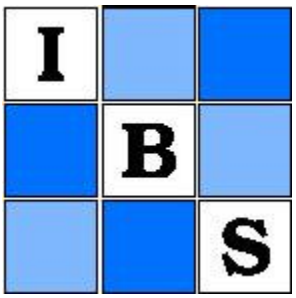

BIOMETRICS 50, 933–944
December 1994

A Random-Effects Ordinal Regression Model for Multilevel Analysis

Donald Hedeker[1,2] and Robert D. Gibbons[2]

[1]Prevention Research Center, School of Public Health, University of Illinois at Chicago, 850 West Jackson Boulevard, Suite 400, Chicago, Illinois 60607, U.S.A.

[2]Department of Psychiatry, College of Medicine, University of Illinois at Chicago, 912 South Wood Street, Chicago, Illinois 60612, U.S.A.

SUMMARY

A random-effects ordinal regression model is proposed for analysis of clustered or longitudinal ordinal response data. This model is developed for both the probit and logistic response functions. The threshold concept is used, in which it is assumed that the observed ordered category is determined by the value of a latent unobservable continuous response that follows a linear regression model incorporating random effects. A maximum marginal likelihood (MML) solution is described using Gauss–Hermite quadrature to numerically integrate over the distribution of random effects. An analysis of a dataset where students are clustered or nested within classrooms is used to illustrate features of random-effects analysis of clustered ordinal data, while an analysis of a longitudinal dataset where psychiatric patients are repeatedly rated as to their severity is used to illustrate features of the random-effects approach for longitudinal ordinal data.

1. Introduction

Models for ordinal response variables are important in many areas of research, since subjects are often classified or may respond on an ordinal, or graded, scale. In biomedical studies, for example, subjects may be classified in terms of exhibiting definite, mild, or no symptomatology of a given disease or condition. Additionally, it is often the case that subjects are observed nested within clusters (i.e., schools, firms, clinics) or are repeatedly assessed across time, and so the use of ordinal regression models which assume that observations are independent (McKelvey and Zavoina, 1975; McCullagh, 1980) is problematic.

For data that are clustered and/or longitudinal, random-effects regression models (RRM) have been developed primarily to model continuous (Laird and Ware, 1982; Bock, 1983; Jennrich and Schluchter, 1986; Liang and Zeger, 1986; Longford, 1987; Goldstein, 1987; Bryk and Raudenbush, 1987) and dichotomous (Stiratelli, Laird, and Ware, 1984; Anderson and Aitkin, 1985; Wong and Mason, 1985; Zeger and Liang, 1986; Gibbons and Bock, 1987; Conaway, 1989; Qu et al., 1992) responses. Although not as developed, an increasing amount of work has focused on random-effects models for ordinal response data. Harville and Mee (1984) describe a mixed-model procedure for the analysis of clustered ordinal outcome data using a Taylor series expansion to approximate the estimation of the random effects. Jansen (1990) utilizes numerical quadrature for a random-effects model for clustered ordinal responses which is suitable when there is only a single random effect. Both Harville and Mee (1984) and Jansen (1990) implement the EM algorithm for parameter estimation, and both comment on the slowness in convergence time for the solution. Ezzet and Whitehead (1991) provide a random-effects model using the Newton–Raphson method; however, their presentation was limited to a model for responses from a crossover trial and allowed for only one random effect. An alternative approach is given by Goldstein (1991), who proposes a log-linear model for categorical response data that can include multiple random effects. However, as noted by McCullagh (1980), ordinal regression models based on the threshold concept (discussed below) do not correspond to a log-linear structure when the number of categories exceeds 2.

Key words: Clustering; Gauss–Hermite quadrature; Logistic regression; Maximum marginal likelihood; Ordered categorical data; Probit regression; Repeated observations; Threshold model.

In this paper, we will develop an ordinal random-effects regression model that is appropriate for either clustered or longitudinal response data. This model will accommodate multiple random effects and, additionally, allow for a general form for model covariates. Assuming either a probit or logistic response function, a maximum marginal likelihood solution is described using multidimensional quadrature to numerically integrate over the distribution of random effects. An iterative Fisher scoring solution is described which, in general, converges much faster than the EM algorithm when applied to random-effects models (Bock, 1989). Additionally, the Fisher scoring solution provides standard errors for all model parameters. Examples of analysis of both clustered and longitudinal data will illustrate features of the random-effects approach for ordinal response data.

2. Random-Effects Ordinal Regression Model

In probit and logistic regression models it is often assumed that there is an unobservable latent variable (y) that is related to the actual response through the "threshold concept." For the dichotomous model, one threshold value is assumed, and for the ordinal model, a *series* of threshold values $\gamma_1, \gamma_2, \ldots, \gamma_{J-1}$, where J equals the number of ordered categories, $\gamma_0 = -\infty$, and $\gamma_J = \infty$. Here, a response occurs in category j ($Y = j$) if the latent response process y exceeds the threshold value γ_{j-1}, but does not exceed the threshold value γ_j.

To describe the model in a general way for data that are either clustered or longitudinal, the terminology of multilevel analysis can be used (Goldstein, 1987). For this, let i denote the level-2 units (clusters in the clustered data context, or subjects in the longitudinal data context), and let k denote the level-1 units (subjects in the clustered data context, or repeated observations in the longitudinal data context). Assume that there are $i = 1, \ldots, N$ level-2 units and $k = 1, \ldots, n_i$ level-1 units nested within each level-2 unit. The random-effects regression model for the latent response strength y_{ik} can be written as follows:

$$y_{ik} = \boldsymbol{x}'_{ik}\boldsymbol{\beta}_i + \boldsymbol{w}'_{ik}\boldsymbol{\alpha} + \varepsilon_{ik}, \tag{1}$$

where $\boldsymbol{w}_{ik}$ is the $p \times 1$ covariate vector and $\boldsymbol{x}_{ik}$ is the design vector for the r random effects, both vectors being for the kth level-1 unit nested within level-2 unit i. Also, $\boldsymbol{\alpha}$ is the $p \times 1$ vector of unknown fixed regression parameters, $\boldsymbol{\beta}_i$ is the $r \times 1$ vector of unknown random effects for the level-2 unit i, and ε_{ik} are the model residuals. The distribution of the random effects is assumed to be multivariate normal with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}_\beta$, and the residuals are assumed to be independently normally distributed with mean 0 and variance σ^2. Since the level-2 subscript i is present for the $\boldsymbol{x}$ vector, not all level-2 units are assumed to have the same number of level-1 observations nested within. Thus, for clustered data, there is no assumption of equal sample sizes within clusters, whereas for longitudinal data, the numbers of repeated observations are not assumed to be equal across subjects.

