

## A Multilevel Thresholds of Change Model for Analysis of Stages of Change Data

Donald Hedeker

Division of Epidemiology and Biostatistics  
and Health Research and Policy Centers  
University of Illinois at Chicago

Robin J. Mermelstein

Health Research and Policy Centers  
and Department of Psychology  
University of Illinois at Chicago

In this article we describe a model for multilevel ordinal response data that allows for non-proportional odds for a subset of explanatory variables. As applied to stages of change data, which are commonly encountered in health promotion research, this model is termed the multilevel thresholds of change model since it focuses on modeling the  $K-1$  thresholds that delineate membership in the  $K$  ordered stages. Explanatory variables can have the same effect across thresholds (i.e., proportional odds) or varying effects across thresholds (i.e., non-proportional odds). In addition to the explanatory variables of the model, random effects are included to account for the multilevel structure of the data (e.g., repeated observations within subjects, or subjects observed within clusters). A maximum marginal likelihood (MML) solution is described using Gauss-Hermite quadrature to numerically integrate over the distribution of normally-distributed random effects. Data from a skin cancer prevention study, in which subjects were repeatedly measured across time and clustered within schools, are used to illustrate the multilevel thresholds of change model.

### *Introduction*

Stage variables are important components to many psychological theories. Perhaps one of the more predominant theories in recent years that utilizes a stage construct is the transtheoretical model (Prochaska & DiClemente, 1983; Prochaska & Velicer, 1997). The transtheoretical model

---

The authors thank Dr. Linda Collins for organizing this special issue of *Multivariate Behavioral Research*. The authors also thank the editor, Dr. Roger Millsap, and an anonymous reviewer for many helpful and constructive comments. Correspondence concerning this article should be addressed to Donald Hedeker, Division of Epidemiology and Biostatistics (M/C 922), School of Public Health, University of Illinois at Chicago, 2121 West Taylor Street, Room 525, Chicago, IL, 60612-7260. Preparation of this manuscript was supported by National Institutes of Mental Health Grant MH56146 and National Cancer Institute Grant CA 62964.

views behavior change as a process that involves progress through a series of stages from precontemplation through termination. One of the basic assumptions of the transtheoretical model is that interventions should be matched to an individual's stage of change. Stage progression is thus an important and logical outcome for any stage-based intervention, and it is critical to use statistical methods that are appropriate and sensitive to evaluating change or movement across the stages. The present article describes a statistical model for analyzing longitudinal movement across stages.

Since the stages of change are ordinal categories representing increasing levels of contemplating behavior change (e.g., precontemplation and contemplation) to actual behavior change (e.g., action), statistical models for ordinal responses have often been used to determine the effects of explanatory variables on stage data. One such model is the ordinal logistic regression model, described as the proportional odds model by McCullagh (1980). In this model, for an ordinal response with  $K$  categories, it is assumed that the effect of an explanatory variable is the same across the  $K-1$  cumulative logits of the model, or proportional across the cumulative odds. For stage data this assumption may not be reasonable. For example, suppose that there are three stages (pre-contemplation, contemplation, and action) and suppose that an intervention is not successful in increasing the proportion of individuals in the action stage, but is successful in moving individuals from the pre-contemplation into the contemplation stage. In this case, the effect of intervention group (i.e., the explanatory variable) may be observed on the first cumulative logit (i.e., the logit comparing the pre-contemplation category versus the contemplation and action categories combined), but not on the second cumulative logit (i.e., the logit comparing the pre-contemplation and contemplation categories combined versus the action category).

To overcome the limitation of the proportional odds model, Peterson and Harrell (1990) developed an ordinal logistic regression model allowing for non-proportional odds for a subset of the explanatory variables. A similar extension of the ordinal probit regression model has been described by Terza (1985). This extended model has recently been applied to stage data and described as the thresholds of change model (TCM; Hedeker, Mermelstein, & Weeks, in press). This formulation focuses on estimating the thresholds that separate the ordered stages and models the influence of explanatory variables on these thresholds by assuming an underlying continuous latent variable related to the  $K$  ordered stages through a series of  $K-1$  threshold values. If the observations in a sample are independent of each other, this thresholds of change model provides a useful approach for analysis of stage data.

When observations are nested within clusters (i.e., classes, schools, firms, clinics) or are repeatedly assessed across time, however, use of statistical models that assume independence of observations is unreasonable. For such multilevel data, (i.e., data that are clustered and/or longitudinal), random-effects regression models have been developed to account for the dependency inherent in the data, while simultaneously estimating model parameters (Bryk & Raudenbush, 1992; Goldstein, 1995). For ordinal responses, several authors have described models including both random and fixed effects. Harville and Mee (1984) and Jansen (1990) both describe ordinal probit models implementing the EM algorithm for parameter estimation. Ezzet and Whitehead (1991) provide a random-intercepts proportional odds model for a crossover trial using the Newton-Raphson method. A random-intercepts proportional odds model is described by Agresti and Lang (1993), however, their approach is limited to within-cluster covariates. Ten Have (1996) describes a random-intercepts model using the complementary log-log link and log-gamma distributed random effects. Molenberghs and Lesaffre (1994) develop a multivariate Dale model for correlated ordinal data using an underlying Plackett distribution. Hedeker and Gibbons (1994) describe both an ordinal probit and logistic model with multiple random effects that allows for both within and between-cluster covariates; software for this approach is also available (Hedeker & Gibbons, 1996).

An alternative method of dealing with correlated data is provided by the generalized estimating equations (GEE) approach described by Liang and Zeger (1986). 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 an extended version (called GEE2) has been developed for situations where the correlation structure of the multilevel data is also of interest (Liang, Zeger, & Qaqish, 1992). For ordinal data, Miller, Davis, and Landis (1993) and Gange, Linton, Scott, DeMets, and Klein (1995) describe GEE-based models for repeated ordinal outcomes. Heagerty and Zeger (1996) extend this approach and also describe a GEE2 formulation for correlated ordinal outcomes.

Most of the models for multilevel ordinal outcomes include the proportional odds assumption (or its equivalent under the probit response function) for model covariates. In the present article, an extension of the random-effects proportional odds model is described to allow for non-proportional odds for a subset of the explanatory variables. This extension follows Peterson and Harrell's (1990) extension of the fixed-effects proportional odds model. To account for the multilevel structure of the data, the proposed model will accommodate multiple random effects and allow for both within and between-cluster covariates. Covariates can be specified

either requiring or relaxing the proportional odds assumption. A maximum marginal likelihood solution is described using multi-dimensional quadrature to numerically integrate over the distribution of random-effects. An iterative Fisher scoring solution is used to provide relatively quick convergence and standard errors for all model parameters. As applied to multilevel stage data, this model allows estimation of the thresholds separating the stages of change, as well as assessing the influence of variables (e.g., time and intervention group) on these thresholds while controlling for the multilevel structure of the data. Data from a skin cancer prevention study, in which subjects were repeatedly measured across time and clustered within schools, are used to illustrate the multilevel thresholds of change model.

### *Multilevel Thresholds of Change Model (MTCM)*

To introduce and motivate application of MTCM, we will utilize the "threshold concept" (Bock, 1975). Here, it is assumed that a continuous distribution of "readiness" of change ( $y$ ) exists in the population, however this "readiness" of change variable is not directly observed. Instead, individuals are classified into "stages" of change depending on their assumed value on the continuous latent "readiness" of change variable. If there are  $K$  ordered stages of change categories, then  $K-1$  strictly increasing thresholds (denoted  $\gamma_1 \dots \gamma_{K-1}$ ) separate individuals into the stages. An individual is staged in category  $k$  ( $Y=k$ ) if the individual's readiness of change value  $y$  exceeds the threshold value  $\gamma_{k-1}$ , but does not exceed the threshold value  $\gamma_k$ . These thresholds can be conceptualized as hurdles of increasing difficulty that separate individuals into the (increasing) stages of change. Estimation of these thresholds is then important in order to characterize a population of individuals in terms of the stages of change. In particular, estimation of the degree to which the thresholds vary across time or group is necessary to compare the distributions (of stages of change data) between multiple timepoints or multiple groups. In order to estimate these thresholds, a distributional form is assumed for the underlying latent readiness of change variable. Convenient choices for this distribution are the normal and logistic distributions leading to, respectively, an ordinal probit regression model and an ordinal logistic regression model (Bock, 1975; Agresti, 1990). In what follows, we will develop the model in terms of the logistic regression formulation and indicate necessary modifications for the probit model.

