



---

Modeling Clustered Ordered Categorical Data: A Survey

Author(s): Alan Agresti and Ranjini Natarajan

Source: *International Statistical Review / Revue Internationale de Statistique*, Vol. 69, No. 3 (Dec., 2001), pp. 345-371

Published by: International Statistical Institute (ISI)

Stable URL: <https://www.jstor.org/stable/1403450>

Accessed: 27-05-2020 18:17 UTC

---

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



JSTOR

*International Statistical Institute (ISI)* is collaborating with JSTOR to digitize, preserve and extend access to *International Statistical Review / Revue Internationale de Statistique*

# Modeling Clustered Ordered Categorical Data: A Survey

Alan Agresti and Ranjini Natarajan

*Department of Statistics, University of Florida, Gainesville, Florida, USA 32611-8545*  
*E-mail: aa@stat.ufl.edu; ranjini@stat.ufl.edu*

## Summary

This article surveys various strategies for modeling ordered categorical (ordinal) response variables when the data have some type of clustering, extending a similar survey for binary data by Pendergast, Gange, Newton, Lindstrom, Palta & Fisher (1996). An important special case is when repeated measurement occurs at various occasions for each subject, such as in longitudinal studies. A much greater variety of models and fitting methods are available than when a similar survey for repeated ordinal response data was prepared a decade ago (Agresti, 1989). The primary emphasis of the review is on two classes of models, *marginal models* for which effects are averaged over all clusters at particular levels of predictors, and *cluster-specific models* for which effects apply at the cluster level. We present the two types of models in the ordinal context, review the literature for each, and discuss connections between them. Then, we summarize some alternative modeling approaches and ways of estimating parameters, including a Bayesian approach. We also discuss applications and areas likely to be popular for future research, such as ways of handling missing data and ways of modeling agreement and evaluating the accuracy of diagnostic tests. Finally, we review the current availability of software for using the methods discussed in this article.

*Key words:* Bayesian inference; Cumulative logit models; Generalized estimating equations, Logit models; Marginal models; Matched pairs; Missing data; Ordinal data; Proportional odds; Random effects; Repeated measures; Square contingency tables.

## 1 Introduction

This article surveys ways of modeling ordered categorical (ordinal) response variables when the data have some type of clustering, such as a cluster of responses for a subject measured repeatedly in a longitudinal study. The development of such models has been rather slow, compared to models for independent observations. Much of this is due to computational complexity, as the major classes of models each present difficulties of this type. For instance, one class of models applies not to the probabilities to which the likelihood refers but to marginal probabilities. For another class of models, it is difficult even to calculate the likelihood. However, the past decade has seen major progress, and it is now possible to handle much more complex models and larger data sets, and a greater variety of model-fitting methods have also been developed.

### 1.1 Example

For a simple illustration of the methods discussed in this article, we use Table 1, a dataset typical of many studies in the pharmaceutical industry. The table shows results of a randomized, double-

blind clinical trial comparing an active hypnotic drug with a placebo in patients with insomnia. The outcome variable is patient response to the question "How quickly did you fall asleep after going to bed?", using categories ( $< 20$ ,  $20 - 30$ ,  $30 - 60$ ,  $> 60$ ) minutes. Patients responded at the beginning and conclusion of a two-week treatment period. Here, each subject forms a cluster, with a natural dependence between the two observations for each subject. These data were used to illustrate various analyses in Agresti (1989), and results quoted there will be compared to ones obtained using models discussed in this paper.

**Table 1**  
*Distribution of time to fall asleep, by treatment and occasion.*

Treatment	Initial occasion	Follow-up occasion			
		$< 20$	$20-30$	$30-60$	$> 60$
Active	$< 20$	7	4	1	0
	$20-30$	11	5	2	2
	$30-60$	13	23	3	1
	$> 60$	9	17	13	8
Placebo	$< 20$	7	4	2	1
	$20-30$	14	5	1	0
	$30-60$	6	9	18	2
	$> 60$	4	11	14	22

Source: Francom, Chunag-Stein & Landis (1989).

## 1.2 Cluster-specific and Marginal Models

In the formulation of models, we refer to the sampling units as *clusters*. In many applications, such as Table 1, each cluster is a set of repeated measurements on a subject. In others, each cluster is a set of subjects expected to be more similar to each other than to other subjects, such as a litter of mice in a teratology study.

As with binary and other forms of categorical data (Pendergast *et al.*, 1996), two major types of model for ordinal responses differ in terms of whether they have *population-averaged* or *cluster-specific* (sometimes called *subject-specific*) effects. The latter models refer to conditional distributions at the cluster (e.g., subject) level, whereas the former models refer to marginal distributions, averaged over clusters in the population. The choice of model affects whether parameter interpretations apply at the cluster or the population level (Zeger, Liang & Albert, 1988; Neuhaus, Kalbfleisch & Hauck, 1991; Ten Have, Landis & Hartzel, 1996). Population-level interpretations are more relevant in epidemiological studies that focus on overall frequency of occurrence in a population. As we will discuss in Section 3.2, approximate relationships exist between population-averaged parameters in marginal models and corresponding parameters in cluster-specific models; however, a cluster-specific model usually does not imply a marginal model of the same form, and a marginal model need not have any simple and meaningful cluster-specific model that implies it.

A third type of model, called a *transition model*, is used in longitudinal studies to describe the distribution of a response conditional on past responses and explanatory variables. An important

special case is Markov chain models. Transition models, unlike models discussed here, take into account the time ordering. This approach has received substantial attention for binary data (e.g., Bonney, 1986; Diggle *et al.* 1994, Ch. 10), but seems to have received little attention so far for ordinal responses (Chuang & Francom, 1986; Albert, 1994).

### 1.3 Link Functions for Ordinal Models

Logistic regression models occupy a central place in the toolkit for analyzing binary data. Likewise, the most popular models for ordinal responses are multi-category generalizations of logistic regression. Currently, the most popular model for ordinal responses uses logits of cumulative probabilities, called *cumulative logits*. For a  $c$ -category ordinal response variable  $Y$  and a set of predictors  $\mathbf{x}$  with corresponding effect parameters  $\beta$ , the model has form

$$\text{logit}[P(Y \leq k)] = \alpha_k + \mathbf{x}'\beta, \quad k = 1, \dots, c - 1. \quad (1)$$

Model (1) assumes an identical effect  $\beta$  of the predictors for each cumulative probability; this case is referred to as a *proportional odds model* (McCullagh, 1980). One can motivate such a model using a regression model for an assumed underlying continuous response with constant variance (Anderson & Philips, 1981). The logit link for cumulative probabilities results from a logistic latent variable. Other latent variables yield other “cumulative links” (i.e., links for cumulative probabilities), such as the probit link for a normal latent variable and a complementary log-log link (i.e.,  $\log\{-\log[1 - P(Y \leq k)]\}$ ) for a latent variable having an extreme value distribution (McCullagh, 1980). Generalizations exist in which some or all of the effects need not be identical for each cumulative probability (Peterson & Harrell, 1990).