2.1 *Multilevel Representation*

Some researchers have described the model given in (1) in terms of a multilevel (Goldstein, 1987) or hierarchical (Bryk and Raudenbush, 1987) structure. For this, the model is partitioned into the following within level-2 unit model (or between level-1 unit model),

$$y_{ik} = \boldsymbol{x}'_{(1)ik}\boldsymbol{b}_i + \boldsymbol{w}'_{(1)ik}\boldsymbol{\alpha}_{(1)} + \varepsilon_{ik},$$

and between level-2 unit model,

$$\boldsymbol{b}_i = \boldsymbol{\mu} + \boldsymbol{w}'_{(2)i}\boldsymbol{\alpha}_{(2)} + \boldsymbol{\delta}_i,$$

where $\boldsymbol{w}_{(1)ik}$ and $\boldsymbol{\alpha}_{(1)}$ represent the fixed level-1 covariates and their effects, $\boldsymbol{w}_{(2)i}$ and $\boldsymbol{\alpha}_{(2)}$ are the fixed level-2 covariates and their effects, and $\boldsymbol{x}_{(1)ik}$ are the level-1 variables allowed to vary at level-2. The level-2 effects $\boldsymbol{b}_i$ are then influenced by an overall mean $\boldsymbol{\mu}$, level-2 covariates $\boldsymbol{\alpha}_{(2)}$, and a unique random component $\boldsymbol{\delta}_i$ distributed normally with mean $\boldsymbol{0}$ and covariance matrix $\boldsymbol{\Sigma}_\beta$. The between level-2 unit model is sometimes referred to as a "slopes as outcomes" model (Burstein, Linn, and Capell, 1978). The multilevel representation shows that just as level-1 covariates are included in the model to explain variation in level-1 outcomes (y_{ik}), level-2 covariates are included to explain variation in level-2 outcomes ($\boldsymbol{b}_i$). In terms of clustered data where there may be only one random effect, which is typically the intercept (the first element of $\boldsymbol{x}'_{ik}$ then equals 1), $\boldsymbol{b}_i$ is a scalar that represents differences due to the clusters and is modeled in terms of cluster-level variables ($\boldsymbol{w}_{(2)i}$), as well as unexplained random cluster-level variation ($\boldsymbol{\delta}_i$). For longitudinal data, the elements of $\boldsymbol{b}_i$ may represent an initial level (the intercept) and a trend across time in the response variable (the value of time then being the second element in $\boldsymbol{x}'_{ik}$); both of these subject-varying coefficients may

then be modeled in terms of subject-level variables ($\boldsymbol{w}_{(2)i}$) to indicate the effect of subject characteristics on the initial level and trend across time of a subject, in addition to unexplained subject-level random variation ($\boldsymbol{\delta}_i$).

To equate the two representations of the model, simply note that combining the between and within level-2 models yields

$$y_{ik} = \mathbf{x}'_{(1)ik}(\boldsymbol{\mu} + \boldsymbol{w}'_{(2)i}\boldsymbol{\alpha}_{(2)} + \boldsymbol{\delta}_i) + \boldsymbol{w}'_{(1)ik}\boldsymbol{\alpha}_{(1)} + \varepsilon_{ik},$$

and then, $\boldsymbol{x}'_{ik} = \boldsymbol{x}'_{(1)ik}$, $\boldsymbol{\beta}_i = \boldsymbol{\mu} + \boldsymbol{\delta}_i$, $\boldsymbol{w}'_{ik} = [\boldsymbol{x}'_{(1)ik} \otimes \boldsymbol{w}'_{(2)i} \vdots \boldsymbol{w}'_{(1)ik}]$, and $\boldsymbol{\alpha} = [\boldsymbol{\alpha}'_{(2)} \vdots \boldsymbol{\alpha}'_{(1)}]$. Typically, some level-2 covariates $\boldsymbol{w}_{(2)i}$ are not thought to influence all of the r level-2 effects $\boldsymbol{b}_i$; in this case, the corresponding elements of the $\boldsymbol{x}'_{(1)ik} \otimes \boldsymbol{w}'_{(2)i}$ partition of the covariate vector $\boldsymbol{w}'_{ik}$ are removed. Notice also that the random effects can either be represented as having a mean $\boldsymbol{\mu}$ or as being deviations from that mean; the $\boldsymbol{\beta}_i$ vector corresponds to the former case, while the $\boldsymbol{\delta}_i$ vector corresponds to the latter.

2.2 *Probit and Logistic Response Functions*

With the above random-effects regression model for the underlying and unobservable variable y_{ik}, the probability, for a given level-2 unit i, that $Y_k = j$ (a response occurs in category j), conditional on $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, is given by the following equation:

$$P(Y_k = j|\boldsymbol{\beta}, \boldsymbol{\alpha}) = \Phi[(\gamma_j - z_k)/\sigma] - \Phi[(\gamma_{j-1} - z_k)/\sigma],$$

where $z_k = \boldsymbol{x}'_k\boldsymbol{\beta} + \boldsymbol{w}'_k\boldsymbol{\alpha}$ and $\Phi(\cdot)$ represents the cumulative standard normal distribution function. Without loss of generality, the origin and unit of z may be chosen arbitrarily. For convenience, let $\gamma_1 = 0$ and $\sigma = 1$. This formulation corresponds to the choice of a probit (more specifically a normit) response function. Alternatively, if the logistic response function is assumed, then the logistic function $\Psi(\cdot)$ replaces $\Phi(\cdot)$ in the conditional probability, with

$$\Psi(\gamma_j - z_k) = \frac{1}{1 + \exp[-(\gamma_j - z_k)]}.$$

For the logistic response function, again, we let $\gamma_1 = 0$; however, the residual variance equals $\pi^2/3$. In the development that follows, we will assume the probit response function; however, we will note the necessary modifications if the logistic response function is assumed instead.