To describe the model in a general way for either clustered or longitudinal data, the terminology of multilevel analysis (Goldstein, 1995) is used. Let  $i$  denote the level-2 units and let  $j$  denote the level-1 units. In the clustered case, the clusters (i.e., schools, hospitals, firms) represent the level-2 units,

while the subjects are the level-1 units nested within the clusters. Alternatively, in the longitudinal or within-subjects context, the level-2 units are the subjects and the level-1 units are the repeated observations. Assume that there are  $i = 1, \dots, N$  level-2 units and  $j = 1, \dots, n_i$  level-1 units nested within each level-2 unit. The mixed-effects regression model for the latent readiness of change  $y_{ij}$  can be written as follows:

$$(1) \quad y_{ij} = \mathbf{x}'_{ij} \boldsymbol{\beta} + \mathbf{w}'_{ij} \mathbf{v}_i + \varepsilon_{ij}$$

where  $\mathbf{x}_{ij}$  is the  $p \times 1$  covariate vector and  $\mathbf{w}_{ij}$  is the design vector for the  $r$  random effects, both vectors being for the  $j$ th level-1 unit nested within level-2 unit  $i$ . Also,  $\boldsymbol{\beta}$  is the  $p \times 1$  vector of unknown fixed regression parameters,  $\mathbf{v}_i$  is the  $r \times 1$  vector of unknown random effects for the level-2 unit  $i$ , and  $\varepsilon_{ij}$  are the model residuals that follow a logistic distribution (or a normal distribution for the probit model).

The distribution of the random effects is assumed to be multivariate normal with mean vector  $\mathbf{0}$  and covariance matrix  $\boldsymbol{\Sigma}_v$ . Assuming the mean vector to be equal to  $\mathbf{0}$ , leads to the random effects being specified as deviations from the model. Typically, then, the variables specified in  $\mathbf{w}$  will also be included in  $\mathbf{x}$ , so that the mean of the random effects is also estimated. Since the level-2 subscript  $i$  is present for the  $\mathbf{w}$  vector, not all level-2 units are assumed to have the same number of level-1 observations nested within. Thus, there is no assumption of equal sample sizes within clusters for clustered data, or the same number of repeated observations per subject for within-subjects data.

### *Logistic and Probit Response Functions*

Assuming a logistic response function, the probability, for a given level-2 unit  $i$ , that  $Y_{ij} = k$  (a response occurs in category  $k$ ), conditional on the random effects  $\mathbf{v}_i$ , is given by:

$$P(Y_{ij} = k | \mathbf{v}_i) = \Psi(\gamma_k - z_{ij}) - \Psi(\gamma_{k-1} - z_{ij})$$

where  $z_{ij} = \mathbf{x}'_{ij} \boldsymbol{\beta} + \mathbf{w}'_{ij} \mathbf{v}_i$  and  $\Psi(\cdot)$  represents the logistic cumulative distribution function (cdf), namely  $\Psi(z) = 1/[1 + \exp(-z)]$ . Since the scale of the latent variable  $y$  is arbitrary it is typical to assume that the logistic distribution is in its standard form (i.e., with mean 0 and variance  $\sigma^2 = \pi^2/3$ ). Additionally, for identification it is assumed that the response model has no intercept (i.e.,  $\boldsymbol{\beta}_0 = 0$ ).<sup>1</sup> In terms of the thresholds, it is assumed that

$\gamma_0 = -\infty$  and  $\gamma_K = \infty$ . For the probit model the normal cdf  $\Phi(\bullet)$  replaces  $\Psi(\bullet)$  in the conditional probability and the assumed standard normal has variance equal to 1.

Denoting the cumulative probabilities for the  $k$  ordered categories ( $k = 1, \dots, K$ ) as:

$$(2) \quad P_{ijk} = Pr(Y_{ij} \leq k \mid \mathbf{v}_i) = \Psi(\gamma_k - z_{ij})$$

the random-effects logistic regression model can be written as:

$$(3) \quad \log \frac{P_{ijk}}{(1 - P_{ijk})} = \gamma_k - (\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{w}'_{ij}\mathbf{v}_i).$$

By including the random effects  $\mathbf{v}_i$ , Equation 3 is an extension of the (fixed-effects) proportional odds model described by McCullagh (1980). Since the regression coefficients  $\boldsymbol{\beta}$  do not depend on  $k$ , the model assumes that the relationship between the explanatory variables and the cumulative logits does not depend on  $k$ . McCullagh (1980) calls this assumption of identical odds ratios across the  $K-1$  thresholds the proportional odds assumption.

#### *Partial Proportional Odds*

As described later, for interpretation purposes in analysis of stage data, it is beneficial to reverse the sign and express the model as:

$$(4) \quad \lambda_{ijk} = \gamma_k + (\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{w}'_{ij}\mathbf{v}_i),$$

with the logit denoted as  $\lambda_{ijk}$ . To allow for a partial proportional odds model the thresholds  $\gamma_k$  are denoted instead as  $\gamma_{k(0)}$ , and the following terms are added to the model:

$$(5) \quad \lambda_{ijk} = \gamma_{k(0)} + \mathbf{u}'_{ij}\boldsymbol{\alpha}_k + (\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{w}'_{ij}\mathbf{v}_i)$$

where,  $\mathbf{u}_{ij}$  is a  $h \times 1$  vector containing the values of observation  $ij$  for the set of explanatory variables for which proportional odds is not assumed. The  $h \times 1$  vector of regression coefficients  $\boldsymbol{\alpha}_k$  is associated with the  $h$  variables in  $\mathbf{u}_{ij}$ . Note that since  $\boldsymbol{\alpha}$  carries the  $k$  subscript, the effects of these  $h$  explanatory variables ( $\mathbf{u}_{ij}$ ) vary across the  $K-1$  cumulative logits. The parameters

---

<sup>1</sup> An alternative specification for identification is that the first threshold  $\gamma_1 = 0$ .

$\lambda_{k(0)}$  might then be considered as baseline thresholds, or threshold values when  $\mathbf{u} = \mathbf{0}$ . As such, the model can be re-expressed as:

$$(6) \quad \lambda_{ijk} = \gamma_{ijk} + (\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{w}'_{ij}\mathbf{v}_i),$$

with

$$(7) \quad \gamma_{ijk} = \gamma_{k(0)} + \mathbf{u}'_{ij}\boldsymbol{\alpha}_k$$

The specification in Equation 7 clearly shows how the thresholds  $\gamma_{ijk}$  can depend on explanatory variables  $\mathbf{u}_{ij}$ , giving rise to the name "thresholds of change model." The subscripts of the thresholds  $\gamma_{ijk}$  indicate that the  $K-1$  thresholds can vary with factors associated with units  $i$  and  $j$ . Since the model also includes random effects to account for the multilevel nature of the data, the model is termed a multilevel thresholds of change model (MTCM). The mixed-effects proportional odds model given in Equation 4 is then seen as the limiting case of this model when  $\gamma_{ijk} = \gamma_{k(0)}$ . In the formulation given in Equation 7, allowing for non-proportional odds can be stated as allowing variables to have heterogeneous threshold effects (i.e., the assumption for the variables included in  $\mathbf{u}_{ij}$ ). Alternatively, a variable with the same effect on all thresholds (i.e.,  $\alpha_1 = \alpha_2 = \dots, \alpha_{K-1}$ ) has the same effect on all cumulative logits, and so is included in the model in  $\mathbf{x}_{ij}$  instead of in  $\mathbf{u}_{ij}$ . As such, the assumption of proportional odds is equivalent to assuming homogeneous threshold effects.

In terms of stage data, the cumulative logit for stage  $k$  contrasts the probability of being in or below stage  $k$  versus being above stage  $k$ . Thus, all other things equal, as the threshold  $\gamma_{ijk}$  is decreased there is a greater probability of being above stage  $k$ , and as  $\gamma_{ijk}$  is increased the probability reduces. An explanatory variable  $u_{ij}$  that lowers a threshold  $\gamma_{ijk}$  (i.e., the coefficient  $\alpha_k$  for the variable is negative) is indicative of greater probability of being classified in the stages above stage  $k$  with increasing values of  $u_{ij}$ . If one visualizes the thresholds as hurdles, interpreting the sign of elements in the coefficient vector  $\boldsymbol{\alpha}_k$  is readily apparent: a negative sign lowers the hurdle (so more can crossover) while a positive sign raises the hurdle (so less can crossover). The magnitude (i.e., scale) of the threshold or hurdle is then given by the assumed logistic distribution.