An alternative ordinal logit model utilizes single-category probabilities rather than cumulative probabilities and hence can be expressed using the canonical parameters for the multinomial distribution, which are the baseline-category logits  $\log[P(Y = k)/P(Y = c)]$ ,  $k = 1, \dots, c - 1$ . One expression of such an ordinal model uses the *adjacent-categories logits*,  $\log[P(Y = k)/P(Y = k + 1)]$ ,  $k = 1, \dots, c - 1$ . Another possibility is the *continuation-ratio* logit model, which uses logits  $\{\log[P(Y = k)/P(Y \geq k + 1)]\}$  or  $\{\log[P(Y = k + 1)/P(Y \leq k)]\}$ . This model form is useful when a sequential mechanism, such as survival through various age periods, determines the response outcome (Tutz, 1991).

When used with a single parameter for each effect (as in the proportional odds form (1) of the cumulative logit model), these model types tend to fit well in similar situations. They all imply that the response distributions are stochastically ordered at the various predictor values. One’s choice of model is usually based less on goodness of fit than on whether one prefers parameter interpretation to refer to individual response categories or instead to groupings of categories or an underlying continuous variable. When such models with the common effect  $\beta$  for each  $k$  fit poorly, possible strategies include (i) trying a link function, such as the log-log or complementary log-log, for which the response curve is nonsymmetric, (ii) adding additional terms, such as interactions, to the linear predictor, (iii) generalizing the model by adding dispersion parameters (McCullagh, 1980; Cox, 1995), and (iv) permitting separate effects for each logit for some of the predictors (Peterson & Harrell, 1990).

Another common approach to analyze ordinal response variables assigns scores to categories and uses ordinary regression or ANOVA methods. Weighted least squares can account for nonconstant variance in the response functions (Grizzle, Starmer & Koch, 1969; Koch, Landis, Freeman, Freeman & Lehnen, 1977). This approach has the advantage of simplicity of interpretation, particularly when it is sufficient to describe effects in terms of simple summaries such as mean location rather than odds ratios based on the individual cell probabilities. This is often the case when  $c$  is large. With models for the mean itself, marginal and cluster-specific models are compatible, and the same effects

occur in each.

### 1.4 Scope of this Article

Agresti (1999) surveyed recent research on modeling ordered categorical data. In this article, we extend that work by providing a more detailed investigation of methods pertaining specifically to models for clustered ordered categorical data. The article complements Pendergast *et al.* (1996), a similar survey for clustered binary data. We review the literature, emphasizing the two major types of models, *marginal models* and *cluster-specific models*, and provide examples of their use. Section 2 introduces marginal models and discusses ways of fitting them using maximum likelihood and the quasi-likelihood method of generalized estimating equations. Section 3 introduces cluster-specific models specified by introducing cluster-specific random effects and discusses maximum likelihood fitting and connections with marginal models. Section 4 summarizes some alternative modeling approaches and ways of estimating parameters, such as the Bayesian approach. Section 5 summarizes applications of the models that have had considerable attention and describes areas that are likely to see substantial research activity in coming years. Section 6 reviews the relatively limited current availability of software for performing these analyses. Finally, we reference much of the literature that has evolved in the past decade for modeling clustered ordered categorical data.

## 2 Marginal Models for a Repeated Ordinal Response

Let  $\pi_k(\mathbf{x}; i)$  denote the probability that the response is category  $k$ , for an observation in cluster  $i$  with values  $\mathbf{x}$  of explanatory variables,  $k = 1, \dots, c$ . Let  $\pi_k(\mathbf{x})$  denote a corresponding “population-averaged” probability, that is, the mean of the probabilities  $\pi_k(\mathbf{x}; i)$  for all clusters in a population of interest at setting  $\mathbf{x}$ . Let  $\gamma_k(\mathbf{x}; i) = \pi_1(\mathbf{x}; i) + \dots + \pi_k(\mathbf{x}; i)$  and  $\gamma_k(\mathbf{x}) = \pi_1(\mathbf{x}) + \dots + \pi_k(\mathbf{x})$ , with  $\gamma_0(\mathbf{x}; i) = \gamma_0(\mathbf{x}) = 0$ .

For the set of predictor values  $\mathbf{x}^* = (\mathbf{x}_1, \dots, \mathbf{x}_n)$  for  $n$  observations in a cluster, let  $\pi_{k_1, \dots, k_n}(\mathbf{x}^*)$  be the marginal probability of the sequence of responses  $(k_1, \dots, k_n)$ . Assuming independent observations within clusters (given the cluster-specific probabilities), this is the mean value of  $[\pi_{k_1}(\mathbf{x}_1; i) \dots \pi_{k_n}(\mathbf{x}_n; i)]$  for all clusters at that combination of predictor values. Thus,  $\pi_{k_1, \dots, k_n}(\mathbf{x}^*)$  is marginal in terms of averaging over clusters, and  $\pi_k(\mathbf{x})$  is a further first-order marginal distribution of this  $n$ -dimensional distribution.

The simplest special case is that of matched pairs without covariates. Then,  $n = 2$  and  $\mathbf{x}^* = (1, 2)$  simply indicates the two responses in each matched pair. For instance, for a  $c \times c$  square table of probabilities  $\{\pi_{kk'}(1, 2)\}$ ,  $\pi_k(1) = \sum_{k'} \pi_{kk'}(1, 2)$  and  $\pi_k(2) = \sum_{k'} \pi_{k'k}(1, 2)$  are row and column marginal totals.