3. Maximum Marginal Likelihood Estimation

Letting $\mathbf{Y}_i$ denote the vector pattern of ordinal item responses from level-2 unit i for the n_i level-1 units nested within, the probability of any pattern $\mathbf{Y}_i$, given $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, is equal to the product of the probabilities of the level-1 responses:

$$\ell(\mathbf{Y}_i|\boldsymbol{\beta}, \boldsymbol{\alpha}) = \prod_{k=1}^{n_i} \prod_{j=1}^{J} [\Phi(\gamma_j - z_{ik}) - \Phi(\gamma_{j-1} - z_{ik})]^{d_{ikj}}, \tag{2}$$

where

$$d_{ikj} = \begin{cases} 1 & \text{if } Y_{ik} = j, \\ 0 & \text{if } Y_{ik} \neq j. \end{cases}$$

Then the marginal density of $\mathbf{Y}_i$ in the population is expressed as the following integral of the likelihood, $\ell(\cdot)$, weighted by the prior density $g(\cdot)$:

$$h(\mathbf{Y}_i) = \int_{\boldsymbol{\beta}} \ell(\mathbf{Y}_i|\boldsymbol{\beta}, \boldsymbol{\alpha}) g(\boldsymbol{\beta})\, d\boldsymbol{\beta},$$

where $g(\boldsymbol{\beta})$ represents the distribution of the $\boldsymbol{\beta}$ vector in the population.

In the estimation of parameters for the random-effects probit regression model, Gibbons and Bock (1987) orthogonally transform the response model to use the maximum marginal likelihood (MML) estimation procedure discussed by Bock and Aitkin (1981) in the context of a dichotomous factor analysis model. Specifically, let $\boldsymbol{\beta} = \boldsymbol{T}\boldsymbol{\theta} + \boldsymbol{\mu}$, where $\boldsymbol{T}\boldsymbol{T}' = \boldsymbol{\Sigma}_\beta$ is the Cholesky decomposition of $\boldsymbol{\Sigma}_\beta$. The reparameterized model is then

$$z_{ik} = \boldsymbol{x}'_{ik}(\boldsymbol{T}\boldsymbol{\theta} + \boldsymbol{\mu}) + \boldsymbol{w}'_{ik}\boldsymbol{\alpha}$$

and the marginal density becomes

$$h(\mathbf{Y}_i) = \int_{\boldsymbol{\theta}} \ell(\mathbf{Y}_i|\boldsymbol{\theta}, \boldsymbol{\alpha})\, g(\boldsymbol{\theta})\, d\boldsymbol{\theta},$$

where $g(\boldsymbol{\theta})$ represents the multivariate standard normal density. In addition to transforming the response model to the multivariate standard normal distribution, a further consequence of transforming from $\boldsymbol{\beta}$ to $\boldsymbol{\theta}$ is that the Cholesky factor $\boldsymbol{T}$, which is a lower-triangular matrix, is estimated instead of the covariance matrix $\boldsymbol{\Sigma}_\beta$. As the Cholesky factor is essentially the square root of the covariance matrix, this then allows more stable estimation of near-zero variance terms.

For the estimation of the p covariate coefficients $\boldsymbol{\alpha}$, the population parameters $\boldsymbol{\mu}$ and $\boldsymbol{T}$ (with r and $r(r+1)/2$ elements, respectively), and the $J-2$ threshold values γ_j ($j = 2, \ldots, J-1$), the marginal log-likelihood for the patterns from the N level-2 units can be written as

$$\log L = \sum_{i}^{N} \log h(\mathbf{Y}_i).$$

Differentiating with respect to the threshold values, we get for a particular $\gamma_{j'}$,

$$\frac{\partial \log L}{\partial \gamma_{j'}} = \sum_{i=1}^{N} h^{-1}(\mathbf{Y}_i) \frac{\partial h(\mathbf{Y}_i)}{\partial \gamma_{j'}},$$

where

$$\frac{\partial h(\mathbf{Y}_i)}{\partial \gamma_{j'}} = \int_{\boldsymbol{\theta}} \sum_{k=1}^{n_i} \sum_{j=1}^{J} d_{ikj} \frac{\phi(\gamma_j - z_{ik})\delta_{jj'} - \phi(\gamma_{j-1} - z_{ik})\delta_{j-1j'}}{\Phi(\gamma_j - z_{ik}) - \Phi(\gamma_{j-1} - z_{ik})} \ell(\mathbf{Y}_i|\boldsymbol{\theta}, \boldsymbol{\alpha})\, g(\boldsymbol{\theta})\, d\boldsymbol{\theta}, \tag{3}$$

$$\delta_{jj'} = \begin{cases} 1 & \text{if } j = j', \\ 0 & \text{if } j \neq j', \end{cases}$$

and $\phi(\cdot)$ represents the standard normal density function. Let $\boldsymbol{\eta}$ represent an arbitrary parameter vector; then for $\boldsymbol{\alpha}$, $\boldsymbol{\mu}$, and the vector $v(\boldsymbol{T})$ that contains the unique elements of the Cholesky factor $\boldsymbol{T}$, we get

$$\frac{\partial \log L}{\partial \boldsymbol{\eta}} = \sum_{i=1}^{N} h^{-1}(\mathbf{Y}_i) \int_{\boldsymbol{\theta}} \sum_{k=1}^{n_i} \sum_{j=1}^{J} d_{ikj} \frac{\phi(\gamma_{j-1} - z_{ik}) - \phi(\gamma_j - z_{ik})}{\Phi(\gamma_j - z_{ik}) - \Phi(\gamma_{j-1} - z_{ik})} \ell(\mathbf{Y}_i|\boldsymbol{\theta}, \boldsymbol{\alpha})\, g(\boldsymbol{\theta}) \frac{\partial z_{ik}}{\partial \boldsymbol{\eta}'} d\boldsymbol{\theta}, \tag{4}$$

where

$$\frac{\partial z_{ik}}{\partial \boldsymbol{\alpha}'} = \boldsymbol{w}'_{ik}, \quad \frac{\partial z_{ik}}{\partial \boldsymbol{\mu}'} = \boldsymbol{x}'_{ik}, \quad \frac{\partial z_{ik}}{\partial (v(\boldsymbol{T}))'} = (\boldsymbol{\theta}' \otimes \boldsymbol{x}'_{ik})\mathbf{J}'_r,$$

and $\mathbf{J}_r$ is the transformation matrix of Magnus (1988) which eliminates the elements above the main diagonal. Note that, for the logistic regression formulation, the logistic function $\boldsymbol{\Psi}(\cdot)$ replaces the normal response function $\boldsymbol{\Phi}(\cdot)$ and the product $\boldsymbol{\Psi}(\cdot) \times (1 - \boldsymbol{\Psi}(\cdot))$ replaces the standard normal density function $\phi(\cdot)$ in the conditional probability (2) and the derivatives (3) and (4).

