Published in final edited form as:

Multivariate Behav Res. 2009 January 1; 44(1): 28–58.

Marginal and Random Intercepts Models for Longitudinal Binary Data With Examples From Criminology

Jeffrey D. Long

University of Minnesota

Rolf Loeber

University of Pittsburgh

David P. Farrington

Cambridge University

Abstract

Two models for the analysis of longitudinal binary data are discussed: the marginal model and the random intercepts model. In contrast to the linear mixed model (LMM), the two models for binary data are not subsumed under a single hierarchical model. The marginal model provides group-level information whereas the random intercepts model provides individual-level information including information about heterogeneity of growth. It is shown how a type of numerical averaging can be used with the random intercepts model to obtain group-level information, thus approximating individual and marginal aspects of the LMM. The types of inferences associated with each model are illustrated with longitudinal criminal offending data based on $N = 506$ males followed over a 22-year period. Violent offending indexed by official records and self-report were analyzed, with the marginal model estimated using generalized estimating equations and the random intercepts model estimated using maximum likelihood. The results show that the numerical averaging based on the random intercepts can produce prediction curves almost identical to those obtained directly from the marginal model parameter estimates. The results provide a basis for contrasting the models and the estimation procedures and key features are discussed to aid in selecting a method for empirical analysis.

Recent advances in statistical methods have made a number of tools available for the analysis of nonnormal longitudinal data. Among these tools are longitudinal extensions of the logistic regression model for binary data (see, e.g., Molenberghs & Verbeke, 2006, chap. 5). A practical problem for applied research is that multiple classes of such models exist and it is not always clear which model should be used. In contrast to methods based on the normality assumption, each binary model class has its unique parameters with specific interpretations. It is important for the applied researcher to understand the key features in order to use the model appropriate for the desired type of inference.

This article addresses fundamental issues of applied analysis using the two most common classes of longitudinal models for binary data: the marginal model and the random intercepts model. To provide motivation and illustration, we consider two real data examples from the Pittsburgh Youth Study (PYS), which is a longitudinal study of inner-city boys that began in 1987 (Loeber et al., 2003). The data consist of the yearly binary classification of violent

offending from official records and self-report collected on the same cohort of males over a 22-year period. Consistent with the theme of the real data examples, we relate aspects of the models to growth curve analysis in criminology research. Discussion focuses on model interpretation and the types of research questions that can be addressed. Incommensurability of the parameters is considered and a numerical averaging method based on the fitted random intercepts model is suggested to address this issue. To provide familiar context, we contrast the models for binary data with the normal-based linear mixed model for longitudinal data.

LINEAR MIXED MODEL

Longitudinal methods based on the normality assumption have the advantage that individual-level and group-level information is subsumed under the same model. As an example, consider one of the most popular models for longitudinal data, the linear mixed model (LMM; Laird & Ware, 1982). Suppose y_{ij} is the response score for the i th person ($i = 1, \dots, N$) at the j th time point ($j = 1, \dots, n_i$), and y_i is a $n_i \times 1$ individual time series vector. The LMM is

$$y_i = X_i \beta + Z_i b_i + e_i, \quad (1)$$

where X_i is a $n_i \times p$ design matrix, β is a $p \times 1$ vector of unknown regression weights known as fixed effects, Z_i is a $n_i \times q$ known matrix linking b_i to y_i , b_i is a $q \times 1$ vector of unobserved random effects, and e_i is a $n_i \times 1$ unobserved random error vector. We assume normality for the random effects and error, $b_i \sim \mathcal{N}(0, G)$, and $e_i \sim \mathcal{N}(0, \sigma_e^2 I_{n_i})$. We also assume the random effects and error are uncorrelated and that the diagonal elements of G are nonnegative so that the LMM has a hierarchical interpretation (Verbeke & Molenberghs, 2001, chap. 3).

In this article we are primarily concerned with models that have no covariates of change other than time transformations. Additionally, the criminal offending data to be considered later have nonlinear trajectories, so we will also be preoccupied with quadratic polynomial growth curves. For such models, the first column of X_i is a ones vector and the remaining columns are the $p - 1 = 2$ polynomial time transformations, t_{ij} and t_{ij}^2 , where t_{ij} is some time metric score (e.g., age). With this background, the quadratic polynomial model based on Equation (1) can be written as

$$y_{ij} = (\beta_0 + b_{i0}) + (\beta_1 + b_{i1}) t_{ij} + (\beta_2 + b_{i2}) t_{ij}^2 + e_{ij}. \quad (2)$$

Here $X_i = Z_i$, so every fixed effect has an associated random effect, but this need not be the case. To ensure proper scaling of the random effects (i.e., zero expected value), we require $q \leq p$.

Equation (2) is an individual-level model as it is a growth curve for the i th person (unit), with the k th person parameter being $\tilde{\beta}_{ik} = \beta_k + b_{ik}$. The collection of $\tilde{\beta}_{ik}$ defines a person's growth curve and e_{ij} indicates the scatter of points about a person's curve. Heterogeneity in growth among people is indexed by the variance of $\tilde{\beta}_{ik}$, for example, $\text{Var}(b_{i0})$ indicates variability in intercepts.

In addition to individual-level information, the LMM also provides information about the group-level growth curve. Taking the expectation of the individual-level parameters in Equation (2),¹ we obtain the group-level parameters, $E(\beta_k + b_{ik}) = \beta_k$. Thus, the β parameters can be interpreted as averages (possibly weighted) of individual-level parameters. Taking the expectation of Equation (2), we have

$$E(y_{ij}) = \beta_0 + \beta_1 t_{ij} + \beta_2 t_{ij}^2, \quad (3)$$

which can be interpreted as the group-level growth curve that results from averaging the individual growth curves. Equation (3) is referred to as a marginal model because $E(y_{ij})$ is analogous to a marginal mean.²

The marginal model of Equation (3) provides insight into the individual-level model as it shows that the collection of random effects for a person represents the individual's deviation from the group-level growth curve. When a person's collection of b_{ik} are all zero, the individual's growth curve is identical to the group-level growth curve. When at least one random effects is not zero, the individual's growth curve in some respect deviates from the group growth curve.

MODELS FOR BINARY RESPONSES

The last section illustrates that the LMM subsumes both an individual-level model and a marginal model. The ability to aggregate over the individual-level model to obtain the marginal model is based on the normality assumption and the fact that y_{ij} is a linear function of the predictors. In the case of binary data, the unity of the individual-level and marginal model is lost because the relation between the response and the predictors is nonlinear. Consequently, different approaches must be taken to develop the two types of models. We begin with the marginal model for binary data as it is a direct extension of the logistic regression model for cross-sectional data.

Marginal Model

Let us for the moment consider a single cross section of time. Suppose for the cross section we have the binary response variable, y_i , which can take the value of 0 or 1. In the context of criminology research, we might have $y_i = 1$ if the i th person committed a criminal offense and $y_i = 0$ if he or she did not. Because of the difficulties in basing a model directly on y_i (see Agresti, 2002, chap. 4), the traditional logistic regression model is based on $E(y_i)$, which can be interpreted as a population proportion (population mean) or as a probability. For example, given the scoring scenario already stated, $E(y_i)$ is the proportion of those committing a criminal offense in the population, known as prevalence (Farrington, 1986). The proportion can also be interpreted as the probability of an offense and denoted as $E(y_i) = Pr(y_i = 1)$.

Now consider the longitudinal case in which the binary response is collected at repeated cross sections. Similar to the LMM, the binary response is denoted as y_{ij} and we have $E(y_{ij}) = Pr(y_{ij} = 1)$. The marginal model is based on this quantity and has the general form

$$E(y_{ij}) = Pr(y_{ij} = 1) = \frac{\exp(\mathbf{X}_i \gamma)}{1 + \exp(\mathbf{X}_i \gamma)}, \quad (4)$$

where $\mathbf{X}_i$ is a $n_i \times p$ design matrix like that in the LMM, and γ is a $p \times 1$ vector of unknown group-level regression coefficients. Similar to the marginal LMM, there are no random effects and Equation (4) is a growth curve for the marginal probabilities. Consistent with the traditional cross-sectional logistic regression model, the right-hand side of Equation (4) “squeezes” the response scale so that the growth curve of probabilities stays within the admissible bounds of 0 and 1.

¹The expectation is analogous to an average if there are no missing data and a weighted average if there are missing data.

²The marginal model is also known as the population-averaged model.

When the time metric is age in years and $Pr(y_{ij} = 1)$ represents prevalence, the growth curve is known as the *age-crime curve* in criminology research (Farrington, 1986). If the beginning of measurement is early in the lives of the participants, the age-crime curve can provide important practical and theoretical information regarding the development of antisocial behavior (Blumstein, Cohen, & Farrington, 1998; Loeber & Hay, 1997; Loeber & LeBlanc, 1990).

