and compare different approaches to dealing with missing data, we use the NELS:88 longitudinal high school testing data. Finally, we carry out a comprehensive simulation study to investigate the performance of different approaches in terms of bias of parameter estimates and standard errors when missing-data assumptions are violated.

Missingness Mechanisms and Modeling Techniques

Models for Growth

Before discussing missing data mechanisms, we begin with a brief description of a model for growth. Typically, heterogeneity, or individual differences, in developmental and educational trajectories is modeled with hierarchical linear models (HLM), also referred to as mixed, mixed effects, random effects, random coefficients, multilevel, or growth curve models (e.g., Bollen & Curran, 2006; Rabe-Hesketh & Skrondal, 2008; Raudenbush & Bryk, 2002; Singer & Willett, 2003). In HLM, repeated measurements are expressed as a function of time and individual differences. Individual differences in the outcome variable when time equals zero, and in rate of change, are modeled by permitting the intercept and slope/slopes of time, respectively, to vary randomly across individuals. The intercept and slope/slopes are, therefore, referred to as random coefficients, random effects, or (latent) growth factors, and they are responsible for the within-person covariances between measurements at different times. The intercept and slope/slopes are, in turn, expressed as linear functions of the individual-specific covariates and residuals. Time-varying covariates may also be used in the models. Generally, growth models are estimated by maximum likelihood or restricted maximum likelihood, which allows the timing and number of observations to differ between individuals.

Here, we consider a multiple-group growth curve model where the response y_{git} for individual i in group g at occasion t is modeled as a linear function of the random coefficients: the intercept, π_{0gi} (the expected response for individual i when time = 0) and, usually, one or more slope coefficients, π_{sgi} ($s = 1, 2, \ldots, S$), which are multiplied by powers of time, a_{git} (e.g., $\pi_{1gi}a_{git}$, $\pi_{2gi}a_{git}{}^2$, $\pi_{3gi}a_{git}{}^3$, ..., $\pi_{Sgi}a_{git}{}^S$). The slope coefficients describe systematic linear or curvilinear change over time in individual i's responses. The times of measurement, a_{git}, are typically based on age or time since enrollment in a program, school, or study; and are usually centered by subtracting a relevant age or time point. Often this is the time of, or age at, the first measurement occasion, making the intercept, π_{0gi}, individual i's expected response at the first measurement occasion, or *initial status*. The times of measurement may be the same for all individuals, or may be unique for each individual (e.g., current age).

We express this multiple-group linear model as having two levels: Level 1 characterizes the time-specific individual responses and Level 2 characterizes the between-person variability in growth trajectories.

Level 1:

$$y_{git} = \pi_{0gi} + \pi_{1gi}a_{git} + \pi_{2gi}z_{git} + \varepsilon_{git}, \tag{1a}$$

Level 2:

$$\begin{aligned} \pi_{0gi} &= \beta_{00g} + \beta_{01g}x_{gi} + u_{0gi}, \\ \pi_{1gi} &= \beta_{10g} + \beta_{11g}x_{gi} + u_{1gi}, \\ \pi_{2gi} &= \beta_{20g} + \beta_{21g}x_{gi} + u_{2gi}, \end{aligned} \tag{1b}$$

where z_{git} is a time-varying covariate for individual i in group g at time t, and x_{gi} is a time-invariant covariate for individual i. It is assumed that $\varepsilon_{git} \sim N(0,\sigma_g^2)$; and that $\mathbf{u}_{gi} \sim N(\mathbf{0}, \mathbf{\Psi}_g)$, where $\mathbf{u}_{gi}$ is a vector of intercept and slope residuals for individual i in group g. It is furthermore assumed that ε_{git} are independent across individuals and time, $\mathbf{u}_{gi}$ are independent across individuals, and ε_{git} and $\mathbf{u}_{gi}$ are independent for all i, g, and t.

Substituting the Level 2 equations into the Level 1 model gives the reduced-form equation:

$$\begin{aligned} y_{git} = {} & \beta_{00g} + \beta_{01g}x_{gi} + \beta_{10g}a_{git} + \beta_{11g}x_{gi}a_{git} + \beta_{20g}z_{git} + \beta_{21g}x_{gi}z_{git} \\ & + u_{0gi} + u_{1gi}a_{git} + u_{2gi}z_{git} + \varepsilon_{git}. \end{aligned} \tag{2}$$

Missing Data Mechanisms and Approaches for Dealing With Missing Data

To discuss missing-data mechanisms and approaches to modeling them, we first introduce some terms. The complete data (if all responses were observed) for individual i are contained in a vector, $\mathbf{y}_{gi}$, which is comprised of two subvectors: the observed data, $\mathbf{y}_{gi}^{\text{obs}}$ and the unobserved data, $\mathbf{y}_{gi}^{\text{miss}}$: $\mathbf{y}_{gi} = \{\mathbf{y}_{gi}^{\text{obs}}, \mathbf{y}_{gi}^{\text{miss}}\}$. Little (1995) and Little and Rubin (2002) describe three broad classes of missing-data processes, or mechanisms. Missing data are considered to be *missing completely at random* (MCAR) when the missingness does not depend on any other observed or unobserved variables in the model. Let y_{git} be an observation for individual i in group g at time t, and d_{git} be an indicator that equals 1 if y_{git} is missing at time t and 0 otherwise. Data are MCAR if:

$$pr(d_{git} = 1|\mathbf{X}_{gi}, \mathbf{y}_{gi}, \mathbf{u}_{gi}) = pr(d_{git} = 1), \tag{3}$$

where $\mathbf{X}_{gi}$ is a matrix of time-varying (z_{gi}) and time-invariant (x_{gi}) covariate values for individual i in group g included in the model for y_{git}. Although rare, true MCAR can occur. For example, if data are missing by design (e.g., only a randomly selected subsample is followed past a certain point in time), it may be assumed to be MCAR. Usually, however, the reason for missing data is unknown and missing data are unlikely to be MCAR. In educational research, for example, individuals may be missing observations because of frequent illness or absence, or school suspension—circumstances that are likely to be related to their test scores ($\boldsymbol{y}_{gi}$).

Mean imputation (imputing the mean value of the observed outcomes at each time point) is an approach that gives consistent estimates of regression coefficients when the data are MCAR, there are no covariates other than time (or functions of time), and the time points are the same across individuals. However, even under these circumstances, estimators of variances, covariances, and conventional standard errors will be inconsistent when mean imputation is used. This occurs for several reasons: First, the imputed data are constant at a given time point, which results in downward bias for the residual variance and consequently for the standard errors of regression coefficients. Second, if the model includes covariates other than time, estimates of regression coefficients will be biased toward zero because the imputed responses at a given time point are the same across covariate values, a problem partly overcome by taking some covariates into account using conditional means or regression imputation. Finally, treating imputed values as real leads to underestimated standard errors because uncertainty regarding the imputed responses is ignored, a problem that can be addressed by using multiple imputation (see, e.g., Enders, 2010; Schafer, 1999).

Covariate-dependent missingness describes a condition in which the missingness can depend on covariates that are included in the model for y_{git}, but does not depend on the outcomes, either missing or observed, after conditioning on the covariates:

$$pr(d_{git} = 1|\mathbf{X}_{gi}, \mathbf{y}_{gi}, \mathbf{u}_{gi}) = pr(d_{git} = 1|\mathbf{X}_{gi}). \tag{4}$$

When missingness depends only on covariates that are in the model, it is not necessary to explicitly model the residual covariance structure of the $\mathbf{y}_{gi}$. Thus, less computationally intensive estimators may be used. An example might be ordinary least squares (OLS) using all available data, with appropriate adjustments for the standard errors, which is known as pooled OLS in econometrics. Pooled OLS is equivalent to generalized estimating equations (GEE; Liang & Zeger, 1986) with continuous responses and an independent working correlation structure. Consistent estimates are also obtained using complete case analysis where the data on individuals with any missing data are discarded (listwise deletion). However, complete case analysis is inefficient (has larger standard errors and lower power) because it uses only a subset of the available data.

Missing data are described as *missing at random* (MAR) when missingness depends only on the observed responses, $\mathbf{y}_{gi}^{\text{obs}}$ and covariates (thus, covariate-dependent missingness can be viewed as a special case of MAR):

$$pr(d_{git} = 1|\mathbf{X}_{gi}, \mathbf{y}_{gi}, \mathbf{u}_{gi}) = pr(d_{git} = 1|\mathbf{X}_{gi}, \mathbf{y}_{gi}^{\text{obs}}). \tag{5}$$

If the missing data are MAR, estimating the correct longitudinal model with maximum likelihood, and utilizing all available data, gives consistent estimates of population values. This approach is sometimes called *full information* maximum likelihood (FIML). Most software that estimates growth models, such as multilevel modeling or structural equation modeling software, utilizes MLE and can handle MAR missingness. When this approach is used, there is no need to impute data unless covariates are missing while the response variable is observed, a situation not considered in this article. It is important to remember that the MAR assumption is more likely to hold if the model includes all the covariates that may be related to missingness (Collins, Schaefer, & Kam, 2001).

The two types of missingness mechanisms, MCAR and MAR, are referred to as *ignorable* or *noninformative*. That is, if MCAR or MAR holds, then the likelihood function can be factorized into the likelihood for the longitudinal data multiplied by the likelihood for the missingness process. This allows consistent estimates when the missing-data process is ignored (Little & Rubin, 2002). Note, however, that of the ignorable missing-data mechanisms, the only one that can be tested is MCAR (Little, 1995). Whether or not MAR holds is therefore an assumption that cannot be assessed empirically.