Fisher's method of scoring can be used to provide the solution to these likelihood equations. For this, provisional estimates for the vector of parameters $\boldsymbol{\Theta}$ on iteration ι are improved by

$$\Theta_{\iota+1} = \Theta_\iota - E\left[\frac{\partial^2 \log L}{\partial \Theta_\iota \partial \Theta'_\iota}\right]^{-1} \frac{\partial \log L}{\partial \Theta_\iota}, \tag{5}$$

where the information matrix, or expectation of the matrix of second derivatives, is given by

$$E\left[\frac{\partial^2 \log L}{\partial \Theta_\iota \partial \Theta'_\iota}\right] = -\sum_{i=1}^{N} h^{-2}(\mathbf{Y}_i) \frac{\partial h(\mathbf{Y}_i)}{\partial \Theta_\iota} \left(\frac{\partial h(\mathbf{Y}_i)}{\partial \Theta_\iota}\right)'.$$

At convergence, the large-sample variance–covariance matrix of the parameter estimates is then obtained as the inverse of the information matrix.

3.1 *Numerical Quadrature*

In order to solve the above likelihood equations, numerical integration on the transformed $\boldsymbol{\theta}$ space can be performed. For this, Gauss–Hermite quadrature can be used to approximate the above integrals to any practical degree of accuracy. In Gauss–Hermite quadrature, the integration is approximated by a summation on a specified number of quadrature points Q for each dimension of the integration; thus, for the transformed $\boldsymbol{\theta}$ space, the summation goes over Q^r points. For the standard normal univariate density, optimal points and weights (which will be denoted B_q and $A(B_q)$, respectively) are given in Stroud and Sechrest (1966). For the multivariate density, the r-dimensional vector of quadrature points is denoted by $\boldsymbol{B}_q' = (B_{q1}, B_{q2}, \ldots, B_{qr})$, with its associated (scalar) weight given by the product of the corresponding univariate weights,

$$A(\boldsymbol{B}_q) = \prod_{h=1}^{r} A(B_{qh}).$$

As the number of random effects r is increased, the terms in the summation (Q^r) increase exponentially in the quadrature solution. Fortunately, as is noted by Bock, Gibbons, and Muraki (1988) in the context of a dichotomous factor analysis model, the number of points in each dimension can be reduced as the dimensionality is increased without impairing the accuracy of the approximations; they indicated that for a five-dimensional solution as few as three points per dimension were necessary to obtain adequate accuracy.

3.2 *Weighted Data*

The above solution can easily be modified to accommodate weighted data, for example, when the same $n_i \times 1$ response pattern $\mathbf{Y}_i$, $n_i \times r$ random-effects design matrix $\boldsymbol{X}_i$, and $n_i \times p$ covariate matrix $\boldsymbol{W}_i$ are observed for a number of level-2 units. In this context, the data can be represented in tabular form, each entry in the table representing a frequency or count of level-2 observations with the same values of $\boldsymbol{Y}_i$, $\boldsymbol{X}_i$, and $\boldsymbol{W}_i$. This can occur, for example, in the longitudinal context if subjects are measured at the same n repeated timepoints and the frequency of each of the possible J^n response patterns is given for a discrete number of treatment groups. For this, let $l = 1, \ldots, N'$ denote the index and total number of distinct occurrences of sets consisting of the same $\mathbf{Y}_i$, $\boldsymbol{X}_i$, and $\boldsymbol{W}_i$, and express the weight for the unique occurrence l as s_l. In the summation over l, the log-likelihood is then written as $\log L = \Sigma_l^{N'} s_l \log h(\mathbf{Y}_l)$. The derivatives are then multiplied by s_l,

$$\frac{\partial \log L}{\partial \gamma_{j'}} = \sum_{l=1}^{N'} s_l\, h^{-1}(\mathbf{Y}_l) \frac{\partial h(\mathbf{Y}_l)}{\partial \gamma_{j'}} \quad \text{and} \quad \frac{\partial \log L}{\partial \boldsymbol{\eta}} = \sum_{l=1}^{N'} s_l\, h^{-1}(\mathbf{Y}_l) \frac{\partial h(\mathbf{Y}_l)}{\partial \eta},$$

as is the expectation of the second derivatives:

$$E\left[\frac{\partial^2 \log L}{\partial \Theta_\iota \partial \Theta_\iota'}\right] = -\sum_{l=1}^{N'} s_l\, h^{-2}(\mathbf{Y}_l) \frac{\partial h(\mathbf{Y}_l)}{\partial \Theta_\iota} \left(\frac{\partial h(\mathbf{Y}_l)}{\partial \Theta_\iota}\right)'.$$

3.3 *Goodness of Fit*

In the case of weighted data, the likelihood ratio test of model fit relative to the general multinomial alternative can be considered, since the N level-2 units can be classified into cells of a contingency table, each cell corresponding to a unique set of response patterns $\mathbf{Y}_i$ and matrices $\boldsymbol{X}_i$ and $\boldsymbol{W}_i$. The chi-square approximation for this likelihood ratio statistic is given as

$$G^2 = 2 \sum_{l=1}^{N'} s_l \log \frac{s_l}{N h(\mathbf{Y}_l)},$$

on $J^n - [(r + p) + (J - 2) + r(r + 1)/2] - 1$ degrees of freedom, where $h(\mathbf{Y}_l)$ is the marginal probability of pattern $\mathbf{Y}_l$ computed from the MML estimates of the $[(r + p) + (J - 2) + r(r + 1)/2]$ parameters. Unfortunately, this test cannot generally be used since either the number of level-1 units n_i is not constant across level-2 units, or the number of possible response patterns J^n (for each unique set of $\boldsymbol{X}_i$ and $\boldsymbol{W}_i$ matrices) is too large relative to the number of level-2 units N, producing near-zero expected values for many of the possible response patterns. Instead, for general application for both weighted and unweighted data, the likelihood ratio χ^2 test can be used to compare the relative fit of the data provided by nested models. For this, suppose that the set of parameters

for model B is a subset of the set for model A. Evaluating the log-likelihood $\log L = \Sigma_i^N \log h(\mathbf{Y}_i)$ (or $\log L = \Sigma_l^{N'} s_l \log h(\mathbf{Y}_l)$ in the case of weighted data) using the estimated parameters of the two models yields $\log L_{\mathrm{A}}$ and $\log L_{\mathrm{B}}$. The significance of the additional terms in model A is determined by comparing $-2(\log L_{\mathrm{A}} - \log L_{\mathrm{B}})$ to a table of the χ^2 distribution with degrees of freedom equal to the number of additional parameters in model A.