One of the most replicated findings regarding the age-crime prevalence curve is that it is nonlinear increasing from late childhood, peaking in adolescence, and dropping afterward (Fergusson & Horwood, 2002; Lacourse, Nagin, Tremblay, Vitaro, & Claes, 2003; Nagin, Farrington, & Moffitt, 1995). Given this backdrop, we primarily discuss marginal models with nonlinear growth curves. The quadratic polynomial is one example:

$$Pr(y_{ij}=1) = \frac{\exp(\gamma_0 + \gamma_1 t_{ij} + \gamma_2 t_{ij}^2)}{1 + \exp(\gamma_0 + \gamma_1 t_{ij} + \gamma_2 t_{ij}^2)}. \quad (5)$$

With the age-crime curve it is common for γ_2 to be less than zero so that the growth curve is concave with respect to time (i.e., $\cap$). The upper panel of Figure 1 shows an example of Equation (5) denoted as a dashed line (the solid lines are discussed later).

The curve in Figure 1 shows asymptote-like behavior as it approaches zero. This is advantageous for modeling age-crime relations as prevalence is expected to be zero at ages prior to the ascent up to the peak and zero at ages after the descent down from the peak (Nagin et al., 1995). The asymptotic behavior does not occur in the logit metric because the model is linear in the predictors,

$$\text{logit}\{Pr(y_{ij}=1)\} = \log_e \left\{ \frac{Pr(y_{ij}=1)}{1 - Pr(y_{ij}=1)} \right\} = \gamma_0 + \gamma_1 t_{ij} + \gamma_2 t_{ij}^2, \quad (6)$$

which produces a parabola as shown in the lower panel of Figure 1, denoted as a dashed line (the solid curves are discussed later).

When the age-crime curve is considered, it is often better to work with the probability metric than the logit metric as the former can be interpreted as prevalence and allows zero values in actual data that one can graph against a prediction curve (odds and logits based on zero prevalence are not defined). The probability and logit metric have a nonlinear but monotonic relation so that among other things, the time value at which the peak occurs is the same (see the upper and lower panels of Figure 1).

In contrast to the LMM, the marginal model for binary data cannot be derived from an individual-level model. An implication is that there is no information available about individual heterogeneity in growth analogous to the variance of the random effects in the LMM. For this reason, the marginal model is most appropriate when the applied researcher wants to make group-level inferences. For example, Equation (5) is useful for modeling change in prevalence over time because this is a group-level probability. The marginal model is not appropriate for studying heterogeneity in growth, nor is it appropriate for tracking changes of individual-level probabilities. For these types of inferences, we must use a random effects model for binary data, which we now discuss.

Random Intercepts Model

The random effects model is an individual-level model and as such describes change for separate people (units). Similar to the LMM, individual growth curves are specified by the inclusion of random effects. However, in the random effects model for binary data only individual-specific information is available. Individual growth curve parameters cannot be averaged over to yield group-level parameters, though as we see, a type of averaging can be used with the individual probabilities.

Multiple random effects can be included in models for binary data, but it is common to include only random intercepts. A practical reason for this is that models with more than random intercepts can be very difficult to estimate with current numerical methods (Diggle, Heagerty, Liang, & Zeger, 2002, chap. 11). Perhaps more compelling is that binary data often provide relatively little information about individual heterogeneity beyond intercept variability. An illustration is provided by the criminal offending data presented later. Individuals in the sample were randomly drawn from the community, and as a consequence the majority of participants had no criminal offenses for the duration of the data collection. The preponderance of time series with zero values provided limited information about individual differences in growth trajectories and models with more than random intercepts could not be estimated (i.e., the computer algorithms did not converge). What information did exist about heterogeneity was probably captured by random intercepts. Because this situation seems to be the rule rather than the exception (Fitzmaurice, Laird, & Ware, 2004, chap. 12), only random intercepts models are considered here.

The random intercepts model in effect “individualizes” the marginal model by incorporating the random intercepts term in its probability definition. Suppose d_i is the random intercepts value for the i th person. Then the expected value of the random intercepts model is $E(y_{ij} | d_i)$, which can also be written as $Pr(y_{ij} = 1 | d_i)$. This is a conditional probability as opposed to the unconditional probability in the marginal model. In the context of the age-crime curve, $Pr(y_{ij} = 1 | d_i)$ is interpreted as the probability that the i th individual will commit an offense at the j th age. In contrast, prevalence is the probability that an offense will occur at the j th age, which is concerned with the aggregate of offenders and nonoffenders, meaning no individual probability statements are made.

The general form of the random intercepts model is

$$E(y_{ij}|d_i) = Pr(y_{ij}=1|d_i) = \frac{\exp(d_i + \mathbf{X}_i\delta)}{1 + \exp(d_i + \mathbf{X}_i\delta)}, \quad (7)$$

where $\mathbf{X}_i$ is as previously defined, β is a vector of unknown individual-level regression coefficients, and we assume $d_i \sim N(0, \sigma_d^2)$. Similar to the LMM, σ_d^2 is an important parameter as it indexes heterogeneity in growth. σ_d^2 is especially useful when examining time-static covariates of change (e.g., sex). The proportional decrease in σ_d^2 when a time-static covariate is added to the model can be used as a measure of effect size.

As a particular example of Equation (7), consider the quadratic polynomial model

$$Pr(y_{ij}=1|d_i) = \frac{\exp(d_i + \delta_0 + \delta_1 t_{ij} + \delta_2 t_{ij}^2)}{1 + \exp(d_i + \delta_0 + \delta_1 t_{ij} + \delta_2 t_{ij}^2)}. \quad (8)$$

The logit form expresses $\text{logit}\{Pr(y_{ij}|d_i=1)\} = \log_e \left\{ \frac{Pr(y_{ij}=1|d_i)}{1 - Pr(y_{ij}=1|d_i)} \right\}$ as a linear function of the predictors (the details are left to the reader).

In contrast to the marginal model, Equation (8) is a growth curve for an individual. Though the δ_k are referred to as fixed effects, they are not analogous to the fixed effects in the LMM. The δ_k cannot be thought of as the group-level parameters that result when individual growth curves are averaged. This is true even for the δ_k ($k > 0$) that are not associated with the random intercepts term. Rather, all the δ_k are inextricably bound to the random intercepts term and represent the growth curve for individuals who have various values of d_i . This point is amplified in the next section.

Contrasting the Models

As an illustration of the random intercepts model and its relation to the marginal model, consider some simulated criminal offending data based on Equation (8). Suppose the random intercepts scores for 3 individuals are $d_1 = -2$, $d_2 = 0$, and $d_3 = 2$, and the parameter values are $\delta_0 = -25$, $\delta_1 = 3$, and $\delta_2 = -0.1$. Substituting these values into Equation (8) and solving for values of t_{ij} in the age range of $[0, 30]$, the individual probability of offending, $Pr(y_{ij} = 1 | d_i)$, can be computed for each person at each age. Curves of these values are shown in the upper panel of Figure 1 denoted as solid lines.

The value of a person's random effect determines the height of his or her growth curve with $d_3 = 2$ having the curve with the highest peak and $d_1 = -2$ having the curve with the lowest peak. Each person has the same rate of change because δ_1 and δ_2 are constant among individuals. Because $E(E(y_{ij} | b_i)) = E(y_{ij})$, the marginal growth curve is obtained by averaging over $Pr(y_{ij} = 1 | d_i)$ at each j . This illustrates yet another interpretation of the marginal probability as the average (perhaps weighted) of the individual-level probabilities (note this does *not* apply

to the growth parameters). In this case, we compute $E(y_{ij}) = \frac{1}{3} \sum_{i=1}^3 E(y_{ij} | d_i)$ for each age to obtain the mean growth curve denoted by the dashed line in the upper panel of Figure 1. This is the marginal model growth curve and if we know the values of γ_k , the group-level probabilities can be computed directly with the right-hand side of Equation (5) by substituting values of t_{ij} .

The lower panel of Figure 1 shows the same growth curves but in the logit metric. The growth curves are vertically shifted, illustrating the random intercepts differences. Variability among the intercepts is indexed by σ_d^2 , which in this case is equal to $\frac{1}{3}(-2^2 + 0^2 + 2^2) = \frac{8}{3}$. σ_d^2 is in the logit metric and expression in the probability metric is involved as the variance changes with the probabilities. It has been noted that even seemingly small values of σ_d^2 can reflect important heterogeneity (Duchateau & Janssen, 2005).

The δ_k of the random effects model are always person-specific, even when considered in isolation. To see this, consider the case of $d_i = 0$, in which the random intercepts term drops out of the right-hand side of Equation (8), yielding

$$Pr(y_{ij}=1|d_i=0) = \frac{\exp(\delta_0 + \delta_1 t_{ij} + \delta_2 t_{ij}^2)}{1 + \exp(\delta_0 + \delta_1 t_{ij} + \delta_2 t_{ij}^2)}. \quad (9)$$

Though the right-hand side of this formula has the exact form as the marginal model it is merely the growth curve for an individual with a random intercepts value of zero, not any type of averaged curve. This is verified in Figure 1 where it can be seen that the $d_i = 0$ growth curve does not correspond to the marginal growth curve.