When the probability that data are missing depends on either (1) the missing values themselves (the values of the missing outcomes had they been observed) or (2) individual values of a latent variable such as a random coefficient that appears in the growth model, the missingness is not ignorable and the missing data are described as *not missing at random* (NMAR; Little & Rubin, 2002). Little (1995) referred to the first case as *outcome-dependent* missingness:

$$pr(d_{git} = 1|\mathbf{X}_{gi}, \mathbf{y}_{gi}, \mathbf{u}_{gi}) = pr(d_{git} = 1|\mathbf{X}_{gi}, \mathbf{y}_{gi}^{\text{obs}}, \mathbf{y}_{gi}^{\text{miss}}). \tag{6}$$

Outcome-dependent missingness might result if a student drops out because of an abrupt drop in achievement, or a patient suddenly becomes too ill to continue an intervention study.

The second case is *random-coefficient-dependent* missingness (Little, 1995; Wu & Carroll, 1988):

$$pr(d_{git} = 1|\mathbf{X}_{gi}, \mathbf{y}_{gi}, \mathbf{u}_{gi}) = pr(d_{git} = 1|\mathbf{X}_{gi}, \mathbf{y}_{gi}^{\text{obs}}, \mathbf{u}_{gi}). \tag{7}$$

Random-coefficient-dependent missingness would result if individual students' probability of dropping out was predicted by their achievement trajectories, or intercept and slope/slopes. When missingness is NMAR, the mechanism of missingness must be explicitly modeled, simultaneously with the growth, to get consistent growth estimates.

There are many ways of modeling nonignorable missingness, but all fall into two basic approaches. In a selection model, the missingness process is modeled directly by assuming that the missing data either depend on the missing outcomes (outcome-dependent missingness) or on the random effects or latent variables (random-coefficient-dependent missingness). In contrast, in a pattern mixture model, the population is viewed as a mixture of subpopulations characterized by different missing data patterns. Each subpopulation has a different set of parameter values for the growth process. These parameters can be estimated separately in each subpopulation and then combined using a weighted mean (Hedeker & Gibbons, 1997). Both modeling approaches make unverifiable assumptions regarding the nature of the missingness mechanism and there is no empirical way of selecting the most appropriate model (Molenberghs, Beunckens, Sotto, & Kenward, 2008), so model choice must be based on theoretical considerations, and sensitivity analyses should be used to assess the impact of different assumptions on the substantive conclusions.

Selection models that assume outcome-dependent missingness have been proposed by Hausman and Wise (1979) and Diggle and Kenward (1994). Dropout is modeled as depending on the current (missing) response as well as on lagged responses. The missing response variable is then integrated out of the likelihood. These models are known as outcome-dependent selection models (Little, 1995) and have recently been applied to education data (Xu & Blozis, 2011) and psychiatric clinical trial data (Enders, 2011).

A second type of selection model is the focus of this article because it may be especially useful when missingness results from students dropping out of school. This model, shown in Figure 1, is known as a *shared-parameter* or *random-effect-dependent* missing data model (Little, 1995; Wu & Carroll, 1988). The random-effect-dependent missing-data model assumes that one or more latent variables (e.g., intercept and slope) are responsible for both growth and dropout processes, and that, conditional on those latent variables (and covariates), the two processes are independent. Similar dual process models were first introduced by Heckman (1979) to deal with sample selection problems in cross-sectional data. In the approach we use, dropout is modeled as a discrete-time survival process in which the conditional probability that dropout for individual i in group g occurs at time t, given that it did not happen before time t, is referred to as the discrete-time hazard (h_{git}):

$$h_{git} = pr(T_{gi} = t \mid T_{gi} \geq t, \pi_{0gi}, \pi_{1gi}, x_{gi}), \tag{8}$$

where T_{gi} is the occasion when person i drops out.

The log-odds, or *logit*, of h_{git} is modeled as a function of individual i's intercept π_{0gi} and slope of time π_{1gi} (and possibly slopes of nonlinear functions of time) in the growth curve model, a covariate x_i (multiple covariates may be used)

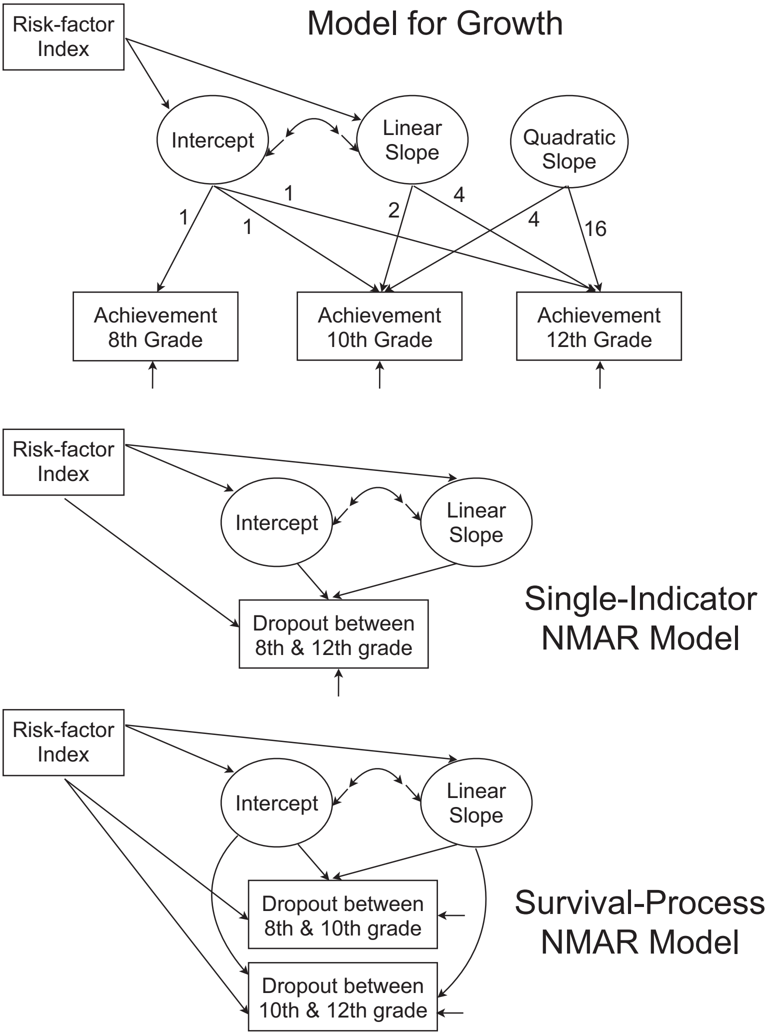

FIGURE 1. *Path diagrams of growth model and two models used in this article for analyzing data with nonignorable missingness (i.e., missing data that are NMAR). The quadratic slope was modeled without variance and is not used in the missing-data models.*

with regression coefficient λ_{hg}, and occasion-specific thresholds τ_{hgt} (or intercepts $-\tau_{hgt}$):

$$\text{logit}(h_{git}) = -\tau_{hgt} + \lambda_{hg1}\pi_{0gi} + \lambda_{hg2}\pi_{1gi} + \lambda_{hg3}x_{gi}. \tag{9a}$$

This model, referred to here as a survival-process MNAR model (see Figure 1), is similar to that used by Wu and Carroll (1988) except that they specified a cumulative probit model for the ordinal drop-out time and did not allow missingness to depend on covariates, given π_{0gi} and π_{1gi}. A similar model was also considered by Enders (2011).

A second, similar, approach is to model the probability of *ever* dropping out. We refer to this as a single-indicator MNAR model (Figure 1). Below, d_{gi} equals 1 if individual i dropped out (note that d_{gi} does not have a time subscript). The logit of the probability that $d_{gi} = 1$ replaces the logit of h_{git} in Equation 9a, and the parameters have a d subscript to distinguish them from those in Equation 9a:

$$\text{logit}(pr(d_{gi} = 1|\pi_{0gi}, \pi_{1gi}, x_{gi}) = -\tau_{dg} + \lambda_{dg1}\pi_{0gi} + \lambda_{dg2}\pi_{1gi} + \lambda_{dg3}x_{gi}. \quad (9b)$$

If data are missing intermittently, the model shown in Equation 9b can be used for time-varying binary indicators d_{git} for the response being missing at time t (e.g., Albert & Follmann, 2009). Alternatively, intermittently missing data can be treated as MAR while dropout is treated as NMAR (e.g., Enders, 2011). The models can be extended to binary or count outcomes using the generalized linear modeling framework (Albert & Follmann, 2009; Follmann & Wu, 1995; ten Have, Kunselman, Pulkstenis, & Landis, 1998).

It is important to note that very strong assumptions underlie all NMAR models, including the random-effect-dependent missing-data model, and consistent estimation relies on these assumptions being correct. In addition to the assumption that the latent variables are normally distributed and that, conditional on latent variables, the outcome and drop-out processes (and their indicators) are independent, consistent estimation requires that the model for dropout is correct. For example, the drop-out indicator regressions must include all relevant covariates.

Latent class versions of NMAR models have also been proposed, including a pattern mixture model where, instead of the observed missing-data patterns, a discrete latent variable (which is influenced by the drop-out time) affects the outcome distribution (Roy, 2003), a random-effects-dependent drop-out model where dropout depends on random effects and latent classes that determine the random-effects means (Beunckens, Molenberghs, Verbeke, & Mallinckrodt, 2008) and a response-dependent drop-out model where a growth mixture model is used to model growth (Muthén et al., 2011). These models are discussed by Muthén et al. (2011) who also extend the Roy (2003) model to include two associated discrete latent class variables, one related to dropout and the other related to growth.

To investigate the potential consequences of incorrectly handling missing data that are NMAR, we conducted two empirical studies. The first compared the effects of different treatments of missing data when modeling trajectories of reading achievement in the National Education Longitudinal Study of 1988 (NELS:88) data set. The second was a simulation study in which data were generated according to a process by which dropout depends on the random coefficients (Equation 9a). All analyses were conducted with Mplus 5.21 (Muthén & Muthén, 1998–2009) using MLE. For the NELS data, robust standard errors were used for all analyses.

