Journal of Probability and Statistical Science
4(1), 1-20, Feb. 2006

Location-Scale Models for Multilevel Ordinal Data: Between- and Within-Subjects Variance Modeling

Donald Hedeker Michael Berbaum Robin Mermelstein
University of Illinois at Chicago

ABSTRACT Mixed-effects logistic regression models are described for analysis of two-level ordinal outcomes, where observations are observed clustered within subjects. Random effects are included in the model to account for the correlation of the clustered observations. This correlation can be the same for all subjects or allowed to vary by groups of subjects. Additionally, whereas the usual logistic model assumes that the covariate effects are the same across the cumulative logits, *i.e.*, proportional odds assumption, we describe two extensions to relax this assumption. The first permits separate covariate effects to be estimated for each of the $C - 1$ cumulative logits, where C = number of ordered categories. The second extension instead allows covariates to influence the scale of the ordinal response, in addition to their usual influence on the location. This latter extension can be more parsimonious and can be used to partition the degree of within- and between-subjects variance. An analysis is presented of a dataset from an adolescent smoking study, highlighting and comparing these extensions of the proportional odds mixed model.

Keywords Categorical data; Multilevel data; Proportional odds assumption; Logistic regression; Clustering; Repeated observations; Complex variance; Mixed-effects models.

1. Introduction

The ordinal logistic regression model, described as the proportional odds model by McCullagh [36], is a popular model for analyzing ordinal outcomes. For multilevel data, where observations are nested within clusters (*e.g.*, classes, schools, clinics) or are repeatedly

☐ Received July 2005, revised December 2005, in final form January 2006.
☐ Donald Hedeker and Michael Berbaum are affiliated with the Division of Epidemiology and Biostatistics, and the Institute for Health Research and Policy, in the School of Public Health; Robin Mermelstein is affiliated with the Institute for Health Research and Policy, and the Department of Psychology; University of Illinois at Chicago, Chicago, IL 60612, USA; website: www.uic.edu/~hedeker/ml.html.

 ISSN 1726-3328

JPSS Vol. 4 No. 1 February 2006 pp. 1-20

assessed across time, mixed-effects regression models are often used to account for the dependency inherent in the data [31, 8, 17], and several authors have developed mixed-effects models for ordinal outcome data. Ezzet and Whitehead [10] and Agresti and Lang [1] describe random-intercepts proportional odds models. Hedeker and Gibbons [24] describe both an ordinal logistic and probit model with multiple random effects. Tutz and Hennevogl [58] propose similar mixed models that additionally allow the model thresholds to be considered as random effects. Besides use of the logit link function, other authors have developed ordinal mixed models utilizing the probit link [22, 30, 39, 51] and complementary log-log link [54] as well.

An alternative method of dealing with correlated data is provided by the generalized estimating equations (GEE) approach described by Liang and Zeger [32]. This approach is most often used when interest is on the fixed effects of the model and the correlation structure of the multilevel data is considered a nuisance, although extended versions (*i.e.*, EGEE and GEE2) have been developed for situations where the correlation structure of the multilevel data is also of interest [20, 33]. For ordinal data, Miller *et al.* [41] and Gange *et al.* [14] describe GEE-based models for repeated ordinal outcomes. Heagerty and Zeger [23] extend this approach and also describe a GEE2 formulation for correlated ordinal outcomes.

Most of the models for ordinal outcomes referenced above include the proportional odds assumption (or its equivalent under the probit or complementary log-log link function) for model covariates. For an ordinal response with C categories, this assumption states that the effect of an explanatory variable is the same across the $C-1$ cumulative logits of the model, or proportional across the cumulative odds. In previous papers [26-27], we have described an extension to the mixed proportional odds model to allow for non-proportional odds for a subset of the explanatory variables. A similar extension is described in Saei and McGilchrist [50], who allow non-proportional time effects in panel studies. These developments follow the extension due to Peterson and Harrell [43] of the fixed-effects proportional odds model. In this model, explanatory variables are allowed to have varying effects on the $C-1$ cumulative logits. Thus, for a particular explanatory variable $C-1$ regression coefficients are estimated. These additional parameters reflect different location effects of the explanatory variables. This extended model has recently been applied successfully in several articles [59, 60, 13, 12], and a similar Bayesian multilevel model is described in Ishwaran [28]. Fielding *et al.* [12] additionally allow the random-effect parameters to have non-proportional effects.

A somewhat different extention of the proportional odds model is described by Tosteson and Begg [57]. Here, in the context of ROC analysis, the *scale* of the effects of explanatory variables are allowed to vary. In other words, the underlying variance of the logistic distribution can vary as a function of model covariates. McCullagh and Nelder [37] refer to this extended model for ordinal data as a generalized "rational" model. Toledano and Gatsonis [56] use this extension in describing GEE analysis of correlated ROC data, while Ishwaran and Gatsonis [29] build upon this approach using Bayesian methods.

For cross-sectional data, Cox [9] brought together these extensions of the proportional odds model into what he termed location-scale cumulative odds models. In this paper, we will describe this approach within a mixed model framework. The inclusion of scale parameters within the mixed model is particularly advantageous because it allows a modeling of both the within- and between-subjects variances. Data from an adolescent smoking study, in which subjects were repeatedly measured across time, are used to illustrate the mixed location-scale cumulative odds models for ordinal data.

2. Mixed Cumulative Logit Model for Ordinal Responses

Let $i = 1, \ldots, N$ denote the level-2 units (subjects), $j = 1, \ldots, n_i$ level-1 units (repeated observations), and $c = 1, 2, \ldots, C$ denote the response categories. Then, Y_{ij} = the ordinal response associated with level-2 unit i and level-1 unit j. As ordinal response models often utilize cumulative comparisons of the ordinal outcome, define the cumulative probabilities for the C categories of the ordinal outcome Y as $P_{ijc} = \Pr(Y_{ij} \le c) = \sum_{k=1}^{c} p_{ijk}$, where p_{ijk} denote the individual category probabilities. The mixed-effects logistic regression model for the cumulative probabilities [24] is given in terms of the cumulative logits λ_{ijc} as

$$\lambda_{ijc} = \log\left[\frac{P_{ijc}}{1 - P_{ijc}}\right] = \gamma_c - \left[\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{\upsilon}_i\right] \quad (c = 1, \ldots, C-1)\,, \tag{1}$$

where $\boldsymbol{x}_{ij}$ is the $p \times 1$ covariate vector, $\boldsymbol{\beta}$ is the $p \times 1$ vector of regression parameters, $\boldsymbol{z}_{ij}$ is the $r \times 1$ vector of random-effect variables, and $\boldsymbol{\upsilon}_i$ is the $r \times 1$ vector of random effects for subject i. The model also includes $C-1$ strictly increasing model thresholds γ_c (*i.e.*, $\gamma_1 < \gamma_2 \ldots < \gamma_{C-1}$).