The only time Equation (9) will have a group-level interpretation and be at least approximately equivalent to a marginal model is when $\sigma_d^2 = 0$. In this case all individuals share the same random intercept value and the same individual growth curve. The concept is formalized in the following approximate relation.

Suppose the estimation procedures described in the next section are used for each model. Then it can be shown (Zeger, Liang, & Albert, 1988) that the parameters of the marginal and random intercepts models have the approximate relation

$$\delta \approx \gamma \sqrt{1 + 0.346\sigma_d^2}. \quad (10)$$

Equation (10) indicates that when the variance of the random intercepts is zero, the random effects parameters (δ_k) are approximately equivalent to the marginal model parameters and thus have a group-level interpretation. However, when the variance is greater than zero, the parameters are not approximately equivalent and the δ_k retain their individual-level interpretation.

Note that Equation (10) refers to the entire vector of parameters. This verifies the point made earlier that none of the δ_k have a group-level interpretation, whether or not they are associated with the random intercepts. Equation (10) is also helpful for explaining why the parameter estimates of the different models can be so different in empirical analysis. The approximate relation indicates the δ_k are larger in absolute value than the γ_k whenever the variance of the random intercepts is greater than zero.

Numerical Averaging Based on the Fitted Random Intercepts Model

The incommensurability of the random effects model and the marginal model is inconvenient for applied analysis. The marginal model provides inferences about the group-level growth curve but has the limitation of not providing information about individual variability. The random intercept model provides this information in the form of σ_d^2 , but only individual-specific conclusions can be made. It would be desirable to have both group-level and individual-level information available in a single model.

In principle, marginal information can be obtained from the random intercepts model by taking the expectation over the random intercepts. In terms of Equation (7) this is

$$E(y_{ij}) = E[E(y_{ij}|d_i)] = E\left\{\frac{\exp(d_i + \mathbf{X}_i\delta)}{1 + \exp(d_i + \mathbf{X}_i\delta)}\right\}. \quad (11)$$

It turns out the last equation on the right is very problematic to work with because it does not have an exponential form or an analytic solution (Fitzmaurice et al., 2004, chap. 13). An alternative is to work with the sample version of $E(y_{ij}|d_i)$ and obtain marginal information from numerical averaging in a similar fashion to what was done in constructing the mean curve in Figure 1 (Lindsey, 1999, chap. 7; Molenberghs & Verbeke, 2006, chap. 16).

Numerical averaging is carried out in the following steps. First, based on empirical data the random intercepts parameters are estimated, including an estimate of σ_d , denoted as $\widehat{\sigma}_d$. Then a large number of simulated random effects values, $\widehat{d}_i$, are drawn from a normal distribution with mean zero and standard deviation $\widehat{\sigma}_d$. We substitute $\widehat{d}_i$, $\widehat{\delta}$, and t_{ij} (where $t_{ij} \equiv t_j$) into Equation (7) and compute the estimated individual probabilities for each participant. For the j th cross section (e.g., age), the average of the estimated individual probabilities is computed yielding the estimated marginal probability, $\widehat{E}(y_{ij})$.

More formally, suppose S is the number of simulated $\widehat{d}_i$ and the selected model is the random effects quadratic polynomial of Equation (8). Then,

$$\widehat{E}(y_{ij}) = \sum_{i=1}^S \frac{\exp(\widehat{d}_i + \widehat{\delta}_0 + \widehat{\delta}_1 t_j + \widehat{\delta}_2 t_j^2)}{1 + \exp(\widehat{d}_i + \widehat{\delta}_0 + \widehat{\delta}_1 t_j + \widehat{\delta}_2 t_j^2)}. \quad (12)$$

As we see in the real data example, Equation (12) can yield marginal probabilities that are nearly identical to those obtained directly from the marginal model.

Therefore, at least some limited marginal information can be obtained from the random intercepts model. The information is limited in that the averaging method does not yield parameters that have group-level interpretations. To obtain approximate group-level parameters, one could use Equation (10) based on the parameter estimates, that is,

$\widehat{\delta}_k = \widehat{\gamma}_k \sqrt{1 + 0.346 \widehat{\sigma}_d^2}$. However, it turns out this is not a very accurate approximation in many instances when $\widehat{\sigma}_d^2 > 0$ (Neuhaus, Kalbfleisch, & Hauck, 1991).

An additional advantage of the numerical averaging is that the variability of the simulated individual probabilities can be examined. For example, boxplots of the estimated probabilities at each timepoint can be constructed. Unlike the conditional variance, $\text{Var}(y_{ij}) = E(y_{ij})(1 - E(y_{ij}))$, the variability based on the estimated probabilities is grounded in the individual-level model.

EXAMPLES: CRIMINAL OFFENDING DATA

To illustrate the methods we present two real data examples. In terms of growth curve form, we consider the quadratic models presented earlier, that is, the marginal model of Equation (5) and the random intercepts model of Equation (8). We consider binary classifications of violent offending from official records and self-report. The data set of the former has a very low proportion of missing data whereas the data set of the latter has a relatively high proportion of missing data. Because missing data are almost inevitable in age-crime analysis spanning many years (Brame & Piquero, 2003), we thought it important to examine the methods under two different missing data scenarios.

METHOD

Participants

Participants were the oldest cohort of males ($N = 506$) in the PYS who were measured beginning in 1987. Details regarding participant characteristics and methods of data collection can be found in Loeber, Farrington, Stouthamer-Loeber, Moffitt, and Caspi (1998) and Loeber et al. (2003). The official records measure was a yearly binary classification of conviction of a violent offense (e.g., strong-arming, attacking to seriously hurt or kill), with 0 = no conviction, 1 =

conviction. The official records measure was culled from various court sources and spanned 9 to 30 years of age. All participants had complete data up to 18 years of age, and 488 participants (96%) had complete data for all ages. Nearly all missing data was due to death and thus was of the dropout variety (i.e., not intermittent).

The self-report measure indexed whether the person committed a violent act at a given age (0 = did not commit the act, 1 = committed the act), based on interviews with the participant, his parent(s), and his teacher(s). The cohort had self-report data for the ages 12–27 years. In contrast to official records, self-report missing was mostly intermittent with the vast majority due to inability to interview the respondents, though 4% of the participants dropped out due to death as previously noted. None of the participants had complete data and 88% had data for half or fewer of the timepoints (i.e., less than or equal to eight timepoints). In terms of cross sections, 50% of the data was missing at 12 years, an average of 12% was missing for 13–24 years, and an average of 76% was missing for 25–27 years.

Data Analysis

The marginal model was estimated using generalized estimating equations (GEE; Liang & Zeger, 1986). GEE is a quasi-likelihood estimation method that accounts for pairwise dependency among the repeated measures through a “working” correlation matrix. For consistency with the random effects model, a working correlation matrix with an exchangeable structure was used, which had a single correlation parameter, ρ . Liang and Zeger show that under certain conditions GEE yields consistent and asymptotically normal estimates of γ and standard errors can be based on a robust “sandwich” estimator, which was used in the analysis. Parameter estimates were obtained with the R function `geeglm` of `geepack` (Halekoh, Höjsgaard, & Yan, 2006), but we note that similar analysis can be performed using SAS PROC GENMOD and other statistical software.

The random intercepts model was estimated using maximum likelihood (ML) methods. Analogous to the compound symmetry structure of LMM, the dependencies due to repeated measures are accounted for through the variance of the random intercepts under the assumption, $d_i \sim N(0, \sigma_d^2)$. Likelihood estimation is generally more difficult for the binary random effects model than the normal-based LMM due to the lack of analytic integral formulas (Diggle et al., 2002, chap. 11). Integrals must be approximated with computer-intensive methods, one of the most popular being Gaussian quadrature, which was used here. In Gaussian quadrature, the accuracy of the integral approximation is controlled by the number of quadrature points, which was 50 in the analysis.³ Based on the normality assumption, asymptotic and nonasymptotic standard errors can be estimated with the former being presented later. The R function and package `glmmML` were used for the analysis (Broström, 2008), but we note that similar estimation can be performed with SAS PROC NLMIXED and other statistical software.

Additional issues regarding estimation are discussed in the following section. R syntax for the analysis is presented in the Appendix along with code for the numerical averaging previously described. Similar SAS syntax is available from Jeffrey D. Long.

RESULTS

Observed prevalence was computed for each variable at each age and is shown in Figures 2 and 3. Official records observed prevalence is denoted by the filled circles in Figure 2 and similarly for self-report in Figure 3. The vertical axes of the two figures have different scales

³A number of quadrature points were tried and the parameter estimates appeared to become stable when greater than 10 point, were used. The value 50 was arbitrarily chosen to provide a high degree of accuracy in the results.

to assess differences (or lack thereof) in the prediction curves to be discussed later. The peak value of the self-report prevalence was higher than official records, which is consistent with research showing official sources of offending tend to underestimate prevalence whereas self-report does not (Espiritu, Huizinga, Crawford, & Loeber, 2001; Loeber et al., 1998).