### 2.1 Model Specification and Complications

This section discusses cumulative logit models for the first-order marginal probabilities. Let  $n_i$  denote the number of observations in cluster  $i$ . Denote the values of the explanatory variables for the  $n_i$  observations in cluster  $i$  by  $\mathbf{x}_i^* = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{in_i})$ ,  $i = 1, \dots, N$ . The models have the form

$$\text{logit}[\gamma_k(\mathbf{x}_{ij})] = \alpha_k + \mathbf{x}_{ij}' \boldsymbol{\beta}, \quad k = 1, \dots, c - 1. \quad (2)$$

Let  $Y_{k_1, \dots, k_{n_i}}(i) = 1$  if cluster  $i$  makes the set of responses  $(k_1, \dots, k_{n_i})$  and let  $Y_{k_1, \dots, k_{n_i}}(i) = 0$  otherwise. Marginally, we assume that these indicators are trials from a multinomial distribution

with parameters  $\{\pi_{k_1, \dots, k_{n_i}}(\mathbf{x}_i^*)\}$ . Then, the log likelihood equals

$$\sum_{i=1}^N \left( \sum_{k_1=1}^c \sum_{k_2=1}^c \cdots \sum_{k_{n_i}=1}^c y_{k_1, \dots, k_{n_i}}(i) \log [\pi_{k_1, \dots, k_{n_i}}(\mathbf{x}_i^*)] \right),$$

where  $y_{k_1, \dots, k_{n_i}}(i)$  denotes a realization of  $Y_{k_1, \dots, k_{n_i}}(i)$ . It is awkward to maximize this log likelihood because model (2) refers to marginal probabilities  $\{\pi_k(\mathbf{x}_{ij})\}$  of the multinomial probabilities  $\{\pi_{k_1, \dots, k_{n_i}}(\mathbf{x}_i^*)\}$  in the log likelihood. Hence, one cannot directly substitute the model formula in the log likelihood function and maximize using standard methods.

Early approaches to this problem used a weighted least squares (WLS) approach (Koch *et al.*, 1977; Landis, Miller, Davis & Koch, 1988; Carr, Hafner & Koch, 1989; Agresti, 1989). For cumulative logit models, this used the delta method to derive a multivariate normal asymptotic distribution for the vector of all the sample marginal cumulative logits, and then minimized the quadratic form comparing those sample logits to the model predicted values, weighting by the inverse of the estimated covariance matrix of the sample logits. The delta method is based on the asymptotic normality of sample cell proportions in a contingency table. Thus, although this method is simple in terms of having a closed-form solution for the parameter estimates, it is severely limited because all predictors must be categorical and the marginal counts must be nonsparse (e.g., ideally all marginal totals having size at least about 5–10).

### 2.2 Quasi Likelihood: Generalized Estimating Equation (GEE) Methodology

The awkwardness of fitting marginal models such as just formulated reflects the lack of a simple multivariate family of distributions for categorical responses. Unlike the multivariate normal family for continuous responses, with categorical responses it is not natural to parameterize in terms of marginal moments and a correlation structure. Rather than attempt to specify fully the joint distribution, one can apply methodology based on a multivariate generalization of quasi likelihood. The quasi likelihood approach specifies a model for the mean and a variance function expressing the dependence of the variance on the mean, and then obtains estimates by solving estimating equations that are score equations under the further assumption of a distribution in the exponential family with that mean and variance (see Wedderburn (1974) and McCullagh & Nelder (1989, Ch. 9)). In the multivariate context discussed here, one also uses a working guess for the correlation structure (Liang & Zeger, 1986; Liang, Zeger & Qaqish, 1992), and the estimates are solutions of *generalized estimating equations*. The method is often referred to as the GEE method. Estimates of model parameters are consistent even if the correlation structure is misspecified. This approach is appealing for categorical data because of not requiring a multivariate distribution, but it has limitations resulting from the lack of a likelihood (Lindsey, 1999) and its subsequent reliance on Wald methods.

The GEE methodology was originally specified for univariate distributions such as the binomial and Poisson. Multinomial generalizations now exist for cumulative logit models (Kenward, Lesaffre & Molenberghs, 1994; Lipsitz, Kim & Zhao, 1994; Lumley, 1996; Mark & Gail, 1994; Williamson, Kim & Lipsitz, 1995) and cumulative probit models (Toledano & Gatsonis, 1996) for clustered ordinal responses.

For the Lipsitz *et al.* approach, let  $Y_k(ij) = 1$  if cluster  $i$  makes response  $k$  ( $k = 1, \dots, c - 1$ ) for the  $j$ th observation. Then for each pair of response categories  $(k, \ell)$  one selects a working correlation matrix  $V_i$  for all pairs of observations  $(j_1, j_2)$  in the cluster. The  $(c - 1) \times (c - 1)$  blocks of  $V_i$  corresponding to the  $Y$ 's for a specific observation  $j$  in the cluster are multinomial covariance matrices with  $a_k(ij) = \pi_k(\mathbf{x}_{ij})(1 - \pi_k(\mathbf{x}_{ij}))$  on the diagonal and  $-\pi_k(\mathbf{x}_{ij})\pi_\ell(\mathbf{x}_{ij})$  on the off-diagonal. The remaining elements of  $V_i$  represent the covariance between pairs  $Y_k(ij_1)$  and  $Y_\ell(ij_2)$ ; for instance, one might choose the exchangeable structure,  $\text{Cov}(Y_k(ij_1), Y_\ell(ij_2)) = \rho_{k\ell} \sqrt{a_k(ij_1)a_\ell(ij_2)}$  for all  $j_1$  and  $j_2$ . The generalized estimating equations for estimating  $\beta$  (including the cut-points  $\alpha$ ) take the

form

$$u(\hat{\beta}) = \sum_{i=1}^N \hat{D}'_i \hat{V}_i^{-1} [y_i - \hat{\pi}_i] = 0$$

where  $y_i$  is the vector of observed responses for the  $i$ th cluster,  $\pi_i$  is the vector of probabilities associated with  $Y_i$ ,  $D'_i = \frac{\partial [ \pi_i ]'}{\partial \beta}$ , and the hats denote the substitution of the unknown parameters with their current estimates. Lipsitz *et al.* (1994) suggested a Fisher scoring algorithm for solving the above equation for  $\beta$  in conjunction with a method of moments update for estimating the correlation parameters  $\rho_{kl}$  at each step of the iteration. A robust covariance matrix of  $\hat{\beta}$  is

$$V_{\hat{\beta}} = \left[ \sum_{i=1}^N D'_i V_i^{-1} D_i \right]^{-1} \left[ \sum_{i=1}^N D'_i V_i^{-1} \text{var}(Y_i) V_i^{-1} D_i \right] \left[ \sum_{i=1}^N D'_i V_i^{-1} D_i \right]^{-1}.$$

This is estimated by substituting  $\hat{\pi}_i$  from the model fit and replacing  $\text{var}(Y_i)$  by the empirical variance-covariance matrix of  $Y_i$ .

Related literature includes applying GEE to the repeated ordinal case with independence estimating equations (Mark & Gail, 1994; Miller, Davis & Landis, 1993), unstructured correlations (Miller *et al.*, 1993), and using a model for global odds ratios (Lumley, 1996; Williamson *et al.*, 1995). More general models allow for dispersion parameters that also depend on covariates, as discussed in Section 5.1 (Toledano & Gatsonis, 1996). Miller *et al.* (1993) showed that under certain conditions the solution of the first iteration in the GEE fitting process is simply the estimate from the WLS approach of Koch *et al.* (1977). For this equivalence, one uses initial estimates based directly on sample values and assumes a saturated association structure that allows a separate correlation parameter for each pair of response categories and each pair of observations in a cluster. In this sense, GEE with unstructured correlation provides an iterated form of WLS. Moreover, in this case, the covariance matrix for the estimates is the same in both approaches.