As written above, a positive coefficient for a regressor indicates that as values of the regressor increase so do the odds that the response is greater than or equal to c. This agrees with the formulation in McCullagh [36], though it is not the only way to write the ordinal model (*e.g.*, see Fielding *et al.* [12]). Also, note that the relationship between the regressors $\boldsymbol{x}$ and the cumulative logits does not depend on c (hence the regression coefficients $\boldsymbol{\beta}$ do not carry the c subscript). McCullagh [36] calls this assumption of identical odds ratios across the $C-1$ cut-offs the proportional odds assumption.

The random effects $\boldsymbol{\upsilon}_i$ are assumed to follow a multivariate normal distribution with mean vector $\mathbf{0}$ and variance-covariance matrix $\boldsymbol{\Sigma}_{\upsilon}$. Standardizing the multiple random effects $\boldsymbol{\upsilon}_i = \boldsymbol{T}\boldsymbol{\theta}_i$, where $\boldsymbol{T}\boldsymbol{T}' = \boldsymbol{\Sigma}_{\upsilon}$ is the Cholesky decomposition, yields

$$\lambda_{ijc} = \gamma_c - \left[\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{T}\boldsymbol{\theta}_i\right] \quad (c = 1, \ldots, C-1)\,, \tag{2}$$

where $\boldsymbol{\theta}_i$ follow a standard multivariate normal. Notice that the random-effects variance terms (*i.e.*, the parameters in the Cholesky decomposition $\boldsymbol{T}$) are now explicitly included in the regression model. Thus, these terms and the regression coefficients are on the same scale, namely, in terms of the logit of response.

Ordinal regression models are often motivated and described using the "threshold concept" [4, 38]. This is also termed a latent variable model for ordinal variables [34]. For this, it is assumed that a continuous latent variable y underlies the observed ordinal response Y. Typically, the continuous latent variable y is assumed to follow either a normal or logistic distribution, leading to ordinal probit or logistic regression, respectively. Under the threshold concept, the observed ordinal outcome $Y_{ij} = c$ if $\gamma_{c-1} \leq y_{ij} < \gamma_c$ for the latent variable (with $\gamma_0 = -\infty$ and $\gamma_C = \infty$). In this article, we will consider the logistic formulation, though we will note the necessary changes for the normal, or probit, formulation.

2.1 Partial Proportional Odds

As noted by Peterson and Harrell [43], violation of the proportional odds assumption is not uncommon. Thus, they described a (fixed-effects) partial proportional odds model in which covariates are allowed to have differential effects on the $C - 1$ cumulative logits. Terza [55] developed a similar extension of the (fixed-effects) ordinal probit model. Hedeker and Mermelstein [26-27] utilize this extension within the context of a mixed ordinal regression model. For this, the model for the $C - 1$ cumulative logits can be written as:

$$\lambda_{ijc} = \gamma_c - \left[\boldsymbol{u}'_{ij}\boldsymbol{\alpha}_c + \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{T}\boldsymbol{\theta}_i\right] \quad (c = 1, \ldots, C-1) \,, \tag{3}$$

where, $\boldsymbol{u}_{ij}$ is a $q \times 1$ vector containing the values of observation ij on the set of q covariates for which proportional odds is not assumed. In this model, $\boldsymbol{\alpha}_c$ is a $q \times 1$ vector of regression coefficients associated with these q covariates. The effects of these q covariates are allowed to vary across the $C - 1$ cumulative logits (note that $\boldsymbol{\alpha}_c$ carries the c subscript).

A caveat about this extension of the proportional odds model should be mentioned. The effects on the cumulative log odds, namely $\boldsymbol{u}'_{ij}\boldsymbol{\alpha}_c$, result in $C - 1$ non-parallel regression lines. These regression lines inevitably cross for some values of $\boldsymbol{u}$, leading to negative fitted values for the response probabilities. For $\boldsymbol{u}$ variables contrasting two levels of an explanatory variable (*e.g.*, gender coded as 0 or 1), this crossing of regression lines occurs outside the range of admissible values (*i.e.*, < 0 or > 1). However, if the explanatory variable is continuous, this crossing can occur within the range of the data, and so, allowing for non-proportional odds can lead to problematic results. For continuous explanatory variables, other than requiring proportional odds, a solution to this dilemma is sometimes possible if the variable has, say, m levels with a reasonable number of observations at each of these m levels. In this case $m - 1$ dummy-coded variables can be created and substituted into the model in place of the continuous variable.

2.2 Scale Parameters

Ordinal regression models including scaling effects are common in ROC analysis, as described by Tosteson and Begg [57]. Here we will add these terms within the mixed model,

$$\lambda_{ijc} = \frac{\gamma_c - (\boldsymbol{u}'_{ij}\boldsymbol{\alpha}_c + \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{T}\boldsymbol{\theta}_i)}{\exp(\boldsymbol{w}'_{ij}\boldsymbol{\tau})}$$

where, $\boldsymbol{w}_{ij} = k \times 1$ vector for the set of k covariates which influence the scale, and $\boldsymbol{\tau}$ are their corresponding effects. Note that the $\boldsymbol{w}$ variables can be the same as the $\boldsymbol{x}$ variables, though, for identification, an intercept cannot be included as one of the $\boldsymbol{w}$ variables. When scaling effects are added to the ordinal model, the result is a non-proportional odds model, using only one additional parameter for each covariate.

2.3 Random Effects Specifications

Several popular random effects specifications are worth noting. First, for many problems it is common to simply specify a random intercept for each subject. For this, $\boldsymbol{z}_{ij} = 1$, $\boldsymbol{\theta}_i = \theta_i$, and the Cholesky $\boldsymbol{T}$ is simply the scalar σ_υ. In such models, σ_υ^2 represents the between-subjects variance and indicates the degree of heterogeneity in the population of subjects.

A slight extension of this random-intercepts model is to allow separate intercepts for groups of subjects. For example, suppose that one was interested in allowing the between-subjects variance to vary across gender groups. For this, in terms of the model, $\boldsymbol{z}_{ij} = [\mathrm{M}_i \;\; \mathrm{F}_i]$ (where M_i and F_i are subject-level indicator variables of male and female gender, respectively), $\boldsymbol{\theta}_i = \theta_i$, and $\boldsymbol{T} = [\sigma_{\upsilon_M} \;\; \sigma_{\upsilon_F}]'$. Notice that there is still only one random effect, but its variance differs for males and females. Models allowing the (scalar) random effect θ to be linearly related to a vector of variables $\boldsymbol{z}_{ij}$ have been termed multilevel models with complex variance structures by Goldstein [17].

