Annu. Rev. Sociol. 2000. 26:441–62

MULTILEVEL MODELING FOR BINARY DATA

Guang Guo
Department of Sociology, University of North Carolina, Chapel Hill, North Carolina 27599; e-mail: guang_guo@unc.edu

Hongxin Zhao
Center for Children and Families, Teachers College, Columbia University, New York, NY 10027

Key Words hierarchical models, contextual analysis, logistic regression, logit, generalized linear mixed models

■ **Abstract** We review some of the work of the past ten years that applied the multilevel logit model. We attempt to provide a brief description of the hypothesis tested, the hierarchical data structure analyzed, and the multilevel data source for each piece of work we have reviewed. We have also reviewed the technical literature and worked out two examples on multilevel models for binary outcomes. The review and examples serve two purposes: First, they are designed to assist in all aspects of working with multilevel models for binary outcomes, including model conceptualization, model description for a research report, understanding of the structure of required multilevel data, estimation of the model via a generally available statistical package, and interpretation of the results. Second, our examples contribute to the evaluation of the approximation procedures for binary multilevel models that have been implemented for general public use.

INTRODUCTION

This article reviews multilevel models for binary outcome variables and their sociological applications. The emphasis of the review is on the practical usage of the models. The review of the applications draws from articles published in *American Sociological Review, American Journal of Sociology*, and *Social Forces* over the past ten years. See DiPrete & Forristal (1994) for an earlier review of multilevel analysis.

The interest in multilevel models for binary data is a natural development of two traditions in sociological analysis. First, sociologists, perhaps more than any other social scientists, are interested in explaining and predicting phenomena that can be characterized by a binary variable. Such phenomena may be the occurrence of discrete events such as dropping out of high school, getting a four-year college education, marrying, giving birth, divorcing, using narcotic drugs, having a business go bankrupt, adopting a new technology, on implementing a new public

0360-0572/00/0815-0441$14.00

policy. Other examples include whether to vote for a Republican or Democratic candidate and whether to say yes or no to an opinion poll. We have counted 85 articles published in *Social Forces*, 61 in *American Sociological Review*, and 39 in *American Journal of Sociology* during 1990 and 1999, which employed standard binary regression techniques in their analysis. Regression models for binary data including logistic regression and probit regression have long been sociologists' standard analytical tools.

The second tradition that prompted an interest in multilevel models is the practice of examining hierarchical social structure. Multilevel models have a natural appeal to sociologists because social structure is often hierarchical. Examples for multilevel social structure are plentiful. In schools, students (level 1) are nested in classes (level 2), and classes are nested in schools (level 3). Individuals are nested in families, and families are nested in communities or neighborhoods. In for-profit and non-profit organizations, employees are the first level and the organization the second. Even prior to the development of the formal statistical methodology for multilevel models, sociologists were engaged in multilevel or contextual analysis. Blalock (1984) reviewed and discussed various theoretical and methodological issues in the literature on multilevel analysis before the statistical models were developed.

Mason et al (1983) were among the first to develop the concepts and methodology for analyzing multilevel data. Further methodological and substantive work by Bryk & Raudenbush (1992) and Goldstein (1987, 1995) has popularized the multilevel models for linear data. It is no surprise that sociologists of education were among the first to apply the methodology to the study of school effects. Using data from the High School and Beyond Study, Gamoran (1992) examined how the characteristics of high school tracking such as selectivity, electivity, inclusiveness, and scope affect students' educational achievement. Roscigno (1998) estimated linear multilevel models of math and reading achievement on race, family/peer influences, class characteristics, and racial composition of school by employing data from the restricted-use National Educational Longitudinal Survey (NELS) and the Common Core of Data. Another common design of multilevel analysis treats metropolitan areas in the United States defined by the Bureau of the Census as level-2 units and individuals as level-1 units. Combining data from the 1% and 5% 1990 Public Use Microdata Samples (PUMS) data from 261 metropolitan areas, Cotter et al (1997) investigated the impact of occupational integration by gender at the metropolitan level on gender earnings equality. Using the same data sources, Cohen (1998) estimated the effects of metropolitan-area black population proportions on earnings inequality between blacks and whites and between men and women. In their longitudinal Beginning School Study in Baltimore, Entwisle et al (1994) examined neighborhood environment as a possible cause of gender gap in math. Xie & Hannum (1996) studied the regional variation in earnings inequality in contemporary urban China, using data at household and city levels.

Considering individuals or some other observations as level-2 units and repeated measures of these level-2 units as level-1 units, we have another application of multilevel linear models: the growth curve model. This model has been used to examine the trajectory of self-reported crime over time among adolescents (Lauritsen 1998), the timing of the influences of poverty on children's cognitive ability and achievement (Guo 1998), the long-term effects of parental divorce on individuals' mental health from age 7 to age 33 (Cherlin et al 1998), and the Kuznets curve or the growth curve of income inequality as measured by the Gini coefficient among US counties from 1970 to 1990 (Nielsen & Alderson 1997).

Social scientists' interest in binary outcome variables and hierarchical social structure made the development of multilevel models for binary data a near certainty. Earlier methodological work on multilevel logit models includes Wong & Mason (1985), Anderson & Aitkin (1985), and Goldstein (1991). Using data from fifteen World Fertility Survey (WFS) countries, Entwisle et al (1986) studied contraceptive behavior of couples as a function of socioeconomic origins at the individual level, of the gross national product per capita (GNP), and of the family planning effort at the country level. Crane (1991) tested the epidemic theory of ghetto and neighborhood effects on dropping out and teenage childearing, drawing data from the 1979 PUMS. Although Crane's analysis does not fall within formal multilevel modeling, his insights into the functional relationship between neighborhood quality and social problems are valuable to multilevel modelers working on the same topic.

More recent years saw an increased number of applications of multilevel models for binary data. Rountree & Land (1996) reported distinctive differences between a general perceived risk of crime and a burglary-specific fear. They based their analysis on a victimization survey collected in Seattle, Washington, in 1990. In the dataset, more than 5000 individuals are clustered into about 300 city-blocks, which are in turn clustered into about 100 census tracts. In an effort to explain the "southern migrant advantage" in family stability, which refers to more stability among black southern families that migrated to northern cities, Tolnay & Crowder (1999) estimated the effects of metropolitan-level distress on urban black family patterns and explored whether group differences in exposure to these contextual conditions can explain the greater stability of migrant families. Their data are from the 1970 Integrated Public Use Microdata Series and 1970 Summary Statistic File Fourth Count. Multilevel data have also been collected in other countries. Using data from Norway, Kalleberg & Mastekaasa (1998) documented and sought to explain the relationships between an organization's size and an individual's interorganizational mobility (i.e., quits and layoffs). Analyzing data collected by the Chinese Academy of Preventive Medicine at the household and village levels, Nee (1996) tested the hypothesis that the shift to a market economy in China caused a redistribution of economic gains among those with political power and those who produce.

WHY MULTILEVEL MODELS?