When marginal models are adopted, the joint dependence structure is usually not the primary focus and is regarded as a nuisance. In such cases with ordinal responses, it seems reasonable to use a simple structure for the associations, rather than to expend much effort modeling it. For instance, one might use an exchangeable correlation structure for the pairs of observations in a cluster, or an autoregressive structure in the longitudinal case. It often makes more sense to parameterize associations for categorical responses in terms of odds ratios rather than correlations; hence, an even simpler structure that might be useful is a common local or global odds ratio. Using a simple structure has the potential for slight efficiency gain over the independence equations and the more general structures that can have large numbers of parameters to characterize associations, as well as less chance of numerical singularities compared to the latter case. An earlier method related in spirit to GEE methods forms a weighted combination of estimates from separate models fitted to margins of a repeated ordinal response (Stram, Wei & Ware, 1988). This approach does not allow for simpler working correlation structures, however.

When the association structure is itself of interest, a GEE2 approach is available for modeling associations using global odds ratios (Heagerty & Zeger, 1996). A disadvantage of GEE2 methods, however, is that estimates of parameters in the marginal model are no longer consistent if one misspecifies the model for the associations.

### 2.3 Maximum Likelihood Fitting of Marginal Models

We mentioned above that it is difficult to fit marginal models using maximum likelihood (ML). Multivariate logistic models have been defined with a one-to-one correspondence between joint multinomial cell probabilities  $\{\pi_{k_1, \dots, k_{n_i}}(\mathbf{x}^*)\}$  and parameters of marginal models as well as higher-

order parameters of the joint distribution (Fitzmaurice & Laird, 1993; Glonek, 1996; Glonek & McCullagh, 1995). One can then use ML to estimate the model parameters, but the correspondence is awkward to specify for more than a few dimensions. One version of this idea (Molenberghs & Lesaffre, 1994) uses a multivariate extension of a bivariate model using global odds ratios discussed by Wahrendorf (1980) and by Dale (1986). In this approach the joint probabilities are decomposed into “main effects” and “interactions”. The main effects correspond to the marginal probabilities while the interactions are the two-way, three-way and  $n_i$ -way generalized cross ratios based on global odds ratios. Molenberghs & Lesaffre (1994) suggested the use of a multivariate distribution based on common values of global odds ratios to determine the joint probabilities from these main effects and interactions. In practice, this approach is also difficult to implement except for small-dimensional problems.

An alternative approach treats a marginal model as a set of constraint equations and uses methods of maximizing (subject to constraints) likelihoods based on a product of multinomials (Haber, 1985; Lang & Agresti, 1994). The method is based on iterative use of Lagrange’s method of undetermined multipliers together with the Newton–Raphson method. One specifies Lagrangian likelihood equations of the form  $h(\boldsymbol{\theta}) = \mathbf{0}$ , where  $\boldsymbol{\theta}$  is a vector containing the sets of multinomial cell probabilities and the Lagrangian multipliers, and solves them using the Newton–Raphson method,

$$\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} - \left( \frac{\partial h(\boldsymbol{\theta}^{(t)})}{\partial \boldsymbol{\theta}} \right)^{-1} h(\boldsymbol{\theta}^{(t)}).$$

In these approaches, it is possible also to model simultaneously the joint distribution or higher-order marginal distributions. For instance, one might use a cumulative logit model for the marginal distributions and a model assuming a common global odds ratio (Glonek, 1996; Glonek & McCullagh, 1995; Molenberghs & Lesaffre, 1994) or a common local odds ratio (Lang & Agresti, 1994) for the pairwise associations.

Although in principle it is possible to use methods such as these to fit marginal models, it is computationally intensive when  $\{n_i\}$  are large or when there are several predictors, especially if some of them are continuous. The difficulty results because the Newton–Raphson part involves inverting a matrix with dimensions larger than the number of cells in the contingency table. A refinement of this approach (Lang & Agresti, 1994; Lang, 1996) uses a matrix to be inverted in the Newton–Raphson step that has much simpler form but the same probability limit. Continuing computational advances are making ML feasible for larger problems, both for constrained ML (Bergsma, 1997; Lang, McDonald & Smith, 1999) and for maximization with respect to joint probabilities expressed in terms of the marginal model parameters and an association model (Heumann, 1997).

An alternative approach unifies marginal modeling of ordinal data with methods for continuous responses. It is based on assuming an underlying multivariate normal distribution, which implies cumulative probit models (Kim, 1995) for the margins. Again, the feasibility diminishes as  $\{n_i\}$  or the number of predictors increases.

## 2.4 Example

We first use marginal cumulative logit models to analyze Table 1. Consider the model

$$\text{logit}[\gamma_k(\mathbf{x})] = \alpha_k + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2, \quad k = 1, 2, 3,$$

where  $\mathbf{x} = (x_1, x_2)$  with  $x_1$  a dummy variable for treatment ( $x_1 = 1$  for active and  $x_1 = 0$  for placebo) and  $x_2$  a dummy variable for occasion ( $x_2 = 1$  for follow-up and  $x_2 = 0$  for initial). Table 2 shows results for fitting this model and the simpler “no interaction” model with  $\beta_3 = 0$ , using ML (with the Lang & Agresti 1994 approach) and using GEE assuming an independence working correlation structure. With the independence working structure, the GEE estimates are identical to



what one would obtain using ML and treating the two observations on a subject as independent, but the standard errors empirically account for the actual dependence.

Results are very similar with the ML and GEE estimation methods, and they are qualitatively similar to those in Table IV of Agresti (1989) for fitting the marginal models using weighted least squares. With either estimation method, there is evidence of interaction. For instance, with the ML approach the Wald statistic for testing  $H_0 : \beta_3 = 0$  is  $(.662/.244)^2 = 7.4$  ( $df = 1, P = .007$ ) and the likelihood-ratio statistic equals 6.9 ( $df = 1, P = .009$ ). The ML estimated occasion effect is  $\hat{\beta}_2 = 1.074$  with placebo and  $\hat{\beta}_2 + \hat{\beta}_3 = (1.074 + 0.662) = 1.736$  with the active drug. With the active drug, for instance, at the follow-up occasion the estimated odds of falling asleep in less time than  $k$  are  $\exp(1.736) = 5.7$  times the estimated odds at the initial occasion. Similarly, the estimated drug effect equals  $\hat{\beta}_1 = 0.046$  initially and  $\hat{\beta}_1 + \hat{\beta}_3 = 0.708$  at the follow-up occasion; initially the response distributions were essentially identical, and at the follow-up occasion, the estimated odds of falling asleep in less time than  $k$  with the active drug are  $\exp(0.708) = 2.0$  times the estimated odds with placebo.