A similar complex variance structure, in some respects, is used for the item response theory (IRT) model of educational testing [35]. Here, the items (level-1) are nested within subjects (level-2), and the indicator variables are at level one. Specifically, $\boldsymbol{z}_{ij} = [\mathrm{I}1_{ij} \;\; \mathrm{I}2_{ij} \;\; \dots \;\; \mathrm{I}n_{ij}]$ is a vector of item indicator variables, $\boldsymbol{\theta}_i = \theta_i$, and $\boldsymbol{T}$ is a vector of item standard deviations, $[\sigma_{\upsilon_{I1}} \;\; \sigma_{\upsilon_{I2}} \;\; \dots \;\; \sigma_{\upsilon_{In}}]'$. The so-called Rasch model assumes that these standard deviations are the same, whereas the more general two-parameter IRT model does not [5]. These standard deviations are akin to the factor loadings of the items on the latent ability factor denoted in the psychometric literature as θ.

While the above specifications only require a single random effect, for longitudinal studies, where repeated observations are nested within subjects, it is common to include multiple random effects (*e.g.*, random subject intercepts and time trends). For example, for dichotomous data, Gibbons and Bock [15] proposed a random intercept and trend model by specifying $\boldsymbol{z}_{ij} = [1 \;\; t]$, where t = value of time. In their model, the random effects vector is $\boldsymbol{\theta}_i = [\theta_{i1} \;\; \theta_{i2}]$, and the Cholesky factor of $\boldsymbol{\Sigma}_\upsilon$ equals:

$$\boldsymbol{T} = \begin{bmatrix} \sigma_{\upsilon_1} & 0 \\ \frac{\sigma_{\upsilon_{12}}}{\sigma_{\upsilon_1}} & \left(\sigma_{\upsilon_2}^2 - \frac{\sigma_{\upsilon_{12}}^2}{\sigma_{\upsilon_1}^2}\right)^{1/2} \end{bmatrix}.$$

Of course, one could have more than two random subject effects to allow for higher-order

JPSS Vol. 4 No. 1 February 2006 pp. 1-20

polynomial trends (*e.g.*, quadratic, cubic, etc.). The model presented in this article allows any of these random-effects specifications, though our example will center on the random-intercepts model and models with complex variance structure.

2.4 Intraclass Correlation and Partitioning of Between- and Within-Cluster Variance

For a random-intercepts model (*i.e.*, $z_i = \mathbf{1}_i$) it is often of interest to express the between-subjects variance in terms of an intraclass correlation (ICC). The intraclass correlation indicates the proportion of unexplained variance at the subject level; in other words, it reflects the magnitude of the between-subject variance. Reference is made to the threshold concept and the underlying latent response tendency that determines the observed ordinal response. For a logistic regression model, the latent response tendency, which is unobserved, is assumed to follow a standard logistic distribution, which has variance equal to $\pi^2/3$. Thus, for the logistic model assuming normally distributed random effects, the ICC equals $\sigma_\upsilon^2/(\sigma_\upsilon^2 + \pi^2/3)$, where the latter term in the denominator represents the variance of the underlying latent response tendency, namely, the within-subjects variance.

If one allows the random intercept to vary across groups of subjects, like males and females as described above, then separate ICCs are obtained for each gender group. Here, $z_{ij} = [\mathrm{M}_i \;\; \mathrm{F}_i]$ (where M_i and F_i are subject-level indicator variables), $\boldsymbol{T} = [\sigma_{\upsilon(M)} \;\; \sigma_{\upsilon(F)}]'$, and $\boldsymbol{\theta}_i = \theta_i$. The between-subjects variance is then different for these two subject groups, and the resulting ICCs are

$$ICC_M = \frac{\sigma_{\upsilon(M)}^2}{\sigma_{\upsilon(M)}^2 + \pi^2/3} \quad \text{and} \quad ICC_F = \frac{\sigma_{\upsilon(F)}^2}{\sigma_{\upsilon(F)}^2 + \pi^2/3} \ .$$

Notice, that these ICCs assume the same within-subjects variance for both groups.

The inclusion of scaling terms in the ordinal mixed model additionally allows an examination of the degree to which the within-subjects variance differs across subject groups. For instance, continuing with the gender example, suppose that a model with $w_{ij} = F_i$ was fit. In this case, the within-cluster variance for males would be $\pi^2/3$, whereas it would be $(\exp\tau)^2\ \pi^2/3$ for females. The resulting intraclass correlations would then be

$$ICC_M = \frac{\hat{\sigma}_{\upsilon(M)}^2}{\hat{\sigma}_{\upsilon(M)}^2 + \pi^2/3} \quad \text{and} \quad ICC_F = \frac{\hat{\sigma}_{\upsilon(F)}^2}{\hat{\sigma}_{\upsilon(F)}^2 + \left[(\exp\tau)^2\ \pi^2/3\right]} \ .$$

Thus, the model allows both heterogeneous between- and within-subjects variance by the inclusion of parameters in $\boldsymbol{T}$ and $\boldsymbol{\tau}$, respectively.

When there are more than two groups, it may be of interest and/or parsimonious to estimate a trend in the variances across the k groups. This makes sense if the grouping variable reflects some kind of ordering of subjects. For example, suppose that a grouping variable g_i is ordered as 0=low, 1=med, and 2=high in terms of some attribute. Then to allow the between-subjects variance to increase across these three groups, one could specify an intercept and the variable

g_i as two independent random effects. Here,

$$T = \begin{bmatrix} \sigma_{\upsilon_0} & 0 \\ 0 & \sigma_{\upsilon_g} \end{bmatrix}, \quad (4)$$

where σ_{υ_0} represents the intercept variation and σ_{υ_g} reflects how the between-subjects variation varies across groups. Note that the between-subjects (BS) variance is equal to a function of these two parameters,

$$\text{BS variance} = \sigma_{\upsilon_0}^2 + g_i^2 \sigma_{\upsilon_g}^2 , \quad (5)$$

which shows that the coding of g_i is important. For example, coding of 0, 1, and 2 for three groups would assume that the BS variance increases across the groups, because variances can never be negative. Alternatively, reverse coding (*i.e.*, 2, 1, 0) could be used if the BS variance decreases across the groups.