Missing Data in the NELS:88 Longitudinal Study

The NELS:88 is a longitudinal study of the academic, vocational, and personal development of a 1988 cohort of 8th graders. A two-stage sampling design was used, with schools randomly sampled from the nation's public and private schools, and students randomly sampled within each school (Curtin, Ingels, Wu, & Heuer, 2002). Sampling weights make the sample nationally representative of 1988 eighth graders. In 1990 and 1992, a subset of the original sample was followed up and new individuals were added to the sample. In each of these years, new sampling weights were calculated to account for changes in the population and for nonresponse, making the 1990 and 1992 data representative of students in 10th and 12th grades in these years, respectively, for cross-sectional analyses. In addition, panel weights were calculated to allow longitudinal study of the subset of the original sample that was followed (Curtin et al., 2002). The panel weights are designed to account for nonresponse and for incomplete follow-up of the original 8th graders, when only those students with complete data are used in the analysis (Curtin et al., 2002). Here, we analyze the entire 1988 sample with all available data up to 1992. The appropriate weights for our sample, therefore, are the 1988 (or baseline) weights. Similar to using panel weights, our approach allows us to study the trajectories of the original eighth graders through their senior year. However, our approach allows us to use all available data instead of dropping individuals who were lost to follow-up, which is required when panel weights are used.

There are two primary causes for missing data in the follow-up years of the NELS:88. The first is failure to follow up due to the increased number of high schools the original eighth-grade cohort attended. Funding limitations did not allow inclusion of all of the high schools, so only a limited number were included (Curtin et al., 2002). The second largest reason for missingness is students dropping out of school. We considered data from students who were lost to follow-up for the first reason to be MAR. However, it is likely that students are dropping out of school because they are doing poorly, or their achievement is lagging behind that of their peers. If this is correct, the probability of dropout may depend on the intercept and slope of the growth model (random-effect-dependent missingness), as shown in Equation 9a or 9b, making the missing data NMAR. Dropout seems unlikely to depend on the test scores themselves. First, the scores are not made available to students in the NELS study, so they cannot influence dropout directly. Second, if achievement affects dropout, this relationship is likely to be captured better

TABLE 1
Descriptive Statistics for NELS:88, by Minority Status

	Minority			White			Total		
Variable	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Reading achievement									
Time 1	6,147	41.37	7.54	15,753	46.97	8.50	21,900	45.54	8.62
Time 2	3,550	45.04	9.11	11,274	51.09	10.02	14,824	49.79	10.16
Time 3	2,580	47.93	9.70	8,772	54.20	10.35	11,352	52.92	10.54
# Risk factors		1.09	1.13		0.55	0.84		0.72	0.97
% Male		.50	.50		.50	.50		.50	.50
% Dropout		.13	.34		.09	.30		.10	.32

by allowing dropout to depend on the latent growth trajectory than by allowing it to depend on the test score, which measures achievement with error.

We used a multiple-group model to analyze reading-test score (scaled ability score, called theta, from the items response model) trajectories of two subgroups of students who had baseline scores: White (n = 16,699) and minority (n = 5,750). The multiple-group approach was used to investigate differences between subgroups because it permits all parameters to differ between groups, allowing tests of mean and variance differences, as well as interactions with a risk-factor index used as a covariate. The minority subsample comprised students of Black, Latino, and Native American ethnicity and, thus, had a higher probability of coming from disadvantaged backgrounds than white students. We did not include the subsample of students identified as Asian/Pacific Islander because the sample was small and, as a single group, difficult to classify with respect to level of advantage. We also omitted students who did not report their ethnicity. In the final analysis sample of 24,449, 10% dropped out of high school. Of these, 3% dropped out between 8th and 10th grades, and 7% between 10th and 12th grades. We used a count of risk factors (e.g., low income, low parent education, sibling dropout, etc.) to help explain individual differences in the intercept (in 8th grade) and slope.

NELS Results

Descriptive statistics for the NELS:88 sample are displayed by group in Table 1. White students had higher mean reading achievement scores than the minority students in all three years. White students also had fewer risk factors and a lower drop-out rate than the minority students. Both groups were about 50% female.

TABLE 2
Results for NELS:88: Maximum Likelihood Parameter Estimates (SE) for Growth Model (MAR) and Dual-Process Survival Model (NMAR) With Risk Index as Covariate

	White		Minority	
Parameter	Growth	Dual Process	Growth	Dual Process
Means				
β_{00} Intercept	48.19 (0.11)	48.19 (0.11)	43.34 (0.20)	43.34 (0.21)
β_{10} Linear slope	2.19 (0.08)	2.21 (0.08)	1.76 (0.13)	1.78 (0.12)
β_{20} Quadratic slope	−0.16 (0.02)	−0.17 (0.02)	−0.11 (0.03)	−0.13 (0.03)
Risk index coefficient				
β_{01} Intercept	−2.26 (0.09)	−2.26 (0.09)	−1.63 (0.10)	−1.62 (0.10)
β_{11} Linear slope	−0.13 (0.02)	−0.16 (0.02)	−0.05 (0.03)	−0.07 (0.03)
Variances				
ψ_{00} Intercept	54.61 (0.93)	54.58 (0.93)	38.65 (1.27)	38.65 (1.27)
ψ_{11} Linear slope	1.20 (0.09)	1.23 (0.09)	0.88 (0.15)	0.92 (0.15)
ψ_{01} Covariance	2.56 (0.21)	2.70 (0.21)	3.26 (0.30)	3.30 (0.30)
σ^2 Level 1 variance	17.47 (0.32)	17.43 (0.32)	17.47 (0.32)	17.43 (0.32)

Note: MAR = missing at random; NMAR = not missing at random.

Model-Based Approaches

We began with the multiple-group linear growth model shown in Equations 1a,1b, and 2 (and Figure 1), using MLE to estimate the model and assuming all missing data were MAR. We fitted the model both with and without the risk-factor index. The model with the risk factor had a better fit (likelihood ratio [LR] statistic = 1293.21, $df = 4$, $p < .001$) so we included the covariate in all analyses (Table 2).

The number of risk factors a student has in his or her life negatively affects both the intercept—reading achievement in eighth grade—and the slope, or improvement in reading across high school (Table 2), suggesting that students with fewer risk factors both begin and end high school with higher reading scores, on average. This is true for both White and minority students, but the effect is slightly stronger, on average, for White students.

In addition to the standard growth model, we estimated the dual-process models for reading scores and dropout, with the drop-out models shown in Equation 9a and 9b and in Figure 1. The single-indicator model (Equation 9b) is for the log-odds of dropping out anytime in high school, and the survival-process model (Equation 9a) is for the log hazard of dropping out in 10th or 12th grade, individually. All coefficients in the dropout model were constrained equal across White and minority students except for the coefficient of the risk index, which was allowed to differ both between

TABLE 3
Results for NELS:88: Estimates for Log-Hazard of Dropout in Dual-Process Model

	Estimate (*SE*)	Odds Ratio	95% CI
Parameter			
τ_{h1} 10th grade	0.37 (0.25)		
τ_{h1} 12th grade	−0.78 (0.23)		
λ_{h11} Intercept	−0.07 (0.01)	0.93	[0.91, 0.94]
λ_{h12} Linear	−0.33 (0.12)	0.72	[0.57, 0.90]
Risk index coefficient			
λ_{gh13} (White, 10th grade)	0.70 (0.04)	2.02	[1.85, 2.20]
λ_{gh13} (Minority, 10th grade)	0.40 (0.04)	1.49	[1.37, 1.62]
λ_{h23} (White, 12th grade)	0.46 (0.04)	1.58	[1.46, 1.70]
λ_{h23} (Minority, 12th grade)	0.25 (0.04)	1.28	[1.18, 1.39]

groups and, in the survival-process model, between time points. The two models resulted in very similar estimates, so only the survival-process estimates are shown (Tables 2 and 3). The NMAR-model estimates were not substantially different from those that resulted from standard MLE, assuming missing data are MAR (Table 2). Although the estimated coefficient of time squared in the minority group was lower (−8.9%) in the MAR model, compared with the (NMAR) survival-process model when not controlling for the risk index (results not shown), this difference disappeared when the risk index was included. This suggests that the risk index is necessary (but not necessarily sufficient) for the MAR assumption to hold.

Ad Hoc Approaches

Unlike the analyses that assumed missingness was MAR, analyses utilizing listwise deletion and mean imputation resulted in some strikingly divergent parameter estimates as compared with the survival-process model. The estimates and standard errors from these two analyses are shown, together with the results from the survival-process model, in Table 4. For all estimates, we calculated the percentage difference from the survival-process model and the largest of these are highlighted in Table 4 (bold if greater than 20% and underlined if greater than 100%). At the first time point, there were no missing values and the estimated intercept means did not differ very much between the survival-process and mean-imputation analyses. The listwise-deletion analysis, however, excluded all students who dropped out in either 10th or 12th grades and the intercept mean was higher than in the model-based approaches, because weaker students were more likely to drop out.