In general, allowing for variables to have heterogeneous threshold effects is not problematic, however, one caveat should be mentioned. For a particular variable  $u$ , the heterogeneous effects on the thresholds (i.e.,  $u_{ij}\alpha_k$ ) result in  $K-1$  non-parallel regression lines for the regression of  $\gamma_k$  on  $u$ . These regression lines inevitably cross for some values of  $u$ , leading to negative

fitted values for the response probabilities. For 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  $> 1$ ). More generally, for categorical explanatory variables with  $m$  categories, if  $m - 1$  dummy-codes or contrasts are included in the model, the crossing of regression lines occurs outside the data range. However, if the explanatory variable is continuous this crossing can occur within the range of the data. It is for this reason that the random effects  $\mathbf{v}_i$ , which are assumed to follow a continuous normal distribution, do not have heterogeneous threshold effects. For continuous explanatory variables, other than requiring homogeneous threshold effects, 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.

*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  $\mathbf{v}_i$  is equal to the product of the probabilities of the level-1 responses:

$$(8) \quad \ell(\mathbf{Y}_i | \mathbf{v}_i) = \prod_{j=1}^{n_i} \prod_{k=1}^K (P_{ijk} - P_{ij,k-1})^{d_{ijk}}$$

$$\text{where } d_{ijk} = \begin{cases} 1 & \text{if } Y_{ij} = k \\ 0 & \text{if } Y_{ij} \neq k \end{cases}$$

As noted by Gibbons and Bock (1987) in the context of a random effects probit model, it is convenient to orthogonally transform the response model. Specifically, let  $\mathbf{v} = \mathbf{T}\boldsymbol{\theta}$ , where  $\mathbf{T}\mathbf{T}' = \boldsymbol{\Sigma}_v$  is the Cholesky decomposition of  $\boldsymbol{\Sigma}_v$ . The reparameterized model is then

$$(9) \quad \lambda_{ijk} = \gamma_{ijk} + \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{w}'_{ij}\mathbf{T}\boldsymbol{\theta}_i,$$

with

$$(10) \quad \gamma_{ijk} = \gamma_{k(0)} + \mathbf{u}'_{ij}\boldsymbol{\alpha}_k,$$

where  $\boldsymbol{\theta}_i$  are distributed according to a multivariate standard normal. A consequence of transforming from  $\mathbf{v}$  to  $\boldsymbol{\theta}$  is that the Cholesky factor  $\mathbf{T}$ , which



is a lower triangular matrix, is estimated instead of the covariance matrix  $\Sigma_{\mathbf{v}}$ . As the Cholesky factor is essentially the square-root of the covariance matrix, this then allows more stable estimation of near-zero variance terms.

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

$$(11) \quad h(\mathbf{Y}_i) = \int_{\theta} \ell(\mathbf{Y}_i | \theta) g(\theta) d\theta$$

where  $g(\theta)$  represents the multivariate standard normal density.

The marginal log-likelihood for the patterns from the  $N$  level-2 units can be written as:

$$\log L = \sum_i^N \log h(\mathbf{Y}_i)$$

Let  $\eta$  represent the parameter vector obtained by stacking  $\beta$  and  $\mathbf{v}(\mathbf{T})$  [the vector, with with  $r(r+1)/2$  elements, which contains the unique elements of the Cholesky factor  $\mathbf{T}$ ]. Differentiating the marginal log-likelihood yields:

$$(12) \quad \frac{\partial \log L}{\partial \eta} = \sum_{i=1}^N h^{-1}(\mathbf{Y}_i) \int_{\theta} \sum_{j=1}^{n_i} \sum_{k=1}^K d_{ijk} \left[ \frac{\partial P_{ijk} - \partial P_{ij,k-1}}{P_{ijk} - P_{ij,k-1}} \right] \ell(\mathbf{Y}_i | \theta) g(\theta) \frac{\partial z_{ij}}{\partial \eta} d\theta,$$

where  $z_{ij} = \mathbf{x}'_{ij} \beta + \mathbf{w}'_{ij} \mathbf{T} \theta$ , and  $\partial z_{ij} / \partial \eta$  contains

$$\frac{\partial z_{ij}}{\partial \beta} = \mathbf{x}_{ij} \text{ and } \frac{\partial z_{ij}}{\partial [\mathbf{v}(\mathbf{T})]} = J_r(\theta \otimes \mathbf{w}_{ij}).$$

Here,  $J_r$  is the transformation matrix of Magnus (1988) that eliminates the elements above the main diagonal.

For estimation of the thresholds  $\gamma_{k(0)}$  and threshold moderators  $\alpha_k$  ( $k = 1, \dots, K-1$ ), let  $\zeta_i$  represent a particular  $\gamma_{k(0)}$  or vector  $\alpha_k$ . Then,

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

with

$$(13) \frac{\partial h(\mathbf{Y}_i)}{\partial \zeta_l} = \int_{\Theta} \sum_{j=1}^{n_i} \sum_{k=1}^K d_{ijk} \left[ \frac{(\partial P_{ijk})a_{kl} - (\partial P_{ij,k-1})a_{k-1,l}}{P_{ijk} - P_{ij,k-1}} \right] \ell(\mathbf{Y}_i | \Theta_i) g(\Theta) \frac{\partial \gamma_{ijk}}{\partial \zeta_l} d\Theta,$$

where

$$a_{kl} = \begin{cases} 1 & \text{if } k = l \\ 0 & \text{if } k \neq l \end{cases}, \quad \frac{\partial \gamma_{ijk}}{\partial \gamma_{l(0)}} = 1, \quad \frac{\partial \gamma_{ijk}}{\partial \alpha_j} = \mathbf{u}_{ij}.$$

Note that for the logistic formulation,  $\partial P_{ijk} = P_{ijk}(1 - P_{ijk})$ . For a probit regression formulation, the logistic function  $\Psi(\bullet)$  is replaced by the normal response function  $\Phi(\bullet)$ , and  $\partial P_{ijk}$  is replaced by the standard normal density function  $\phi(\bullet)$ , in the conditional probability in Equation 8 and the derivatives in Equations 12 and 13.

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  $\Theta$ , on iteration  $t$  are improved by:

$$(14) \quad \Theta_{t+1} = \Theta_t - \varepsilon \left[ \frac{\partial^2 \log L}{\partial \Theta_t \partial \Theta_t'} \right]^{-1} \frac{\partial \log L}{\partial \Theta_t}$$

where, following Bock and Lieberman (1970), the information matrix, or expectation of the matrix of second derivatives, is given by

$$\varepsilon \left[ \frac{\partial^2 \log L}{\partial \Theta_t \partial \Theta_t'} \right] = - \sum_{i=1}^N h^2(\mathbf{Y}_i) \frac{\partial h(\mathbf{Y}_i)}{\partial \Theta_t} \left( \frac{\partial h(\mathbf{Y}_i)}{\partial \Theta_t} \right)'$$

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

### Numerical Quadrature

In order to solve the above likelihood equations, numerical integration on the transformed  $\Theta$  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  $\Theta$  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  $B_q' = (B_{q1}, B_{q2}, \dots, B_{qr})$ , with its associated (scalar) weight given by the product of the corresponding univariate weights,

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

Although, here we consider the random effects to be normally-distributed, other distributional forms can be used. For example, if a rectangular or uniform distribution is assumed, then  $Q$  points may be set at equal intervals over an appropriate range (for each dimension) and the quadrature weights are then set equal to  $1/Q$ . Other distributions are possible: Bock and Aitkin (1981) discuss the possibility of empirically estimating the random-effect distribution.

For models with few random effects the quadrature solution is relatively fast and computationally tractable. In particular, if there is only one random effect in the model, there is only one additional summation over  $Q$  points relative to the fixed effects solution. As the number of random effects  $r$  is increased, however, the terms in the summation ( $Q^r$ ) increases 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 sufficient to obtain adequate accuracy.

### *Estimation of Random Effects and Marginal Probabilities*

In some cases, it may be of interest to estimate values of the random effects  $\theta_i$  within the sample. A reasonable choice for this is the expected "a posteriori" (EAP) or empirical Bayes estimator  $\bar{\theta}_i$  (Bock & Aitkin, 1981). For the univariate case, this estimator  $\bar{\theta}_i$ , given the vector of stage outcomes  $Y_i$  is given by:

$$(15) \quad \bar{\theta}_i = E(\theta_i | Y_i) = \frac{1}{h(Y_i)} \int_0^1 \theta_i \ell(Y_i | \theta_i) g(\theta) d\theta.$$

The variance of this estimator is obtained similarly as:

$$(16) \quad V(\bar{\theta}_i | \mathbf{Y}_i) = \frac{1}{h(\mathbf{Y}_i)} \int_{\theta} (\theta_i - \bar{\theta}_i)^2 \ell(\mathbf{Y}_i | \theta) g(\theta) d\theta.$$

Upon convergence, these quantities can be obtained using one additional round of quadrature and the converged values of  $h(\mathbf{Y}_i)$ , which vary by  $i$  units, and  $\ell(\mathbf{Y}_i|\theta)$ , which vary by  $i$  units and quadrature points. They may then be used, for example, to evaluate the response probabilities for particular level-2 units. Also, Ten Have (1996) suggests how these empirical Bayes estimates might be used in performing residual diagnostics.