For estimating a trend in the within-subjects (WS) variance across subject groups, a log-linear representation is often used in ordinary multiple regression [21, 2], and this is similar to the implementation of scaling terms in the model described here. For example, in terms of a grouping variable g_i, the current model posits

$$\text{WS variance} = \left[\pi/\sqrt{3}\,\exp(g_i \tau)\right]^2 . \quad (6)$$

Notice that, this representation allows for both increasing and decreasing WS variance across groups. Namely, if $\tau > 0$ then variance increases over groups, while if $\tau < 0$ then it decreases over the groups. Including trends in both the BS and WS variance across groups, the ICC for a particular group, coded g_i, would equal:

$$\text{ICC} = \frac{\sigma_{\upsilon_0}^2 + g_i^2 \sigma_{\upsilon_g}^2}{\sigma_{\upsilon_0}^2 + g_i^2 \sigma_{\upsilon_g}^2 + \pi^2/3\,[\exp(g_i \tau)]^2} . \quad (7)$$

More generally, this modeling of the WS variance allows it to depend on multiple variables, namely,

$$\text{WS variance} = \left[\pi/\sqrt{3}\,\exp(\boldsymbol{w}_{ij}' \boldsymbol{\tau})\right]^2 . \quad (8)$$

This regression-like structure allows estimation of separate WS variance by group, as well as more complicated multiple regression-like forms for the WS variance. By combining modeling of the WS variance with the inclusion of random effects in the mixed model (to model the BS variance), a wide variety of heterogeneous variance models can be estimated and compared.

3. Estimation

Parameter estimation can be solved using maximum likelihood (ML) or variants of ML. Such solutions are iterative ones that can be numerically quite intensive. Here, the approach is sketched; further details can be found in Hedeker and Gibbons [24], Hedeker and Mermelstein [26], and Fahrmeir and Tutz [11]. Let $\mathbf{Y}_i$ denote the vector of responses from subject i. The probability of any response pattern $\mathbf{Y}_i$ (of size n_i), conditional on the random effects $\boldsymbol{\theta}$, is equal to the product of the probabilities of the level-1 responses:

$$\ell(\mathbf{Y}_i \mid \boldsymbol{\theta}_i) = \prod_{j=1}^{n_i} \prod_{c=1}^{C} \Pr(Y_{ij} = c \mid \boldsymbol{\theta}_i) , \tag{9}$$

where

$$\Pr(Y_{ij} = c \mid \boldsymbol{\theta}_i) = \Psi(\lambda_{ijc}) - \Psi(\lambda_{ij,c-1}) , \tag{10}$$

and $\Psi(\cdot)$ represents the logistic cumulative distribution function (cdf). The assumption that a subject's responses are independent given the random effects (and therefore can be multiplied to yield the conditional probability of the response vector) is known as the *conditional independence* assumption. The marginal density of $\mathbf{Y}_i$ in the population is expressed as the following integral of the conditional likelihood $\ell(\cdot)$

$$h(\mathbf{Y}_i) = \int_{\boldsymbol{\theta}} \ell(\mathbf{Y}_i \mid \boldsymbol{\theta}_i) \; f(\boldsymbol{\theta}) \; d\boldsymbol{\theta} , \tag{11}$$

where $f(\boldsymbol{\theta})$ represents the distribution of the random effects, namely the standard multivariate normal density. Whereas (9) represents the conditional probability, (11) indicates the unconditional probability for the response vector of subject i. The marginal log-likelihood from the sample of N subjects is then obtained as $\log L = \sum_i^N \log h(\mathbf{Y}_i)$. Maximizing this log-likelihood yields ML estimates, which are sometimes referred to as maximum marginal likelihood estimates because they are obtained by maximizing the marginal likelihood.

3.1 Integration Over the Random-Effects Distribution

In order to solve the likelihood solution, integration over the random-effects distribution must be performed. As a result, estimation is much more complicated than in models for continuous normally-distributed outcomes where the solution can be expressed in closed form. Various approximations for evaluating the integral over the random-effects distribution have been proposed in the literature; many of these are reviewed in Rodríguez and Goldman [49]. Perhaps the most frequently used methods are based on first- or second-order Taylor expansions. Marginal quasi-likelihood (MQL) involves expansion around the fixed part of the model, whereas penalized or predictive quasi-likelihood (PQL) additionally includes the random part in its expansion [18]. Unfortunately, these procedures yield estimates of the regression coefficients and random effects variances that are biased towards zero in certain situations, especially for the first-order expansions [7]. To remedy this, Raudenbush *et al.* [48] proposed an approach that uses a combination of a fully multivariate Taylor expansion and a Laplace approximation. This method yields accurate results and is computationally fast. Also, as opposed to the

MQL and PQL approximations, the deviance obtained from this approximation can be used for likelihood-ratio tests.

Numerical integration can also be used to perform the integration over the random-effects distribution [6]. Specifically, if the assumed distribution is normal, Gauss-Hermite quadrature can approximate the above integral to any practical degree of accuracy. Additionally, like the Laplace approximation, the numerical quadrature approach yields a deviance that can be readily used for likelihood-ratio tests. The integration is approximated by a summation on a specified number of quadrature points for each dimension of the integration. An issue with the quadrature approach is that it can involve summation over a large number of points, especially as the number of random-effects is increased. To address this, methods of adaptive quadrature have been developed that use a few number of points per dimension that are adapted to the location and dispersion of the distribution to be integrated [45].

3.2 Software

Several computer programs provide maximum likelihood estimates of the mixed-effects proportional odds model, including SAS PROC NLMIXED, Stata [53], HLM [47], MLwiN [46], LIMDEP [19], GLLAMM [44], Mplus [42], and MIXOR [25]). Some of these programs, however, only allow random-intercepts models, and few allow the inclusion of threshold interactions and scaling parameters described here. Also, these programs differ in how the integration over the random effects is performed. For the analyses presented in this article, SAS PROC NLMIXED, which utilizes adaptive quadrature for integration of the random effects, was used. This is perhaps the most reliable, though computationally demanding, solution. The syntax for the analyses presented in this article is available from the first author on request.

4. Illustration: Adolescent Smoking Study

This study, described in Mermelstein *et al.* [40], included 100 adolescents in 8th and 10th grades, with varying amounts of cigarette smoking experience. Adolescents were divided into three groups based on their reported lifetime smoking levels: 1) those who had smoked less than 6 cigarettes in their lifetimes (n = 18), representing very novice smokers; 2) those who had smoked between 6 and 99 cigarettes in their lifetimes (n = 48), representing a group of irregular or experimental smokers; and 3) those who had smoked 100 or more cigarettes during their lifetimes (n = 34), representing more regular smokers. Lifetime smoking levels were assessed at a baseline measurement wave.