The GEE results for the marginal model are shown in Table 1 and the ML results for the random intercepts model are shown in Table 2. The left-hand side of the tables shows the official records results and the right-hand side shows the self-report results. Each Wald-type Z value is computed as the parameter estimate divided by its estimated standard error (SE) and is used to test the null hypothesis that the parameter in question is zero. For the case of σ_d , the Z test tends to yield too large a p value (it tends to be “conservative”) because the null parameter value is at the lower limit and the sampling distribution is not normal (Stram & Lee, 1994). To address this problem, a larger Type I error rate such as $\alpha = 0.10$ (rather than $\alpha = 0.05$) may be used when evaluating the variance component (Berkhof & Snijders, 2001).

The last row of Table 1 contains the estimates of the constant value in the working correlation matrix (ρ). The estimate for official records was not statistically significant ($|Z| < 1.64 < 1.96$), whereas that for self-report was significant. The last row of Table 2 contains estimates of σ_d , which were statistically significant for both official records and self-report.

As for the growth curve parameter estimates, Table 1 and Table 2 indicate that $|\widehat{\delta}_k| > |\widehat{\gamma}_k|$ for every k . The standard errors associated with the $\widehat{\delta}_k$ also were larger except for the case of the official records quadratic coefficient. The Z values of the growth curve parameter estimates in both tables were relatively large ($|Z_k| > 6$, $p_k < 0.0001$), indicating rejection of the null hypothesis in question.

Predicted prevalence curves were computed based on the parameter estimates. For the marginal model, the GEE estimates were substituted into Equation (5) along with integer values of t_j ($t_{ij} = t_j$). For the random intercepts model, the numerical averaging method of Equation (12) was used based on the ML estimates and $S = 10,000$ simulated d_i values. The predicted prevalence curves for each model are shown in Figure 2 (official records) and Figure 3 (self-report). In both figures, the prediction curve based on the GEE estimates is denoted by a dashed line whereas the curve based on the ML estimates and numerical averaging is denoted by a solid line.

DISCUSSION

The real data examples illustrate analysis based on the marginal model and the random intercepts model. GEE was used to estimate the former whereas ML based on Gaussian quadrature was used to estimate the latter. Two longitudinal binary response variables were analyzed: violent offending from official records and self-report measured on the same $N = 506$ participants over the ages of 9–30 years (self-report spanned 12–27 years). The official records offending data was complete for almost all the participants whereas the self-report offending had incomplete data for all the participants.

Features of the two models can be seen by comparing Tables 1 and 2. Consistent with Equation (10), each random intercepts parameter estimate ($\widehat{\delta}_k$) was larger in absolute value than the corresponding marginal parameter estimate ($\widehat{\gamma}_k$). Note this difference was true for the linear and quadratic terms illustrating that the introduction of the random intercept causes an “offset” of the entire vector of the $\widehat{\delta}_k$ relative to the marginal model. Interestingly, the larger $\widehat{\delta}_k$ were accompanied by larger standard error estimates in all but one case (official records quadratic), which produced Z values that were similar across the models. Such consistency in Z values has

been shown in other contexts (Jansen, Beuchens, Molenberghs, Verbeke, & Mallinckrodt, 2006; Neuhaus, 1993; Neuhaus et al., 1991; Zeger et al., 1988).

Though Z values are provided for all the growth curve parameter estimates in Tables 1 and 2, it should be noted that only the Z value of the quadratic coefficient are interpreted. The quadratic coefficient is the only coefficient invariant to arbitrary linear transformations of the time scale (Griepentrog, Ryan, & Smith, 1982; Morrell, Pearson, & Brant, 1997). For both models and both variables, the estimate of the quadratic parameter ($\widehat{\gamma}_2$ or $\widehat{\delta}_2$) had a negative sign, indicating all the growth curves were concave with respect to age. The corresponding Z values were relatively large, indicating the need for concave growth curves as opposed to linear growth curves. Additional analysis not presented indicated that cubic terms were not required.

Perhaps the most interesting result of the analysis was the agreement of the prevalence prediction curves, shown in Figures 2 and 3. The curves were nearly identical for both official records and self-report. The agreement indicates that in at least some situations, group-level information like that obtained from the marginal model can also be obtained from the random intercepts model using numerical averaging.

The agreement of the prevalence prediction curves is especially intriguing because the two variables had different missing data conditions and the estimation procedures treat missing data in different ways. As explained by Liang and Zeger (1986), GEE uses a type of pairwise deletion and thus requires the assumption of missing completely at random (MCAR) for valid inference (for a discussion of missing data mechanisms, see Little & Rubin, 1989). In contrast, the ML estimation of the random intercepts model requires the less stringent assumption of missing at random (MAR; Jansen et al., 2006). Assumptions about missing data are especially pertinent in age-crime research as some proportion of dropout is likely if the repeated measures span more than a few years (Brame & Piquero, 2003; Manski & Nagin, 1998). Because criminal offending is correlated with incarceration and premature death (Loeber & Stouthamer-Loeber, 1998), dropout is often predictable from the responses at previous timepoints meaning that MAR might be a more plausible assumption than MCAR.

With the aforementioned in mind, we computed prevalence aggregating over ages 9–18 for those who dropped out and those who did not (all participants had complete data for 9–18 years of age). The dropout group had a prevalence rate of 0.03 compared with a rate of 0.01 for those with complete data. Thus, dropout was correlated with prevalence, indicating missingness could have been MAR for official records. We hasten to add that there is no formal test of the MAR assumption and the true state of affairs is unknown for this analysis. Despite the possibility of MAR, the rate of dropout was very low (4%) and the sample size relatively large ($N = 506$), which may explain why the different treatments of missing data yielded similar results. Analytic and simulation studies suggest that GEE and ML can yield similar results under such conditions (Fitzmaurice, Laird, & Rotnitsky, 1993; Touloumi, Babiker, Pocock, & Darbyshire, 2001).

Self-report had a substantially greater proportion of missing data. The missing data mechanism might have been a combination of MAR and MCAR, with the latter dominating. Because self-report was collected on the same participants as official records, the small proportion of dropout beginning at 18 years (4%) was possibly MAR. The much larger proportion of nondropout missingness (up to 76% at some ages) was due to the inability to contact informants. Ability to contact may be influenced by quasi-random factors such as informant illness, and thus, an MCAR mechanism might have been responsible for the majority of missingness. When the missing mechanism is MCAR and there is a large sample size, pairwise methods and ML can yield similar results even for relatively large proportions of missing data (Shin, Davison, & Long, 2009). This might explain the agreement of predicted prevalence among the methods

for the self-report data. Further research is needed, perhaps in the form of simulation studies, to help determine situations in which GEE and ML with numerical averaging might yield substantially different results.

Additional differences in the estimation methods are illustrated by the results in the last row of Tables 1 and 2. The last row of Table 1 shows the estimates of ρ , and the last row of Table 2 shows the estimates of σ_d . The estimate of ρ for official records was not statistically significant whereas the estimate of σ_d was significant. This result highlights an important interpretation difference between the models in this analysis.

Both the marginal model as realized in GEE and the random intercepts model account for dependencies due to repeated measures through a single parameter. But the underlying model for each method is very different. A key feature of the GEE method is that ρ is treated as a nuisance parameter as the working correlation matrix makes no attempt at providing a full or even accurate specification of the multivariate distribution of the repeated measures (Crowder, 1995; Fitzmaurice, 1995). The working correlation matrix is a means of accounting for dependencies without having to specify the large number of parameters required when basing the marginal model on discrete multivariate distributions, such as the multinomial (Molenberghs & Verbeke, 2006, chap. 7).

In contrast to ρ of the marginal model, σ_d of the random intercepts model is integral, providing an index of individual heterogeneity and a basis for the specification of the full underlying multivariate distribution (Larsen, Petersen, Budtz-Jorgensen, & Endahl, 2000; Neuhaus, 2001). For official records, the values of $\hat{\rho}$ and $\hat{\sigma}_d$ and their associated Z values reflect an important difference in the results. The relatively small value of $\hat{\rho}$ and its associated Z value suggests that the parameter might be eliminated from the model, resulting in a diagonal working correlation matrix. Such a matrix implies the dependencies between timepoints can be ignored, meaning the traditional cross-sectional logistic regression model might be adequate.⁴

Based on the result for $\hat{\rho}$, we analyzed the official records data as a traditional logistic regression using the R function glm. We obtained the γ -parameter estimates $\hat{\gamma} = [-21.4736, 2.0403, -0.0579]'$, which were nearly identical to the GEE estimates in Table 1. In addition, we obtained the standard error estimates, $SE_{\hat{\gamma}} = [2.5152, 0.2871, 0.0081]'$, which were also very close to the GEE estimates. These results indicate that from the GEE viewpoint, there is very little difference in statistical information when the repeated measures of the official records are completely ignored. This is consistent with analytic and simulation studies that show the GEE model may have no advantage over the traditional logistic regression model for small values of ρ (Fitzmaurice, 1995; McDonald, 1993; Sutradhar & Das, 1999).