An advantage of the ML approach over the GEE approach is that one has a likelihood function, thus permitting both likelihood-ratio tests of effects as well as tests of goodness of fit. For instance, the deviance statistic equals 15.0 for the no interaction model ( $df = 7$ ) and 8.0 for the interaction model ( $df = 6$ ). Although the data are sparse and conclusions are tentative, the fit appears adequate for the interaction model and better than that for the no interaction model. The  $df = 6$  value for the interaction model represents the modeling of 12 marginal logits (for the three cumulative probabilities at each of the four combinations of treatment and occasion) with a model containing 6 parameters.

Table 2 does not show  $\{\hat{\alpha}_k\}$ , which are of secondary interest, needed for obtaining fitted values but not for interpreting treatment effects. For each method it is unnecessary to impose constraints on the fitting process to reflect their ordering; it has always been our experience that the ordinary estimates are properly ordered (e.g., for the ML fit with the interaction model,  $\hat{\alpha}_1 = -2.27, \hat{\alpha}_2 = -0.96, \hat{\alpha}_3 = 0.32$ ). Although the response variable is actually a grouped continuous variable with known cutpoints,  $\{\hat{\alpha}_k\}$  are not based on those values since the model-fitting does not assume a parametric shape for an underlying continuous distribution.

**Table 2**  
*Results of fitting cumulative logit models (with standard errors in parentheses) to Table 1.*

Model	Effect	Marginal ML	GEE	Random Effects
No Interaction	Treatment	0.402 (.198)	0.394 (.200)	0.606 (.307)
	Occasion	1.342 (.136)	1.394 (.138)	2.100 (.230)
Interaction	Treatment	0.046 (.236)	0.034 (.238)	0.058 (.366)
	Occasion	1.074 (.162)	1.038 (.168)	1.602 (.283)
	Treatment×Occasion	0.662 (.244)	0.708 (.244)	1.081 (.380)

Note: Random effects fit used PROC NL MIXED in SAS, based on adaptive Gauss–Hermite quadrature with 30 quadrature points.

### 3 Random Effects and Mixed Models for Ordinal Responses

Cluster-specific models usually represent cluster effects by a random effects term in the model. Those random effects can help to describe extra variability occurring in a variety of situations, due to factors such as subject heterogeneity, unobserved covariates, and other conditions that lead to overdispersion. The repeated responses are typically assumed to be independent, given the random effect, but variability in the random effects induces a marginal nonnegative association between pairs of responses after averaging over the random effects.

For binary repeated measurement data, the basic model has a logit link with a linear predictor that contains random effects having normal distributions with unknown variances. The model form extends to ordinal logits. For instance, for the *random intercept* form of proportional odds model, the  $k$ th cumulative logit for response  $j$  in cluster  $i$  has form

$$\text{logit}[\gamma_k(\mathbf{x}_{ij}; i)] = u_i + \alpha_k + \mathbf{x}'_{ij} \boldsymbol{\beta}, \quad k = 1, \dots, c - 1, \quad (3)$$

where  $\{u_i\}$  are *iid* from a  $N(0, \sigma^2)$  distribution. More generally, it is often sensible to treat slopes as well as intercepts as random. Thus, it is useful to have the more general form

$$\text{logit}[\gamma_k(\mathbf{x}_{ij}; i)] = \alpha_k + \mathbf{x}'_{ij} \boldsymbol{\beta} + \mathbf{z}'_{ij} \mathbf{u}_i, \quad k = 1, \dots, c - 1, \quad (4)$$

where  $\mathbf{z}_{ij}$  refers to a vector of explanatory variables for the random effects (which may overlap with some of the  $\mathbf{x}_{ij}$ ), and  $\{\mathbf{u}_i\}$  are *iid* from a multivariate  $N(\mathbf{0}, \boldsymbol{\Sigma})$  distribution.

#### 3.1 ML Model Fitting

Since the random effects are unobserved, to obtain the likelihood function one constructs the usual product of multinomials that would apply if they were known and then integrates out the random effects. Given the data, the likelihood function depends on the fixed effects parameters and the parameters of the random effects distributions. Except in rare cases, this integral does not have closed form and it is necessary to use some approximation for the likelihood function. One can then maximize the approximated likelihood using a variety of standard methods.

Algorithmic approaches for approximating the integral that determines the likelihood include Gauss–Hermite quadrature, an automated Monte Carlo EM algorithm, and a pseudo-likelihood algorithm that utilizes Taylor series and Laplace approximations. When feasible, in our opinion the best method is Gauss–Hermite quadrature, which replaces the integral by an approximating finite sum. The precision of approximation improves as the number of elements in the sum increases. An adaptive version of Gauss–Hermite quadrature (e.g., Pinheiro & Bates, 1995) uses the same weights and nodes for the finite sum as Gauss–Hermite quadrature, but to increase efficiency it centers the nodes with respect to the mode of the function being integrated and scales them according to the estimated curvature at the mode. Using the transformed Gauss–Hermite nodes, the contribution to the likelihood by the  $i$ th cluster is approximated by a finite summation taken over  $K$  quadrature points.

Because of the “curse of dimensionality,” multivariate adaptive quadrature is currently computationally feasible only for integral dimensions up to about 5 or 6. For models with higher-dimensional integrals, more feasible methods use Monte Carlo methods, which use  $K$  randomly sampled nodes to approximate integrals. An important issue is then the number of nodes to sample to approximate adequately the integrals. Booth & Hobert (1999) proposed an automated Monte Carlo EM algorithm for generalized linear mixed models that assesses the Monte Carlo error in the current parameter estimates and increases the number of nodes if the error exceeds the change in the estimates from the previous iteration.

Once one has approximated the likelihood function, large-sample inference proceeds in the usual manner. Maximizing the log likelihood yields the ML estimates. Their asymptotic covariance matrix

is based on estimating the curvature at the mode of the integrand by inverting the negative of the second derivative matrix of the integrand evaluated at the estimated mode. In practice, one sequentially increases  $K$  until the changes are negligible in both the resulting parameter estimates and standard errors. Inference about variance components requires some care. Under the null hypothesis that a variance component  $\sigma^2 = 0$ , complications result from being on the boundary of the parameter space. For instance, the likelihood-ratio statistic does not have the usual chi-squared asymptotic distribution, but rather has a mixture distribution since with null probability approximately equal to  $1/2$ ,  $\hat{\sigma} = 0$  and the likelihood-ratio statistic equals 0. One can obtain ordinary chi-squared distributions using the score test (Li & Lin, 2000), since the score (i.e., derivative of the log likelihood with respect to  $\sigma$ , evaluated at  $\sigma = 0$ ) is not constrained to 0 when  $\sigma = 0$ .