The multilevel models we review in this article are statistical multilevel models, which allow not only independent variables at any level of a hierarchical structure, but also at least one random effect above level one. In a particular analysis, multilevel modeling offers a number of the following advantages. Some influential sociological work was conceptualized as multilevel analysis but analyzed by traditional models (Hogan & Kitagawa 1985, Sampson 1991, Billy & Moore 1992, Brooks-Gunn et al 1993). Traditional linear or nonlinear models, however, do not enjoy all the advantages we describe.

First, a multilevel model provides a convenient framework for studying multilevel data. Such a framework encourages a systematic analysis of how covariates measured at various levels of a hierarchical structure affect the outcome variable and how the interactions among covariates measured at different levels affect the outcome variable. One of the frequently examined cross-level interaction effects is how the macro context affects the impact of a covariate at the micro level. For example, Entwisle et al (1986) tested the idea that the strength of the effect of maternal education on fertility depends on the characteristics of a country such as gross national product (GNP) and the intensity of family planning efforts.

Second, multilevel modeling corrects for the biases in parameter estimates resulting from clustering. In contrast to the popular belief, ignoring multilevel structure can result in biases in parameter estimates as well as biases in their standard errors. The more highly correlated the observations are within clusters, the more likely that ignoring clustering would result in biases in parameter estimates.

Third, multilevel modeling provides correct standard errors and thus correct confidence intervals and significance tests. When observations are clustered into higher-level units, the observations are no longer independent. Independence is one of the most basic assumptions underlying traditional linear and binary regression models. When the clustering structure in the data is ignored and the independence assumption is violated, the traditional linear and binary models tend to underestimate the standard errors. The following is an intuitive argument for this statement. The observations in the same cluster tend to be more similar in their outcome measures if clustering matters regarding the outcome measures. Similarity within a cluster implies that we can, to some extent, predict the outcome of an observation if we know the outcome of another observation in the same cluster. This suggests that not every observation provides an independent piece of information and that the total amount of information contained in a sample with clustering is less than that in a sample without clustering.

The estimation of standard errors is not merely a technical issue. The size of a standard error can uphold or overturn an important conclusion. Using linear regression techniques and ignoring the fact that the students are grouped into teachers and classes, Bennett (1976) showed that in Great Britain elementary school students benefitted more from a formal style of teaching. The results were widely known and became quite influential until Aitkin et al (1981) demonstrated that, once the grouping of the students is taken into consideration in a multilevel model, the results obtained by Bennett concerning teaching styles were no longer statistically significant.

When all variations at levels higher than one are captured by observed variables, multilevel data can be analyzed by traditional linear or nonlinear models. In such a case, conditional on the observed variables, the observations in the same cluster are no longer dependent, and the standard errors obtained by traditional models are correct.

Fourth, estimates of the variances and covariances of random effects at various levels enable investigators to decompose the total variance in the outcome variable into portions associated with each level. Using the 1987 National Survey of Maternal and Child Health in Guatemala, Pebley et al (1996) modeled a binary variable of whether the child has received a complete set of immunizations as a function of observed variables at the individual, family, and community levels and unobserved variables at the family and community levels. After controlling for observed variables, they showed that the variance due to families is about five times larger than that due to communities.

THE MULTILEVEL LINEAR MODEL

To provide a familiar starting point, we begin with a review of the multilevel linear model. Our review focuses on a few specific multilevel models that sociologists are likely to estimate. For a description of the multilevel linear model in its most general form, see Mason et al (1983), Goldstein (1987, 1995), and Bryk & Raudenbush (1992). We first consider a simple two-level model with a single explanatory variable,

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_j + e_{ij}, \tag{1}$$

where y_{ij} is the outcome variable for the ith unit at level one and the jth unit at level two, β_0 is the intercept, x_{ij} is the explanatory variable, β_1 is its effect, u_j is a random effect accounting for the random variation at level two, and e_{ij} is the level-one random effect. The parameters for the random effects are $E[u_j] = E[e_{ij}] = 0$, $var(u_j) = \sigma_u^2$, $var(e_{ij}) = \sigma_e^2$, $cov(u_j, e_{ij}) = 0$, and $cov(u_j, u_{j'}) = 0$ for $j \neq j'$. The within-cluster or intraclass correlation after controlling for the explanatory variable can be obtained from $\rho = \sigma_u^2/(\sigma_u^2 + \sigma_e^2)$. Equation (1) can also be considered as a random effect model for panel data or a growth curve model. In both cases, i and j would index time points and individuals, respectively, and x_{ij} would be a time-varying covariate. A linear growth model requires that a linear term be added to Equation (1) and a quadratic growth model requires an additional quadratic term.

We next extend the simple two-level model to a three-level model with random coefficients,

$$y_{ijk} = \beta_0 + \beta_1 x_{ijk} + u_{1jk} x_{ijk} + v_{0k} + u_{0jk} + e_{0ijk}, \tag{2}$$

where k indexes level 3, v_{0k} and u_{0jk} are the random intercepts for level three and level two, respectively, x_{ijk} is an observed explanatory variable at level one, and u_{1jk} is x_{ijk}'s random effect at level two. Other parameters of the model include $E[v_{0k}] = E[u_{0jk}] = E[e_{0ijk}] = 0$, $var(v_{0k}) = \sigma_{v0}^2$, $var(u_{0jk}) = \sigma_{u0}^2$,

$var(u_{1jk}) = \sigma^2_{u1}$, $var(e_{0ijk}) = \sigma^2_{e0}$, and $cov(u_{0jk}, u_{1jk}) = \sigma_{u01}$. The model assumes again that the random effects across different levels and the random effects across different clusters in the same level are uncorrelated. More complex models can be constructed by adding more observed variables to Equation (2) and allowing cross-level interactions.

The multilevel models have not only the familiar regression parameters β_0 and β_1 but also the unknown random parameters u_{0jk}, u_{1jk}, and v_{0k}. Viewing the multilevel model as a special case of the mixed model, statisticians mostly estimate the model parameters via the generalized least squares (GLS), which minimizes $(y-X\beta)'V^{-1}(y-X\beta)$. Because of the unknown random parameters in V, however, either maximum likelihood or restricted maximum likelihood is generally used first to estimate the variances and covariances of $\boldsymbol{u}$ and e under the assumption that they are normally distributed (Thompson 1971, Harville 1977, Laird & Ware 1982). Mason et al (1983), Goldstein (1986), Raudenbush & Bryk (1986), and Longford (1987) have also studied the estimation of the multilevel linear models.