Turning to the ML results for official records (Table 2), the value of $\hat{\sigma}_d$ and its associated Z value suggest statistically significant heterogeneity of intercepts and correlation between timepoints. This is at odds with the GEE results as it suggests that the traditional logistic regression model is not appropriate for the data. Though the model that generated the data is unknown, we feel the random intercepts model has greater appeal in this case. One problem with the GEE model is that the possible value of ρ is limited by the values of the marginal probabilities (Chaganty & Joe, 2004; Demidenko, 2004, chap. 7). Because the observed official records prevalence values were relatively low ($\max \hat{E}(y_{ij}) < 0.04$), this may have caused the small value of $\hat{\rho}$ rather than a lack of true correlation. Longitudinal data by nature induces

⁴If the working correlation matrix is an identity matrix (not just diagonal), then GEE is identical to the traditional logistic regression model estimated with ML (Fitzmaurice et al., 2004, chap. 10).

correlation between timepoints. The random intercepts model accounts for this explicitly without constraints dictated by the growth curve probabilities.

WHICH MODEL SHOULD BE USED?

Having introduced and discussed two distinct models for binary longitudinal data, a natural question arises as to which model should be used in practice. One answer to this question is to analyze the data with both models as was done in this article (for an example in a social science journal see Pellegrini & Long, 2007). One benefit of using both models is that group- and individual-level information is provided. The mean growth curve parameter estimates are available directly and so is the estimate of the variance of the intercepts. Another benefit is the analyst obtains an idea of the robustness of results under different estimation methods and different missing data assumptions. Consistency of results across the models suggests we might have greater confidence in their validity relative to the situation in which only one or the other method is used.

There are some pitfalls in presenting results based on both models. A justification for the inclusion of both models is usually required, which can be extensive depending on the level of sophistication of the audience. Perhaps more problematic is the discussion that might be required if the results disagree. As we have seen, issues and assumptions regarding the models and their estimation must be discussed to account for differences in results. Such discussion might require substantial manuscript space.

In the event the analyst wants to use a single method, the key features of the models and their estimation procedures should be considered. A summary of these features appears in Table 3. Some of these features have already been discussed, but it is worthwhile to highlight the potential advantages and disadvantages of the methods so that an informed decision can be made.

Marginal Model and GEE

An advantage of the marginal model is that it provides a group-level growth curve, which is often of interest in social science research. For example, prevalence rather than individual probabilities is typically examined in criminology research as the age-crime curve is used to make inferences about aggregates of people (Farrington, 1986). Because GEE provides estimates of the group-level growth curve, the estimates can be directly substituted into Equation (5) to compute the estimated prevalence as shown in Figures 2 and 3. The GEE estimates can also be used to compute such substantively interesting quantities as the instantaneous rate of change in prevalence at a particular timepoint and the age at which the peak of the curve occurs. For the latter, we take the first derivative of Equation (5) with respect to t_{ij} , set it to zero, and solve for t_{ij} , which yields $t_{ij} = \frac{-\gamma_1}{2\gamma_2}$. The GEE estimates may be substituted into this equation to find the age at the peak of the predicted prevalence curve. Doing so with the GEE estimates in Table 1 yields age 17.6 years for official records and age 16.8 years for self-report.

On the negative side of things, the marginal model has been criticized because its growth curve is based on a series of dependent but unconnected cross-sections (Lindsey, 1999, chap. 11; Lindsey & Lambert, 1998; Neuhaus, 2001). The cross sections are unconnected as there are no individual growth curves and no means of tracking a person over time. Because it is not possible to locate people in relation to the prevalence at any timepoint, individuals are simply units over which to aggregate. Not only is this inconsistent with the hierarchical formulation of the LMM but it may also call into question the benefit of having a longitudinal design. If the goal is to estimate prevalence for several cross sections of age, then it is more efficient to draw random samples from different age cohorts rather than follow the same individuals as

they age. In fact, a number of age-crime studies have relied on this type of cross-sectional design, for example, Sampson and Laub (2003).

As we saw with the official records variable, the GEE results can be nearly identical to traditional logistic regression that completely ignores the repeated measures (i.e., treats them as independent). Therefore, from the GEE perspective, a longitudinal design may have no advantage over a cross-sectional design, especially when the repeated measures have weak dependence. An applied researcher may be uncomfortable with a model and estimation procedure that do not necessarily take advantage of the longitudinal nature of their data. On the other hand, GEE does have advantages when the dependencies are stronger, such as greater efficiency (Chaganty & Joe, 2004). Furthermore, nonstatistical considerations, e.g., history effects, are usually of greater importance in the selection of a longitudinal design than statistical criteria. Therefore, the possible equivalence or near equivalence of the GEE results to traditional logistic regression may be irrelevant.

On a blatantly practical level, the marginal model has the advantage that it is relatively easy to estimate with GEE. The GEE method uses a type of iterative weighted least squares, which provides a quick and stable algorithm. GEE does not require any special demands on the user other than the specification of the growth curve and the structure of the working correlation matrix. The latter is usually selected from among a small number of alternatives (e.g., exchangeable, autocorrelation) built into statistical software. In addition, GEE requires no distributional assumptions (e.g., normality) and its estimators have desirable large-sample properties such as consistency and normality (assuming some weak regularity conditions and correct specification of the growth curve; Liang & Zeger, 1986).

As mentioned, GEE does require MCAR for valid inferences with missing data, making it less attractive than ML methods. To address this problem, versions of GEE have been suggested that use weights based on the probability of dropout (e.g., Robins, Rotnitzky, & Zhao, 1994). A difficulty with these weighting methods is they generally require correct specification of both the growth curve and the dropout process. If either is not correctly specified, the weighting can produce more biased estimates than regular GEE (Beunckens, Sotto, & Molenberghs, 2008). A more promising solution is to use GEE with multiple imputation (MI). This amounts to an analysis assuming MAR as MI is performed under this assumption and the GEE analysis is based on complete data (with no assumptions regarding missing data required). There is evidence based on simulation studies that this approach works well in terms of providing relatively low bias and relatively high precision (Beunckens et al., 2008; Jansen et al., 2006).

Another possible disadvantage of GEE is that it provides no formal statistical index of global fit.⁵ This means that model selection is less flexible than with ML methods. Specifically, there is no statistical comparison for nonnested models such as models with the same number of parameters. For example, suppose we want to compare the quadratic polynomial model with a log model formed by replacing t_{ij} with $\log(t_{ij})$, that is, $\text{logit}\{P(r_{ij} = 1)\} = \gamma_0 + \gamma_1 \log_e(t_{ij}) + \gamma_2 [\log_e(t_{ij})]^2$. The log model defines a concave but asymmetric curve that might be appropriate when the empirical prevalence curve is positively skewed. Because GEE is not based on a likelihood function, there is no formal means of comparing the fit of the quadratic polynomial and the log models. In fairness, comparison of nonnested models with the same number of parameters is probably less likely than comparisons of nested models. Nested models can be easily tested using the Z values based on the GEE estimates or with a Wald-type omnibus chi-squared statistic (Rotnitzky & Jewell, 1990).

⁵Some ad hoc procedures have been suggested; see, for example, Pan (2001).

Finally, the adequacy of the working correlation as a basis for analyzing longitudinal binary data has been questioned by a number of researchers (Chaganty & Joe, 2004; Crowder, 1995; Demidenko, 2004, chap. 7; McDonald, 1993; Sutradhar & Das, 1999). Much of the criticism has focused on the range restrictions of p imposed by the marginal probabilities, as previously discussed. One alternative is to model the dependencies using conditional odds ratios, which are not constrained by the marginal probabilities (e.g., Fitzmaurice, Laird, & Lipsitz, 1994).

Another increasingly popular alternative is a latent variable approach based on longitudinal extensions of the probit regression model (e.g., Muthén, 1996). In the latent variable approach, dependencies are modeled with correlations among normally distributed latent variables and thus are not bounded by the observed marginal probabilities. Similar to GEE, estimation of these models is typically not feasible with full ML as this involves integration over multivariate normal probabilities (Le Cessie & Van Houwelingen, 1994). Consequently, so-called limited-information or quasi-likelihood methods are typically used (Molenberghs & Verbeke, 2006, chap. 7).

Random Intercepts Model and ML

The random intercepts model has the advantage of providing a growth curve for individuals and information about heterogeneity of growth. Because of its individual-level nature, the random intercepts model has been widely used in settings such as biomedical research where focus is often on person-specific effects, such as individual response to medication (Diggle et al., 2002, chap. 7; Lindsey & Lambert, 1998; Neuhaus, 1992). In fact, the random intercepts model has been strongly recommended over the marginal model in such contexts because it provides growth curves that correspond to actual empirical trajectories (Lindsey & Lambert, 1998). An illustration is provided in Figure 1, where it can be seen that the the marginal growth curve is a trajectory that is not followed by any of the individuals and thus is a kind of statistical creation.