The random effects literature for ordinal data so far considers primarily cumulative logit and probit models. An exception is Coull & Agresti (2000), who fitted random effects models for clustered data with continuation-ratio logits using a binomial model with multivariate random effects, and Hartzel, Liu & Agresti (2001) who used adjacent-categories logits. An early approach for parameter estimation utilized best linear unbiased prediction of parameters of an underlying continuous model (Harville & Mee, 1984). Recent work considers extensions with heterogeneous variances for the latent variables (Jaffrézic, Robert-Granié & Foulley, 1999). Model fitting has used a variety of approaches for approximating the integral over the random effects that determines the likelihood function, including Gauss–Hermite numerical integration (Ezzet & Whitehead, 1991; Hedeker & Gibbons, 1994; Jansen, 1990; Tutz & Hennevogel, 1996; Hartzel *et al.*, 2001), Laplace approximation (Sheiner, Beal & Dunne, 1997), and a combination of Monte-Carlo and EM algorithms (Natarajan, 1995; Hartzel *et al.*, 2001). A rare exception in which the log likelihood has closed form occurs for a complementary log-log link with the log of a gamma or inverse Gaussian distribution as the random effects distribution (Crouchley, 1995; Ten Have, 1996).

### 3.2 Connections with Marginal Models

The fixed effects  $\beta$  in a model such as (4) with random effects refer to a *conditional* effect, given the random effect. That is,  $\beta$  has a cluster-specific interpretation, referring to the change in the link function for a given subject or cluster. This is not normally the same as the marginal (i.e., population-averaged) effect  $\beta$  in the corresponding marginal model (2).

Specifically, consider the generalization of model (4) with an arbitrary link function  $g$ . Then,

$$\gamma_k(\mathbf{x}_{ij}; i) = g^{-1}(\alpha_k + \mathbf{x}'_{ij} \beta + \mathbf{z}'_{ij} \mathbf{u}_i).$$

The marginal cumulative probabilities satisfy

$$\gamma_k(\mathbf{x}_{ij}) = E \left[ g^{-1}(\alpha_k + \mathbf{x}'_{ij} \beta + \mathbf{z}'_{ij} \mathbf{u}_i) \right],$$

with expectation taken with respect to the normal random effects distribution. Then, the same link applied with the marginal model implies that

$$g \left[ \gamma_k(\mathbf{x}_{ij}) \right] = g \left( E \left[ g^{-1}(\alpha_k + \mathbf{x}'_{ij} \beta + \mathbf{z}'_{ij} \mathbf{u}_i) \right] \right).$$

Generally, the right-hand side does not equal  $\alpha_k + \mathbf{x}'_{ij} \beta$  with the same parameter values as in the cluster-specific model. The exceptions are the identity link (i.e., the model  $\gamma_k(\mathbf{x}_{ij}; i) = \alpha_k + \mathbf{x}'_{ij} \beta + \mathbf{z}'_{ij} \mathbf{u}_i$ ) and a general link but with elements of  $\mathbf{u}_i$  having zero variance. In fact, if the cluster-specific model holds with a particular link, the implied marginal model will not normally hold with that same link. An exception is the probit link, in which case the cluster-specific cumulative probit model with normal random effect does imply a marginal model of probit form, but with different parameters. In the case of a univariate random intercept (i.e., a *probit-normal model*), the marginal model has effect

equal to the cluster-specific effect multiplied by  $[1 + \sigma^2]^{-1/2}$  (Zeger *et al.*, 1988).

Approximate connections exist between cluster-specific and marginal models when the variance components are small. If the cluster-specific model holds with the logit link, then to the first order in a Taylor series expansion one can show that the marginal model is approximately of logit form. For the special case of a random intercept model, the fixed effect parameters for the marginal model equal those for the random effects model multiplied by approximately  $[1 + .59\sigma^2]^{-1/2}$ . Thus, the estimate of  $\beta$  from the marginal model is typically smaller in absolute value than the estimate from the cluster-specific model. The discrepancy between the two increases as  $\sigma$ , and hence the correlation between observations within a cluster, increases. Neuhaus *et al.* (1991) and Zeger *et al.* (1988) provided approximate relationships between the two types of estimates for various models.

### 3.3 Example

For Table 1 we now consider the model

$$\text{logit}[\gamma_k(\mathbf{x}; i)] = u_i + \alpha_k + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2, \quad k = 1, 2, 3,$$

with  $(x_1, x_2)$  defined as in Section 2.4. Table 2 shows results for this model and the simpler one without interaction, based on using adaptive Gauss–Hermite quadrature with 30 quadrature points. Results are qualitatively similar to the marginal model, but estimates and standard errors are on the order of 50% larger. This reflects the relatively large variance component ( $\hat{\sigma} = 1.85$  for the no-interaction model and  $\hat{\sigma} = 1.90$  for the model permitting interaction) and the consequent strong association between the responses at the two occasions.

## 4 Other Approaches

The previous two sections have focused on estimating marginal and random effects models for clustered ordinal data within a traditional frequentist structure and assuming normality for random effects. This section presents other approaches for estimation in models for clustered ordinal data.

### 4.1 Bayesian Modeling

Bayesian methods have gained increasing popularity in the last decade, in part due to the development of computational algorithms such as the Gibbs sampler, and more generally Markov Chain Monte Carlo (MCMC). One of the first papers to illustrate the application of MCMC techniques for ordinal data was written by Albert & Chib (1993). This paper focused on models for independent ordinal responses using a family of  $t$ -distributions to generate the link function. One can approximate the cumulative probit and logit links with certain settings. Posterior calculations were then implemented under a diffuse normal prior distribution for the regression coefficients  $\beta$  and a uniform prior for the threshold cut-points  $\alpha_k$ . The ideas contained in this paper have subsequently been generalized to models that allow for clustering in the ordinal measurements.