Both listwise deletion and mean imputation resulted in considerable, and sometimes hard-to-predict, differences in the estimated slopes, as well as the

TABLE 4

Results for NELS:88: Estimates (SE) Using Ad Hoc Approaches and Panel Weights to Deal With Missing Data

	White				Minority			
Parameter	Dual Process	Listwise Deletion	Mean Imputation	Panel Weights	Dual Process	Listwise Deletion	Mean Imputation	Panel Weights
Means								
β_{00} Intercept	48.19 (0.11)	49.17 (0.15)	48.15 (0.11)	48.43 (0.18)	43.34 (0.21)	44.79 (0.29)	43.37 (0.20)	45.03 (0.33)
β_{10} Linear slope	2.21 (0.08)	2.29 (0.09)	2.01 (0.07)	2.29 (0.12)	1.78 (0.12)	1.63 (0.13)	**2.96 (0.13)**	1.67 (0.15)
β_{20} Quad. slope	−0.17 (0.02)	−0.18 (0.02)	**−0.12 (0.02)**	−0.18 (0.03)	−0.13 (0.03)	**−0.09 (0.03)**	**−<u>0.26 (0.03)</u>**	**−0.09 (0.03)**
Risk Index coef.								
β_{01} Intercept	−2.26 (0.09)	−2.06 (0.15)	−2.18 (0.09)	−2.28 (0.19)	−1.62 (0.10)	−1.49 (0.15)	−1.53 (0.10)	−1.54 (0.18)
β_{11} Linear slope	−0.16 (0.02)	**−0.11 (0.03)**	**<u>0.15 (0.02)</u>**	**−0.09 (0.03)**	−0.07 (0.03)	**−0.01 (0.04)**	**<u>0.21 (0.03)</u>**	**−0.03 (0.04)**
Variances								
ψ_{00} Intercept	54.58 (0.93)	53.22 (1.15)	49.29 (0.87)	53.13 (1.40)	38.65 (1.27)	42.91 (2.02)	32.20 (1.16)	41.24 (2.29)
ψ_{11} Linear slope	1.23 (0.09)	1.15 (0.09)	1.28 (0.07)	1.04 (0.11)	0.92 (0.15)	0.87 (0.14)	**1.53 (0.11)**	**0.62 (0.15)**
ψ_{01} Covariance	2.70 (0.21)	**1.97 (0.22)**	**−<u>2.63 (0.21)</u>**	**2.04 (0.27)**	3.30 (0.30)	**2.60 (0.36)**	**−<u>2.35 (0.29)</u>**	2.92 (0.37)
σ^2 Residual	17.43 (0.32)	16.92 (0.33)	**20.51 (0.32)**	17.33 (0.50)	17.43 (0.32)	16.92 (0.33)	**20.51 (0.32)**	17.33 (0.50)

Note: Bold indicates estimate >20% difference relative to dual-process survival model. Bold and underlined indicates estimate >100% difference, relative to dual-process survival model.

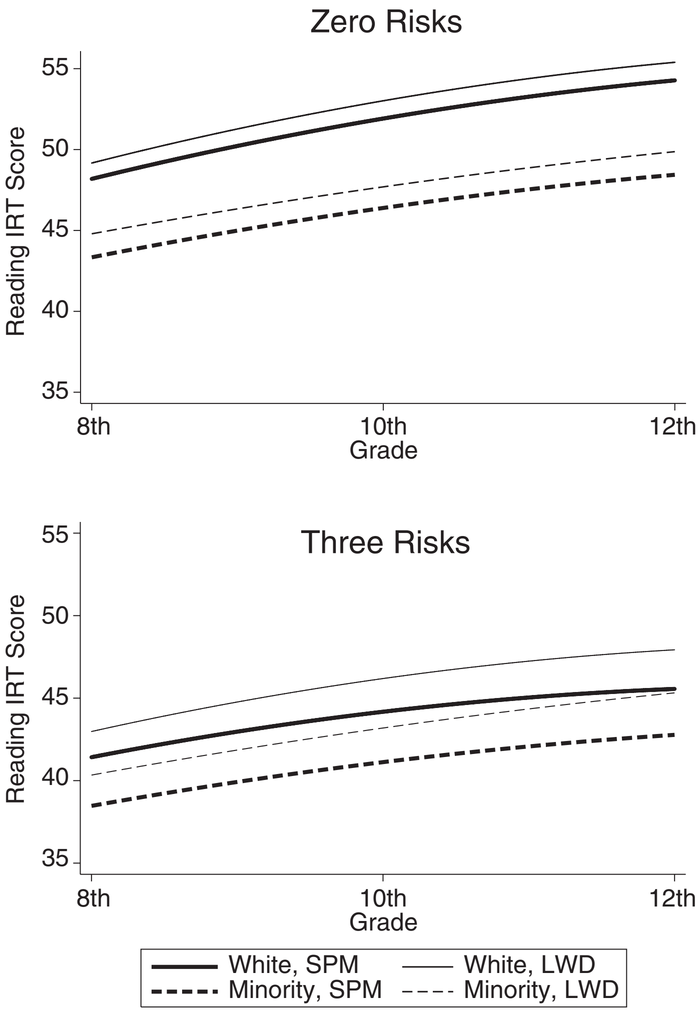

FIGURE 2. *NELS:88 reading achievement trajectories, comparing listwise-deletion (LWD) with the survival-process model (SPM) results. When there are no risk factors, the intercept for LWD is slightly higher but the slopes are not affected. As risks increase, the slopes are increasingly more positive, relative to SPM.*

coefficients from the regression of growth factors on the risk index (Table 4). Listwise deletion resulted in an overall positive difference in the mean trajectories, relative to the survival-process model, but the magnitude of the difference varied with the number of risk factors. This is shown in Figure 2. With no risks,

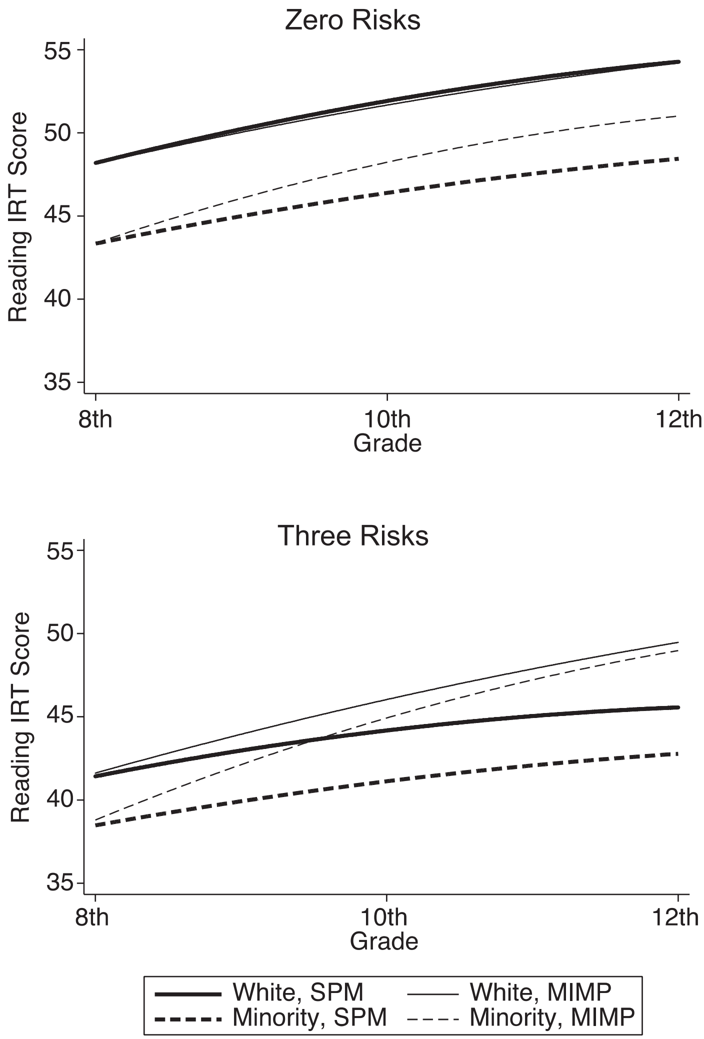

FIGURE 3. *NELS:88 reading achievement trajectories: comparing mean-imputation (MIMP) with the survival-process model (SPM) results. Minority slopes are more affected by the mean-imputation approach, and as risks increase, the difference grows greater.*

the lines from the two analyses are approximately parallel, with the listwise-deletion estimates higher across high school. However, using listwise deletion gives the appearance that students with more risks are not only starting higher but also have a more positive trajectory. This hides the effect of risk factors on the progress of the student. The survival-process model suggests that more risks actually decrease the slope, especially for White students (Figure 2, Table 4).

Mean imputation resulted in even greater differences in the estimated average trajectories (Figure 3). With no risk factors, mean imputation, as opposed to the survival-process model, gives the appearance that the minority students are improving more rapidly. Based on this analysis, it could be argued that the minority students with no risks are, on average, catching up with their White peers and narrowing the minority achievement gap. When students come in with more risk factors, the slope difference is even greater, giving the appearance that the achievement gap may close on its own during high school.

NELS:88 Panel Weights

A final set of analyses utilized the panel weights supplied with the NELS:88 data. Many people using the NELS:88 data for longitudinal research would utilize listwise deletion, and correct for nonresponse using the panel weights, as suggested in the accompanying use manuals (Curtin et al., 2002). The results are shown in Table 4 and are similar to the estimates that use listwise deletion with baseline weights. Panel weights are therefore not a substitute for MLE of all available data.

Simulation Study

Method

To conduct a controlled investigation of the behavior of models in which the missing-data-mechanism assumptions are violated, we utilized simulated data. Unlike real data, such as the NELS:88 data set, simulated data permit comparisons of model estimates with known population values. In addition, the mean of the estimated standard errors can be compared with the standard deviation of the estimates to assess bias in standard error estimates.

We simulated five equally spaced waves of continuous data using the growth model shown in Equations 1a, 1b, and 2, but without groups or covariates, z_{git} and x_{gi}. Missing data were generated according to the survival-process model in Equation 9a, again without groups or covariates, and the log-hazard of dropout at each time point was negatively related to individual values of the intercept via a single coefficient for all drop-out indicators, and the slope, by way of a second coefficient common across all drop-out indicators. We utilized five analysis models: (a) a traditional growth model (HLM); (b) a dual-process single-indicator model that includes a single-indicator variable for dropping out at any time after the first time point, as shown in Equation 9b; (c) the survival-process model that generated the data; (d) listwise deletion; and (e) mean imputation. Intercept and linear slope parameters were specified to resemble the NELS:88 reading achievement data.

We analyzed 1,000 data sets for each cell of a $2 \times 2 \times 2$ design:

- two sample sizes ($n = 300$ and $n = 1{,}000$),

- two percentages of missing data (10% and 40% dropout by the last time point), and
- two levels of dependence (weak and strong) of the drop-out process on the intercept and slope. The coefficients for the intercept and slope were set to −0.1 and −0.2 for weak dependence and −0.5 and −1.4 for strong dependence.