MULTILEVEL MODELS FOR BINARY DATA

We first consider a two-level model for binary outcomes with a single explanatory variable. Conceptually, this model is equivalent to model (1) except for the outcome variable. Suppose we have data consisting of students (level one) grouped into schools (level two). We observe y_{ij}, a binary response for student i in school j and x_{ij}, an explanatory variable at the student level. We define the probability of the response equal to one as $p_{ij} = Pr(y_{ij} = 1)$ and let p_{ij} be modeled using a logit link function. The standard assumption is that y_{ij} has a Bernoulli distribution. Then the two-level model can be written as

$$\log[p_{ij}/(1 - p_{ij})] = \beta_0 + \beta_1 x_{ij} + u_j \quad (combined\ model) \tag{3}$$

where u_j is the random effect at level two. Without u_j, (3) would be a standard logistic regression model. Conditional on u_j, y_{ij}s are assumed to be independent. As in the case of multilevel linear models, u_j is assumed to be normally distributed, with the expected value 0 and the variance σ^2_u. Model (3) is often described alternatively in the literature on multilevel models by Equations (4)

$$\log[p_{ij}/(1 - p_{ij})] = \beta_{0j} + \beta_1 x_{ij} \quad (level\ 1\ model) \tag{4}$$

and (5)

$$\beta_{0j} = \beta_0 + u_j \quad (level\ 2\ model). \tag{5}$$

Relative to Equations (4) and (5), Equation (3) is the so-called combined model.

The multilevel model for binary outcomes can also be derived through a latent variable conceptualization. We assume that there exists a latent continuous variable y_{ij}^* underlying y_{ij}. We observe only our binary response variable y_{ij} directly, but not y_{ij}^*. We know, however, $y_{ij}^* > 0$ if $y_{ij} = 1$ and $y_{ij}^* \leq 0$ if $y_{ij} = 1$. A multilevel model for y_{ij}^* equivalent to (3) can be written as

$$y_{ij}^* = \beta_0 + \beta_1 x_{ij} + u_j + e_{ij}. \tag{6}$$

Conditional on the random effect u_j at level two, either a logit multilevel model such as (3) or a probit multilevel model can be derived from (6) depending on whether we assume that e_{ij} in (6) has a standard logistic distribution or a normal distribution. This conceptualization illustrates the close connections between the multilevel models for linear data and those for binary data. Later in this paper, we use this result to calculate intra-cluster correlations for binary data.

Conditional on u_j or assuming that u_j were observed, the conditional density function for cluster j for model (3) is identical to that for the logistic regression

$$f(\boldsymbol{y}_j \mid \boldsymbol{x}_j, u_j) = \prod_{i=1}^{n_j} \frac{\exp[y_{ij}(\beta_0 + \beta_1 x_{ij} + u_j)]}{1 + \exp(\beta_0 + \beta_1 x_{ij} + u_j)}, \tag{7}$$

where $\boldsymbol{y}_j$ and $\boldsymbol{x}_j$, respectively, denote the responses and the explanatory variables in cluster j. The standard strategy for estimating the model parameters in the literature is to assume that u_j, is normally distributed and to integrate out the unobserved random effect u_j,

$$f(\boldsymbol{y}_j \mid \boldsymbol{x}_j) = \int f(\boldsymbol{y}_j \mid \boldsymbol{x}_j, u_j) g(u_j) du_j, \tag{8}$$

where $g(.)$ represents the normal density function. The resulting unconditional density $f(\boldsymbol{y}_j \mid \boldsymbol{x}_j)$, however, does not have a closed expression. Maximum likelihood estimation has to resort to approximation procedures such as numerical integration. Anderson & Aitkin (1985) estimated a model similar to (3) via the EM algorithm, and the solution still requires numerical integration.

Model 3 is almost the simplest possible multilevel model for binary data. Greater challenges arise in the estimation of more general models with multiple random effects. Equation (9) describes a three-level model with a single explanatory variable that has both a fixed effect and a random effect,

$$\log[p_{ijk}/(1-p_{ijk})] = \beta_0 + \beta_1 x_{ijk} + u_{1jk} x_{ijk} + v_{0k} + u_{0jk} \quad (\textit{combined model}), \tag{9}$$

where i, j, and k index, respectively, levels 1, 2, and 3, v_{0k} and u_{0jk} are the random intercepts for level 3 and level 2, respectively, and u_{1jk} is the random coefficient for the explanatory variable x_{ijk}. To fix the idea, we could let levels 1, 2, and 3 represent students, classes, and schools. Again, Equation (9) is the combined model.

Alternatively, model (9) can be described by the multiple equation system

$$\begin{aligned} \log[p_{ij}/(1-p_{ij})] &= \beta_{0jk} + \beta_{1j}x_{ij} \quad \textit{(level 1 model)} \\ \beta_{0jk} &= \beta_{0k} + u_{0jk} \quad \textit{(level 2 model)} \\ \beta_{1j} &= \beta_1 + u_{1j} \quad \textit{(level 2 model)} \\ \beta_{0k} &= \beta_0 + v_{0k} \quad \textit{(level 3 model)}. \end{aligned} \tag{10}$$

The conditional density for (9) is still identical to that for the logistic regression; but with three random effects in the model, the unconditional density is a high-dimensional integral. Numerical integration over the high-dimensional integral is the most straightforward solution. As computing gets less and less expensive, numerical integration has become feasible. Traditionally, however, software packages for multilevel models for binary data tend to use other approximate methods.

Social scientists usually do not need to be concerned with the estimation method of a statistical model because the estimation procedures of most of the statistical models (e.g., linear regression, logit models, and log-linear models) routinely used are well established. But this is not true for multilevel models for binary data. Marginal quasi-likelihood or MQL (Goldstein 1991, Goldstein & Rasbash 1996) and penalized quasi-likelihood or PQL (Breslow & Clayton 1993) are the two prevailing approximation procedures. Both MQL and PQL rely on the Taylor expansion to achieve the approximation. Rodriguez & Goldman (1997) compared four approximation estimation procedures (first-order MQL, second-order MQL, first-order PQL, and second-order PQL) with the maximum likelihood achieved through high-dimensional numerical integration and the method of Gibbs sampling. The second-order MQL and PQL are expected to yield more accurate estimates than the first-order ones because they use some of the second-order terms in the Taylor expansion. The maximum likelihood method based on numerical integration and Gibbs sampling are treated as standards.

Using a sample collected in Guatemala in 1987, Rodriguez & Goldman (1997) estimated a three-level model of complete immunization among children receiving any immunization. Their sample consists of 2159 children from 1595 families in 161 communities. They reported large differences among the estimates from different estimation methods. For instance, the estimated variances of the random effect at the family level are respectively .40, .52, .53, 3.06, 5.38, and 6.76 for MQL-1, MQL-2, PQL-1, PQL-2, ML, and Gibbs sampling. The differences among the estimated fixed effects are also large. The odds ratios of receiving a complete set of immunization among children whose fathers have had a primary education to those whose fathers have had no education are, respectively, 1.34, 1.32, 1.40, 1.26, 1.55, 1.72, and 1.80 for MQL-1, MQL-2, PQL-1, PQL-2, ML, and Gibbs sampling. The same article reported much larger differences across different estimation procedures from an analysis that examined the usage of modern prenatal care. The conclusion is that all approximation methods (MQL-1, MQL-2, PQL-1, and PQL-2) underestimate the random as well as fixed effects and that the underestimations of MQL-1, MQL-2, and PQL-1 are severe.