Tan, Qu, Mascha & Schubert (1999) described a Bayesian hierarchical model for modeling the performance of a group of anaesthesiology residents on oral practice examinations (OPEs). The OPE outcome was an ordinal test score consisting of four categories (“Definite Not Pass”, “Probable Not Pass”, “Probable Pass”, “Definite Pass”) which was determined by two board certified anaesthesiologists randomly selected from a pool of 12 experts. Furthermore, in order to assess the validity of the OPE, residents were required to take multiple examinations. The authors used a proportional odds model for the analysis, while accounting for the correlation within residents and raters through the use of a two-dimensional random effect vector. Their preference for this cluster-specific strategy was motivated by their focus on accurately estimating the correlation parameters. Johnson (1996)

described a similar model for studying inter-rater reliability in situations where the repeated ordinal measurements are ratings of items by multiple raters. His approach differs from that of Tan *et al.* (1999) in the choice of link function (a probit link was used primarily for computational convenience) and by allowing the category cut-offs  $\alpha_k$  to be specific to each rater. (This idea has previously been utilized by Farewell (1982) in a frequentist context with a cumulative complementary log-log link function in order to allow for variation in the values regarded as category boundaries on an underlying scale.) More recently, Chen (1998) and Chen & Shao (1999) have proposed extensions of these models to cases where the correlated outcomes may be a combination of binary and ordinal measurements.

We will now describe the basic ingredients of a Bayesian treatment for modeling clustered ordinal data. As with the classical frequentist approach, the cumulative probability  $\gamma_k(\mathbf{x}_{ij}; i)$  is modeled conditionally on a cluster-specific random effect  $\mathbf{u}_i$ , according to the specification on the right-hand-side of (4). This formulation is typically completed by assuming that  $\{\mathbf{u}_i\}$  arise from a  $N(\mathbf{0}, \Sigma)$  distribution, although other distributions may be easily accommodated. The remaining stages of the model specify prior distributions on the regression coefficients  $\beta$ , the cut-points  $\alpha_k$ 's and the components of variance in  $\Sigma$ . Chen (1998) described a procedure for constructing a prior when historical information is available. However, in the absence of such information, the usual strategy is to employ "non-informative" prior distributions. The choices suggested by the authors above include a uniform or diffuse normal prior for  $\beta$  and the  $\alpha_k$ 's, and an inverted Wishart distribution (or inverted gamma for scalar variances) for  $\Sigma$ . The hyperparameters of these distributions are then modeled using uniform distributions (e.g., Johnson (1996)), or more typically seemingly innocuous values are chosen for them.

The main advantage offered by these priors is computational, as they allow for particularly straightforward MCMC implementation using software such as BUGS (Spiegelhalter, Thomas, Best & Gilks, 1996). However, there is no sound theoretical basis for these choices, and in fact recent research for binary data models shows that such priors may actually lead to badly behaved posterior distributions, particularly for  $\Sigma$  (Natarajan & Kass, 2000). Hence an important area for future research appears to be the development of alternate non-informative priors in hierarchical models for clustered ordinal data. Two ideas that may prove fruitful in this context are the approximate uniform shrinkage prior proposed by Natarajan & Kass (2000) for generalized linear mixed models, and the conditional means prior suggested by Bedrick, Christensen & Johnson (1996) for generalized linear models. Although these prior distributions are conceptually different, they are similar in that both are induced by prior distributions placed on transformation of the parameters. In the former, the reparameterization corresponds to the weight given to the prior mean of  $\mathbf{u}_i$  in its posterior update, while in the latter it is the cumulative probability of the categories. The benefits of working with such transformations are two-fold. Firstly, the transformed scale is often more meaningful from the perspective of prior specification. Secondly, the priors on the transformed scale are themselves proper (i.e., integrate to a finite number) and hence ensure propriety of the resulting posterior distributions.

The primary advantage of a Bayesian approach over the classical viewpoint is that one does not need to rely on asymptotic arguments in order to make inferences, which can be beneficial when the sample sizes are small. Further, since Bayesian computations typically involve generation of samples from the posterior distribution, rather than the high-dimensional maximization often required for maximum likelihood, the Bayesian approach is often less awkward from a computational perspective. Thus, we believe that the development and dissemination of Bayesian methodology for this class of models should be an important direction for future work.

## 4.2 Semi-parametric Random Effects Modeling

In the random effects models discussed in Section 3, advantages of assuming a normal distribution for the random effects include permitting a variety of covariance structures for them and connections with Gaussian mixed models. However, the normality assumption can rarely be checked very carefully. An attractive alternative that avoids this parametric assumption is to assume a mixing distribution of unspecified form for the random effect. Aitkin (1999) discussed this approach for binary data. One assumes a discrete distribution with unknown support size, mass points, and probabilities. Joint estimation of the fixed effects parameters and the discrete mixing distribution can be implemented with an EM algorithm. Hartzel *et al.* (2001) utilized this approach for summarizing heterogeneity among centers in ordinal odds ratios describing results of clinical trials.

Inference with the semi-parametric approach presents complications and suggests some interesting research problems. The complications are partly due to the unknown number of mass points for the mixing distribution. This number of parameters describing the mixing distribution may be different for two models, for instance under null and alternative hypotheses about fixed effects. Hence, the dimension of the parameter vector is also unknown for a given model or a model comparison. Similarly, work is needed on developing inferences about the mixing distribution. For instance, one might want to compare the heterogeneity model to one not containing the random effect, essentially testing that the variance component equals 0. This entails testing that the masses are on the boundary of the parameter space and precludes the ordinary use of the likelihood-ratio test. Another complication that needs addressing is that the estimated mixing distribution can have positive probability at  $\pm\infty$ , particularly when all observations in some clusters have the same response. This can influence standard error estimates obtained through inverting the observed information matrix.

A traditional nonparametric way to handle cluster-specific terms in logit models for binary response data is *conditional maximum likelihood*. With this approach, one treats those terms as fixed nuisance parameters rather than random effects and eliminates them from the likelihood by conditioning on their sufficient statistics. Examples of such methods include Rasch models (Rasch, 1961) for item response modeling and logistic regression for case-control studies (Breslow & Day, 1980). The conditional ML approach has not received nearly as much attention for ordinal responses. It is limited to models for which sufficient statistics exist, which are models using canonical link functions such as the logit for binomial response data. It is also limited to within-cluster effects rather than between-cluster effects, since the distribution relating to the latter effects disappears in the conditioning process. For multinomial responses, the canonical link functions are the baseline-category logits in which each logit pairs a particular category with a baseline, or logits that are equivalent to these. For ordinal response data, the  $(c - 1)$  adjacent-category logits are an equivalent set. For examples of the use of this approach for within-subject effects in multinomial models, see Conaway (1989). The conditional approach is not relevant for cumulative logit models, because of the lack of reduced sufficient statistics for those logits.