3.4 *Computer Implementation*

In terms of computer programming, the procedure described in this article has been implemented for use in the MIXOR program (Hedeker, unpublished technical report, Prevention Research Center, School of Public Health, University of Illinois at Chicago, 1993). [This DOS-based program can currently be obtained from Ann Hohmann, Ph.D., M.P.H., NIMH Services Research Branch, 5800 Fishers Lane, Room #10C-06, Rockville, Maryland 20857, U.S.A.]. The program starts by reading in for each level-2 unit the $n_i \times 1$ response vector $\mathbf{Y}_i$, the $n_i \times r$ random-effect design matrix $\boldsymbol{X}_i$, and the $n_i \times p$ matrix of covariates $\boldsymbol{W}_i$. Provisional starting values for the model parameters must be specified prior to the start of the iterative procedure. These can be estimated using a fixed-effect ordinal regression solution for coefficient vectors $\boldsymbol{\mu}$ and $\boldsymbol{\alpha}$, and threshold values γ_j ($j = 2, \ldots, J - 1$). Starting values for the Cholesky factor $\boldsymbol{T}$ of the random-effects covariance matrix can be specified arbitrarily as a diagonal matrix, with each diagonal element set equal to some fraction of the assumed residual variance value. At each iteration and for each level-2 unit, the solution goes over the Q^r quadrature points, with summation replacing the integration over the random-effect distribution. The conditional probabilities $\ell(\mathbf{Y}_i|\boldsymbol{\theta}, \boldsymbol{\alpha})$ are obtained substituting the random-effect vector $\boldsymbol{\theta}$ by the current r-dimensional vector of quadrature points $\boldsymbol{B}_q$. The marginal density for each level-2 unit is then approximated as

$$h(\mathbf{Y}_i) \approx \sum_{q}^{Q^r} \ell(\mathbf{Y}_i|\boldsymbol{B}_q, \boldsymbol{\alpha}) A(\boldsymbol{B}_q).$$

At each iteration, computation of the first derivatives and information matrix then proceeds summing over level-2 units and quadrature points. In the summation over the Q^r quadrature points, substitutions are made in the equations for the first derivatives and information matrix as follows: the $\boldsymbol{\theta}$ random-effect vector is substituted by the current vector of quadrature points $\boldsymbol{B}_q$, and the evaluation of the multivariate standard normal density $g(\boldsymbol{\theta})$ is substituted by the current quadrature weight $A(\boldsymbol{B}_q)$. Following the summation over level-2 units and quadrature points, parameters are corrected using (5), and the entire procedure is repeated until convergence. With 10 quadrature points per random-effect dimension, for the problems described below, convergence (corrections of less than .0001 for all parameters) was obtained within 10 and 20 iterations for models with one and two random effects, respectively.

4. Clustered Data Example

The Television School and Family Smoking Prevention and Cessation Project (TVSFP) study (Flay et al., 1988) was designed to test independent and combined effects of a school-based social-resistance curriculum and a television-based program in terms of tobacco use prevention and cessation. The initial study sample consisted of seventh-grade students who were pretested in January 1986. Students who took the pretest completed an immediate post-intervention questionnaire in April 1986, a 1-year follow-up questionnaire (April 1987), and a second year follow-up (April 1988). The study involved students of schools from Los Angeles and San Diego. Randomization to various design conditions was at the school level, while much of the intervention was delivered to students within classrooms.

For this illustration of the random-effects ordinal probit regression model, a subset of the TVSFP data was used. We concentrated on students from 28 Los Angeles schools, where the schools were randomized to one of four study conditions: (a) a social-resistance classroom curriculum, (b) a media (television) intervention, (c) a social-resistance classroom curriculum combined with a mass-media intervention, and (d) a no-treatment control group. These conditions form a 2 × 2 design of social-resistance classroom curriculum (CC = yes or no) by mass-media intervention (TV = yes or no). A tobacco and health knowledge scale (THKS) score was one of the primary study outcome variables, and the one chosen for this analysis. The scale consisted of seven questionnaire items used to assess student tobacco and health knowledge. A student's score on this scale was defined as the sum of the items that the student answered correctly. Only data from the pretest and post-intervention timepoints were analyzed, so subjects were included if they had complete data on the tobacco and health knowledge scale at these two timepoints. In all, there were 1,600 students

from 135 classrooms and 28 schools who met these criteria. The resulting dataset was unbalanced with a range of between 1 and 13 classrooms per school, and between 2 and 28 students per classroom. The frequency distribution of the post-intervention THKS total scores suggested four ordinal classifications corresponding to 0–1, 2, 3, and 4–7 correct responses. Student frequencies for these categories of the THKS, broken down by condition subgroups, are given in Table 1.

Table 1
Tobacco and Health Knowledge Scale post-intervention scores: Subgroup frequencies (and percentages)

Subgroup		THKS score				
CC	TV	0–1	2	3	4–7	Total
no	no	117 (27.8)	129 (30.6)	89 (21.1)	86 (20.4)	421
no	yes	110 (26.4)	105 (25.2)	91 (21.9)	110 (26.4)	416
yes	no	62 (16.3)	78 (20.5)	106 (27.9)	134 (35.3)	380
yes	yes	66 (17.2)	86 (22.5)	114 (29.8)	117 (30.5)	383
Total		355 (22.2)	398 (24.9)	400 (25.0)	447 (27.9)	1,600

Two ordinal probit regression models were fit to these data. Results from these analyses are given in Table 2. In both cases, the post-intervention THKS score was modeled in terms of baseline THKS score, dummy-coded (no = 0 and yes = 1) effects of CC and TV, and the CC by TV interaction. The first column of Table 2 lists results for the ordinal probit regression analysis of the student-level data ignoring the clustering of students and treating each student as an independent observation. This analysis clearly indicates the positive effect of the social-resistance classroom curriculum as well as the television part of the intervention. However, the interaction of CC by TV is also observed to be statistically significant; thus, student-level analysis suggests that while TV intervention is effective in increasing THKS scores for those not receiving the CC component, it has a slight negative effect on those exposed to both components.