To obtain model-based estimated marginal probabilities, a distinction is made between subject-specific and marginal effects (Neuhaus, Kalbfleisch, & Hauck, 1991). In random-effects models the covariate effects of the model are estimated adjusting for the random effect  $\mathbf{v}_i$  and so have been called cluster or subject-specific effects by Neuhaus et al. This contrasts to the population-averaged effects that are estimated by marginal procedures like the generalized estimating equation (GEE) approach of Liang and Zeger (1986). Observed proportions are marginal estimates indicating the estimated probability of a response in category  $k$  for a group of subjects. For non-linear models like the logistic model used here, marginal and subject-specific estimates are not on the same scale, and so cannot be directly compared. However, marginal estimates can still be obtained from the subject-specific estimates, but an additional step is required.

In the model considered here, subject-specific probabilities can be estimated for specific values of covariates and  $\theta_i$  by applying the logistic transform  $\Psi(z_{ijk}) = 1/[1 + \exp(-z_{ijk})]$  with  $\hat{z}_{ijk} = \hat{\gamma}_{k(0)} + \mathbf{u}'_{ij}\hat{\alpha}_k + \mathbf{x}'_{ijk}\hat{\beta} + \mathbf{w}'_{ij}\hat{\mathbf{T}}\theta_i$ . Denoting these subject-specific probabilities as  $\hat{P}_{ss}$ , marginal probabilities  $\hat{P}_m$  can also be obtained by integrating over the random-effect distribution:

$$(17) \quad \hat{P}_m = \int_{\theta} \hat{P}_{ss} g(\theta) d\theta.$$

Again, numerical quadrature can be used for this. These estimated marginal probabilities can then be compared to the observed marginal proportions to assess model fit, either for the whole sample or stratified by covariates.

Another approximation that can be used when there is only a single random subject effect in the model, described in Diggle, Liang, and Zeger (1994, page 142), does not require use of numerical integration. Here, the coefficients  $\gamma_{k(0)}$ ,  $\alpha_k$ , and  $\beta$  from the random-effects model are divided by  $\sqrt{c^2\sigma_v^2 + 1}$  where  $c = 16\sqrt{3}/(15\pi)$ . These "marginalized" coefficients can then be directly used with  $\Psi(z_{ijk})$  to produce marginal proportions. In most cases, results using either of these methods are in close agreement, though the

accuracy of the quadrature method can be increased by using more points. In the example below, we illustrate model fit using both methods.

### *Computer Implementation*

In terms of computer programming, the procedure described in this article has been implemented for use in an extended-version of MIXOR (Hedeker & Gibbons, 1996) named MIXORE.<sup>2</sup> The program starts by reading in for each level-2 unit the  $n_i \times 1$  vector of stage outcomes  $\mathbf{Y}_i$ , the  $n_i \times r$  random-effect design matrix  $\mathbf{W}_i$ , the  $n_i \times p$  matrix of covariates  $\mathbf{X}_i$  with homogeneous effects on the thresholds, and the  $n_i \times h$  matrix of covariates  $\mathbf{U}_i$  that are allowed to have heterogeneous threshold effects. Provisional starting values for the model parameters must be specified prior to the start of the iterative procedure. These are estimated by the program using an approximate fixed-effects ordinal regression solution for coefficient vector  $\boldsymbol{\beta}$  and thresholds  $\gamma_{k(0)}$  ( $k = 1, \dots, K - 1$ ). Starting values for the Cholesky factor  $\mathbf{T}$  of the random-effects covariance matrix are 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  $Q'$  quadrature points, with summation replacing the integration over the random-effect distribution. The conditional probabilities  $\ell(\mathbf{Y}_i | \boldsymbol{\theta}_i)$  are obtained substituting the random-effect vector  $\boldsymbol{\theta}_i$  by the current  $r$ -dimensional vector of quadrature points  $B_q$ . The marginal density for each level-2 unit is then approximated as:

$$h(\mathbf{Y}_i) \approx \sum_q^{Q'} \ell(\mathbf{Y}_i | B_q) A(B_q).$$

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  $Q'$  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  $B_q$ , and the evaluation of the multivariate standard density  $g(\boldsymbol{\theta})$  is substituted by the current quadrature weight  $A(B_q)$ . Following the summation over level-2 units and quadrature points, parameters are corrected using Equation 14, and the entire procedure is repeated until convergence. With 10 quadrature points for the one dimensional examples described in the

<sup>2</sup> This program can be obtained from <http://www.uic.edu/~hedeker/mix.html>.

following, convergence (corrections of less than .0001 for all parameters) was typically obtained within 20 iterations.

### *Examples of MTCM*

#### *Overview of Data set and Variables*

The data for this example come from "Eclipse," a school-based skin cancer prevention program designed to reduce unprotected sun exposure and to increase sun protection among high school students. The project was conducted in 10 Chicago-area suburban high schools, all with predominantly Caucasian populations. The 10 schools were randomly assigned to one of two treatment conditions: Basic ( $N = 1772$ ) or Enhanced ( $N = 1374$ ). The primary classroom intervention took place during the 1994-1995 academic year. Students in the Basic condition received a one class session covering skin cancer risk factors and protection. Students in the Enhanced condition received the same information as those in the Basic condition, but in addition, received a personalized risk assessment, along with newsletters during the summer of 1995 to help promote sun protection. Data were collected at four timepoints: baseline (Fall, 1994) and three follow-ups (Post 1, immediately following the classroom intervention; Post 2, fall, 1995; and Post 3, fall, 1996). Data reported in this article are from the baseline, Post 2, and Post 3 data collections (all at equivalent times of the year).

#### *Participants*

Participants were 3146 (49.3% female) high school students. At the time of the baseline survey, 74.7% were in the tenth grade and 25.3% were in the ninth grade. Ethnic representation was 78.7% white, 10.2% Asian/Pacific Islander, 7.7% Hispanic, 1.0% African-American, and 2.4% other.

#### *Measures*

At all timepoints, students completed a questionnaire that contained a consistent core of questions: a) demographics and measures of predisposing risk factors (e.g., race, gender, skin type, hair color); b) past and current sun exposure, risk, and protective behaviors; c) future intentions of sun exposure and protective behaviors; d) knowledge about skin cancer, risk factors, and protection; e) related attitudes (e.g., perceived susceptibility, concern, positive attitudes about the sun) and self-efficacy for sun protection.

The dependent variable in this example is stage of change for sunscreen use, which was measured using a combination of items asking about frequency of sunscreen use and whether students intended to protect themselves consistently in the future. Participants were classified into three ordered stages: a) precontemplators were participants who reported not currently regularly using sunscreen (frequency of never, rarely, or sometimes) and who did not intend to do so in the future; b) contemplators were participants who also reported not currently regularly using sunscreen, but who did report that they planned to do so in the future; and c) action participants were students who were currently regularly using sunscreen (frequency of either often or always) regardless of whether they were considering increasing their future use. Because of the timing of the measurement (early fall after the summer), we were unable to further classify participants into a preparation stage. Also, because the prevalence of the maintenance stage was so low, for the purposes of the present study, we grouped participants who could be considered to be in a maintenance stage (those who had consistently used sunscreen for more than one summer) into the action stage.

Table 1 presents the distribution of participants by stage and condition at each measurement wave. Table 2 lists the observed cumulative odds and logits (corresponding to the three stages) across the three timepoints broken down by condition.