For inclusion in the analyses reported here, all of these adolescents had smoked at least one cigarette during a 7-day data collection period which followed the baseline wave. Data collection occurred via hand-held (Palm Pilot) computers, which participants carried at all times during a week-long data collection wave. Participants were trained to complete a series of questions on the hand held computers immediately after smoking a cigarette. Of interest here are the responses to questions asking about subjective physiological sensations immediately after and prior to smoking a cigarette. Specifically, the subjects rated their subjective physiological sensations in terms of two items, `Sick` and `Buzz`, on a 10-point scale. The definition of these items varied slightly for the before and after assessments. Specifically, for the after assessment, they were prompted as:

- `Sick`: Think about how you feel right now: Do you feel sick?
- `Buzz`: Think about how you feel right now: Do you feel buzzed? whereas, for the before assessment the prompts were:
- `Sick`: Now think about the time just before you smoked: I felt sick.
- `Buzz`: Now think about the time just before you smoked: I felt buzzed.

For all four questions, individuals were instructed to rate their answers from 1 (not at all) to 10 (very). An individual's subjective physiological sensation was operationalized as their average based on the two items (*i.e.*, `Sick` and `Buzz`), and was calculated for both the before and after cigarette assessments. These were then expressed as a change score (after - before) to represent the changes in the level of subjective physiological sensations attributable to smoking a cigarette. The distribution of these change scores suggested a categorization in terms of five ordered categories: -2 for change score < -1, -1 for change score < 0 and ≥ -1, 0 for change score $= 0$, 1 for change score > 0 and ≤ 1, and 2 for change score > 1.

Because participants could smoke more than one cigarette during the 7-day data collection period, there were multiple observations per subject. In all, there was a total of 517 observations clustered within these 100 subjects. A question of interest is whether the changes in physiological sensations vary between smoking groups (defined by level of smoking experience). In particular, we will address this in terms of level of change (location), change of scale (within-subjects variance), and change of the intraclass correlation (between-subjects variance).

Table 1 presents the crosstabulation of smoking group by the ordinal outcome. It is important to realize that the data in this table are not independent, since subjects can have multiple outcomes, so this table provides a rough impression of the data. What the table makes clear is that the highest smoking group has a greater proportion of responses in the 0 category. This is consistent with models of the development of dependence that postulate that the physiological effects of smoking (like feeling sick or buzzed) are diminished as one advances in their smoking career (*i.e.*, tolerance develops). Thus, it appears that the variance of the ordinal outcome may not be the same across these groups.

To provide a further impression of the data, the cumulative odds and logits are presented in Table 2. The table also presents the difference in these cumulative logits between the smoking groups, using the lowest smoking group as the reference cell. If the proportional odds assumption were reasonable, then the logit differences would be the same across the cumulative comparisons. However, this does not seem to be the case, especially when comparing the lowest and highest smoking groups.

Table 1 Changes in physiological sensations by smoking group: frequencies and column proportions

smoking group	physiological sensations change -2	-1	0	1	2	total
< 6 cigs	8 (.160)	6 (.120)	14 (.280)	8 (.160)	14 (.280)	50
6-99 cigs	17 (.093)	28 (.153)	43 (.235)	36 (.197)	59 (.322)	183
100+ cigs	14 (.049)	45 (.159)	109 (.384)	56 (.197)	60 (.211)	284
total	39 (.075)	79 (.152)	166 (.321)	100 (.193)	133 (.257)	517

Table 2 Cumulative odds (logits) by smoking group

smoking group	cumulative comparison $\frac{(2 \text{ to } 5)}{1}$	$\frac{(3 \text{ to } 5)}{(1 \text{ to } 2)}$	$\frac{(4 \text{ to } 5)}{(1 \text{ to } 3)}$	$\frac{5}{(1 \text{ to } 4)}$
< 6 cigs	42/8 = 5.25 (1.66)	36/14 = 2.57 (.94)	22/28 = .79 (-.24)	14/36 = .39 (-.94)
6-99 cigs	166/17 = 9.76 (2.28)	138/45 = 3.07 (1.12)	95/88 = 1.08 (.08)	59/124 = .48 (-.74)
100+ cigs	270/14 = 19.29 (2.96)	225/59 = 3.81 (1.34)	116/168 = .69 (-.37)	60/224 = .27 (-1.32)
logit difference compared to < 6 cigs group				
6-99 cigs	.62	.18	.32	.20
100+ cigs	1.30	.40	-.13	-.38

To investigate this more formally, we examined a number of ordinal mixed logistic regression models. These are summarized in Table 3. In terms of the fixed effects, models included thresholds (four parameters for the five categories), group contrasts (two parameters for the three groups), and in some cases group by threshold interactions (eight parameters). For the random effects, we considered three possibilities: a random (subject) intercept (1 parameter);

a random intercept that varied linearly with the standard deviation of the random effect across smoking groups (2 parameters); and a random intercept with a separate variance for each smoking group (3 parameters). These three correspond to models of the between-subjects variation as homogeneous, linearly changing, or heterogeneous across the three smoking groups. Similarly, for the scaling terms, we considered three possibilities: no scaling terms; linearly changing scaling across smoking groups (1 parameter); or separate scaling terms across smoking groups (2 parameters). These correspond to models of the within-subjects variation as homogeneous, linear, or heterogenous across smoking groups. For the group contrasts, the first smoking group (low number of cigarettes) was the reference cell, and the linear trend term was coded as $-1, 0,$ and 1 for the three smoking groups. For each model, Table 3 presents values of the model deviance (−2 log likelihood), Akaike's information criterion (AIC; [3]), and Schwarz's Bayesian information criterion (BIC; [52]).

Table 3 Ordinal logistic mixed model comparisons

Model	fixed	p*	random	r*	scale	s*	deviance	AIC	BIC
Ia	T + G	6	I	1			1524.9	1538.9	1557.1
Ib	T + G	6	I	1	S	1	1515.7	1531.7	1552.6
Ic	T + G	6	I	1	G	2	1514.4	1532.4	1555.8
IIa	T + G	6	I + S	2			1520.4	1536.4	1557.3
IIb	T + G	6	I + S	2	S	1	1513.5	1531.5	1554.9
IIc	T + G	6	I + S	2	G	2	1511.6	1531.6	1557.7
IIIa	T + G	6	I + G	3			1520.4	1538.4	1561.9
IIIb	T + G	6	I + G	3	S	1	1513.5	1533.5	1559.5
IIIc	T + G	6	I + G	3	G	2	1511.6	1533.6	1562.2
IVa	T + G + TG	12	I	1			1510.1	1536.1	1570.0
IVb	T + G + TG	12	I	1	S	1	1509.4	1537.4	1573.9
IVc	T + G + TG	12	I	1	G	2	1508.9	1538.9	1578.0
V	T + G + TG	12	I + S	2			1509.5	1537.5	1574.0
VI	T + G + TG	12	I + G	3			1508.9	1538.9	1578.0

T = thresholds; G = group contrasts; S = linear trend across group; I = intercept

p* = total number of fixed-effects parameters

r* = total number of random-effects variance parameters

s* = total number of scaling parameters

deviance = $-2 \log L$; AIC = deviance + 2q*, (q*=p*+r*+s*)

BIC = deviance + q* $\log N$, (N = number of subjects)

Since Model Ia is the proportional odds mixed model (including a random subject effect), comparing it to non-proportional odds models yields tests of the proportional odds assumption for the group effects. Comparing models Ib and Ic to it yielded likelihood ratio test values of 9.14 (1 df) and 10.49 (2 df); both clearly rejecting proportional odds in favor of the models with a scaling term ($p < .01$ for both). Similarly, comparing model IVa to model Ia yields 14.77 (6 df), rejecting the proportional odds model in favor of the model with group by threshold interactions ($p < .025$). Thus, there is clear evidence to reject proportional odds for the group effect.