Data were simulated in Stata, Version 10 (StataCorp, 2007), and analyzed using Mplus 5.1 (Muthén & Muthén, 1998–2009).

In evaluating the simulation results, we assessed bias in parameter estimates as the percentage difference from the true parameter value, bias in the estimated model-based standard errors as the percentage difference from the standard deviation of the estimates, and coverage (the proportion of 95% confidence intervals that contained the true parameter). In the growth model analyses, we also report the average chi-square statistic (LR test comparing the fitted model with saturated model), as well as the percentage of the analyses with LR-test values above the critical value for the model degrees of freedom. Only models that converged properly were included in the summaries. This was done to cull out improper parameter estimates and standard error values. We generated extra simulations, allowing us to use the first 1,000 cases that converged properly when calculating the estimate summaries.

Results

Convergence. All models that ignored the missing data mechanism, including those that utilized ad hoc approaches, converged. The two approaches that explicitly modeled the missing-data process resulted in some analyses in which convergence was a problem. In a few cases, Mplus used a different ML estimator (*MLF*) that substitutes the outer product of the gradients for a noninvertible Hessian matrix to estimate standard errors (Muthén, 1998–2004, p. 32). The survival-process drop-out models resulted in more problems with convergence because they utilize a binary drop-out indicator at each time point that is zero before dropout and is missing for the time points following dropout. In particular, strong dependence of dropout on the random coefficients, combined with a small percentage of dropouts (10%), led to 174 data sets, of the 1,000, with convergence problems. Of these, 54 failed to converge and the rest resulted in either parameters fixed at boundary values or substitution of the MLF estimator. Fewer convergence problems occurred in the survival-process model when dropout was 40%; 22 failed to converge at all and only 5 had other problems. The single-indicator drop-out model had fewer problems with convergence than the survival-process model; 13 of the 1,000 replications failed to converge when the sample was small and drop-out dependence was strong. In nearly all cases, the parameters that caused the problems were those associated with the drop-out indicators.

Across all models, the results from the samples of 300 and of 1,000 were very similar, except for small differences in the standard errors, so only the large-

TABLE 5

Mean Estimates, Estimated Bias, and Coverage; and Mean Fit Statistics for Growth Model (MAR): Simulated Samples of 1,000

		10% Missing Data						40% Missing Data					
		Small Coefficients			Large Coefficients			Small Coefficients			Large Coefficients		
Parameter	True Value	*M*	% Bias	Cov.	*M*	% Bias	Cov.	*M*	% Bias	Cov.	*M*	% Bias	Cov.
Means													
Intercept	44.00	44.00	0.00	.95	44.01	0.03	.95	44.00	0.00	.95	44.05	0.11	.95
Slope	2.50	2.53	1.11	.91	2.62	4.92	.35	2.64	5.61	.39	3.12	24.64	.00
Variances													
Intercept	68.00	67.96	−0.06	.95	67.51	−0.72	.94	67.90	−0.15	.95	67.30	−1.03	.93
Slope	2.00	1.98	−1.13	.94	1.81	−9.39	.63	1.91	−4.62	.88	1.61	−19.65	.19
Covariance	3.50	3.29	−5.98	.91	2.06	−41.08	.18	2.77	−20.86	.73	−0.86	−124.46	.00
Residual	6.97	6.97	0.04	.94	6.99	0.24	.94	6.97	0.05	.94	7.02	0.70	.94
Fit statistics ($df = 14$)													
Ave. LR(% rejected)			13.96 (6.3%)			14.52 (6.4%)			14.09 (5.3%)			17.86 (16.3%)	
CFI (% $\geq$ .95)			1.00 (0%)			1.00 (0%)			1.00 (0%)			1.00 (0%)	
TLI (% $\geq$ .95)			1.00 (0%)			1.00 (0%)			1.00 (0%)			1.00 (0%)	
RMSEA (% $\leq$ .05)			0.01 (0%)			0.01 (0%)			0.01 (0%)			0.01 (0.1%)	

Note: MAR = missing at random. Small coefficients are −0.1 and −0.2 for intercept and slope, respectively; large coefficients are −0.5, and −1.4. Cov. is 95% coverage. LR = likelihood ratio statistic (model chi-square).

sample results are shown in the tables. Any differences in results between the two sample sizes are reported in the text.

Model-based approaches. The first growth model used MLE with an MAR assumption. Table 5 shows parameter estimates and fit statistics from this analysis, along with the true parameter values. The first part of the table contains estimates, percent bias, and coverage for the parameters: intercept and slope means, variances, and the intercept–slope covariance. The lower part of the table shows averages of fit statistics and the percentage of LR statistics (model chi-square) rejected.

The intercept mean and variance (Table 5) showed little bias under any conditions, which was expected because there are no missing values at the first time point, where the intercept is located. This was not, however, always the case for the slope parameters. Under the best of circumstances, when dropout was 10% and dependence was weak, coverage for the slope mean and the slope–intercept covariance were just slightly less than the nominal coverage value of 95% and parameter bias was low. However, when the dependence was strong, dropout was 40%, or both, positive bias in the slope mean estimate became substantial (in the worst case, greater than 20%) and coverage dropped precipitously. The slope variance showed similar degrees of negative bias and its coverage was also poor. The covariance between the intercept and slope suffered the greatest negative bias, exceeding 100% (and switching sign) when both dropout and dependence were high.

A researcher assessing model fit relies on statistics, such as the LR statistic comparing the estimated model with the saturated model. When estimating the correct model, the average LR statistic should be close to the degrees of freedom (14 in this model) and approximately 5% of the models should have LR statistics greater than the critical value of 23.7. Table 5 shows that the LR test rejected at substantially higher rates than expected only when 40% of the data were missing and dependence on the random coefficients was strong. Even then, only 16% of models were rejected. Furthermore, a researcher looking at other fit statistics commonly used in structural equation modeling, such as Comparative Fit Index (CFI) and root mean squared error of approximation (RMSEA), would probably conclude that the fit is adequate, even if the model is rejected using the LR test (which is known to be overly sensitive to minor misfit with large samples). If missingness depends on random coefficients but is treated as ignorable, it is likely that the fit statistics will not indicate a problem even when estimates are severely biased and coverage is low. The RMSEAs for the small samples ($n = 300$) were slightly more likely to suggest misfit with the approximately 5% to 7% of RMSEA point estimates greater than .05.

We found very little standard error bias in the growth model. That is, most of the standard errors for the growth model analyses were acceptably close to the standard deviations of the estimates, although, in the smaller samples, some standard errors were slightly elevated (bias close to 7%) when dependence and the percentage of missing data were both high. Ignoring the nonignorable missing-data mechanism does not seem to have a strong effect on the standard errors in this case. However, as mentioned previously, coverage is poor whenever the parameter is estimated with bias.

The second set of analyses included a single drop-out indicator that equaled one if the individual had dropped out at any time (Equation 9b). This approach is easier to implement than the survival-process model that generated the data, making it an appealing option. Table 6 gives the results of these analyses. When 10% of the data were missing and the dependence of missingness on random coefficients was low, results showed no notable bias and the coverage was adequate. This was even true for variances and covariances, which did not have adequate coverage when dropout was assumed to be MAR. When dropout was 40% or dependence was strong, the covariance was negatively biased by approximately 10% and coverage was also somewhat low (.89), but the means of the other parameter estimates were within 5% of their true values (Table 6). Coverage, in all of these cases, was adequate for the smaller sample. Finally, with 40% dropout and strong dependence, all of the slope parameters estimates were biased and their coverage was poor. Nonetheless, both estimates and coverage were better than those produced by growth models that ignored the dropout. Standard error averages were, again, all close to the *SD* of the estimates (no bias greater than 6% in the small samples, and none greater than 5% in the large samples).

The analyses that utilized the data-generating model (dual growth plus survival model; Equation 9a) consistently resulted in parameter estimates that were close to the true values, and coverage ranged from .93 to .97. The *SE*s also did not show any substantial bias. This was the case for both large and small samples.

Ad hoc approaches. The results of the listwise-deletion analyses are shown in Table 7. Listwise deletion was the only approach that resulted in biased estimates for the intercept mean. Because dropouts were more likely to have low intercepts, the variance of the intercept was reduced and its mean was biased upward (Table 7). When more data were missing or the dependence was stronger, or both, the slope variance and intercept–slope covariance estimates were also negatively biased, and the slope mean was positively biased.

Finally, Table 8 shows the results of imputing test-score means to fill in missing values. Intercept means were unbiased, but slope means were biased upward because the sample mean was substituted for what would have been lower scores. In more severe missing-data cases, the slope bias was quite large—nearly 65% when 40% dropped out and dependence was high. The effects of mean imputation on the random-coefficient variances were harder to predict. Despite the fact

TABLE 6

Estimates, Bias, Coverage, and Fit Statistics for Single-Indicator Dropout Model: Simulated Samples of 1,000

		10% Missing Data						40% Missing Data					
		Small Coefficients			Large Coefficients			Small Coefficients			Large Coefficients		
Parameter	True Value	*M*	% Bias	Cov.	*M*	% Bias	Cov.	*M*	% Bias	Cov.	*M*	% Bias	Cov.
Means													
Intercept	44.00	44.00	0.00	.95	44.01	0.02	.95	44.00	0.00	.95	44.03	0.07	.95
Slope	2.50	2.50	0.05	.95	2.52	0.90	.93	2.55	1.84	.90	2.71	8.59	.25
Variances													
Intercept	68.00	67.96	−0.06	.94	67.64	−0.53	.94	67.85	−0.23	.95	67.40	−0.88	.94
Slope	2.00	2.00	0.21	.96	1.94	−2.78	.92	1.95	−2.41	.93	1.71	−14.26	.50
Covariance	3.50	3.44	−1.58	.95	3.15	−9.95	.89	3.11	−11.08	.89	1.76	−49.61	.32
Residual	6.97	6.97	−0.01	.94	6.97	−0.04	.94	6.96	−0.10	.94	6.96	−0.08	.94

Note: Small coefficients are −0.1 and −0.2 for intercept and slope, respectively; large coefficients are −0.5, and −1.4. Cov. is 95% coverage.