Table 2
THKS post-intervention (ordinal) scores: Ordinal probit estimates (standard errors)

Term	Ordinal probit	Ordinal probit RRM
Intercept μ	.0419 (.073)	.0622 (.091)
Baseline THKS α_1	.2472*** (.022)	.2427*** (.024)
Social resistance α_2	.5095*** (.079)	.5093*** (.112)
Television α_3	.1532** (.075)	.1237 (.100)
Interaction α_4	−.2312** (.109)	−.1937 (.150)
Threshold γ_2	.7347*** (.033)	.7606*** (.036)
Threshold γ_3	1.4388*** (.041)	1.4875*** (.044)
Class s.d. σ_β		.2616 (.045)
log L	−2,127.76	−2,117.72

***$p < .01$ **$p < .05$ *$p < .10$

The next column of Table 2 list results obtained from analyses using the random-effects ordinal probit regression model allowing for the nesting of students within classrooms. Results from this random-effects analysis are somewhat different from those obtained from ordinal probit regression analysis at the student-level. Notice that, unlike ordinary student-level analysis, random-effects analysis indicates that neither the TV effect nor the interaction of CC by TV is statistically significant at even the $p < .10$ level. Additionally, the variability attributable to the classes is highly significant and when expressed as an intraclass correlation equals .064, reflecting the degree of non-independence for this clustered dataset. Finally, the likelihood-ratio χ^2_1 equals 20.08, which clearly supports the significance of including the random classroom effect in the model.

5. Longitudinal Data Example

To illustrate application of the random-effects ordinal probit regression model to longitudinal data, we examined data collected in the NIMH Schizophrenia Collaborative Study on treatment-related changes in overall severity. Specifically, we examined Item 79 of the Inpatient Multidimensional Psychiatric Scale (IMPS; Lorr and Klett, 1966). Item 79, "Severity of Illness," was scored as: 1 = normal, not at all ill; 2 = borderline mentally ill; 3 = mildly ill; 4 = moderately ill; 5 = markedly ill; 6 = severely ill; and 7 = among the most extremely ill. Previously, we have analyzed these data both assuming a continuous scale for these seven ordered response categories using random-effects regression (Gibbons et al., 1988) and also dichotomizing responses using random-effects binary probit regression (Gibbons and Hedeker, 1994). For the present illustration of the ordinal random-effects model, we recoded the seven ordered categories into four: (1) normal or borderline mentally ill, (2) mildly or moderately ill, (3) markedly ill, and (4) severely or among the most extremely ill. In this study, patients were randomly assigned to receive one of four medications: placebo, chlorpromazine, fluphenazine, or thioridazine. Since our previous analyses revealed similar effects for the three anti-psychotic drug groups, they were combined in the present analysis. Focusing on patients who were assessed at four study timepoints, there were 64 placebo and 249 drug patients measured at baseline (week 0) and weeks 1 and 6. Additionally, all of these placebo patients and most of the drug patients ($n = 244$) were assessed at week 3, with the 5 remaining drug patients having been assessed at either week 2 ($n = 2$) or week 4 ($n = 3$). Finally, again based on previous analysis, to linearize the relationship of the IMPS79 scores over time, a square root transformation of time was chosen.

Table 3
Severity of illness across time: Ordinal probit RRM estimates (standard errors)

	Random intercepts		Random intercepts and slopes	
Term	Parameter	Estimate	Parameter	Estimate
Intercept	μ	3.264 (.245)	μ_{β_1}	3.927 (.326)
Time	α_1	−.454 (.081)	μ_{β_2}	−.528 (.151)
Drug	α_2	.234 (.227)	α_1	.333 (.289)
Drug by time	α_3	−.663 (.086)	α_2	−.849 (.165)
Threshold	γ_2	1.792 (.084)	γ_2	2.206 (.126)
Threshold	γ_3	3.081 (.108)	γ_3	3.735 (.164)
variance terms				
Intercept (Cholesky)	σ_β	1.117 (.077)	σ_{β_1}	1.476 (.150)
Covariance (Cholesky)			$\sigma_{\beta_1\beta_2}/\sigma_{\beta_1}$	−.303 (.091)
Time (Cholesky)			$\sqrt{\sigma^2_{\beta_2} - (\sigma^2_{\beta_1\beta_2}/\sigma^2_{\beta_1})}$	.655 (.073)
log L		−1,321.78		−1,295.49