Turning to the AIC and BIC values, it is clear that models with scaling terms are preferred, relative to models with threshold interactions. In particular, model Ib has the lowest BIC value and model IIb has the lowest AIC value. Both include a linear scaling term across the smoking groups. Table 4 lists the estimates from these two models, as well as from the proportional odds model.

Table 4 Ordinal logistic mixed model estimates and standard errors (se)

term	Model Ia estimate	se	Model Ib estimate	se	Model IIb estimate	se
Fixed effects						
threshold 1	-2.84	.42	-2.51	.44	-2.64	.51
threshold 2	-1.44	.40	-1.25	.41	-1.35	.49
threshold 3	.23	.39	.19	.40	.13	.48
threshold 4	1.31	.39	1.12	.40	1.08	.48
MID cigs	.24	.44	.21	.44	.16	.52
HI cigs	-.06	.45	-.07	.43	-.14	.50
Variance terms						
subject sd	1.11	.18	.87	.17	1.07	.21
Linear CIG sd (LO=-1, MID=0, HI=1)					-.33	.22
Scaling terms						
Linear CIG (LO=-1, MID=0, HI=1)			-.26	.09	-.22	.09
$-2 \log L$	1524.9		1515.7		1513.5	

In terms of the fixed effects, there is no evidence of group-related differences from any of the models. It is interesting to note, however, that the fixed-effects estimates and standard errors do change a fair amount, especially if one compares models Ia and IIb. Both models Ib and IIb indicate that the within-subjects variance diminishes as smoking level is increased. Also, the linear trend in between-subjects variation is also negative, suggesting that this variation also diminishes as smoking level is increased. Thus, there is evidence that physiological sensation variation, both within- and between-subjects, diminishes as smoking level increases.

JPSS Vol. 4 No. 1 February 2006 pp. 1-20

It is interesting to compare the intraclass correlation estimates from these three models. For Model Ia (the proportional odds model), we obtain:

$$ICC = \frac{(1.11)^2}{(1.11)^2 + \pi^2/3} = .27$$

This is the estimate for all three smoking groups, since this model assumes that the ICCs for these smoking groups are the same. However, Model IIb yields:

$$\text{Lo Cigs } ICC = \frac{(1.07 + .33)^2}{(1.07 + .33)^2 + \left[(\exp(.22))^2 \, \pi^2/3\right]} = .28$$

$$\text{Mid Cigs } ICC = \frac{(1.07)^2}{(1.07)^2 + \pi^2/3} = .26$$

$$\text{Hi Cigs } ICC = \frac{(1.07 - .33)^2}{(1.07 - .33)^2 + \left[(\exp(-.22))^2 \, \pi^2/3\right]} = .21$$

which indicates that the ICC diminishes as smoking group increases. In fact, from a stratified analysis (not shown), ICC estimates of .43, .21, and .21 are obtained for the three smoking groups, low to high, respectively.

Finally, Table 5 presents estimates from the model allowing separate between- and within-variance parameters for all three groups (Model IIIc).

Table 5 Ordinal logistic mixed model estimates and standard errors (se)

term	Model Ia estimate	Model Ia se	Model IIIc estimate	Model IIIc se
Fixed effects				
threshold 1	-2.84	.42	-2.59	.67
threshold 2	-1.45	.40	-1.32	.52
threshold 3	.23	.39	.14	.46
threshold 4	1.31	.39	1.08	.51
MID cigs	.24	.44	.18	.50
HI cigs	-.06	.45	-.12	.48
Variance terms				
subject sd	1.11	.18		
LO subject sd			1.47	.58
MID subject sd			1.01	.32
HI subject sd			.73	.22
Scaling terms				
MID Cig			.07	.22
HI Cig			-.26	.21
$-2 \log L$	1524.9		1511.6	

Again, for comparison purposes, estimates from the proportional odds model are also included in the table. Model IIIc supports the notion that the between-subjects variation decreases approximately linearly as smoking level is increased. Additionally, in terms of the within-subjects variation, it is primarily the highest smoking group that has diminished variation relative to the lowest smoking group. The intraclass correlation coefficient estimates for this model are obtained as:

$$\begin{aligned}
\text{Lo Cigs } ICC &= \frac{(1.47)^2}{(1.47)^2 + \pi^2/3} = .40 \\
\text{Mid Cigs } ICC &= \frac{(1.01)^2}{(1.01)^2 + \left[(\exp(.07))^2\, \pi^2/3\right]} = .21 \\
\text{Hi Cigs } ICC &= \frac{(.73)^2}{(.73)^2 + \left[(\exp(-.26))^2\, \pi^2/3\right]} = .22
\end{aligned}$$

which agree very closely to the estimates obtained from the stratified analysis. Figure 1 (on page 16) graphically depicts the between- and within-subjects variance estimates from model IIIc. Note that the former are in terms of the standard normal distribution, whereas the latter are standard logistic. The figure clearly shows the reduced variation for the highest smoking level group, both in terms of the between- and within-subjects variance. Thus, as smoking level is increased the change in physiological sensation data within subjects are less correlated and less spread out.

5. Discussion

We have presented an ordinal mixed model that allows for explanatory variables to have varying effects on the cumulative logits, explanatory variables to have scaling effects, and a variety of forms for the random effects. The latter two features allow examination of the degree to which both the between- and within-cluster variance vary across levels of the explanatory variables. In this example we have expressed these variances in terms of a subject grouping variable, the smoking history of the subject, to address an interesting question in smoking research. Namely, we have been able to show that for changes in physiological sensations both the between- and within-subject variances diminish as the level of smoking history increases. This supports the notion of habituation to the physiological effects of cigarette smoking.