The evaluation clearly demonstrates that the parameters of the multilevel model for binary data can be severely downwardly biased. Social scientists who work with the models must be aware of the possibility that the estimates they are obtaining from the commercial packages such as HLM and MLn can be seriously inaccurate. Unfortunately, we usually do not know the extent of the biases or even whether the biases are substantively important in a particular analysis. For any particular analysis, the estimates from PQL-2 or even PQL-1 can be accurate enough. The huge differences across different estimation procedures reported by Rodriguez & Goldman are likely to be the exception rather than the rule in typical social science work. In their analysis on the use of prenatal care, the variance of the unobserved family effect estimated by PQL-2 (7.56) is about five time as large as that estimated by PQL-1 (1.56), and the same variance estimated by the Gibbs sampling (112.3) is about 72 times as large as that estimated by PQL-1. In the same analysis, the odds ratio by PQL-2 (6.89) of those whose mothers have had at least a secondary education to those whose mothers have had no education is about 2.6 times as large as that by PQL-1 (2.66) and the same odds ratio by the Gibbs sampling (424) is about 160 times as large as that by PQL-1.

The huge differences across different estimation methods in Rodriguez & Goldman's analysis could be related to two factors. First, the approximation procedures tend not to work well when the observations in a cluster are highly correlated. In the analysis on use of prenatal care, the intrafamily correlation estimated by the Gibbs sampling is 0.98 after controlling for observed variables. This is extremely high. Recall that the correlation between identical twins with respect to IQ is typically estimated to be between 0.65 and 0.75.

Second, in both analyses, a large proportion of family clusters contains one observation. The proportions of single-observation clusters in the two analyses are at least 24% and 32%, respectively. In both analyses, the sizes of family clusters are small, averaging 1.35 for one analysis and 1.60 for the other.

Another way of understanding Rodriguez & Goldman's work is to compare their results with those from previous work on random effects models for multivariate event history data (Guo & Rodriguez 1992, Guo & Grummer-Strawn 1993) and count data (Guo 1996). These random effects models are random-intercept-only or variance-component two-level multilevel models for event history data and count data; but unlike the multilevel models for binary data, the likelihood function for these models has a closed expression. Therefore, no approximation procedure is needed for estimating these models. The applications of these two models generally reported small-to-moderate variances of random effects and small-to-moderate changes between the fixed effects estimated by conventional models and those estimated by multilevel models. The size of the random effects and the changes in the fixed effects are considerably larger (but never approach the magnitudes reported by Rodriguez & Goldman) when the observations in a cluster are highly correlated (twins) than when the observations in a cluster are moderately correlated (siblings).

To further evaluate the approximate estimation procedures, we have estimated a number of two-level logit models using individual/community data from the National Longitudinal Study of Adolescent Health (called Add Health) (Bearman et al 1997) and the generally available estimation procedures. The results from PQL-1 differ very little from those from PQL-2.

THE EXAMPLES

The purpose of the examples is twofold: to illustrate the multilevel models for binary data and to compare a variety of estimation procedures that are generally available to the research community. Add Health is a school-based study of the health-related behaviors of adolescents in grades 7–12 in the United States. Our sample was drawn from the first wave of the in-home survey of the Add Health study. The interviews were carried out from April through December of 1995. Adolescents and block groups are level-1 and level-2 units, respectively. The block group is the smallest geographic area for which the Census Bureau publishes sample data. In 1990, block groups averaged 452 housing units or 1100 individuals. In comparison, the more familiar census tract usually contains between 2500 and 8000 individuals. The block group variables in Add Health have been created from the Census of Population and Housing, 1990.

In this article, we present two examples. The outcome variable for the first example is grade retention coded as 1 if the individual has ever repeated a grade and 0 otherwise. After excluding observations whose information is missing on any of the variables used in the analysis, the working sample consists of 13,900 adolescents. Of the 13,900 adolescents, 2,916 or 21% had ever repeated a grade by the time of the in-home survey in 1995. In the sample, each block group contains about 154 adolescents on average, with 89% of the block groups containing at least 45 adolescents and with no census block containing fewer than 2 adolescents. The basic input dataset for multilevel analysis looks almost identical to the standard input dataset. The only difference is that the multilevel dataset has an ID for block group. This is in addition to the individual ID. This block group ID tells the computer the community membership of each individual.

We first estimated an intercept-only model that predicts the probability of having ever repeated a grade. The multilevel model is described by Equation (11). Readers interested in a multiple-equation description of the model should refer to Equations (3), (4), and (5). The Appendix gives the SAS and MLn codes for the model. See Littell et al (1996: Ch. 11) for more examples on the SAS codes. The estimates of

$$\log[p_{ij}/(1-p_{ij})] = \beta_0 + u_j \tag{11}$$

parameters and standard errors are presented in Table 1. The ML estimate from the standard logit model of the ratio of repeaters to nonrepeaters is $\exp(-1.326) = .265$, which is the same as the sample ratio of 2,916 repeaters to 10,984 nonrepeaters. In comparison, the same ratio is estimated to be $\exp(-1.476) = .228$ from the multilevel model by PQL-2. Failing to take into account the clustering within block groups, the standard logit model has overestimated the ratio by about 15%. The parameters under *random effect* in Table 1 are the estimated variances of the random intercepts. To understand the random effect, one can imagine a unique effect for each block group in addition to the fixed intercept of −1.476, which is the average of all block groups. The addition of the block-group specific effects makes

TABLE 1 Parameters and standard errors of an intercept-only logit model and an intercept-only multilevel model predicting the probability of ever repeating a grade: Add Health

	Logit	Multilevel models				
		SAS	MLn			
	Logit	GLIM-MIX	MQL-1	MQL-2	PQL-1	PQL-2
Fixed effect						
Intercept	−1.326 (.021)	−1.457 (.076)	−1.354 (.070)	−1.471 (.075)	−1.457 (.076)	−1.476 (.077)
Random effect						
Intercept		.428 (.078)	.359 (.063)	.389 (.073)	.426 (.075)	.439 (.077)
Intra-block correlation (ρ)		.11	.10	.11	.11	.12
Deviance		13430.6				
Extradispersion		.976				
−2logL	14280.0					
N	13,900	13,900	13,900	13,900	13,900	13,900

the model more accurate than the fixed intercept only model. In a random effect model, the block-group specific effects are assumed to be distributed normally for the purpose of estimation. The estimate of the random effect does increase as we go from MQL-1 to MQL-2, to PQL-1, and to PQL-2; but the increases are much smaller than those Rodriguez & Goldman (1997) reported, and the increases between PQL-1 and PQL-2 are minuscule.