Since the first threshold separates precontemplation from the contemplation and action stages, it is termed the contemplation threshold. Similarly, the second threshold is termed the action threshold since it separates those below action to those in the action stage. Inspection of Table 2 indicates that for both the basic and enhanced conditions, the contemplation threshold increases with time, though the increase is somewhat more pronounced for the

Table 1  
Distribution of Participants by Stage and Condition Across Time.

|          | Basic Condition  |                |                | Enhanced Condition |                |                |
|----------|------------------|----------------|----------------|--------------------|----------------|----------------|
|          | Precontemplation | Contemplation  | Action         | Precontemplation   | Contemplation  | Action         |
| Baseline | 37.0%<br>(656)   | 45.8%<br>(812) | 17.2%<br>(304) | 38.1%<br>(524)     | 48.3%<br>(663) | 13.6%<br>(187) |
| Post 2   | 37.9%<br>(593)   | 35.6%<br>(557) | 26.5%<br>(415) | 38.8%<br>(480)     | 36.3%<br>(449) | 24.9%<br>(308) |
| Post 3   | 41.8%<br>(658)   | 34.8%<br>(548) | 23.4%<br>(369) | 40.4%<br>(451)     | 37.5%<br>(419) | 22.1%<br>(247) |

Table 2

Stage of Change by Condition Across Time: Observed Cumulative Odds (and logits)

|          | Basic Condition                         |                                     | Enhanced Condition                     |                                     |
|----------|---|-------------------------------------|--|-------------------------------------|
|          | 1 vs 2-3*<br>Contemplation<br>Threshold | 1-2 vs 3<br>Action<br>Threshold     | 1 vs 2-3<br>Contemplation<br>Threshold | 1-2 vs 3<br>Action<br>Threshold     |
| Baseline | $\frac{656}{1116} = .59$<br>(-.53)      | $\frac{1468}{304} = 4.83$<br>(1.57) | $\frac{524}{850} = .62$<br>(-.48)      | $\frac{1187}{187} = 6.35$<br>(1.85) |
| Post 2   | $\frac{593}{972} = .61$<br>(-.49)       | $\frac{1150}{415} = 2.77$<br>(1.02) | $\frac{480}{757} = .63$<br>(-.46)      | $\frac{929}{308} = 3.02$<br>(1.10)  |
| Post 3   | $\frac{658}{917} = .72$<br>(-.33)       | $\frac{1206}{369} = 3.27$<br>(1.18) | $\frac{451}{666} = .68$<br>(-.39)      | $\frac{870}{247} = 3.52$<br>(1.25)  |

\* 1 = precontemplation, 2 = contemplation, 3 = action

basic condition (i.e., -.53 to -.49 to -.33 for basic and -.48 to -.46 to -.39 for enhanced). Alternatively, the action threshold is reduced across time; again, this change is perhaps more pronounced for the enhanced condition, relative to the basic condition (i.e., 1.57 to 1.02 to 1.18 for basic and 1.85 to 1.10 to 1.25 for enhanced). In other words, over time, more participants are moving out of contemplation and into both precontemplation and action.

As indicated by Table 1, not all subjects were observed at all timepoints. In particular, some attrition occurred across time since 89% and 86% of the original sample were measured at Post 2 and 3, respectively. Since estimation of model parameters is based on a full-likelihood approach, the missing data are assumed to be "ignorable" conditional on both the model covariates and the observed (stages of change) responses (Laird, 1988). In longitudinal studies, ignorable nonresponse falls under Rubin's (1976) "missing at random" (MAR) assumption, in which the missingness depends only on observed data.

In some cases, the MAR assumption may not be reasonable and more sophisticated modeling of the missing data may be required. One approach for dealing with non-ignorable missing data, proposed by Little (1993, 1994,



1995), falls under the rubric of “pattern-mixture models.” In this regard, Little (1995) and Hedeker and Gibbons (1997) describe applications of mixed-effects pattern-mixture models for longitudinal data in which the MAR assumption is too restrictive. In these models, subjects are divided into groups depending on their missing-data pattern. These groups then can be used, for example, to examine the effect of the missing-data pattern on the outcome(s) of interest. Thus, one can examine both the degree to which the missing data patterns differ in terms of outcome, and the degree to which the missing data pattern moderates the other fixed effects in the model (e.g., treatment group). Although not presented here, the pattern-mixture approach can be utilized within the MTCM; for further details see Hedeker and Gibbons (1997).

*Longitudinal Example*

These data were analyzed using a MTCM including effects for time, plus a random subject effect to account for the repeated observations within individuals. For now, the clustering of students within schools is ignored. For the time effects, baseline is treated as the reference cell and comparisons with each of the two post-intervention timepoints are made to it. Represented in matrix form, the model assuming homogeneous time effects (on the  $K - 1$  cumulative logits) for subject  $i$  and cumulative logit  $k$  is given by:

$$\begin{matrix} \begin{bmatrix} \lambda_{i1k} \\ \lambda_{i2k} \\ \lambda_{i3k} \end{bmatrix} \\ \lambda_i \end{matrix} = \begin{matrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ \mathbf{I}_i \end{matrix} \begin{matrix} \begin{bmatrix} \gamma_{i1k} \\ \gamma_{i2k} \\ \gamma_{i3k} \end{bmatrix} \\ \boldsymbol{\gamma}_{ik} \end{matrix} + \begin{matrix} \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \\ \mathbf{X}_i \end{matrix} \begin{matrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \\ \boldsymbol{\beta} \end{matrix} + \begin{matrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \\ \mathbf{W}_i \end{matrix} [\sigma_v] \quad \begin{matrix} [\theta_i] \\ \theta_i \end{matrix}$$

with

$$\begin{matrix} \begin{bmatrix} \gamma_{i1k} \\ \gamma_{i2k} \\ \gamma_{i3k} \end{bmatrix} \\ \boldsymbol{\gamma}_{ik} \end{matrix} = \begin{matrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \\ 1_i \end{matrix} \begin{matrix} [\gamma_{k(0)}] \\ \boldsymbol{\gamma}_{k(0)} \end{matrix}$$

This matrix representation is for a subject with data at all three timepoints. For subjects with incomplete data across time, the corresponding rows of  $\lambda_i$ ,  $\mathbf{I}_i$ ,  $\boldsymbol{\gamma}_{ik}$ ,  $\mathbf{X}_i$ ,  $\mathbf{W}_i$ , and  $1_i$  would be deleted. In this model,  $\beta_1$  and  $\beta_2$  represent, respectively, the Post 2 and Post 3 differences, relative to baseline, on the cumulative logits. As previously described, these parameters are adjusted for the random subject effects, the population variance of the latter is

represented by  $\sigma_v$  in the model. The  $K-1$  thresholds of the model are given by  $\gamma_{k(0)}$ , where  $\gamma_{1(0)}$  represents the contemplation threshold and  $\gamma_{2(0)}$  represents the action threshold. Both time effects  $\beta_1$  and  $\beta_2$  are assumed to have the same effect on these two thresholds, and thus on the two cumulative logits.

To allow for heterogeneous time effects on the thresholds, the model is represented as:

$$\begin{bmatrix} \lambda_{i1k} \\ \lambda_{i2k} \\ \lambda_{i3k} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \gamma_{i1k} \\ \gamma_{i2k} \\ \gamma_{i3k} \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \quad [\sigma v] \quad [\theta_i]$$

$\lambda_i \qquad \qquad \qquad \mathbf{I}_i \qquad \qquad \qquad \gamma_{ik} \qquad \qquad \qquad \mathbf{W}_i \qquad \mathbf{T} \quad \theta_i$

with

$$\begin{bmatrix} \gamma_{i1k} \\ \gamma_{i2k} \\ \gamma_{i3k} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} [\gamma_{k(0)}] + \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_{k1} \\ \alpha_{k2} \end{bmatrix}$$

$\gamma_{ik} \qquad \qquad \mathbf{1}_i \qquad \gamma_{k(0)} \qquad \qquad \mathbf{U}_i \qquad \qquad \alpha_k$

Here,  $\alpha_{11}$  and  $\alpha_{21}$  represent the Post 2 to baseline differences on the contemplation and action thresholds, respectively. Similarly,  $\alpha_{12}$  and  $\alpha_{22}$  represent the Post 3 to baseline differences on these two thresholds. As a result of the coding that is used for the two time contrasts in  $\mathbf{U}$ , the parameters  $\gamma_{1(0)}$  and  $\gamma_{2(0)}$  represent the contemplation and action thresholds, respectively, at baseline (i.e., when all values of the  $\mathbf{u}$  variables equal 0).

Table 3 lists MTCM results assuming homogeneous and heterogeneous time effects. A likelihood-ratio test clearly supports the model that allows the effects of time to vary across the two thresholds ( $\chi^2_2 = 132.7, p < .001$ ). The conclusions also vary between the two models. The homogeneous effects model indicates a significant Post 2 ( $\hat{\beta}_1 = -.327$ ) and a marginally significant Post 3 ( $\hat{\beta}_2 = -.100$ ) reduction on both thresholds compared to baseline. Allowing for differential effects on the thresholds indicates that at Post 2 the significant reduction is only in terms of the action threshold ( $\hat{\alpha}_{21} = -.894$ ), while at Post 3 there is actually a significant increase in the contemplation threshold ( $\hat{\alpha}_{12} = .232$ ) in addition to a significant reduction in the action threshold ( $\hat{\alpha}_{22} = -.655$ ). Thus, relative to baseline, more subjects are classified in action at Post 2, and in both precontemplation and action at Post 3.