Even if the average growth curve is a statistical creation, it is appealing to a number of social scientists because its estimation is often the goal of growth curve analysis.⁶ The lack of an average growth curve may be one reason the random intercepts model has not been used extensively in the social sciences. As illustrated in our analysis, it is possible to obtain group-level information with the random intercepts model through numerical averaging. In this respect, the random intercepts model is more closely aligned with the LMM than the marginal model. The ability to provide both group-level and individual-level information seems an attractive quality of the random intercepts model.

There is some inconvenience regarding the group-level information obtained from the random intercepts model. Numerical averaging requires an extra step in the data analysis and routines for carrying it out are not part of standard statistical software (though the programming is relatively straightforward; see the Appendix). Numerical averaging is a post hoc procedure and as such it does not affect the interpretation of the model parameters. This means the random intercept parameters cannot be used directly to compute a predicted prevalence curve; nor can the parameters be used to compute instantaneous rate of change or age at peak of the prevalence curve (these can be directly computed for individual curves). Rather, these things must be obtained from the predicted prevalence, if possible.⁷

⁶Statistics based on aggregates are, of course, used extensively to summarize information at the group level regardless if they represent empirical impossibilities (e.g., 3.5 children per household).

⁷The age at peak can be determined by an inspection of the predicted prevalence values, but it is not clear if the instantaneous rate of change can be computed.

An advantage of the random intercepts model is that its random effects structure is the basis for a full specification of the multivariate dependencies between timepoints. Cross sections are linked together by the individual growth curves and people can be traced over time. In this respect, the random intercepts model is a “full-fledged” longitudinal model that takes advantage of the repeated measurements inherent in longitudinal data. Accordingly, ML methods can be used meaning that global fit indices can be computed, such as Akaike's information criteria. These fit indices make it possible to statistically compare nonnested models such as models with equal numbers of parameters, a feat that is not possible with GEE. Nested models can be tested in the same way as with GEE, using specific or omnibus tests of point estimates. Similar to GEE, the ML estimators have favorable large-sample properties such as consistency and normality, but the assumption of normality for the random intercepts is typically required. In theory, small sample methods can be based on ML theory, but to date, these methods are in their infancy and not widely available in statistical software (SAS PROC NLMIXED may be an exception; see Littell et al., 2006, chap. 14). As for missing data, valid inference is possible under the MAR assumption making it consistent with LMM analysis.

A practical disadvantage of ML estimation is that it is much more demanding on the user than GEE. Gaussian quadrature requires that starting values and the number of quadrature points be supplied. Numerical algorithms can be very sensitive to starting values and parameter estimates may not be computed if values are very far afield of the final estimates. Starting values for fixed effects can be obtained from a cross-sectional logistic regression solution as shown in the R code of the Appendix. A starting value for the variance of the random intercepts is more difficult to come by, which often means a number of values must be tried. Care must also be taken in the selection of the number of quadrature points. Severe inaccuracies can occur when the number is too small, but substantial computing time and resources might be required when the number is large (Jansen et al., 2006; Lesaffre & Spiessens, 2001).

COVARIATES OF CHANGE

A final issue important to mention is covariates of change. Our discussion to this point has dealt with models that do not have covariates of change other than polynomial transformations of age. Here we consider covariates other than time transformations, which we refer to as correlates of change. There are two types of correlates of change: time-static, which have constant values over time, and time-varying. In age-crime studies, an example of a time-static correlate is family's socioeconomic status at first measurement (FSES), and an example of a time-varying correlate is yearly drug use (see, e.g., Fabio et al., 2006). Similar to LMM, both types of correlates may be included in the marginal model or the random effects model through the X_i matrix. In the case of time-static correlates, columns of single correlates and their products with the polynomial transformations are added to the X_i matrix previously described (i.e., added to the matrix consisting of a column of ones and columns of polynomial time transformations). In the case of time-varying correlates, single correlates may appear with or without the time transformations and the product terms, depending on the nature of the research questions (see, e.g., Diggle et al., 2002, chap. 12).

When at least one correlate of change is time-varying, analysis with the marginal model using GEE can be quite involved as a nontrivial assumption about the conditional expectations must be investigated (Pepe & Anderson, 1994). If the assumption does not appear plausible, then an independent working correlation structure should be adopted to avoid bias in the growth curve estimates (Diggle et al., 2002, chap. 12). As previously noted, such an analysis ignores the longitudinal nature of the data and might be unattractive to applied researchers. Because a correctly specified random effects structure automatically meets the assumption about the conditional expectations, the random intercepts model may be preferable when modeling time-varying correlates (Lindsey & Lambert, 1998).

The issue with time-varying correlates may not weigh heavily as the majority of correlates studied in the social sciences appear to be of the time-static type. With static correlates, no additional assumptions are required for using the marginal model, and the features listed in Table 3 and discussed earlier may be considered when selecting a model. Similar to what was previously discussed, the marginal and random intercepts model parameters have different interpretations when static correlates are included. As a consequence, differences in growth curves based on correlates have different meanings in the two models. We illustrate the differences with a basic example.

Generally, when there are static correlate effects growth curves are conditional on the levels of the correlate(s). For the marginal model the conditional growth curves are based on the prevalence of people who share the same level of the static correlate(s). For example, suppose FSES is our single static correlate, coded as 0 = “low,” and 1 = “high,” and additional columns of X_i consist of FSES and its product with the linear and quadratic age transformations. Then the marginal model with FSES is

$$Pr(y_{ij}=1) = \frac{\exp(\gamma_0 + \gamma_1 t_{ij} + \gamma_2 t_{ij}^2 + \gamma_3 \text{FSES}_i + \gamma_4 t_{ij} \text{FSES}_i + \gamma_5 t_{ij}^2 \text{FSES}_i)}{1 + \exp(\gamma_0 + \gamma_1 t_{ij} + \gamma_2 t_{ij}^2 + \gamma_3 \text{FSES}_i + \gamma_4 t_{ij} \text{FSES}_i + \gamma_5 t_{ij}^2 \text{FSES}_i)} \quad (13)$$

Solving Equation (13) for FSES = 0 provides the prevalence curve for low FSES with parameters $\gamma_0, \gamma_1, \gamma_2$, and solving for FSES = 1 provides the prevalence curve for high FSES with parameters $\gamma_0 + \gamma_3, \gamma_1 + \gamma_4, \gamma_2 + \gamma_5$. The growth curves of the groups are identical when $\gamma_3 = \gamma_4 = \gamma_5$, which provides a natural null hypothesis of no difference in prevalence curves based on FSES. Because the prevalence curve of each group is based on aggregating among those who share the same level of FSES, the γ_k ($k > 2$) index differences between average growth curves. This is directly analogous to the types of effects that are indexed in LMMs with static correlates.

In contrast, the random intercepts model with FSES is

$$Pr(y_{ij}=1|d_i) = \frac{\exp(d_i + \delta_0 + \delta_1 t_{ij} + \delta_2 t_{ij}^2 + \delta_3 \text{FSES}_i + \delta_4 t_{ij} \text{FSES}_i + \delta_5 t_{ij}^2 \text{FSES}_i)}{1 + \exp(d_i + \delta_0 + \delta_1 t_{ij} + \delta_2 t_{ij}^2 + \delta_3 \text{FSES}_i + \delta_4 t_{ij} \text{FSES}_i + \delta_5 t_{ij}^2 \text{FSES}_i)} \quad (14)$$

Similar to the marginal model, we can solve Equation (14) for each value of FSES to obtain two growth curves. However, because the random intercepts model is a model for an individual, differences between the two growth curves are not similar to those in the marginal model or the LMM. The growth curve for FSES = 0 is a growth curve for an individual with low FSES. Likewise, the growth curve for FSES = 1 is a growth curve for an individual with high FSES. It follows that δ_3, δ_4 , and δ_5 represent differences between low and high FSES *for an individual*. The question is whether the same individual is referenced for both levels of FSES.

Some researches answer the aforementioned question in the affirmative and consider the δ_k ($k > 2$) to index differences that would occur if the same individual had experienced low and high FSES (Diggle et al., 2002, chap. 7; Zeger et al., 1988). This type of effect has no direct counterpart in LMM. No individual actually experiences both levels of FSES (or the correlate would be time-varying), so the differences between levels are considered an abstraction or a type of extrapolation (Zeger, Liang, & Albert, 1991). Such a view seems to discourage the use of the random intercepts model when comparisons among levels of a static covariate is the goal. An applied researcher probably would not want to estimate parameters that index effects

that might never occur. For example, if sex is the correlate there is not even a theoretical possibility that an individual can experience both its levels.

Other researchers point out that, similar to the parameters in the marginal model, the δ_k ($k > 2$) are estimated based on individuals who actually do have different levels of FSES, which must allow for some type of between-group interpretation (Galbraith, 1991; Lindsey & Lambert, 1998). In fact, the δ_k ($k > 2$) can be viewed as indexing differences similar to those of their corresponding γ_k , with averaging over individuals within levels of FSES being implicit rather than explicit (Galbraith, 1991). This view is buttressed by the findings previously noted that the inferential statistics (e.g., Z values) of marginal and random intercepts models tend to be similar in practice. This similarity would not exist if the parameters of the models were indexing radically different correlate effects (Neuhaus, 1993; Pendergast et al., 1996).