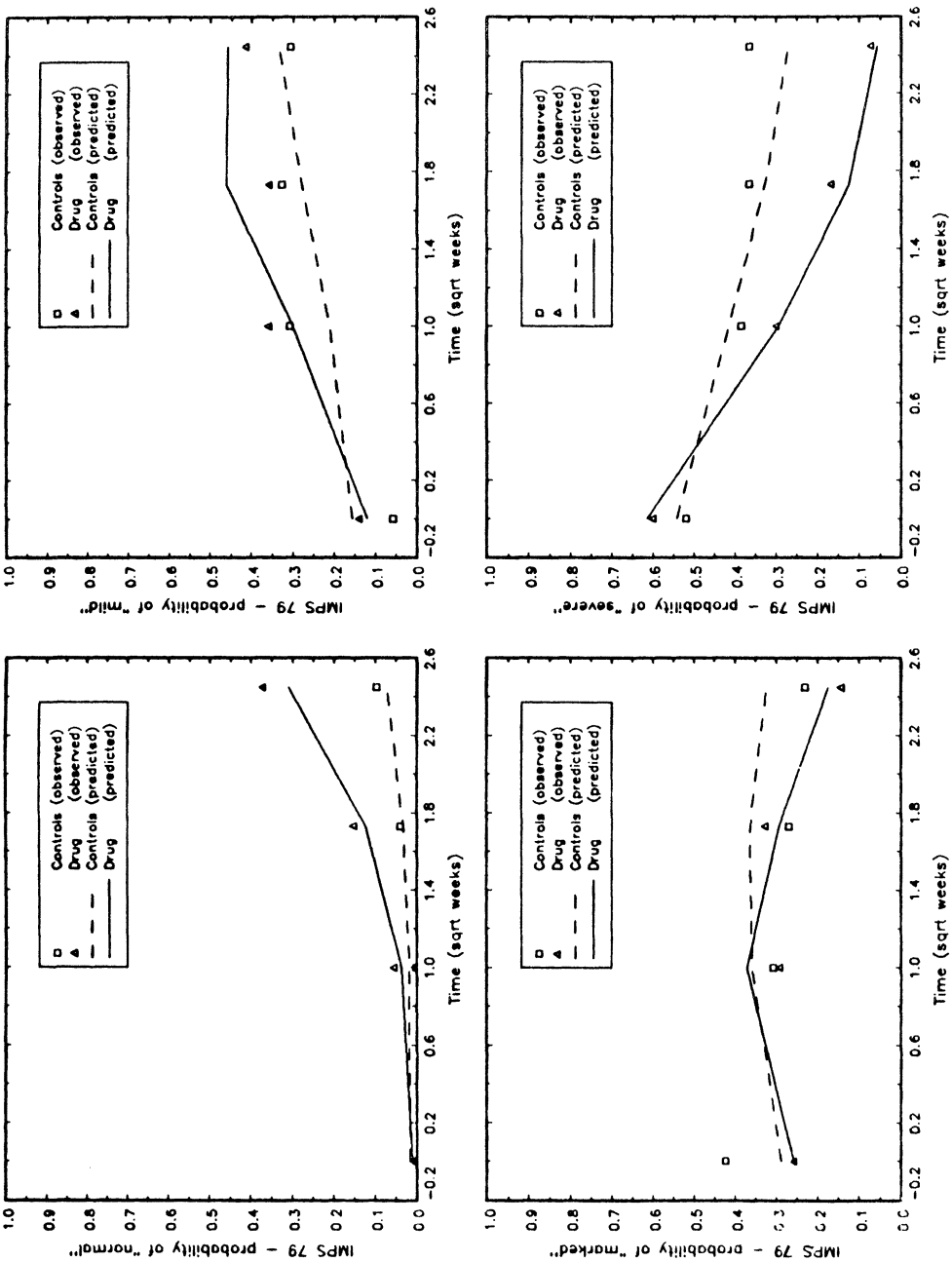

Figure 1. Plot of IMPS79 severity by time: Predicted and observed proportions.

Two random-effects ordinal probit regression models were fit to these data. Results from these analyses are given in Table 3. For both models, the repeated IMPS score was modeled in terms of a dummy-coded drug effect (placebo = 0 and drug = 1), a time effect (square root of week) and a drug by time interaction. In terms of the random effects, in the first model we simply fit a random-intercepts model, whereas in the second model we allowed patients to vary in terms of both their intercept and their trend across time (random intercepts and slopes).

The results from the two models given in Table 3 generally agree in terms of indicating that the treatment groups do not significantly differ at baseline, the placebo group does improve over time, and that the drug group has greater improvement over time, relative to the placebo group. However, contrasting the two models, there is clear evidence of significant variation in both the individual intercept and linear time-trend (likelihood ratio $\chi^2_2 = 52.58$). Also, significant negative association between the intercept and linear time terms is indicated, suggesting that those patients with the highest initial severity show the greatest improvement across time (e.g., largest negative time-trends). This latter finding could be a result of a "floor effect," in that patients with low initial severity scores cannot exhibit large negative time-trends due to the limited range in the ordinal outcome variable. These results generally agree with the results reported earlier (Gibbons et al., 1988) when these data were analyzed using a model assuming a continuous response variable. It can be argued, however, that the current analysis is more appropriate for rating scale-type data like these, where assuming an ordinal level of measurement is more reasonable than the interval level of measurement assumed by the continuous-variable model. Finally, the fit of the model to the actual data is depicted in Figure 1. This figure shows reasonably good model fit to the proportions observed across the four response categories for the two treatment groups over the timepoints of the study. It is important to note that these plots do not represent the degree of model fit to the unique patterns of ordinal responses across time $\mathbf{Y}_i$ (there being $4^4 = 256$ possible response patterns for each treatment group), but rather, the fit of the model to the observed proportion of subjects within each group responding in each of the ordered categories at each of the specific timepoints. As such, plotting the estimated curves (which are based on each subject's response vector $\mathbf{Y}_i$) against the observed proportions (which ignore the time-relatedness of the subject's responses) provides only a visual heuristic for the degree of model fit.

6. Discussion

A random-effects ordinal regression model is proposed for the analysis of multilevel ordinal response data, using either the probit or logistic response function. Maximum marginal likelihood methods are used to estimate the model parameters. For this solution, Gauss–Hermite quadrature is utilized to numerically integrate over the distribution of random effects. For multilevel data with two levels, this model allows for multiple random effects at the second level (of level-1 variables). Fixed covariates can be included in the model at either level of the data, and the general model can be partitioned into between and within level-2 unit submodels.

As noted, the solution via quadrature can involve summation over a large number of points when the number of random effects is increased. An issue, then, is the number of necessary quadrature points to ensure accurate estimation of the model parameters. As Jansen (1990) noted in the unidimensional quadrature solution, the estimation is affected very little when the number of points is 5 or greater. Also, as suggested by Bock et al. (1988) in the context of a dichotomous factor analysis model, the number of points in each dimension can be reduced as the dimensionality is increased. These authors noted that as few as three points per dimension were necessary for a five-dimensional solution. In the present examples, we used 10 quadrature points per dimension for both the uni- and bidimensional solutions. More work on this issue in the context of random-effects regression models with more than two random effects is currently under way.

As there is more readily available software for measurement data, it is important to note advantages of using the ordinal model presented here, rather than simply analyzing ordinal responses as measurements using the available software. As is well known, the probit or logistic specifications take into account the ceiling and floor effects of the dependent variable, whereas linear models for measurement data clearly do not. As McKelvey and Zavoina (1975) point out, due to the ceiling and floor effects of the dependent variable, values of the residuals and regressors will be correlated when linear models for measurement data are applied to ordinal outcomes, which can result in biased estimates of the regression coefficients. Furthermore, as Winship and Mare (1984) note, the advantage of ordinal regression models in accounting for ceiling and floor effects of the dependent variable is most critical when the dependent variable is highly skewed, or when groups, defined by different covariate values, are compared which have widely varying skewness in the dependent variable.

The examples presented clearly illustrate the importance of the random-effects approach for both clustered and longitudinal ordinal outcome data. For clustered data, RRM is useful in accounting for variability attributable to data clustering, whereas for longitudinal data, RRM can accommodate individual trends across time. A further extension of the model is under way to allow for three-level data in order to accommodate clustered data where the clustered subjects are also repeatedly measured across time.