The estimate from the SAS' GLIMMIX macro is particularly close to, but not the same as, the estimate from MLn's PQL-1. The GLIMMIX macro is based on Wolfinger & O'Connell's (1993) pseudo-likelihood (PL), which is the same as Breslow & Clayton's (1993) PQL-1 except that PL explicitly estimates the extra-dispersion parameter ϕ. PQL-1 sets ϕ to one. In this sense, PL is a slight generalization of PQL-1. By adding an additional parameter ϕ in the conditional variance $\phi[\pi_{ij}(1-\pi_{ij})]/n_{ij}$, the GLIMMIX macro or PL takes into consideration both underdispersion when ϕ is substantially smaller than 1, and overdispersion when ϕ is substantially greater than 1. Underdispersion or overdispersion can lead to unreliable estimates of standard errors. In Table 1, the extradispersion parameter is estimated to be 0.976, and we conclude that there is no evidence for extradispersion.

The intra-block-group correlations are estimated by $\rho = \sigma_u^2/(\sigma_u^2 + \sigma_e^2)$, where $\sigma_e^2 = \pi^2/3$ is the variance of the standard logistic distribution (Table 1). These correlations are computed on the logit scale, that is, the correlation is between y_{ij}^* and $y_{i'j}^*$, where $i \neq i'$ and y_{ij}^* and $y_{i'j}^*$ are the unobserved latent variables described earlier in the latent variable formation of the multilevel model for binary data. According to the estimate from PQL-2 in Table 1, the intra-block-group correlation in terms of the latent variable representing grade retention is 0.12.

Table 2 presents parameters and standard errors from a logit model predicting the probability of having repeated a grade and its equivalent multilevel model. We describe the multilevel model as follows:

$$\begin{aligned}\log[p_{ij}/(1-p_{ij})] = \beta_0 &+ \beta_1\, age_{ij} + \beta_2 female_{ij} + \beta_3\, black_{ij} \\ &+ \beta_4\, income_{ij} + \beta_5\, education_{ij} + \beta_6\, below15k_j \\ &+ u_{1j}\, income_{ij} + u_{2j} below15k_j + u_{0j}. \qquad (12)\end{aligned}$$

See Appendix for the SAS and MLn programs for the model. We have included six observed covariates in the model with five at the individual level and one at the block group level. The observed covariate, gender, is coded as 1 if the adolescent is female and 0 if the individual is male. Ethnicity is coded as 1 if the adolescent is African American and 0 otherwise. Mother's education is length of schooling in years. Family income is in thousands of US dollars. Proportion <\$15,000 is the proportion of the households in a block group with an annual income of less than \$15,000. Age is in years measured at the time of the survey.

The multilevel model has one random intercept and two random coefficients, with one for family income and the other for proportion below \$15,000. We were unsuccessful in obtaining estimates through MLn's MQL-2 and PQL-2 for this model. Neither converged after more than 2,000 iterations. We came across the same problems when estimating a similar model using college plan as the outcome variable. The second-order approximation procedures seem to have difficulty converging when estimating multiple random coefficients.

The differences in parameter estimates between MQL-1 and PQL-1 appear rather small. The estimates from GLIMMIX are again very similar to those from MLn's PQL-1 except for the random coefficient for proportion <\$15,000. The estimates from the standard logit model differ moderately from those from the multilevel models. The standard logit model has overestimated the effects of African American and proportion <\$15,000 and underestimated the effect of age. The standard logit model has also underestimated the standard error of the effect of proportion <\$15,000.

We now interpret the estimates from GLIMMIX. The results from GLIMMIX have been obtained after taking into account unobserved block-group specific random effects. The parameters of observed variables can be interpreted much the same way as those from the standard logit model. Thus, everything else being equal, female adolescents are about $100\% - \exp(-.661)*100 = 48.4\%$ less likely to have repeated a grade than male adolescents. Similarly, the likelihood of having repeated a grade for those adolescents with an annual family income of \$50,000 is about $100\% - \exp[-.011(50-15)]*100 = 32\%$ lower than those with an annual family income of \$15,000. This fixed effect of family income is the average effect of income across block groups. One can imagine a block-group-specific effect of

TABLE 2 Parameters and standard errors of a logit model and a multilevel model (with one random intercept and two random coefficients) predicting the probability of ever repeating a grade: Add Health

	Logit	Multilevel models		
		SAS	MLn	
	Logit	GLIM-MIX	MQL-1	PQL-1
Fixed effect				
Intercept	−3.614 (.258)	−4.000 (.268)	−3.921 (.269)	−4.000 (.272)
Female	−.647 (.045)	−.661 (.045)	−.638 (.045)	−.661 (.046)
African-American	.433 (.052)	.308 (.066)	.301 (.066)	.310 (.067)
Mother education	−.156 (.009)	−.146 (.009)	−.144 (.009)	−.147 (.009)
Family income	−.009 (.001)	−.011 (.001)	−.009 (.002)	−.011 (.001)
Age	.281 (.014)	.305 (.014)	.294 (.014)	.305 (.014)
Proportion <$15,000	1.215 (.146)	1.056 (.193)	1.070 (.190)	1.059 (.193)
Random effect				
Intercept		.224 (.059)	.239 (.060)	.222 (.054)
Family income		.00005 (.00002)	.00006 (.00005)	.00005 (.00002)
Proportion <$15,000		.083 (.319)	.026 (.217)	.046 (.220)
Deviance		12054.2		
Extradispersion		.968		
−2logL	12656.0			
N	13,900	13,900	13,900	13,900

family income that serves as a correction for the fixed average slope of family income. The variance of the block-group specific effects is estimated to be .00005.

Many researchers prefer to interpret the model in terms of predicted probabilities rather than odds ratios. The predicted probability for adolescent i in block group j is

$$\hat{p}_{ij} = \exp(\boldsymbol{x}_{ij}\hat{\beta} + \hat{u}_{1j}\, income_{ij} + \hat{u}_{2j}\, below15k_j + \hat{u}_{0j})/ \\ (1 + \exp(\boldsymbol{x}_{ij}\hat{\beta} + \hat{u}_{1j}\, income_{ij} + \hat{u}_{2j}\, below15k_j + \hat{u}_{0j})), \quad (13)$$

TABLE 3 Parameters and standard errors of an intercept-only logit model and an intercept-only multilevel model predicting the probability of having a college plan: Add Health

	Logit	Multilevel models				
		SAS	MLn			
	Logit	GLIM-MIX	MQL-1	MQL-2	PQL-1	PQL-2
Fixed effect						
Intercept	.991 (.019)	.952 (.057)	.913 (.053)	.956 (.054)	.951 (.057)	.960 (.057)
Random effect						
Intercept		.225 (.045)	.193 (.036)	.198 (.037)	.224 (.041)	.229 (.042)
Intra-block correlation (ρ)		.06	.06	.06	.06	.07
Deviance		16184.7				
Extradispersion		.988				
−2logL	16750.9					
N	13,849	13,849	13,849	13,849	13,849	13,849

where

$$x_{ij}\hat{\beta} = \hat{\beta}_0 + \hat{\beta}_1\, age_{ij} + \hat{\beta}_2\, female_{ij} + \hat{\beta}_3\, black_{ij} + \hat{\beta}_4\, income_{ij} + \hat{\beta}_5\, education_{ij} + \hat{\beta}_6\, below15k_j.$$