For these models assuming normally-distributed random-effects, the estimated subject variance can be expressed as an approximate intraclass correlation,  $\sigma_v^2 / (\sigma_v^2 + \sigma^2)$ , where  $\sigma^2$  represents the variance of the latent continuous readiness of change variable. For the logistic distribution, the

Table 3  
Multilevel Thresholds of Change Model Estimates (Standard Errors): Homogeneous and Heterogeneous Time Effects on Thresholds.

| Parameter                               | Homogeneous Effects | Heterogeneous Effects |
|---|---------------------|-----------------------|
| Contemplation threshold $\gamma_{1(0)}$ | -.556*<br>(.054)    | -.777*<br>(.055)      |
| Action threshold $\gamma_{2(0)}$        | 2.065*<br>(.061)    | 2.456*<br>(.073)      |
| Post 2 $\beta_1$                        | -.327*<br>(.056)    |                       |
| Post 2 contemplation $\alpha_{11}$      |                     | .026<br>(.065)        |
| Post 2 action $\alpha_{21}$             |                     | -.894*<br>(.085)      |
| Post 3 $\beta_2$                        | -.100**<br>(.055)   |                       |
| Post 3 contemplation $\alpha_{12}$      |                     | .232*<br>(.064)       |
| Post 3 action $\alpha_{22}$             |                     | -.655*<br>(.084)      |
| Subject sd $\sigma_v$                   | 1.706<br>(.047)     | 1.727<br>(.048)       |
| -2 log L                                | 17203.0             | 17070.3               |

\* $p < 0.01$ . \*\* $p < 0.10$ .

standard variance  $\sigma^2$  equals  $\pi^2/3$  (Agresti, 1990). The estimated intraclass correlations for the models in Table 3 are then .47 and .48 for the proportional odds and partial proportional odds models, respectively. As expected, the stage data are highly correlated within subjects. It should be noted that there are some concerns in using the standard errors in constructing a hypothesis test for the random-effect variance term (i.e., the subject standard deviation  $\sigma_v$ ), particularly when the variance is near-zero and the number of subjects is small (Bryk & Raudenbush, 1992). As a result, we do not indicate statistical significance for the random-effect variance parameters in the tables.

To examine the degree of model fit to the observed proportions across time, we calculated model-based estimated marginal proportions across time as described above. These are plotted, along with the observed marginal proportions, for both homogeneous and heterogeneous MTCM in Figure 1 (pages 448-449). Figures 1a-1c were obtained using 10 point quadrature while Figures 1d-1f were obtained using the approximate marginalization method of Diggle et al. (1994), both described above.

As can be seen, the two methods of obtaining marginal estimates yield very similar results, though the quadrature method does slightly better. The figure also illustrates the better fit of the heterogeneous relative to the homogeneous MTCM. Thus, an improvement in model fit is observed by allowing the Time effects to vary across the two thresholds.

To examine whether there is an condition effect on the stage outcomes, Group and Group by Time were added to both models. The results of these analyses are listed in Table 4. Again, the likelihood-ratio test supports the heterogeneous MTCM that allows the effects of Time, Group, and Group by Time to vary across the two thresholds ( $\chi^2_5 = 138.6, p < .001$ ). For both models there is no evidence of significant Group by Time interaction, however, for the partial proportional odds model there is evidence of a Group difference at Baseline in terms of the action threshold ( $z = .362/.138 = 2.62, p < .01$ ). Specifically, at Baseline, the action threshold is elevated for the enhanced group relative to the basic group. This can be seen in Table 2 where the corresponding observed logits equal 1.85 and 1.57 for the enhanced and basic groups, respectively. There is no evidence for a group effect at the other timepoints.

### *Clustered Example*

To assess the degree of clustering attributable to the nesting of students within schools we examined models at each timepoint separately. These analyses include a random school effect to account for the clustering of students within schools in addition to the Group effect, which is allowed to vary across thresholds. Table 5 (page 450) lists the results of these analyses.

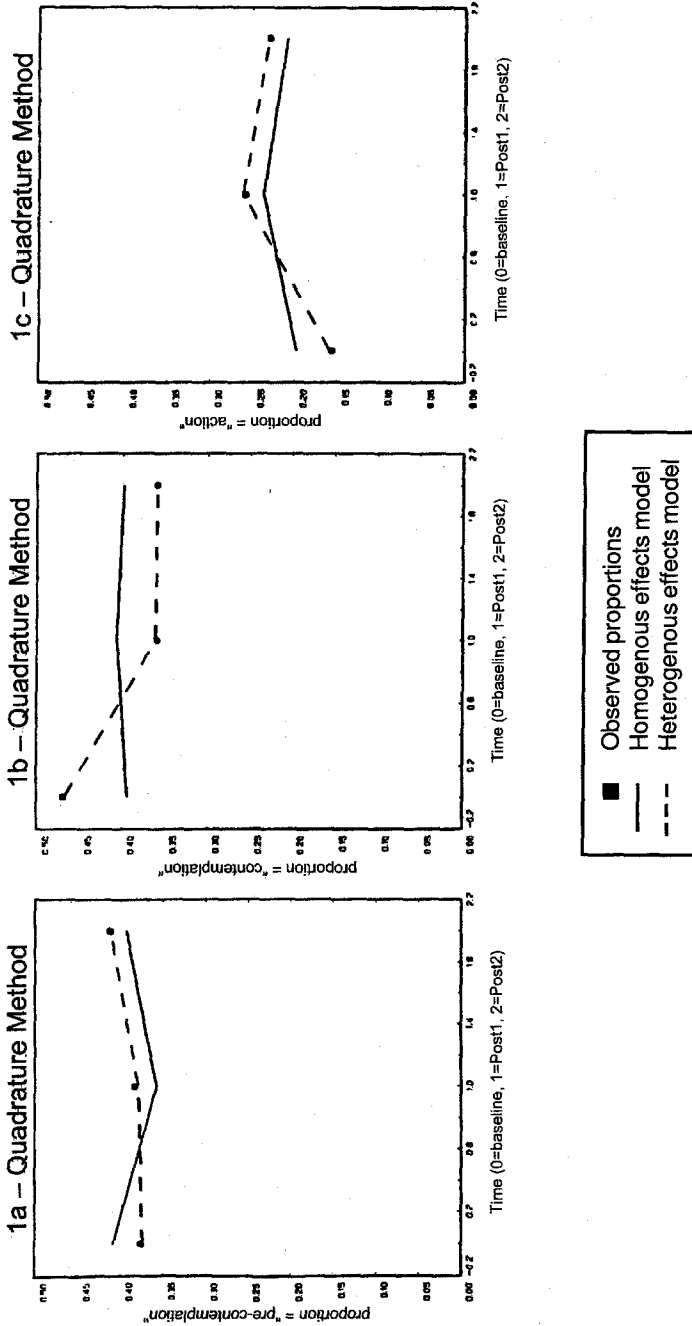
At baseline, there is evidence of a significantly different Group effect on the thresholds by the likelihood-ratio test ( $\chi^2_1 = 4.9, p < .05$ ). As in the longitudinal analysis, at baseline, the action threshold is elevated for the enhanced group relative to the basic group. ( $z = .306/.138 = 2.22, p < .05$ ). In terms of a clustering effect, the school standard deviation is relatively small and expressed as an intra-school correlation equals .004. Thus, accounting for the Group effect, less than 1% of the variation is attributable to the

Table 4  
Multilevel Thresholds of Change Model Estimates (Standard Errors): Homogeneous and Heterogeneous Time, Group, and Group by Time Effects.

| Parameter                          | Homogeneous Effects |                  | Heterogeneous Effects |                  |
|------------------------------------|---------------------|------------------|-----------------------|------------------|
|                                    | Contemplation       | Action           | Contemplation         | Action           |
| Threshold                          | -.625*<br>(.070)    | 1.997*<br>(.075) | -.811*<br>(.073)      | 2.305*<br>(.091) |
| Post 2                             | -.297*<br>(.074)    |                  | .032<br>(.086)        | -.790*<br>(.111) |
| Post 3                             | -.025<br>(.073)     |                  | .301*<br>(.086)       | -.540*<br>(.109) |
| Group (0 = basic,<br>1 = enhanced) | .157<br>(.106)      |                  | .078<br>(.111)        | .362*<br>(.138)  |
| Group × Post 2                     | -.069<br>(.113)     |                  | -.014<br>(.131)       | -.254<br>(.171)  |
| Group × Post 3                     | -.174<br>(.111)     |                  | -.165<br>(.129)       | -.278<br>(.170)  |
| Subject <i>SD</i> $\sigma_v$       | 1.706<br>(.047)     |                  | 1.727<br>(.048)       |                  |
| -2 log <i>L</i>                    |                     | 17199.4          |                       | 17060.8          |

\* $p < 0.01$ .

clustering of students within schools. At the subsequent timepoints, the clustering effect is even smaller (intra-school correlation equals .0003 and .0007, respectively) and there is no evidence of significant Group-related effects. It should be noted that with only 10 schools in the sample, there is not much information available for estimation of the population school variance term.