Acknowledgements

The authors wish to thank Dr R. Darrell Bock for many valuable discussions, comments, and suggestions during the preparation of this paper. The authors also thank Ohidul Siddiqui for computer programming assistance, Dr Brian Flay for use of the clustered data, Drs Nina Schooler and John Davis for use of the longitudinal data, and to several anonymous referees and the editor for their helpful comments. This work was supported by National Institutes of Mental Health Grant MH44826-01A2, University of Illinois at Chicago Prevention Research Center Developmental Project–CDC Grant R48/CCR505025, and National Institute of Drug Abuse Grant DA06307.

Résumé

Un modèle de régression ordinal à effets aléatoires est proposé pour l'analyse de données en classes ou de réponses ordinales longitudinales. Ce modèle est développé pour les fonctions-réponses probit et logistiques. Le concept de seuil est utilisé par l'intermédiaire duquel on postule que la classe ordonnée observée est déterminée par la valeur d'une réponse continue sous-jacente non observable qui suit un modèle linéaire de régression incluant des effets aléatoires. Une solution du maximum de vraisemblance marginal est décrite en utilisant la quadrature de Gauss–Hermite pour l'intégration numérique de la distribution des effets aléatoires. L'analyse d'un ensemble de données où des étudiants sont affectés dans des classes est utilisée pour illustrer les caractéristiques des analyses à effets aléatoires de données ordinales regroupées alors qu'une analyse d'un ensemble de données longitudinales où des patients en psychiatrie sont classés plusieurs fois suivant leur sévérité est utilisée pour illustrer les caractéristiques de l'approche "effets aléatoires" sur des données ordinales longitudinales.

References

Anderson, D. and Aitkin, M. (1985). Variance components models with binary response: Interviewer variability. *Journal of the Royal Statistical Society, Series B* **47,** 203–210.

Bock, R. D. (1983). The discrete Bayesian. In *Modern Advances in Psychometric Research*, H. Wainer and S. Messick (eds). Hillsdale, New Jersey: Earlbaum.

Bock, R. D. (1989). Measurement of human variation: A two-stage model. In *Multilevel Analysis of Educational Data*, R. D. Bock (ed.). San Diego, California: Academic Press.

Bock, R. D. and Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: An application of the EM algorithm. *Psychometrika* **46,** 443–459.

Bock, R. D., Gibbons, R. D., and Muraki, E. (1988). Full-information item factor analysis. *Applied Psychological Measurement* **12,** 261–280.

Bryk, A. S. and Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin* **101,** 147–158.

Burstein, L., Linn, R. L., and Capell, I. (1978). Analyzing multi-level data in the presence of heterogeneous within-class regressions. *Journal of Educational Statistics* **4,** 347–389.

Conaway, M. R. (1989). Analysis of repeated categorical measurements with conditional likelihood methods. *Journal of the American Statistical Association* **84,** 53–61.

Ezzet, F. and Whitehead, J. (1991). A random effects model for ordinal responses from a crossover trial. *Statistics in Medicine* **10,** 901–907.

Flay, B., Brannon, B., Johnson, C., Hansen, W., Ulene, A., Whitney-Saltiel, D., Gleason, L., Sussman, S., Gavin, M., Kimarie, G., Sobol, D., and Spiegel, D. (1988). The Television School and Family Smoking Prevention and Cessation Project: I. Theoretical Basis and Program Development. *Preventive Medicine* **17,** 585–607.

Gibbons, R. D. and Bock, R. D. (1987). Trend in correlated proportions. *Psychometrika* **52,** 113–124.

Gibbons, R. D. and Hedeker, D. (1994). Application of random-effects probit regression models. *Journal of Consulting and Clinical Psychology* **62,** 285–296.

Gibbons, R. D., Waternaux, C., Hedeker, D., and Davis, J. M. (1988). Random regression models: A comprehensive approach to the analysis of longitudinal psychiatric data. *Psychopharmacology Bulletin* **24,** 438–443.

Goldstein, H. (1987). *Multilevel Models in Educational and Social Research*. New York: Oxford University Press.

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

Harville, D. A. and Mee, R. W. (1984). A mixed-model procedure for analyzing ordered categorical data. *Biometrics* **40,** 393–408.

Jansen, J. (1990). On the statistical analysis of ordinal data when extravariation is present. *Applied Statistics* **39,** 75–84.

Jennrich, R. I. and Schluchter, M. D. (1986). Unbalanced repeated-measures models with structured covariance matrices. *Biometrics* **42,** 805–820.

Laird, N. M. and Ware, J. H. (1982). Random effects models for longitudinal data. *Biometrics* **38,** 963–974.

Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73,** 13–22.

Longford, N. T. (1987). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. *Biometrika* **74,** 817–827.

Lorr, M. and Klett, C. J. (1966). *Inpatient Multidimensional Psychiatric Scale: Manual*. Palo Alto, California: Consulting Psychologists Press.

Magnus, J. R. (1988). *Linear Structures*. London: Charles Griffin.

McCullagh, P. (1980). Regression models for ordinal data (with Discussion). *Journal of the Royal Statistical Society, Series B* **42,** 109–142.

McKelvey, R. D. and Zavoina, W. (1975). A statistical model for the analysis of ordinal level dependent variables. *Journal of Mathematical Sociology* **4,** 103–120.

Qu, Y., Williams, G. W., Beck, G. J., and VanderBrug-Medendorp, S. (1992). Latent variable models for clustered dichotomous data with multiple subclusters. *Biometrics* **48,** 1095–1102.

Stiratelli, R., Laird, N. M., and Ware, J. H. (1984). Random-effects models for serial observations with binary response. *Biometrics* **40,** 961–971.

Stroud, A. H. and Sechrest, D. (1966). *Gaussian Quadrature Formulas*. Englewood Cliffs, New Jersey: Prentice Hall.

Winship, C. and Mare, R. D. (1984). Regression models with ordinal variables. *American Sociological Review* **49,** 512–525.

Wong, G. Y. and Mason, W. M. (1985). The hierarchical logistic regression model for multilevel analysis. *Journal of the American Statistical Association* **80,** 513–524.

Zeger, S. L. and Liang, K. Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* **42,** 121–130.

Received September 1992; *revised April* 1993; *accepted July* 1993.