An analyst can obtain $\hat{u}_{0j}$, $\hat{u}_{1j}$, and $\hat{u}_{2j}$ from GLIMMIX by adding the option "solution" in the line that specifies the random component of the model. Predicted probabilities can be calculated in a number of ways. For instance, for adolescents who are from block group 403($j = 403$), aged 15, white, with an annual family income of $30,000, with a mother having a high school education, and living in a community with 20% of the families having an income of less than $15,000, the predicted probability of having repeated a grade is estimated to be about 8% for females and 14% for males. The random effects $\hat{u}_{0j}$, $\hat{u}_{1j}$, and $\hat{u}_{2j}$ for block group 403 are −.505, .0014, and −.094, respectively. Another common practice is to simulate the predicted probabilities. Suppose we are again interested in gender differences. The first step in the simulation is to compute $\hat{p}_{ij}$ for all *ij*s. The computation is done twice. Both times, all individuals are allowed to retain their own characteristics except gender. The first time, gender is set to male and the second time female. The second step is to compute the average over $\hat{p}_{ij}$s. This is also done twice, once for males and once for females. The predicted probabilities for males and females can then be compared.

TABLE 4 Parameters and standard errors of a logit model and a multilevel model (with one random intercept and two random coefficients) predicting the probability of having a college plan: Add Health

	Logit	Multilevel models		
		SAS	MLn	
	Logit	GLIM-MIX	MQL-1	PQL-1
Fixed effect				
Intercept	.684	.898	.930	.896
	(.219)	(.227)	(.228)	(.229)
Female	.421	.423	.414	.423
	(.039)	(.039)	(.039)	(.039)
African-American	.286	.365	.364	.364
	(.049)	(.061)	(.061)	(.061)
Mother education	.132	.121	.118	.121
	(.008)	(.008)	(.008)	(.008)
Family income	.005	.005	.004	.005
	(.001)	(.001)	(.001)	(.001)
Age	−.121	−.122	−.119	−.121
	(.011)	(.012)	(.012)	(.012)
Proportion <$15,000	−.277	−.600	−.628	−.601
	(.135)	(.189)	(.188)	(.189)
Random effect				
Intercept		.115	.119	.114
		(.031)	(.034)	(.031)
Family income		.00002	.00004	.00002
		(.00001)	(.00001)	(.00001)
Proportion <$15,000		.264	.247	.251
		(.236)	(.234)	(.233)
Deviance		15473.8		
Extradispersion		.984		
−2logL	15940.5			
N	13,849	13,849	13,849	13,849

Assuming that the three random effects in (12) are uncorrelated, we can calculate the intra-block-group correlation using

$$\rho(income_{ij},\ lowinc_j) = \frac{\sigma_{u0}^2 + \sigma_{u1}^2\ income_{ij}^2 + \sigma_{u2}^2\ lowinc_j^2}{\sigma_{u0}^2 + \sigma_{e0}^2 + \sigma_{u1}^2\ income_{ij}^2 + \sigma_{u2}^2\ lowinc_j^2}, \tag{14}$$

where lowinc is proportion <$15,000. Now the intra-cluster correlation ρ is a function of family income and neighborhood poverty. For example, the intra-block-group correlation ρ is estimated to be .12 for adolescents who have an annual family income of \$70,000 and who live in a block group with 0% of families having an income below \$15,000. The ρ is estimated to be .08 for adolescents whose family income is \$15,000 and who live in a community with 75% of families having an income below \$15,000. We obtained all the parameter estimates necessary for the calculation from Table 2.

The deviance statistic D reported by the GLIMMIX macro can be used to carry out a likelihood ratio test for hypothesis testing. The deviance is defined as

$$D = 2(\ln f(y \mid \tilde{\theta}) - \ln f(y \mid \hat{\theta})), \tag{15}$$

where $\ln f(y \mid \tilde{\theta})$ is the loglikelihood for the saturated model or the observed data and $\ln f(y \mid \hat{\theta})$ is the loglikelihood for the model of interest. Suppose we have two models of interest: model 1 with p_1 parameters and model 2 with p_2 parameters. The two models are nested with $p_2 > p_1$. To test if model 2 has improved the explanatory power of model 1 significantly, we use $D_1 - D_2$, the difference between the deviance for model 1 and that for model 2, as a likelihood ratio statistic, which has an approximate χ_2 distribution with $p_2 - p_1$ degrees of freedom. A likelihood ratio test of the model in Table 2 against the model in Table 1 ($13430.6 - 12054.2 = 1376.4$ with 8 degrees of freedom) shows that the addition of the six fixed effects and two random coefficients has significantly improved the fit of the model.

In Tables 3 and 4, we present the results from the second example that runs parallel to the first. Also based on Add Health, the second example uses having a college plan as the binary outcome variable. Having a college plan was constructed from the responses to the question "do you want to attend college?" "Yes" is coded as one and all the other responses are coded as 0. Our analysis sample consists of 13,849 adolescents, 9,792 or 71% of whom are coded as one. The comments we can make from the second example concerning different estimation methods are generally similar to those we made based on the first example, except that the coefficient of proportion <\$15,000 in the second example (Table 4) is seriously underestimated by the standard logit model.

SUMMARY, CONCLUSIONS, AND FINAL REMARKS

Many social scientists have applied multilevel models for binary data. We have reviewed some of the work published over the past ten years. We attempt to provide a brief description of the hypothesis tested, the hierarchical data structure analyzed, and the multilevel data source for each piece of work we have reviewed. We have also reviewed the technical literature and worked out two examples on multilevel models for binary outcomes. The review and examples serve two purposes. First, they are designed to assist in all aspects of working with multilevel models for binary outcomes, including model conceptualization, model description for a research report, understanding of the structure of required multilevel data, estimation of the model via a generally available statistical package, and interpretation of the results. Second, our examples contribute to the evaluation of the approximation procedures for binary multilevel models that have been implemented for general public use.

Our examples have further demonstrated the tendency for the standard logit model to seriously bias the parameter estimates of observed covariates when analyzing multilevel data. The differences in estimates across the different estimation procedures for the binary multilevel models, however, are much smaller in our examples than those reported by Rodriguez & Goldman (1995, 1997). The differences between PQL-1 and PQL-2 in our example are minimal (Tables 1 and 3). We have estimated other models than those presented in Tables 1 and 3, in which we included a number of observed covariates and a random intercept, and we succeeded in getting estimates from MQL-2 and PQL-2. The differences between PQL-1 and PQL-2 are always very small. This is consistent with Goldstein & Rasbash's (1996) observation that in the more common case where variances in a multilevel model for binary data do not exceed about 0.5, the first order PQL model can be expected to perform well. Our conclusion is that while an analyst should always be aware of the possibility that his or her estimates are seriously biased, MLn's PQL-1 and PQL-2 and SAS' GLIMMIX are likely to be adequate for most of the projects undertaken in social sciences. Additional work is needed to determine more precisely the relationship between bias size and factors such as level of within-cluster correlation, proportion of clusters that has a single observation, average size of clusters, number of clusters, and so on. Before an entirely reliable and practical estimation procedure is developed, such a systematic evaluation of the approximate procedures will yield more precise recommendations for researchers.