**Figure 1**  
Plots of Model Fit—Quadrature Method versus Marginalization Method

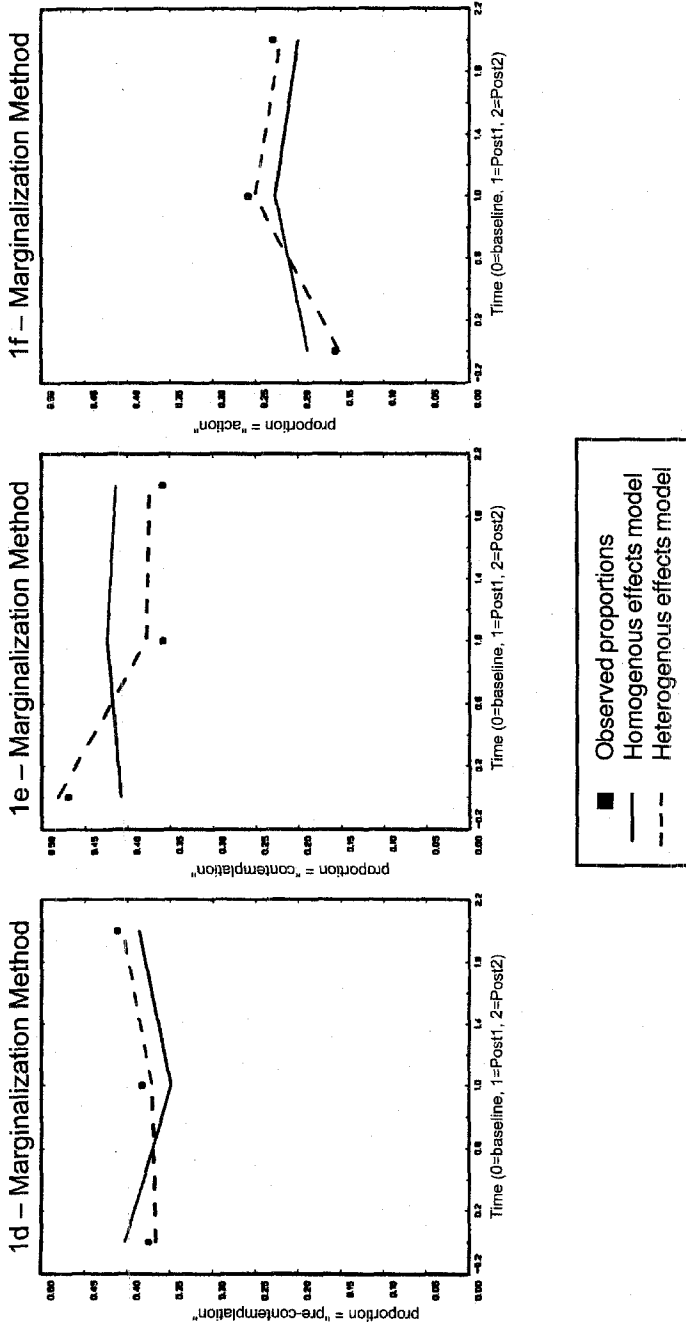


Figure 1 (cont.)

Plots of Model Fit—Quadrature Method versus Marginalization Method

Table 5

Multilevel Thresholds of Change Model Estimates (Standard Errors). Homogeneous and Heterogeneous Group Effects on Thresholds at Baseline: Students Clustered by Schools Model.

| Parameter                                 | Homogeneous Effects | Heterogeneous Effects |
|---|---------------------|-----------------------|
| Contemplation threshold $\gamma_{1(0)}$   | -.571*<br>(.089)    | -.541*<br>(.105)      |
| Action threshold $\gamma_{2(0)}$          | 1.634*<br>(.078)    | 1.569*<br>(.084)      |
| Group (0 = basic, 1 = enhanced) $\beta_1$ | .144<br>(.138)      |                       |
| Group contemplation $\alpha_{11}$         |                     | .079<br>(.212)        |
| Group action $\alpha_{21}$                |                     | .306**<br>(.138)      |
| School SD $\sigma_v$                      | .114<br>(.092)      | .115<br>(.129)        |
| -2 log L                                  | 6367.7              | 6362.8                |

\* $p < 0.01$ . \*\* $p < 0.05$ .

### Discussion

We have described a random-effects thresholds of change model for statistical analysis of multilevel stages of change data. A main feature of this model is its focus on the thresholds that separate the ordered stages. Just as the stages are ordered, the thresholds are also ordered, each of increasing magnitude. By assuming an underlying distribution for a latent readiness to change variable, the probability of crossing each threshold can be determined for the population of subjects. The most common choices for this underlying distribution are the logistic and normal, leading respectively, to ordinal logistic and probit regression models.

Explanatory variables can exert their influence on these thresholds, and the effects can be assumed to be the same or to vary across thresholds. Assuming homogeneous effects is practical for continuous explanatory



variables, however, for categorical explanatory variables differential effects are realistic and obtainable. This latter feature is especially attractive in health promotion research since one can estimate the influence of an intervention (or other grouping of subjects) on each threshold separately. In our example from a skin cancer prevention study, we observed differential effects due to time and intervention group on the thresholds separating the three stages of sunscreen use (precontemplation, contemplation, and action).

The model presented in this article used time-related contrasts as explanatory variables to examine changes in stage membership across time. We further included interactions with the time-related contrasts to examine whether changes in stage membership across time varied for different groups of individuals (e.g., treatment group). Another approach that has been used for longitudinal stage data is to define the stages of change as a dynamic latent variable (Velicer, Martin, & Collins, 1996). This latter approach, termed latent transition analysis, allows a more complete view of the transitions between particular stages, though it is more limited in the number of explanatory variables that can be accommodated. A software program for latent transition analysis is described by Collins, Wugalter, and Rousculp (1991).

Though the sample size in the example presented in this article is quite large, this is not always the case. The necessary sample size for application of the MTCM is an important issue, however it is not easy to provide global recommendations. An important consideration is the numbers of observations in the  $K$  response categories as they are broken down by the covariates and covariate interactions of the model. These numbers may vary quite a bit depending on whether the covariate is at level 1 or 2. Consider the simplest case of one covariate with two categories (e.g., gender) and a response variable with  $K$  categories. The data then form a  $2 \times K$  crosstabulation table. Estimating heterogeneous threshold effects for gender then requires observations in each gender group at each of the  $K - 1$  comparisons across the response variable (i.e., category 1 versus 2 through  $K$  combined, categories 1 and 2 combined versus 3 through  $K$  combined, . . . , categories 1 through  $K-1$  combined versus  $K$ ). In this situation, observations in the extreme categories (i.e., 1 and  $K$ ) are most critical, since gender effects on thresholds 1 and  $K-1$  cannot be estimated if either gender group has no observations in response categories 1 and  $K$ , respectively. Since allowing higher-order interactions to have heterogeneous threshold effects splits the data up even further, this may not always be possible or may require collapsing of covariate or response categories. Thus, data sparseness can restrict applicability of the MTCM or limit the number of covariates (and/or interactions) that are allowed to have heterogeneous threshold effects. A further point regarding sample size is that the significance tests that are formed by taking the ratio of the parameter estimate to its standard error are based on asymptotic

statistical theory. In this regard though, many other statistical techniques using maximum likelihood estimation also invoke asymptotic theory for hypothesis testing (e.g., logistic regression, log-linear models, and structural equation models). For more details on asymptotic theory as applied to mixed-effects models like the MTCM, see Longford (1993).

The examples illustrate the utility of the random-effects approach for correlated (i.e., clustered or longitudinal) stage data. In particular, random-effects models are useful in accounting for variability attributable to data clustering, while concurrently estimating effects of model covariates. The variance terms are also estimated indicating the degree of data clustering. The present model can be applied to either clustered or longitudinal data, however, it cannot handle data that are both clustered and longitudinal. A more appropriate analysis of the example dataset would account for both the variation attributable to subjects and schools. An extension of the model is underway to accommodate clustered data where the clustered subjects are also repeatedly measured across time. This development follows the approach described in the three-level model for dichotomous response data (Gibbons & Hedeker, 1997).