The example highlighted the utility of including scaling terms relative to threshold interactions in the model. The former can be parsimonious since only one parameter per covariate is added to the model. Alternatively, adding threshold interaction terms requires $C-1$ parameters per covariate. Since both approaches generalize the proportional odds assumption in different ways, substantive interpretation is also important in deciding which of these approaches to utilize if proportional odds is rejected. For example, in Hedeker and Mermelstein [27], where the ordinal weekly smoking outcome of 0 / 1-5 days / 6-7 days abstinence was longitudinally modeled, threshold interactions were helpful in examining whether covariates had the same effects on partial abstinence (logit 1: 0 vs 1-7 days abstinent) and *nearly*-full abstinence (logit 2: 0-5 versus 6-7 days abstinent). Instead of just analyzing one of these two definitions of abstinence, as is often done in smoking research, the ordinal model with threshold interactions allows covariate effects to be estimated for both, and to test the degree to which these covariate effects are the same across these alternate definitions of abstinence.

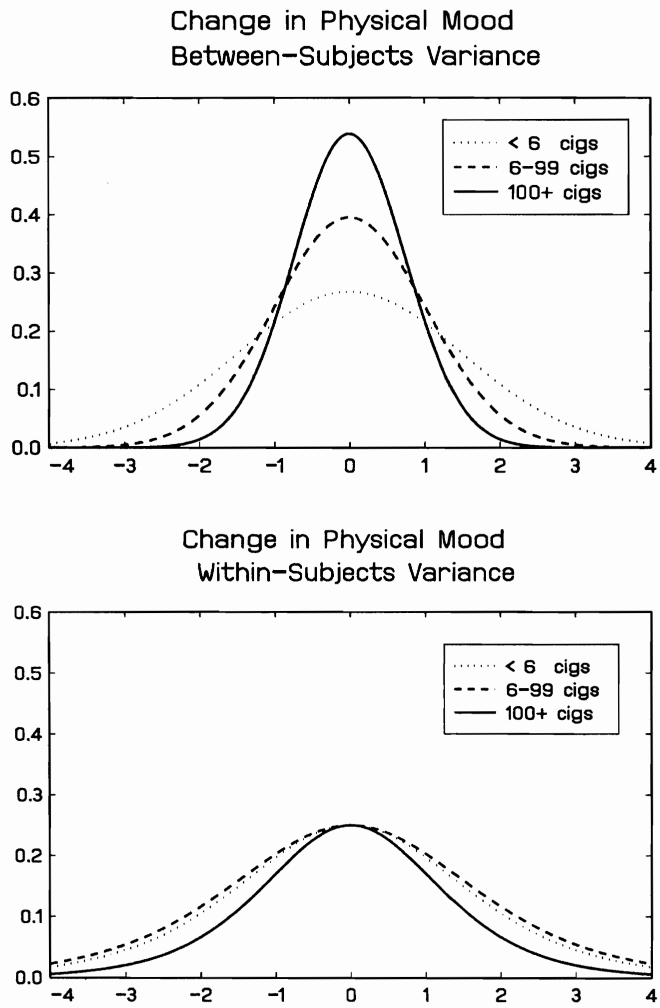

Figure 1 Model IIIc: between- and within- subjects variance estimates

The example presented clearly illustrates the usefulness of the random-effects approach for within-subjects ordinal outcome data. Additionally, this model can be used for clustered data, where subjects are observed nested within some higher-level units (*e.g.*, schools, clinics, etc.). A further extension of the model is underway to allow for such three-level data in order to accommodate clustered data where the clustered subjects are also repeatedly measured across time. For this, the three-level model for dichotomous data described by Gibbons and Hedeker

[16] can be generalized along the lines presented in this article.

Acknowledgements

The authors thank the Managing Editor, Editor-in-Chief, and an anonymous reviewer for several helpful and constructive comments. The authors also thank Dr. Karen Bandeen-Roche for organizing a special session on mixed models at the 2003 ENAR meeting where portions of this paper were presented. Correspondence concerning this article should be addressed to Donald Hedeker, Division of Epidemiology and Biostatistics (M/C 923), School of Public Health, University of Illinois at Chicago, 1603 West Taylor Street, Room 955, Chicago, IL, 60612-4336, email: hedeker@uic.edu. Preparation of this manuscript was supported by National Institutes of Mental Health Grant MH56146, National Cancer Institutes Grant CA80266, and by a grant from the Tobacco Etiology Research Network, funded by the Robert Wood Johnson Foundation.

References

[1] Agresti, A. and Lang, J. B. (1993). A proportional odds model with subject-specific effects for repeated ordered categorical responses, *Biometrika*, **80**, 527-534.

[2] Aitkin, M. (1987). Modelling variance heterogeneity in normal regression using GLIM, *Applied Statistics*, **36**, 332-339.

[3] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, Second International Symposium on Information Theory, Petrov, 267-281, B. N. and Csaki, F. (eds.), Akademiai Kiado, Budapest.

[4] Bock, R. D. (1975). Multivariate Statistical Methods in Behavioral Research, McGraw-Hill, New York.

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

[6] Bock, R. D. and Lieberman, M. (1970). Fitting a response model for n dichotomously scored items, *Psychometrika*, **35**, 179-197.

[7] Breslow, N. E. and Lin, X. (1995). Bias correction in generalised linear mixed models with a single component of dispersion, *Biometrika*, **82**, 81-91.

[8] Bryk, A. S. and Raudenbush, S. W. (1992). Hierarchical Linear Models: Applications and Data Analysis Methods, Sage Publications, Inc., Newbury Park, CA.

[9] Cox, C. (1995). Location-scale cumulative odds models for ordinal data: a generalized non-linear model approach, *Statistics in Medicine*, **14**, 1191-1203.

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

[11] Fahrmeir, L. and Tutz, G. T. (2001). Multivariate Statistical Modelling Based on Generalized Linear Models, 2nd ed., Springer-Verlag, New York.

[12] Fielding, A., Yang, M., and Goldstein, H. (2003). Multilevel ordinal models for examination grades, *Statistical Modelling*, **3**, 127-153.

[13] Freels, S. A., Warnecke, R. B., Johnson, T. P., and Flay, B. R. (2002). Evaluation of the effects of a smoking cessation intervention using the multilevel thresholds of change model, *Evaluation Review*, **26**, 40-58.

[14] Gange, S. J., Linton, K. L. P., Scott, A. J., DeMets, D. L., and Klein, R. (1995). A comparison of methods for correlated ordinal measures with ophthalmic applications, *Statistics in Medicine*, **14**, 1961-1974.

[15] Gibbons, R. D. and Bock, R. D. (1987). Trend in correlated proportions, *Psychometrika*, **52**, 113-124.

[16] Gibbons, R. D. and Hedeker, D. (1997). Random-effects probit and logistic regression models for three-level data, *Biometrics*, **53**, 1527-1537.

[17] Goldstein, H. (1995). Multilevel Statistical Models, 2nd ed., Halstead Press, New York.

[18] Goldstein, H. and Rasbash, J. (1996). Improved approximations for multilevel models with binary responses, *Journal of the Royal Statistical Society, Series B*, **159**, 505-513.