TABLE 7
Estimates, Bias, Coverage, and Fit Statistics for Listwise Deletion: Simulated Samples of 1,000

		10% Missing Data						40% Missing Data					
		Small Coefficients			Large Coefficients			Small Coefficients			Large Coefficients		
Parameter	True Value	*M*	% Bias	Cov.	*M*	% Bias	Cov.	*M*	% Bias	Cov.	*M*	% Bias	Cov.
Means													
Intercept	44.00	44.73	1.66	.24	45.43	3.26	.00	46.94	6.68	.00	48.75	10.80	.00
Slope	2.50	2.57	2.84	.74	2.67	6.67	.11	2.79	11.62	.01	3.05	22.12	.00
Variances													
Intercept	68.00	63.26	−6.97	.65	52.38	−22.98	.00	54.08	−20.48	.02	36.29	−46.64	.00
Slope	2.00	1.96	−2.08	.93	1.79	−10.44	.57	1.87	−6.72	.82	1.57	−21.47	.13
Covariance	3.50	3.02	−13.82	.80	1.66	−52.61	.00	2.10	−39.85	.22	−0.20	−105.66	.00
Residual	6.97	6.97	−0.01	.94	6.97	−0.02	.94	6.97	−0.05	.94	6.98	+0.10	.94
Fit statistics ($df = 14$)													
Ave. LR (% rejected)			14.03 (5.2%)			14.14 (6.0%)			14.22 (5.3%)			14.19 (5.2%)	
CFI (% ≥ .95)			1.00 (0%)			1.00 (0%)			1.00 (0%)			1.00 (0%)	
TLI (% ≥ .95)			1.00 (0%)			1.00 (0%)			1.00 (0%)			1.00 (0%)	
RMSEA (% ≤ .05)			0.01 (0%)			0.01 (0%)			0.01 (0.2%)			0.01 (0.1%)	

Note: Small coefficients are −0.1 and −0.2 for intercept and slope, respectively; large coefficients are −0.5, and −1.4. Cov. is 95% coverage. LR = likelihood ratio statistic (model chi-square).

TABLE 8

Estimates, Bias, Coverage, and Fit Statistics for Mean Imputation: Simulated Samples of 1,000

		10% Missing Data						40% Missing Data					
		Small Coefficients			Large Coefficients			Small Coefficients			Large Coefficients		
Parameter	True Value	*M*	% Bias	Cov.	*M*	% Bias	Cov.	*M*	% Bias	Cov.	*M*	% Bias	Cov.
Means													
Intercept	44.00	43.97	−0.06	0.95	44.24	0.54	0.84	43.99	−0.01	0.94	45.12	2.53	0.00
Slope	2.50	2.75	10.14	0.01	3.00	20.06	0.00	3.52	40.98	0.00	4.12	64.85	0.00
Variances													
Intercept	68.00	68.05	0.08	0.95	59.96	−11.83	0.19	64.05	−5.81	0.73	42.68	−37.23	0.00
Slope	2.00	2.31	15.49	0.45	2.45	22.55	0.10	2.88	44.13	0.00	2.49	24.46	0.01
Covariance	3.50	0.20	−94.32	0.00	−1.89	−154.00	0.00	−6.37	−281.96	0.00	−6.14	−275.55	0.00
Residual	6.97	8.41	20.71	<0.01	9.57	37.31	0.00	10.81	55.16	0.00	12.28	76.20	0.00
Fit statistics (df = 14)													
Ave. LR (% rejected)			27.49 (88.2%)			145.28 (100%)			114.29 (100%)			996.34 (100%)	
CFI (% ≥ .95)			1.00 (0%)			.97 (2.7%)			.97 (1.6%)			.72 (100%)	
TLI (% ≥ .95)			1.00 (0%)			.98 (0%)			.98 (0%)			.80 (100%)	
RMSEA (% ≤ .05)			.03 (20.5%)			.10 (100%)			.08 (100%)			.26 (100%)	

Note: Small coefficients are −0.1 and −0.2 for intercept and slope, respectively; large coefficients are −0.5, and −1.4. Cov. is 95% coverage. LR = likelihood ratio statistic (model chi-square).

that no data were missing at the first time point, intercept variances were negatively biased if the proportion of missingness or dependence were high, and, in general, the slope variances were biased upward. The covariance was severely negatively biased; in the worst case, it exceeded -275%. Finally, neither ad hoc approach showed any substantial standard error bias when compared with standard deviations of the estimates.

Discussion

NELS:88

This study investigated the results of assuming that missing data due to student dropout are ignorable in an educational setting. Our specific concern was the effect of school dropout on modeling of learning trajectories. We began by modeling high school reading achievement trajectories using the NELS:88 reading achievement data. We felt that the missing data of individuals who dropped out of school might be NMAR if their dropout resulted from poor progress during high school. We found little evidence of nonignorable missingness after adding a risk-factor index as a covariate.

Estimating an NMAR missing-data model can be viewed as a type of sensitivity analysis (e.g., Diggle & Kenward, 1994); changes in parameter values that result from adding the missingness-model suggest that the dropout is NMAR. Unchanged estimates suggest, but do not guarantee, that the missing data are MAR, conditional on the risk-factor covariate.

Although ad hoc approaches to dealing with missing data have been criticized many times in the past (e.g., Enders, 2001; Peugh & Enders, 2004), such approaches are still found in recent educational research papers. Because of this, we also analyzed the data using two of the more common ad hoc methods traditionally found in the educational literature: listwise deletion and mean imputation. The results when these approaches were used differed from the model-based approaches even when NELS:88 panel survey weights, which are designed to correct for nonresponse, were used. In general, both approaches gave an appearance of greater academic growth during high school than the model-based methods. In our analyses, these differences became increasingly severe as the number of risk factors increased (see Figures 2 and 3). These methods also resulted in large differences in the estimated variances and covariances (Table 4).

Simulated Data

The simulation was useful for two reasons. With simulated NMAR data, we were able to assess bias in parameter estimates when an incorrect approach was used to deal with the missing data. In addition, the simulation allowed us to explore what proportion of the data had to be missing, as well as how strong the dependence of the missingness on the random coefficients had to be, before treating an NMAR missing-data process as MAR would begin to result in serious bias. We found that the limits depend on the research questions. For example, a researcher solely interested in the mean trajectories might be able to ignore as much as 40% nonignorable missing data, as long as the dependence of the missingness on the random coefficients is low or, equivalently, the MAR assumption is not severely violated. When the MAR assumption is more severely violated, however, the slope mean estimate will be biased if the proportion of missingness is high. Moreover, if variances or covariances are of substantive interest, significant bias begins to show up with as little as 10% missingness when dependence is strong, or with 40% missingness when dependence is weak.

A negatively biased intercept–slope covariance, in particular, might be of concern when initial scores are actually positively correlated with slopes, as was the case in the NELS:88 analysis. The positive correlation suggested that students who start at an academic disadvantage are likely to end school at an even greater disadvantage, and could potentially benefit from a targeted intervention. If the covariance bias is so great that the direction of the correlation is reversed, as occurred in several of the simulation analyses, there might be no impetus to intervene—the gap would appear to be closing on its own.

As we did in the NELS:88 analyses, we also estimated the simulated models using listwise deletion and mean imputation. Both of these methods assume that data are MCAR. The effect of violating this assumption can be seen in the results of the simulation study. Biases can be severe—as great as −275% when mean imputation was used under the most demanding of conditions (40% missing, strong dependence on the random effects).

Conclusions

These analyses showed that treating missingness as ignorable when, in fact, it depends on the random coefficients, can result in biased estimates, especially for the estimated variances and covariances. When ad hoc methods such as listwise deletion and mean imputation are used, the bias can be quite severe and may not be entirely predictable. Many other ad hoc methods were not discussed in this article, but may also cause seriously distorted estimates. Examples are last observation carried forward and, similar to the listwise deletion approach used in this article, deleting all individuals without some arbitrary minimum number of time points, such as two or three. Unfortunately, the latter approach is quite common because many researchers seem to think that individuals with only one observation cannot be included in the analysis.

Because there is no way to know the nature of the missing-data process, we recommend using maximum likelihood estimators and investigating the sensitivity of the estimates to the MAR assumption by fitting NMAR models, as we did here with the NELS:88 analysis. This approach has been recommended by Beunckens, Molenberghs, Thijs, and Verbeke (2007), Molenberghs et al.

(2004), Xu and Blozis (2011), and others. However, NMAR models, themselves, depend on unverifiable assumptions and should not be relied on without checking for sensitivity to violations of their assumptions (Kenward, 1998; Little, 1995; Verbeke, Molenberghs, Thijs, Lesaffre, & Kenward, 2001). In particular, the selection models of this article are sensitive to distributional assumptions (see, e.g., Little, 1994, and references therein), such as the assumption of multivariate normality of the latent variables, as well as the conditional independence assumption and the requirement that the missing-data model is correct. We also strongly suggest that NMAR models be based on theory, as was the case in our analysis, and not used in an exploratory fashion (see Molenberghs et al., 2008), as the wrong model can yield misleading results. We join with other researchers in recommending that individuals designing longitudinal studies get as much information as possible on variables that have been shown to predict missingness (Collins et al., 2001; Diggle & Kenward, 1994; Little, 1995; Schafer & Graham, 2002); the more this type of information is used in the analysis, the more likely the missing data are to be MAR.

Finally, there exists controversy over whether test scores from dropouts should even be considered missing. Under similar circumstances, Zhang and Rubin (2003) present an alternative model based on an argument that data missing due to dropout are not MNAR, but undefined. This is a complex issue, and will probably depend on the research question.