We have shown through the examples how the multilevel logit model can be interpreted in terms of odds ratios and predicted probabilities, how the intra-cluster correlation can be calculated for a simple two-level random-intercept model and a model with a random intercept as well as random coefficients, and how hypothesis testing can be carried out for nested models using deviance provided by SAS GLIMMIX. These results will remain useful even if an entirely new estimation procedure is to replace all the procedures currently used.

Finally, our review of sociological applications of multilevel analysis reveals a few inadequacies regarding the presentation of analysis results. First, it is essential that researchers get across the basic structure of the hierarchical data and model. State clearly and in a prominent place how many levels the data have, what they are, how many units there are at level 2 and level 3, how many observations each level-2 or level-3 unit has on average, and what is the proportion of level-2 units that has only one observation. Equations are very useful in conveying the structure of hierarchical data and the model if they are used correctly. Define the elements including the subscripts in the equations immediately before or after the equations. Equations convey much more substantive information about the analysis in multilevel analysis than, say, in analysis that uses the techniques of sample selection. Second, researchers need to pay special attention to the presentation of random effects. It is a good idea to always present random effects in the table in which the results are presented even if they are not statistically significant. Random effects are necessarily part of a multilevel model. In the table, label random effects clearly. Let the audience know which random effect(s) is a random intercept and the level at which it varies. Label which random effect(s) is a random coefficient and the level at which it varies. A more precise presentation of random effects is to use, in the table, the mathematical symbols that are used in the equations for the multilevel model. Third, researchers probably want to present the intra-cluster correlation to give a sense of how the observations are correlated within clusters. Fourth, because of the lack of entirely reliable estimation methods, researchers should report the specific estimation method used, not just the statistical package, so that the analysis can be replicated and compared if necessary.

ACKNOWLEDGMENTS

We acknowledge the support of the W. T. Grant Foundation to Guo through the Faculty Scholars Program and a postdoctoral fellow grant to Zhao from NICHD (HD08189). We are grateful to Jianmin Wang, Jeremy Reynolds, and Erin Leahey for helpful comments.

APPENDIX

(1) SAS codes for the intercepts-only model

```
%INCLUDE 'D:/DATA/GLM/GLMM612.SAS' /NOSOURCE;
%GLIMMIX(DATA=TEMP3, PROCOPT=COVTEST, STMTS=%STR(
  CLASS COMMID;
  MODEL REPEAT=/SOLUTION;
  RANDOM INTERCEPT/SUB=COMMID;),
  ERROR=BINOMIAL, LINK=LOGIT);
```

The intercepts-only multilevel model described in (10) is estimated by SAS' GLIMMIX macro contained in glmm612.sas. The first line shows the location of the SAS macro. The second line shows the location of the SAS data file. After the first two lines, the syntax of the SAS codes is the same as that in PROC MIXED. COMMID is block group ID and CLASS COMMID declares COMMID as a categorical variable. REPEAT is the binary outcome variable and SOLUTION is an option asking SAS to print out the estimates. INTERCEPT after RANDOM indicates that the model has a random intercept. SUB=COMMID tells SAS that the observations within the same block group are subject to the same random effect. The last line of the SAS codes indicates that we are estimating a multilevel model or generalized linear mixed model for a binary outcome using the logit transformation.

(2) MLn codes for the intercepts-only model

$$
\begin{aligned}
&repeat_{ij} \sim Binomial(denom_{ij}, \pi_{ij})\\
&repeat_{ij} = \pi_{ij} + e_{0ij}\ bcons*\\
&logit(\pi_{ij}) = \beta_{1j}\ cons\\
&\beta_{1j} = \beta_1 + u_{1j}\\
&[u_{1j}] \sim N(0, \Omega_u) : \Omega_u = \left[\sigma^2_{u1}\right]\\
&bcons* = \ bcons[\pi_{ij}(1-\pi_{ij})/denom_{ij}]^{0.5}\\
&[e_{0ij}] \sim (0, \Omega_e) : \Omega_e = [1]
\end{aligned}
$$

The analyst does not need to create the whole program shown here. The window version of MLn provides the general setting and the analyst fills in the rest. Before "writing" the program, the analyst must create three columns of ones, name them "bcons", "cons", and "denom", and add them to the input dataset. The right-hand side of the third equation defines the model. The intercepts-only model has only one β, but this β changes by j because it has a random component u_{1j} as defined in line four. The parameter β_1 is the fixed intercept. In line five, the random effect u_{1j} is assumed to have a normal distribution with $E[u_{1j}] = 0$ and $var(u_{1j}) = \sigma^2_{u1}$.

(3) SAS codes for the multiple random effects model

```
%INCLUDE 'D:/DATA/GLM/GLMM612.SAS' /NOSOURCE;
%GLIMMIX(DATA =TEMP3, PROCOPT=COVTEST, STMTS=%STR(
 CLASS COMMID;
 MODEL REPEAT=FEMALE BLACK AGE MEDUC INCOME LOW-INC/
   SOLUTION;
 RANDOM INTERCEPT INCOME LOWINC/SUB=COMMID;),
 ERROR=BINOMIAL, LINK=LOGIT);
```

In line four, we have added the six observed covariates to the model. In line five, we have added two random coefficients, one for INCOME and the other LOWINC both varying by block groups.

(4) MLn codes for the multiple random effects model

$$
\begin{aligned}
&repeat_{ij} \sim Binomial(\ denom_{ij}, \pi_{ij})\\
&repeat_{ij} = \pi_{ij} + e_{0ij}\ bcons*\\
&logit(\pi_{ij}) = \beta_{1j}\ cons + \beta_2\ age_{ij} + \beta_{3j}\ lowinc_j + \beta_4\ black_{ij}\\
&\quad + \beta_{5j}\ income_{ij} + \beta_6\ meduc_{ij} + \beta_7\ male_{ij}\\
&\beta_{1j} = \beta_1 + u_{1j}
\end{aligned}
$$

$$\beta_{3j} = \beta_3 + u_{3j}$$

$$\beta_{5j} = \beta_5 + u_{5j}$$

$$\begin{bmatrix} u_{1j} \\ u_{3j} \\ u_{5j} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} \sigma_{u1}^2 & & \\ 0 & \sigma_{u3}^2 & \\ 0 & 0 & \sigma_{u5}^2 \end{bmatrix}$$

$$bcons* = \; bcons[\pi_{ij}(1 - \pi_{ij})/denom_{ij}]^{0.5}$$

$$[e_{0ij}] \sim (0, \Omega_e) : \Omega_e = [1]$$

In line three, we have added six observed covariates and their coefficients, among which those for lowinc and income are allowed to vary by j. Lines four, five, and six are level-2 equations defining β_{1j}, β_{3j}, and β_{5j}. Line seven defines the random components of level-2 models. In this particular model, we set the covariances of the random effects (u_{1j}, u_{3j}, and u_{5j}) to zeros under the assumption that these random effects are uncorrelated.