[19] Greene, W. H. (1998). LIMDEP Version 7.0 User's Manual (revised edition), Econometric Software, Inc., Plainview, N.Y.

[20] Hall, D. B. and Severini, T. A. (1998). Extended generalized estimating equations for clustered data, *Journal of the American Statistical Association*, **93**, 1365-1375.

[21] Harvey, A. C. (1976). Estimating regression models with multiplicative heteroscedasticity, *Econometrica*, **44**, 461-465.

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

[23] Heagerty, P. J. and Zeger, S. L. (1996). Marginal regression models for clustered ordinal measurements, *Journal of the American Statistical Association*, **91**, 1024-1036.

[24] Hedeker, D. and Gibbons, R. D. (1994). A random-effects ordinal regression model for multilevel analysis, *Biometrics*, **50**, 933-944.

[25] Hedeker, D. and Gibbons, R. D. (1996). MIXOR: a computer program for mixed-effects ordinal probit and logistic regression analysis, *Computer Methods and Programs in Biomedicine*, **49**, 157-176.

[26] Hedeker, D. and Mermelstein, R. J. (1998). A multilevel thresholds of change model for analysis of stages of change data, *Multivariate Behavioral Research*, **33**, 427-455.

[27] Hedeker, D. and Mermelstein, R. J. (2000). Analysis of longitudinal substance use outcomes using random-effects regression models, *Addiction (Supplement 3)*, **95**, S381-S394.

[28] Ishwaran, H. (2000). Univariate and multirater ordinal cumulative link regression with covariate specific cutpoints, *Canadian Journal of Statistics*, **28**, 715-730.

[29] Ishwaran, H. and Gatsonis, C. A. (2000). A general class of hierarchical ordinal regression models with applications to correlated ROC analysis, *Canadian Journal of Statistics*, **28**,

731-750.

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

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

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

[33] Liang, K.-Y. and Zeger, S. L., and Qaqish, B. (1992). Multivariate regression analyses for categorical data, *Journal of the Royal Statistical Society, Series B*, **54**, 3-40.

[34] Long, J. S. (1997). Regression Models for Categorical and Limited Dependent Variables, Sage Publications, Thousand Oaks, CA.

[35] Lord, F. M. (1980). Applications of Item Response Theory to Practical Testing Problems, Erlbaum, Hillside, N.J.

[36] McCullagh, P. (1980). Regression models for ordinal data (with discussion), *Journal of the Royal Statistical Society, Series B*, **42**, 109-142.

[37] McCullagh, P. and Nelder, J. A. (1989). Generalized Linear Models, 2nd ed., Chapman and Hall, New York.

[38] 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.

[39] Mehta, P. D. and Neale, M. C., and Flay, B. R. (2004). Squeezing interval change from ordinal panel data: latent growth curves with ordinal outcomes, *Psychological Methods*, **9**, 301-333.

[40] Mermelstein, R., Hedeker, D., Flay, B., and Shiffman, S. (2002). Situational versus intra-individual contributions to adolescents' subjective mood experience of smoking, *Annual Meeting for the Society for Research on Nicotine and Tobacco*, Savannah, GA.

[41] Miller, M. E., Davis, C. S., and Landis, J. R. (1993). The analysis of longitudinal polytomous responses: generalized estimating equations and connections with weighted least squares, *Biometrics*, **49**, 1033-1044.

[42] Muthén, B. and Muthén, L. (2001). Mplus User's Guide, Muthén & Muthén, Los Angeles.

[43] Peterson, B. and Harrell, F. E. (1990). Partial proportional odds models for ordinal response variables, *Applied Statistics*, **39**, 205-217.

[44] Rabe-Hesketh, S., Pickles, A., and Skrondal, A. (2001). GLLAMM Manual, Institute of Psychiatry, King's College, University of London.

[45] Rabe-Hesketh, S., Skrondal, A., and Pickles, A. (2002). Reliable estimation of generalized linear mixed models using adaptive quadrature, *The Stata Journal*, **2**, 1-21.

[46] Rasbash, J., Steele, F., Browne, W., and Prosser, B. (2004). A User's Guide to MLwiN, version 2.0, University of London.

[47] Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., and Congdon, R. (2004). HLM 6: Hierarchical Linear and Nonlinear Modeling, Scientific Software International, Inc., Chicago.

[48] Raudenbush, S. W., Yang, M.-L., and Yosef, M. (2000). Maximum likelihood for generalized linear models with nested random effects via high-order, multivariate Laplace approximation, *Journal of Computational and Graphical Statistics*, **9**, 141-157.

[49] Rodríguez, G. and Goldman, N. (1995). An assessment of estimation procedures for multilevel models with binary responses, *Journal of the Royal Statistical Society, Series A*, **158**, 73-89.

[50] Saei, A. and McGilchrist, C. A. (1998). Longitudinal threshold models with random components, *Journal of the Royal Statistical Society, Series D (The Statistician)*, **47**, 365-375.

[51] Saei, A., Ward, J., and McGilchrist, C. A. (1996). Threshold models in a methadone programme evaluation, *Statistics in Medicine*, **15**, 2253-2260.

[52] Schwarz, G. (1978). Estimating the dimension of a model, *The Annals of Statistics*, **6**, 461-464.

[53] StataCorp. (1999). Stata: release 6.0, Stata Corporation, College Station, TXJ.

[54] Ten Have, T. R. (1996). A mixed effects model for multivariate ordinal response data including correlated discrete failure times with ordinal responses, *Biometrics*, **52**, 473-491.

[55] Terza, J. V. (1985). Ordinal probit: a generalization, *Communications in Statistical Theory and Methods*, **14**, 1-11.

[56] Toledano, A. Y. and Gatsonis, C. (1996). Ordinal regression methodology for ROC curves derived from correlated data, *Statistics in Medicine*, **15**, 1807-1826.

[57] Tosteson, A. N. and Begg, C. B. (1988). A general regression methodology for ROC curve estimation, *Medical Decision Making*, **8**, 204-215.

[58] Tutz, G. and Hennevogl, W. (1996). Random effects in ordinal regression models, *Computational Statistics and Data Analysis*, **22**, 537-557.

[59] Wakefield, M. A., Chaloupka, F. J., Kaufman, N. J., Orleans, C. T., Barker, D. C., and Ruel, E. E. (2001). Effect of restrictions on smoking at home, at school, and in public places on teenage smoking: cross sectional study, *British Medical Journal*, **321**, 333-337.

[60] Xie, H., McHugo, G., Sengupta, A., Hedeker, D., and Drake, R. (2001). An application of the thresholds of change model to the analysis of mental health data, *Mental Health Services Research*, **3**, 107-114.