As longitudinal research is increasingly used, missing data will become a more common issue and missing-data problems will require more attention than they have in the past. Data collection for the NELS:88 was well funded and few individuals were lost to follow-up. States enacting longitudinal student tracking systems are likely to have greater problems with missing data, as the students who drop out of school will most likely not be assessed. Students moving to another state may also be lost to follow-up. It is our hope that the results we have shown here will help alert individuals who are engaged in such data collection about the need for careful attention to the information that is gathered, and will inform those who analyze the data to the need to use appropriate methods for dealing with the missing data in longitudinal analyses.

Appendix: Mplus Syntax

```
Mplus Growth Model Syntax (MAR)

TITLE: Nels:88 Conditional Growth Model with Ethnicity Groups

DATA: FILE IS nels.txt;
```

Data file is in folder with Mplus input file

```
VARIABLE:
NAMES ARE id sch_id sstratid byqwt minority
male race asian latino black amerind white unknown
byrisk birthyr dropout dof1 dof2 rdirtd1 rdirtd2 rdirtd3
rdmimp1 rdmimp2 rdmimp misread missrace;
```

Names of all of the variables in the data

```
IDVAR IS id;
```

This tells Mplus to include ID variable in any individual-level output

```
USEVARIABLES ARE rdirtd1 rdirtd2 rdirtd3 byrisk;
```

Names of the variables in the analysis

```
MISSING ARE all (-99);
```

`MISSING ARE` *identifies missing-data code (-99)*

```
STRATIFICATION IS sstratid;
CLUSTER IS sch_id;
WEIGHT IS byqwt;
```

These next three commands handle the complex survey design, including weights

```
CLASSES IS ethnicty(2);
KNOWNCLASS IS ethnicity(minority=0 minority=1)
```

These two commands tell Mplus to estimate this multiple-group model as a mixture model with two classes that are idenfified by the variable, ethnicty

```
ANALYSIS:
TYPE IS COMPLEX MIXTURE;
```

`COMPLEX` *means complex survey design*
`MIXTURE` *indicates a mixture model*

```
MODEL:
%OVERALL%
int slope quad | rdirtd1@0 rdirtd2@2 rdirtd3@4;
quad@0;
rdirtd1(1); rdirtd2(1); rdirtd3(1);
int ON byrisk;
slope ON byrisk;
```

`%OVERALL%` *is required for a mixture model. Under it is the full model for both classes. The variance of the quadratic slope is fixed to zero* `(q@0)` *and the residual variances are fixed equal across the three timepoints by the number in parentheses* `(1)`

Intercept and linear slope (`int` *and* `slope`*) are regressed on the risk-factors variable (*`byrisk`*)*

```
%ethnicty#1%
int; slope; int WITH slope;
[int slope quad];
int ON byrisk;
slope ON byrisk;

%ethnicty#2%
int; slope; int WITH slope;
[int slope quad];
int ON byrisk;
slope ON byrisk;
```

These two sections specify what is constrained to be equal, versus free to differ, between groups. Parameters in square brackets refer to means; those outside of brackets refer to variances. Both are permitted to vary across groups.

```
OUTPUT: TECH1 TECH3 TECH4 STANDARDIZED SAMPSTAT RESIDUAL;

SAVEDATA:
RESULTS = results.nels.risk.txt;
FILE IS data.nels.risk.txt;
SAVE = FSCORES;
```

`RESULTS =` *saves the parameter estimates into the text file named. The order of the estimates is in the output file.* `Save = FSCORES` *saves the individual intercepts and slopes (factor scores) along with the variables in the analysis.*
`IS, ARE`*, and "=" are interchangeable in Mplus.*
Mplus code and variable names are not case sensitive.
All commands must end with a semicolon.

Mplus Growth Model Syntax with Dropout (NMAR) Model

```
TITLE: Nels:88 Conditional Growth Model with Ethnicity Groups and Dropout

DATA: FILE IS nels.txt;

VARIABLE:
NAMES ARE id sch_id sstratid byqwt minority
male race asian latino black amerind white unknown
byrisk birthyr dropout dof1 dof2 rdirtd1 rdirtd2 rdirtd3
rdmimp1 rdmimp2 rdmimp misread missrace;
IDVAR IS id;
USEVARIABLES ARE rdirtd1 rdirtd2 rdirtd3 byrisk dropout;
CATEGORICAL is dropout;
MISSING ARE all (-99);
STRATIFICATION IS sstratid;
CLUSTER IS sch_id;
WEIGHT IS byqwt;
CLASSES IS ethnicty(2);
KNOWNCLASS IS ethnicity(minority=0 minority=1);

ANALYSIS:
TYPE IS COMPLEX MIXTURE;
ALGORITHM = INTEGRATION;
STARTS = 50 25;
PROCESSOR = 8(STARTS);

MODEL:
%OVERALL%
int slope quad | rdirtd1@0 rdirtd2@2 rdirtd3@4;
quad@0;
rdirtd1(1); rdirtd2(1); rdirtd3(1);
int ON byrisk;
slope ON byrisk;
dropout ON byrisk;
[dropout$1](6);

%ethnicty#1%
int; slope; int WITH slope;
int ON byrisk;
slope ON byrisk;
dropout ON byrisk;
%ethnicty#2%
int; slope; int WITH slope;
int ON byrisk;
slope ON byrisk;
dropout ON byrisk;

OUTPUT: TECH1 TECH3 TECH4 STANDARDIZED SAMPSTAT RESIDUAL;
SAVEDATA:
RESULTS = results.nels.risk.dropout.txt;
FILE IS data.nels.risk.dropout.txt;
SAVE = FSCORES;
```

`dropout` *is an indicator that the individual dropped out at any point in high school*

Observed categorical outcomes (endogenous variables) must be described as such (→ `CATEGORICAL is dropout;`)

`ALGORITHM = INTEGRATION` *is needed with a categorical distal outcome (*`dropout`*).*

`STARTS =` *tells Mplus to do multiple starts from random places in the parameter space as a precaution against converging to a local maximum. It will do 50 starts up to 10 iterations and continue the best 25 of those to convergence.*

`PROCESSOR =` *tells Mplus how many processors (or cores) available in the computer and* `(STARTS)` *tells it to spread the random starts among processors (for added speed).*

`[dropout$1]` *refers to the threshold of the binary variable (fixed equal across groups)*

`dropout` *was regressed on the risk index (*`byrisk`*) to control for direct effects of risk factors on the probability of dropping out.*

```
Mplus Growth Model Syntax with Survival-Process (NMAR) Model

TITLE: Nels:88 Conditional Growth Model with Ethnicity Groups and SP Dropout
DATA: FILE IS nels.txt;
VARIABLE:
NAMES ARE id sch_id sstratid byqwt minority
male race asian latino black amerind white unknown
byrisk birthyr dropout dof1 dof2 rdirtd1 rdirtd2 rdirtd3
rdmimp1 rdmimp2 rdmimp misread missrace;
IDVAR IS id;
USEVARIABLES ARE rdirtd1 rdirtd2 rdirtd3 byrisk dof1 dof2;
CATEGORICAL is dof1 dof2;
MISSING ARE all (-99);
STRATIFICATION IS sstratid;
CLUSTER IS sch_id;
WEIGHT IS byqwt;
CLASSES IS ethnicty(2);
KNOWNCLASS IS ethnicity(minority=0 minority=1);
ANALYSIS:
TYPE IS COMPLEX MIXTURE;
ALGORITHM = INTEGRATION;
STARTS = 50 25;
PROCESSOR = 8(STARTS);

MODEL:
%OVERALL%
int slope quad | rdirtd1@0 rdirtd2@2 rdirtd3@4;
quad@0;
rdirtd1(1); rdirtd2(1); rdirtd3(1);
int ON byrisk;
slope ON byrisk;
dropout ON byrisk;
dof1 ON int slope byrisk;
dof2 ON int slope byrisk;
[dof1$1](8); [dof2$1](9);
%ethnicty#1%
int; slope; int WITH slope;
int ON byrisk;
slope ON byrisk;
dof1 ON slope(3);
dof2 ON slope(3);
dof1 on int(2);
dof2 on int(2);
dof1 on byrisk(6);
dof2 on byrisk(7);
%ethnicty#2%
int; slope; int WITH slope;
int ON byrisk;
slope ON byrisk;
dropout ON byrisk;
dof1 ON slope(3);
dof2 ON slope(3);
dof1 on int(2);
dof2 on int(2);
dof1 on byrisk(61);
dof2 on byrisk(71);

OUTPUT: TECH1 TECH3 TECH4 STANDARDIZED SAMPSTAT RESIDUAL;
SAVEDATA: RESULTS = results.nels.risk.surv.txt;
FILE IS data.nels.risk.surv.txt;
SAVE = FSCORES;
```

`dof1` *and* `dof2` *are indicators that the individual dropped out before 10th or 12th grades, respectively.*

Regressions of `dof1` *and* `dof2` *on the growth parameters were assumed to be equal for 10th and 12th grades, but were tested across group. The regressions on the risk index was not assumed to be equal across time or across groups*

Declaration of Conflicting Interests

The authors declared no conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through grant R305A070254 to University of California, Berkeley. The opinions expressed are those of the authors and do not represent the views of the Institute or the U.S. Department of Education.

References

Albert, P. S., & Follmann, D. A. (2009). Shared-parameter models. In G. Fitzmaurice, M. Davidian, G. Verbeke, & G. Molenberghs (Eds.), *Longitudinal data analysis* (pp. 433–452). Boca Raton, FL: CRC Press.

Beunckens, C., Molenberghs, G., Thijs, H., & Verbeke, G. (2007). Incomplete hierarchical data. *Statistical Methods in Medical Research*, *16*, 457–492. doi:10.1177/0962280206075310

Beunckens, C., Molenberghs, G., Verbeke, G., & Mallinckrodt, C. (2008). A latent-class mixture model for incomplete longitudinal Gaussian data. *Biometrics*, *64*, 96–105. doi:10.1111/j.1541-0420.2007.00837.x

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective*. Hoboken, NJ: Wiley-Interscience.