Visit the Annual Reviews home page at www.AnnualReviews.org

LITERATURE CITED

Aitkin M, Anderson D, Hinde J. 1981. Statistical modeling in school effectiveness studies (with discussion). *J. R. Statis. Soc. A* 149:1–43

Anderson DA, Aitkin M. 1985. Variance component models with binary response: interviewer variability. *J. R. Statist. Soc. B* 47:203–10

Bearman PS, Jones J, Udry JR. 1997. *The National Longitudinal Study of Adolescent Health: Research Design* [WWW Document]. URL: *http://www.cpc.unc.edu/projects/addhealth/design.html*

Bennett N. 1976. *Teaching Styles and Pupil Progress*. London: Open Books

Billy JOG, Moore DE. 1992. A multilevel analysis of marital and nonmarital fertility in the U.S. *Soc. Forces* 70:977–1011

Blalock H. 1984. Contextual-effects models: theoretical and methodological issues. *Annu. Rev. Sociol.* 10:353–72

Breslow NE, Clayton DG. 1993. Approximate inference in generalized linear mixed models. *J. Am. Statist. Assoc.* 88:9–25

Brooks-Gunn J, Duncan GJ, Klebanov PK, Sealand N. 1993. Do neighborhoods influence child and adolescent development? *Am. J. Socio.* 99:353–95

Bryk AS, Raudenbush SW. 1992. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Newbury Park, CA: Sage

Cherlin AJ, Chase-Lansdale PL, McRae C. 1998. Effects of parental divorce on mental health throughout the life course. *Am. Sociol. Rev.* 63:239–49

Cohen P. 1998. Black concentration effects on black-white and gender inequality: multilevel analysis for U.S. metropolitan areas. *Soc. Forces* 77:207–29

Cotter DA, JoAnn D, Hermsen JM, Kowalewski BM, Vanneman R. 1997. All women benefit: the macro-level effect of occupational integration on gender earnings equality. *Am. Sociol. Rev.* 62:714–34

Crane J. 1991. The epidemic theory of ghettos and neighborhood effects on dropping out and teenage childbearing. *Am. J. Sociol.* 96:1226–59

DiPrete TA, Forristal JD. 1994. Multilevel models: methods and substance. *Annu. Rev. Sociol.* 20:331–57

Entwisle B, Mason WM, Hermalin AL. 1986. The multilevel dependence of contraceptive use on socioeconomic development and family planning program strength. *Demography* 23:199–216

Entwisle DR, Alexander KL, Olson LS. 1994. The gender gap in math: its possible origins in neighborhood effects. *Am. Sociol. Rev.* 59:822–38

Gamoran A. 1992. The variable effects of high school tracking. *Am. Sociol. Rev.* 57:812–28

Goldstein H. 1987. *Multilevel Models in Educational and Social Research.* London: Griffin.

Goldstein H. 1991. Nonlinear multilevel models with an application to discrete response data. *Biometrika* 78:45–51

Goldstein H. 1995. *Multilevel Statistical Models.* London: Arnold; New York: Halstead. 2nd ed.

Goldstein H, Rasbash J. 1996. Improved approximations for multilevel models with binary responses. *J. Roy. Statist. Soc. A.* 159:505–13

Guo G. 1998. The timing of the influences of cumulative poverty on children's cognitive ability and achievement. *Soc. Forces* 77:257–88

Guo G. 1996. Negative multinomial regression models for clustered event counts. *Sociol. Methodol.* 26:113–32

Guo G, Rodríguez G. 1992. Estimating a multivariate proportional hazards model for clustered data using the em algorithm, with an application to child survival in Guatemala. *J. Am. Statist. Assoc.* 87:969–76

Guo G, Grummer-Strawn L. 1993. Child mortality among twins in developing countries. *Pop. Stud.* 47(3):1–16

Harvill DA. 1977. Maximum likelihood approaches to various component estimation and to related problems. *Journal of the American Statistical Association.* 72:320–38

Hogan DP, Kitagawa EM. 1985. The impact of social status, family structure, and neighborhood on the fertility of black adolescents. *Am. J. Sociol.* 90:825–55

Kalleberg A, Mastekaasa A. 1998. Organizational size, layoffs, and quits in Norway. *Soc. Forces* 76:1243–73

Laird NM, Ware JH. 1982. Random-effects models for longitudinal data. *Biometrics* 38:963–74

Lauritsen J. 1998. The age-crime debate: assessing the limits of longitudinal self-report data. *Soc. Forces* 77:127–55

Littell RC, Milliken GA, Stroup WW, Wolfinger RD. 1996. *SAS System for Mixed Models.* Cary, NC: SAS Inst.

Mason WM, Wong GM, Entwistle B. 1983. Contextual analysis through the multilevel linear model. *Sociol. Methodol.* 13:72–103

Nee V. 1996. The emergence of a market society: changing mechanisms of stratifcation in China. *Am. J. Sociol.* 101:908–49

Nielsen F, Alderson AS. 1997. The Kuznets curve and the great U-turn: income inequality in U.S. counties, 1970 to 1990. *Am. Sociol. Rev.* 62:12–33

Pebley AR, Goldman N, Rodriguez G. 1996. Prenatal and delivery care and childhood immunization in Guatemala: do family and community matter? *Demography* 33:231–47

Rodriguez G, Goldman N. 1995. An assessment of estimation procedures for multilevel models with binary response. *J. R. Statist. Soc. A.* 158:73–89

Rodriguez G, Goldman N. 1997. *Multilevel models with binary response: a comparison of estimation procedures.* Paper Pres. Pop. Assoc. Am. Annu. Meet. Washington, DC, March 27–29, 1997

Roscigno VJ. 1998. Race and the reproduction of educational disadvantages. *Soc. Forces* 76:1033–60

Rountree PW, Land KC. 1996. Perceived risk versus fear of crime: empirical evidence of conceptually distinct reactions in survey data. *Soc. Forces* 74:1353–76

Sampson RJ. 1991. Linking the micro- and macrolevel dimensions of community social organization. *Soc. Forces* 70:43–64

Tolnay SE, Crowder KD. 1999. Regional origin and family stability in northern cities: the role of context. *Am. Sociol. Rev.* 64: 97–112

Wolfinger R, O'Connell M. 1993. Generalized liner models: a pseudo-likelihood approach. *J. Statist. Comput. Simul.* 48:233–43

Wong GY, Mason WM. 1985. The hierarchical logistic regression model for multilevel analysis. *J. Am. Statist. Assoc.* 80: 513–23

Xie Y, Hannum E. 1996. Regional variation in earnings inequality in reform-era urban China. *Am. J. Sociol.* 101:950–92