A distinct advantage of the random intercepts model is that, similar to the LMM, the variability in the random intercepts can be interpreted as individual differences due to omitted static correlates (Pendergast et al., 1996). In theory, modeling this variability can help prevent attenuation of the regression coefficients due to random error (Fuller, 1987, chap. 1). In addition, this provides a basis for an effect size measure for the static correlate(s). The variance of the intercepts can be estimated without the static correlate(s) in the model, and then the proportional decrease in the variance after adding the static correlate(s) can be assessed yielding a proportion-reduction-in-error measure.

When static correlates are included, explicit averaging with the random effects model can be performed along the lines already outlined. This is necessary if one wants to obtain average growth curves for the levels of the correlate(s) as the random intercepts parameter estimates cannot be used directly for this purpose (the marginal model parameters estimates can be used directly). After obtaining parameter estimates one can simulate a large number of random effects and then take averages of the individual probability estimates within levels of the correlate(s) at each timepoint. For example, one could average over the individual probabilities for those with FSES = 0 and then separately for those with FSES = 1 to yield average growth curves for low and high FSES. In this way, a type of marginal information can be obtained from the random intercepts model that considers correlate effects.

SUMMARY

In summary, two models for longitudinal binary data were presented: the marginal model and the random intercepts model. In contrast to LMM, the models are incommensurate in that the marginal model parameters pertain to group-level growth curves whereas the random intercepts model parameters pertain to individual-level growth curves. It was shown how group-level prediction curves can be obtained from the random intercepts model using numerical averaging based on a fitted model. The real data examples showed that the numerical averaging can produce almost identical predicted curves to those based directly on marginal model parameter estimates. The ability of numerical averaging to approximate the marginal model growth curves appears to be unaffected by missing data, but more research on this topic is warranted.

A number of key features of the models and their associated estimation methods was discussed. Based on the discussion, we recommend where possible to report results from both the marginal model and the random intercepts model. When this is not practical, one or the other model must be selected. If only group-level information is of interest to the analyst we lean toward using the marginal model with MI if there is missing data. If individual-level information is of interest in part or whole, we recommend using the random intercepts model with numerical averaging.

Acknowledgments

The work on this article was supported by Grant 96-MU-FX-0012 from the Office of Juvenile Justice and Delinquency Prevention, Grant 50778 from the National Institute of Mental Health, and Grant 411018 from the National Institute of Drug Abuse. Points of view or opinions in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Justice, the National Institute of Mental Health, and the National Institute of Drug Abuse.

APPENDIX

```
#####
# R code for GEE and ML binary models: Official Records Violence Example.
# R, geepack, and glmmML must be installed (see http://www.r-project.org/).
# Data set is assumed to be in stacked format (y_i and other matrices are stacked).
# Variables in data set ORviol: ID, Age (in years), Wave (= Age - 8 = 1, ..., 30),
#   viol (binary official records violence).
#####
# Obtain GEE solution
#####
library(geepack)                                # Note regression-like formula below
GEE1 <- (geeglm(viol~1+Age+I(Age^2),data=ORviol,
               family="binomial",corstr="exchangeable",id=ID,waves=Wave))
summary(GEE1)                                    # Prints summary
#####
# Obtain ML solution
#####
# Get starting values from cross-sectional logistic regression:
ML0 <- glm(viol~Age+I(Age^2),family=binomial,data=ORviol) # glm = General Linear Model
library(glmmML)
ML1 <- glmmML(viol~Age+I(Age^2),family=binomial,data=ORviol,cluster=ID,
              start.coef=coef(ML0),start.sigma=2,method="ghq",n.points=50) # 50 quadrature points
summary(ML1)
#####
# Numerical averaging based on the ML solution
#####
Age.alone <- seq(9,30,1)                         # Integer ages (age 9 to 30)
X <- matrix(nrow=5000,ncol=22)                   # Matrix to hold estimated probabilities (wide format)
delta <- coef(ML1)                               # delta parameter estimates
RE <- rnorm(10000,0,ML1$sigma)                   # 10,000 random intercepts
for(j in 1:22){                                  # Compute the individual probabilities, store in X
  for(i in 1:5000){
    X[i,j] <- RE[i]+delta[1]+delta[2]*Age.alone[j]+delta[3]*Age.alone[j]^2
    X[i,j] <- exp(X[i,j])/(1+exp(X[i,j]))
  }
}
ML.prev <- apply(X,2,mean)                        # Numerical averaging by columns of X
#####
# Predicted prevalence based on GEE solution
#####
gamma <- coef(GEE1)
GEE.prev <- exp(gamma[1]+gamma[2]*Age.alone+gamma[3]*Age.alone^2)/
              (1+exp(gamma[1]+gamma[2]*Age.alone+gamma[3]*Age.alone^2))
#####
# Plot observed and predicted prevalence curves
#####
# Observed prevalence:
OBS.prev <- tapply(ORviol$Age,ORviol$viol,na.rm=T) # Compute conditional mean
plot(Age.alone,OBS.prev,cex=1.5,ylab="Probability", # Plot observed prevalence
     xlab="Age",pch=19,xaxt="n",ylim=c(0,.04))
axis(1,at=9:30,labels=c(seq(9,30,1)))
lines(Age.alone,ML.prev,lwd=2)                   # Add ML predicted curve
lines(Age.alone,GEE.prev,lty=2,lwd=2)           # Add GEE predicted curve
legend(24,.1,c("OBS","ML","GEE"),lty=c(0,1,2),pch=c(19,-1,-1),lwd=c(2,2,2))
#####
```


REFERENCES

- Agresti, A. Categorical data analysis. Wiley; Hoboken, NJ: 2002.
- Berkhof J, Snijders TAB. Variance component testing in multilevel models. *Journal of Educational and Behavioral Statistics* 2001;26:133–152.
- Beunckens C, Sotito C, Molenberghs G. A simulation study comparing weighted estimating equations with multiple imputation based estimating equations for longitudinal binary data. *Computational Statistics & Data Analysis* 2008;52:1533–1548.
- Blumstein A, Cohen J, Farrington DP. Criminal career research: Its value for criminology. *Criminology* 1998;26:1–35.
- Brame R, Piquero AR. Selective attrition and the age-crime relationship. *Journal of Quantitative Criminology* 2003;19:107–127.
- Broström, G. The glmmML package [Computer software manual]. 2008. Retrieved from <http://cran.rproject.org/web/packages/glmmML/index.html>
- Chaganty NR, Joe H. Efficiency of generalized estimating equations for binary responses. *Journal of the Royal Statistical Society* 2004;66:851–860.
- Crowder M. On the use of a working correlation matrix in using generalized linear models for repeated measures. *Biometrika* 1995;82:407–410.
- Demidenko, E. Mixed models: Theory and applications. Wiley; New York: 2004.
- Diggle, PJ.; Heagerty, P.; Liang, KY.; Zeger, SL. Analysis of longitudinal data. 2nd ed.. Oxford University Press; New York: 2002.
- Duchateau L, Janssen P. Understanding heterogeneity in generalized mixed and frailty models. *The American Statistician* 2005;59:143–146.
- Espiritu, RC.; Huizinga, D.; Crawford, A.; Loeber, R. Epidemiology of self-reported delinquency. In: Loeber, R.; Farrington, DP., editors. *Child delinquents: Development, intervention, and service needs*. Sage; Thousand Oaks, CA: 2001. p. 47–66.
- Fabio A, Loeber R, Balasubramani GK, Roth J, Fu W, Farrington DP. Why some generations are more violent than others: Assessments of age, period, and cohort effects. *American Journal of Epidemiology* 2006;164:151–160. [PubMed: 16775045]
- Farrington, DP. Age and crime. In: Tonry, M.; Morris, N., editors. *Crime and justice*. University of Chicago Press; Chicago: 1986. p. 189–250.
- Fergusson DM, Horwood LJ. Male and female offending trajectories. *Development and Psychopathology* 2002;14:159–177. [PubMed: 11893091]
- Fitzmaurice GM. A caveat concerning independence estimating equations with multivariate binary data. *Biometrics* 1995;51:309–317. [PubMed: 7766784]
- Fitzmaurice GM, Laird NM, Lipsitz SR. Analyzing incomplete longitudinal binary responses: A likelihood-based approach. *Biometrics* 1994;50:601–612. [PubMed: 7981387]
- Fitzmaurice GM, Laird NM, Rotnitsky AG. Regression models for discrete longitudinal responses. *Statistical Science* 1993;8:284–299.
- Fitzmaurice, GM.; Laird, NM.; Ware, JH. Applied longitudinal analysis. Wiley; New York: 2004.
- Fuller, WA. Measurement error models. Wiley; New York: 1987.
- Galbraith JI. Interpreting a regression coefficient. *Biometrics* 1991;47:1593–1595. [PubMed: 1786332]
- Griepentrog GL, Ryan JM, Smith LD. Linear transformations of polynomial regression models. *The American Statistician* 1982;36:171–174.
- Halekoh U, Højsgaard S, Yan J. The R package geepack for generalized estimating equations. *Journal of Statistical Software* 2006;15:1–11.
- Jansen I, Beuchens C, Molenberghs G, Verbeke G, Mallinckrodt C. Analyzing incomplete discrete longitudinal clinical trial data. *Statistical Science* 2006;21:52–69.
- Lacourse E, Nagin D, Tremblay RE, Vitaro F, Claes M. Developmental trajectories of boys' delinquent group membership and facilitation of violent behaviors during adolescence. *Development and Psychopathology* 2003;15:183–197. [PubMed: 12848441]
- Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982;38:963–974. [PubMed: 7168798]