As mentioned, the GEE approach of Liang and Zeger (1986) can also be applied for correlated data, both where focus is on the fixed effects alone (GEE1), and where interest is on both the fixed effects and the correlation structure of the multilevel data (GEE2). For modeling correlated ordinal data, Heagerty and Zeger (1996) describe both GEE1 and GEE2 estimation procedures. Their development also includes the possibility of non-proportional odds for covariate effects by incorporating threshold by covariate interactions. Thus, the MTCM presented here could be recast as a GEE model by applying Heagerty and Zeger's work. An advantage of the GEE approach is that it is often simpler to implement. Also, parameter estimates are on the marginal scale and statistical tests for model parameters are robust to misspecification of the correlation structure. Alternatively, by explicitly including random effects in the model, parameter estimates from random-effects models depend on the distribution specified for the random effects, and the statistical tests are valid only for the assumed correlation structure (e.g., compound symmetry structure for a model with a single random subject effect). As described in this article, marginal estimates can be obtained from a random-effects model, though additional work is required. Also, the quadrature solution does allow specification of the random-effect distribution to be varied. An advantage of the random-effects approach is that estimates of the random effects themselves can be obtained, and these are often of interest in behavioral research (see Bock, 1989). Another consideration concerns the assumption made about the missing data: the

random-effects approach using full-likelihood estimation procedures assumes the missing data are missing at random (MAR) as opposed to the more restrictive missing completely at random (MCAR) assumption of the GEE-based approach. In this regard, Kenward, Lesaffre, and Molenberghs (1994) caution use of non-likelihood based methods (e.g., GEE) for analysing longitudinal data when missing data are not MCAR. Further discussion on the differences between MAR and MCAR can be found in Little and Rubin (1987), and further discussion about the differences between the random-effects and GEE approaches can be found in Zeger, Liang, and Albert (1988) and Neuhaus, Kalbfleisch, and Hauck (1991).

Although we have illustrated model usage applied to stage data, it can be applied more generally to many types of correlated ordinal data. Ordinal responses are frequently encountered in behavioral research, especially when dealing with rating scale data (Bock, 1975). Similarly, in epidemiologic studies, health status is often measured on an ordinal scale from "absent" to "mild" to "severe" (Armstrong & Sloan, 1989). The present model extends the random-effects ordinal regression model of Hedeker and Gibbons (1994) by relaxing the proportional odds assumption for explanatory variables. Since ordinal outcomes are common in many research areas and, as noted by Peterson and Harrell (1990), examples of non-proportional odds are not difficult to find, the present model represents a useful extended approach for analysis of correlated ordinal data.

## References

- Agresti, A. (1990). *Categorical data analysis*. New York: Wiley.
- Agresti, A. & Lang, J. B. (1993). A proportional odds model with subject-specific effects for repeated ordered categorical responses. *Biometrika*, 80, 527-534.
- Armstrong, B. G. & Sloan, M. (1989). Ordinal regression models for epidemiologic data. *American Journal of Epidemiology*, 129, 191-204.
- Bock, R. D. (1975). *Multivariate statistical methods in behavioral research*. New York: McGraw-Hill.
- Bock, R. D. (1989). Measurement of human variation: A two stage model. In: R. D. Bock (Ed.), *Multilevel analysis of educational data* (pp. 319-342). New York: Academic Press.
- Bock, R. D. & 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., & Muraki, E. (1988). Full-information item factor analysis. *Applied Psychological Measurement*, 12, 261-280.
- Bock, R. D. & Lieberman, M. (1970). Fitting a response model for  $n$  dichotomously scored items. *Psychometrika*, 35, 179-197.
- Bryk, A. S. & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage Publications, Inc.
- Collins, L. M., Wugalter, S. E., & Rousculp, S. S. (1991). *LTA user's manual*. Los Angeles: J.P. Guilford Laboratory of Quantitative Psychology, University of Southern California.

D. Hedeker and R. J. Mermelstein

- Diggle, P., Liang, K.-Y., & Zeger, S. L. (1994). *Analysis of longitudinal data*. New York: Oxford University Press.
- Ezzet, F. & Whitehead, J. (1991). A random effects model for ordinal responses from a crossover trial. *Statistics in Medicine*, 10, 901-907.
- Gange, S. J., Linton, K. L. P., Scott, A. J., DeMets, D. L., & Klein, R. (1995). A comparison of methods for correlated ordinal measures with ophthalmic applications. *Statistics in Medicine*, 14, 1961-1974.
- Gibbons, R. D. & Bock, R. D. (1987). Trend in correlated proportions. *Psychometrika*, 52, 113-124.
- Gibbons, R. D. & Hedeker, D. (1997). Random-effects probit and logistic regression models for three-level data. *Biometrics*, 53, 1527-1537.
- Goldstein, H. (1995). *Multilevel statistical models* (2nd edition). New York: Halstead Press.
- Harville, D. A. & Mee, R. W. (1984). A mixed-model procedure for analyzing ordered categorical data. *Biometrics*, 40, 393-408.
- Heagerty, P. J. & Zeger, S. L. (1996). Marginal regression models for clustered ordinal measurements. *Journal of the American Statistical Association*, 91, 1024-1036.
- Hedeker, D. & Gibbons, R. D. (1994). A random-effects ordinal regression model for multilevel analysis. *Biometrics*, 50, 933-944.
- Hedeker, D. & Gibbons, R.D. (1996). MIXOR: A computer program for mixed-effects ordinal regression analysis. *Computer Methods and Programs in Biomedicine*, 49, 157-176.
- Hedeker, D. & Gibbons, R. D. (1997). Application of random-effects pattern-mixture models for missing data in longitudinal studies. *Psychological Methods*, 2, 64-78.
- Hedeker, D., Mermelstein, R. J., & Weeks, K. A. (in press). The thresholds of change model: an approach to analyzing stages of change data. *Annals of Behavioral Medicine*.
- Jansen, J. (1990). On the statistical analysis of ordinal data when extravariation is present. *Applied Statistics*, 39, 75-84.
- Kenward, M. G., Lesaffre, E. & Molenberghs, G. (1994). An application of maximum likelihood and generalized estimating equations to the analysis of ordinal data from a longitudinal study with cases missing at random. *Biometrics*, 50, 945-953.
- Laird, N. M. (1988). Missing data in longitudinal studies. *Statistics in Medicine*, 7, 305-315.
- Liang, K.-Y. & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73, 13-22.
- Liang, K.-Y. Zeger, S. L., & Qaqish, B. (1992). Multivariate regression analyses for categorical data. *Journal of the Royal Statistical Society, Series B*, 54, 3-40.
- Little, R. J. A. (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association*, 88, 125-133.
- Little, R. J. A. (1994). A class of pattern-mixture models for normal incomplete data. *Biometrika*, 81, 471-483.
- Little, R. J. A. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association*, 90, 1112-1121.
- Little, R. J. A. & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: Wiley.
- Longford, N. T. (1993). *Random coefficient models*. New York: Oxford University 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.
- Miller, M. E., Davis, C. S., & Landis, J. R. (1993). The analysis of longitudinal polytomous responses: generalized estimating equations and connections with weighted least squares. *Biometrics*, 49, 1033-1044.

- Molenberghs, G. & Lesaffre, E. (1994). Marginal modeling of correlated ordinal data using a multivariate Plackett distribution. *Journal of the American Statistical Association*, 89, 633-644.
- Neuhaus, J. M., Kalbfleisch, J. D., & Hauck, W. W. (1991). A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data. *International Statistical Review*, 59, 25-35.
- Peterson, B. & Harrell, F. E. (1990). Partial proportional odds models for ordinal response variables. *Applied Statistics*, 39, 205-217.
- Prochaska, J. O. & DiClemente, C. C. (1983). Stages and processes of self-change of smoking: toward an integrative model of change. *Journal of Consulting and Clinical Psychology*, 51, 390-395.
- Prochaska, J. O. & Velicer, W. F. (1997). The transtheoretical model of health behavior change. *American Journal of Health Promotion*, 12, 38-48.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63, 581-592.
- Stroud, A. H. & Sechrest, D. (1966). *Gaussian Quadrature Formulas*. Englewood Cliffs, NJ: Prentice Hall.
- 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.
- Terza, J. (1985). Ordinal probit: a generalization. *Communications in Statistics*, 14, 1-12.
- Velicer, W. F., Martin, R. A., & Collins, L. M. (1996). Latent transition analysis for longitudinal data. *Addiction*, 91 (Supplement), S197-S209.
- Zeger, S. L., Liang, K.-Y. & Albert, P. A. (1988). Models for longitudinal data: a generalized estimating equation approach. *Biometrics*, 44, 1049-1060.