Borman, G. D., Slavin, R. E., Cheung, A. C. K., Chamberlain, A. M., Madden, N. A., & Chambers, B. (2007). Final reading outcomes of the National Randomized Field Trial of Success for All. *American Educational Research Journal*, *44*, 701–731. doi:10.3102/0002831207306743

Bouffard, T., Vezeau, C., Roy, M., & Lengelé, A. (2011). Stability of biases in self-evaluation and relations to well-being among elementary school children. *International Journal of Educational Research*, *50*, 221–229. doi:10.1016/j.ijer.2011.08.003

Cai, J., Wang, N., Moyer, J. C., Wang, C., & Nie, B. (2011). Longitudinal investigation of the curricular effect: An analysis of student learning outcomes from the LieCal Project in the United States. *International Journal of Educational Research*, *50*, 117–136. doi:10.1016/j.ijer.2011.06.006

Collins, L. M., Schaefer, J. L., & Kam, C.-M. (2001). A comparison of restrictive and inclusive strategies in modern missing data procedures. *Psychological Methods*, *6*, 330–351. doi:10.1037//1082-989X.6.4.330

Corriveau, K. H., Harris, P. L., Meins, E., Fernyhough, C., Arnott, B., Elliott, L., . . . de Rosnay, M. (2009). Young children's trust in their mother's claims: Longitudinal links with attachment security in infancy. *Child Development*, *80*, 750–761.

Curtin, T. R., Ingels, S. J., Wu, S., & Heuer, R. (2002). *National education longitudinal study of 1988: Base-year to fourth follow-up data file user's manual (NCES 2002-323)*. Washington, DC: U.S. Department of Education, National Center for Educational Statistics. Retrieved from http://nces.ed.gov/surveys/nels88/design.asp

D'Agostino, J. V. (2000). Instructional and school effects on students' longitudinal reading and mathematics achievements. *School Effectiveness and School Improvement*, *11*, 197–235.

Diggle, P., & Kenward, M. G. (1994). Informative drop-out in longitudinal data-analysis. *Applied Statistics-Journal of the Royal Statistical Society Series C*, *43*, 49–93.

Enders, C. K. (2001). The performance of the full information maximum likelihood estimator in multiple regression models with missing data. *Educational and Psychological Measurement*, *61*, 713–740.

Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: Guilford.

Enders, C. K. (2011). Missing not at random models for latent growth curve analyses. *Psychological Methods*, *16*, 1–16. doi:10.1037/a0022640

Follmann, D., & Wu, M. (1995). An approximate generalized linear model with random effects for informative missing data. *Biometrics*, *51*, 151–168.

Grimm, K. J., Steele, J. S., Mashburn, A. J., Burchinal, M., & Pianta, R. C. (2010). Early behavioral associations of achievement trajectories. *Developmental Psychology*, *46*, 976–983. doi:10.1037/a0018878

Hausman, J. A., & Wise, D. A. (1979). Attrition bias in experimental and panel data: The Gary Income Maintenance experiment. *Econometrica*, *47*, 455–473.

Heckman, J. J. (1979). Causal inference and nonrandom samples. *Journal of Educational Statistics*, *14*, 159–168.

Hedeker, D., & Gibbons, R. D. (1997). Application of random-effects pattern-mixture models for missing data in longitudinal studies. *Psychological Methods*, *2*, 64–78.

Jeličić, H., Phelps, E., & Lerner, R. M. (2009). Use of missing data methods in longitudinal studies: The persistence of bad practices in developmental psychology. *Developmental Psychology*, *45*, 1195–1199. doi:10.1037/a0015665

Kenward, M. G. (1998). Selection models for repeated measurements with non-random dropout: An illustration of sensitivity. *Statistics in Medicine*, *17*, 2723–2732.

Kieffer, M. J. (2011). Converging trajectories: Reading growth in language minority learners and their classmates, kindergarten to grade 8. *American Educational Research Journal*, *48*, 1187–1225. doi:10.3102/0002831211419490

Kim, Y.-S., Petscher, Y., Schatschneider, C., & Foorman, B. (2010). Does growth rate in oral reading fluency matter in predicting reading comprehension achievement? *Journal of Educational Psychology*, *102*, 652–667. doi:10.1037/a0019643

Ladd, H. F., & Lauen, D. L. (2009). Status vs. growth: The distributional effects of school accountability policies. In National Center for the Analysis of Longitudinal Data in Eductional Research (Ed.), *Working Paper,* No. 21. Washington, DC: Urban Institute.

Liang, K., & Zeger, S. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*, 13.

Little, R. J. A. (1994). Discussion of Diggle, P., & Kenward, M. G. Informative drop-out in longitudinal data analysis. *Applied Statistics-Journal of the Royal Statistical Society Series C, 43*, 85.

Little, R. J. A. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association*, *90*, 1112–1121.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). Hoboken, NJ: John Wiley.

Marcotte, D. E., Bailey, T., Borkoski, C., & Kienzl, G. S. (2005). The returns of a community college education: Evidence from the National Education Longitudinal Survey. *Educational Evaluation and Policy Analysis*, *27*, 157.

McNeil, M. (2009). Data systems set for crucial fund infusion. *Education Week*, *28*, 13.

McNiece, R., Bidgood, P., & Soan, P. (2004). An investigation into using national longitudinal studies to examine trends in educational attainment and development. *Educational Research*, *46*, 119–136.

McNiece, R., & Jolliffe, F. (1998). An investigation into regional differences in educational performance in the National Child Development Study. *Educational Research*, *40*, 17–30. doi:10.1080/0013188980400102

Metzger, A., Dawes, N., Mermelstein, R., & Wakschlag, L. (2011). Longitudinal modeling of adolescents' activity involvement, problem peer associations, and youth smoking. *Journal of Applied Developmental Psychology*, *32*, 1–9. doi:10.1016/j.appdev.2010.09.005

Molenberghs, G., Beunckens, C., Sotto, C., & Kenward, M. G. (2008). Every missingness not at random model has a missingness at random counterpart with equal fit. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, *70*, 371–388.

Molenberghs, G., Thijs, H., Jansen, I., Beunckens, C., Kenward, M. G., Mallinckrodt, C. H., & Carroll, R. J. (2004). Analyzing incomplete longitudinal clinical trial data. *Biostatistics*, *5*, 445–464.

Montague, M., Enders, C., & Cavendish, W. (2011). Academic and behavioral trajectories for at-risk adolescents in urban schools. *Behavioral Disorders, 36*, 141–156.

Muthén, B. (1998–2004). *Mplus technical appendices*. Los Angeles, CA: Muthén & Muthén.

Muthén, B., Asparouhov, T., Hunter, A. M., & Leuchter, A. F. (2011). Growth modeling with nonignorable dropout: Alternative analyses of the STAR*D antidepressant trial. *Psychological Methods*, *16*, 17–33. doi:10.1037/a0022634

Muthén, B., & Muthén, L. K. (1998–2009). Mplus (Version 5.1) [Computer software]. Los Angeles, CA: Muthén & Muthén.

Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research*, *74*, 525–556.

Rabe-Hesketh, S., & Skrondal, A. (2008). *Multilevel and longitudinal modelling using stata* (2nd ed.). College Station, TX: Stata Press.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Rivkin, S. G. (2007). Value-added analysis and education policy. In National Center for Educational Statistics (Ed.), *Brief*. Washington, DC: Urban Institute.

Roy, J. (2003). Modeling longitudinal data with nonignorable dropouts using a latent dropout class model. *Biometrics*, *59*, 829–836.

Schafer, J. L. (1999). Multiple imputation: A primer. *Statistical Methods in Medical Research*, *8*, 3.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*, 147–177.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence* (644 pp.). New York, NY: Oxford University Press.

StataCorp. (2007). Stata Statistical Software: Release 10. College Station, TX: Author.

ten Have, T. R., Kunselman, A. R., Pulkstenis, E. P., & Landis, J. R. (1998). Mixed effects logistic regression models for longitudinal binary response data with informative drop-out. *Biometrics*, *54*, 367–383.

U.S. Department of Education. (2009). Race to the top program: Executive summary. Retrieved from http://www2.ed.gov/programs/racetothetop/executive-summary.pdf

Verbeke, G., Molenberghs, G., Thijs, H., Lesaffre, E., & Kenward, M. G. (2001). Sensitivity analysis for nonrandom dropout: A local influence approach. *Biometrics*, *57*, 7–14.

Wu, M., & Carroll, R. (1988). Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics*, *44*, 175–188.

Xu, S., & Blozis, S. A. (2011). Sensitivity analysis of mixed models for incomplete longitudinal data. *Journal of Educational and Behavioral Statistics*. Advance online publication. doi:10.3102/1076998610375836

Zhang, J., & Rubin, D. (2003). Estimation of causal effects via principal stratification when some outcomes are truncated by ''death.'' *Journal of Educational and Behavioral Statistics*, *28*, 353.

Authors

BETSY J. FELDMAN is a research scientist at the University of Washington working with Partners for Our Children in the School of Social Work; Partners for Our Children, School of Social Work, University of Washington, Box 359476, Seattle, WA 98195; e-mail: bjfeld@washington.edu. Her research focuses primarily on latent-variable models, especially longitudinal modeling of problem behaviors.

SOPHIA RABE-HESKETH is a professor at the Graduate School of Education and Graduate Group in Biostatistics, University of California, Berkeley, and at the Institute of Education, University of London; 3659 Tolman Hall, Graduate School of Education, University of California, Berkeley, CA 94720-1670; e-mail: sophiarh@berkeley.edu. Her research interests include multilevel and latent variable modeling.

Manuscript received March 30, 2011
First revision received August 25, 2011
Second revision received December 19, 2011
Accepted January 4, 2012