- Larsen K, Petersen JH, Budtz-Jorgensen E, Endahl L. Interpreting parameters in the logistic regression model with random effects. *Biometrics* 2000;56:909–914. [PubMed: 10985236]
- Le Cessie S, Van Houwelingen JC. Logistic regression for correlated binary data. *Applied Statistics* 1994;43:95–108.
- Lesaffre E, Spiessens B. On the effect of the number of quadrature points in a logistic random-effects model. *Applied Statistics* 2001;50:325–335.
- Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986;73:13–22.
- Lindsey, JK. Models for repeated measurements. 2nd ed.. Oxford University Press; New York: 1999.
- Lindsey JK, Lambert P. On the appropriateness of marginal models for repeated measures in clinical trials. *Statistics in Medicine* 1998;17:447–469. [PubMed: 9496722]
- Littell, RC.; Milliken, GA.; Stroup, WW.; Wolfinger, RD.; Schabenburger, O. SAS for mixed models. SAS Press; Cary, NC: 2006.
- Little RJA, Rubin DB. The analysis of social science data with missing values. *Sociological Methods & Research* 1989;18:292–326.
- Loeber R, Farrington DP, Stouthamer-Loeber M, Moffitt TE, Caspi A. The development of male offending: Key findings from the first decade of the Pittsburgh youth study. *Studies in Crime and Crime Prevention* 1998;7:141–172.
- Loeber, R.; Farrington, DP.; Stouthamer-Loeber, M.; Moffitt, TE.; Caspi, A.; White, HR., et al. The development of male offending: Key findings from fourteen years of the Pittsburgh youth study. In: Thornberry, TP.; Krohn, MD., editors. Taking stock of delinquency: An overview of findings from contemporary longitudinal studies. Kluwer Academic/Plenum; New York: 2003. p. 93-136.
- Loeber R, Hay DG. Key issues in the development of aggression and violence from childhood to early adulthood. *Annual Review of Psychology* 1997;48:371–410.
- Loeber, R.; LeBlanc, M. Toward a developmental criminology. In: Morris, MTN., editor. Crime and justice: A review of research. Vol. Vol. 12. University of Chicago Press; Chicago: 1990. p. 375-473.
- Loeber R, Stouthamer-Loeber M. Development of juvenile aggression and violence: Some common misconceptions and controversies. *American Psychologist* 1998;53:242–259. [PubMed: 9491750]
- Manski CF, Nagin DS. Bounding disagreements about treatment effects: A case study of sentencing and recidivism. *Sociological Methodology* 1998;28:99–137.
- McDonald BW. Estimating logistic regression parameters for bivariate binary data. *Journal of the Royal Statistical Society, Series B* 1993;55:391–397.
- Molenberghs, G.; Verbeke, G. Models for discrete longitudinal data. Springer; New York: 2006.
- Morrell CH, Pearson JD, Brant LJ. Linear transformations of linear mixed-effects models. *The American Statistician* 1997;51:338–343.
- Muthén, BO. Growth modeling with binary responses. In: Eye, AV.; Clogg, C., editors. Categorical variables in developmental research: Methods of analysis. Academic; San Diego, CA: 1996. p. 37-54.
- Nagin DS, Farrington DP, Moffitt TE. Life course trajectories of different types of offenders. *Criminology* 1995;33:111–139.
- Neuhaus JM. Statistical methods for longitudinal and clustered designs with binary responses. *Statistical Methods in Medical Research* 1992;1:249–273. [PubMed: 1341660]
- Neuhaus JM. Estimation efficiency and tests of covariate effects with clustered binary data. *Biometrics* 1993;49:989–996. [PubMed: 8117909]
- Neuhaus JM. Assessing change with longitudinal and clustered binary data. *Annual Review of Public Health* 2001;22:115–128.
- Neuhaus JM, Kalbfleisch JD, Hauck WW. A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data. *International Statistical Review* 1991;59:25–35.
- Pan W. Model selection in estimating equations. *Biometrics* 2001;57:529–534. [PubMed: 11414579]
- Pellegrini AD, Long JD. An observational study of early heterosexual interaction at middle school dances. *Journal of Research on Adolescence* 2007;17:613–638.
- Pendergast JF, Gange SJ, Newton MA, Lindstrom MJ, Palta M, Fisher MR. A survey of methods for analyzing clustered binary response data. *International Statistical Review* 1996;64:89–118.

- Pepe MS, Anderson GL. A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data. *Communications in Statistics, Simulation and Computation* 1994;23:939–951.
- Robins JM, Rotnitzky AG, Zhao LP. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association* 1994;89:846–866.
- Rotnitzky A, Jewell N. Hypothesis testing of regression parameters in semiparametric generalized linear models for cluster correlated data. *Biometrika* 1990;77:485–497.
- Sampson RJ, Laub JH. Life course desisters? Trajectories of crime among delinquent boys followed to age 70. *Criminology* 2003;41:555–592.
- Shin T, Davison ML, Long JD. Effects of missing data methods in structural equation modeling with nonnormal longitudinal data. *Structural Equation Modeling* 2009;16:1–30.
- Stram DO, Lee JW. Variance components testing in the longitudinal mixed effects model. *Biometrics* 1994;50:1171–1177. [PubMed: 7786999]
- Sutradhar BC, Das K. On the efficiency of regression estimators in generalized linear models for longitudinal data. *Biometrika* 1999;86:459–465.
- Touloumi G, Babiker AG, Pocock SJ, Darbyshire JH. Impact of missing data due to drop-outs on estimators for rates of change in longitudinal studies: A simulation study. *Statistics in Medicine* 2001;20:3715–3728. [PubMed: 11782028]
- Verbeke, G.; Molenberghs, G. *Linear mixed models for longitudinal data*. Springer-Verlag; New York: 2001.
- Zeger SL, Liang KY, Albert PS. Models for longitudinal data: A generalized estimating equation approach. *Biometrics* 1988;44:1049–1060. [PubMed: 3233245]
- Zeger SL, Liang KY, Albert PS. The interpretation of a regression coefficient: Response. *Biometrics* 1991;47:1596–1596.

FIGURE 1. Simulated individual growth curves (solid lines) and mean growth curve (dashed line). Upper panel is probability metric and lower panel is logit metric.

FIGURE 2.

Official records observed violence prevalence (filled circles), and predicted growth curves from the GEE solution (dashed line) and maximum likelihood solution (solid line).

FIGURE 3. Self-report observed violence prevalence (filled circles), and predicted growth curves from the GEE solution (dashed line) and maximum likelihood solution (solid line).

TABLE 1

Results of GEE Analysis for Official Records and Self-Report

Parameter	Official Records			Self-Report		
	Estimate	SE	Z	Estimate	SE	Z
γ_0	-21.3867	2.2245	-9.6140	-10.8483	1.4814	-7.3229
γ_1	2.0302	0.2625	7.7347	1.0334	0.1677	6.1639
γ_2	-0.0576	0.0077	-7.5094	-0.0308	0.0046	-6.6344
ρ	0.0169	0.1188	0.1422 [†]	0.1445	0.0673	2.1472

all other Z values $p < .05$.[†] One-tailed $p = .44$;

TABLE 2

Results of ML Analysis for Official Records and Self-Report

Parameter	Official Records			Self-Report		
	Estimate	SE	Z	Estimate	SE	Z
δ_0	-22.5980	2.3702	-9.5341	-13.2369	1.5409	-8.5905
δ_1	2.0910	0.2694	7.7611	1.2043	0.1729	6.9641
δ_2	-0.0594	0.0076	-7.8102	-0.0359	0.0048	-7.5355
σ_d	1.2271	0.1346	9.1164	1.7066	0.1032	16.5402

Note. All Z values $p < .001$.

TABLE 3

Taxonomy of Model Features

Feature	Marginal Model	Random Intercepts Model
Growth curve type	Group	Individual
Heterogeneity of growth?	No	Yes
Dependencies due to RM	Treated as nuisance	Modeled through random intercepts
Estimation procedure	GEE	ML
Estimation requirements	Minimal	Starting values and quadrature values
Distributional assumptions	None	Normality (of random intercepts)
Large N results?	Yes	Yes
Missing data assumption	MCAR	MAR
Global fit index	No	Yes
Predicted prevalence	Computed from growth curve	Requires numerical averaging