For binary responses with the Rasch item-response model, Tjur (1982) showed that the conditional ML approach provides identical estimates as obtained by fitting a certain loglinear model for the cross classification of the repeated responses. Specifically, when each cluster has the same number of responses  $n$  and there are no covariates, the conditional ML estimates comparing the item parameters are determined by ordinary ML estimates of main effect parameters in the quasi-symmetry model (Bishop, Fienberg & Holland, 1975) for the  $2^n$  contingency table that cross classifies the responses. Analogous results apply for an adjacent-categories logit model. Conditional ML estimates for the model that expresses this logit as an additive function of a subject effect and an item parameter are related to ML estimates for a corresponding quasi-symmetric loglinear model of ordinal form having quantitative main effects with fixed scores for the ordered response categories (Agresti, 1993a, 1993b). Applications of ordinal generalized Rasch models and corresponding loglinear models include cross-over studies and randomized clinical trials with matched-pairs responses (Kenward &

Jones, 1991; Agresti, 1993b).

In the case of matched pairs without covariates, simple special cases occur for semi-parametric modeling of ordinal repeated measures. Denote cell counts in a square contingency table for outcome  $k$  at the first response and  $k'$  at the second by  $\{n_{kk'}\}$ , and denote their expected values by  $\{\mu_{kk'}\}$ . The cumulative logit model (3) for a square table simplifies to

$$\text{logit} [\gamma_k(x_{ij}; i)] = u_i + \alpha_k + x_{ij}\beta, \quad k = 1, \dots, c - 1, \quad j = 1, 2,$$

where  $x_{i1} = 0$  and  $x_{i2} = 1$ . Simple estimates then exist (McCullagh, 1977; Agresti & Lang, 1993a) of the effect  $\beta$  comparing the margins. One such estimate, based on summarizing the conditional ML estimates from the  $(c - 1)$  possible collapsings of the response, is

$$\tilde{\beta} = \log \left[ \frac{\left\{ \sum_{k < k'} (k' - k)n_{kk'} \right\}}{\left\{ \sum_{k > k'} (k - k')n_{kk'} \right\}} \right].$$

For a corresponding adjacent-categories logit model

$$\text{log} [\pi_k(x_{ij}; i) / \pi_{k+1}(x_{ij}; i)] = u_i + \alpha_k + x_{ij}\beta$$

with a cluster-specific intercept and shift  $\beta$  for each subject between margin 2 and margin 1, the conditional ML estimate of  $\beta$  is identical to the ordinary ML estimate for logit model

$$\text{log}(\mu_{kk'} / \mu_{k'k}) = \beta(k' - k),$$

fitted treating the  $c(c - 1)/2$  pairs  $(n_{kk'}, n_{k'k})$  of cell counts as independent binomial variates (Agresti, 1993a, 1993b). Goodman (1979, 1994) discussed other models pertaining to matched pairs with an ordinal response, and S. Tomizawa has written numerous papers (e.g., Tomizawa, 1995) on this topic.

The approach of using an arbitrary discrete mixing distribution instead of a normal distribution has much in common with traditional latent class and latent variable modeling. Again, for this approach the main emphasis is on describing within-subject effects rather than effects of explanatory variables on the clustered responses. For the ordinal response case, see for example Samejima (1969); Rost (1985); Masters (1985); Croon (1990); Croon (1993), and Heinen (1993).

### 4.3 Alternative Model Forms

We have emphasized models using the logit link. Results with this link function are very similar to those using a probit link. That link has not been as common in the statistics literature, but it arises naturally with cumulative probabilities from latent models assuming underlying normality (Harville & Mee, 1984; McCullagh, 1980). The complementary log-log and log-log links are natural ones for underlying distributions that are extremely skewed, such as extreme value (Crouchley, 1995; Ten Have, 1996).

Many researchers have difficulty interpreting effects, such as odds ratios, resulting from non-linear link functions. For purposes of conveying information, it may be useful in some applications to provide results in a manner more widely understood, such as with means. For instance, having fitted an ordinal model, one could report estimated means for its fitted distributions or for the sample distributions based on a reasonable scoring of response categories. For Table 1 with response scores {10, 25, 45, 75} for time to falling asleep, for instance, the initial means were 50.0 for the active drug group and 50.3 for the placebo, and the follow-up means were 28.8 for the active drug group and 37.3 for the placebo.

In principle, when one is willing to score the ordinal categories, one could model the marginal means and report effects relating to differences of means in the context of such models. Rather than use ordinary regression, one should treat the responses as multinomial rather than normal. Koch *et al.* (1977) did this using weighted least squares methods. This approach breaks down for sparse data

or small marginal counts, but one could use maximum likelihood subject to whatever constraints may be natural to impose on the marginal distributions.

### 5 Applications and Areas for Future Research

This section describes some recent applications of modeling clustered ordered categorical data. It also discusses possible problems for future work.

#### 5.1 Receiver Operating Characteristic (ROC) Curves

The Receiver Operating Characteristic (ROC) curve is a graphical technique for evaluating the accuracy of diagnostic tests in discriminating between two populations of individuals, typically labelled “diseased” and “non-diseased”. The ROC curve plots the false positive rate (proportion of non-diseased individuals classified as diseased) on the  $x$ -axis versus the true positive rate (proportion of diseased individuals classified as diseased) on the  $y$ -axis. A typical ROC curve is concave in shape with each point representing a false-positive and true-positive pair calculated using a particular threshold to discriminate between diseased and non-diseased individuals. Diagnostic tests having area under the ROC curve close to 1.0 are preferred, as such a value can be obtained only when the true positive rate of the test is high even for small false positive rates. Hence, such tests have the desirable property that they correctly diagnose diseased subjects a large proportion of the time, while maintaining a small probability of misclassifying non-diseased individuals.

Obtaining model-based estimates of the ROC curve has been an active area of research in the diagnostic radiology literature. See, for instance, Toledano & Gatsonis (1995), Toledano & Gatsonis (1996), and references therein. The diagnostic test results in these applications are recorded on an ordinal scale depicting an expert rater’s level of suspicion regarding the presence of disease. To explain the connection between ROC curves and ordinal regression models, we first discuss the case of independent observations. In this situation, a single ordinal rating is observed for each of several subjects. The measurement on the  $i$ th individual is assumed to arise from a cumulative probit link model:

$$\Phi^{-1} [\gamma_k (\mathbf{x}_{1i}, \mathbf{x}_{2i}; i)] = \frac{\alpha_k - \mathbf{x}'_{1i} \beta}{\exp(\mathbf{x}'_{2i} \delta)}, \tag{5}$$

where  $\Phi^{-1}$  is the inverse cumulative distribution function of the standard normal density, and  $\mathbf{x}_{1i}$  and  $\mathbf{x}_{2i}$  are covariate vectors associated with the location parameter  $\beta$  and scale parameter  $\delta$  respectively. This model is a member of a general class of ordinal models proposed by McCullagh (1980), with links including the probit and logit. The denominator accounts for the possibility that the dispersion of the response varies according to covariates.

For simplicity, here, suppose the true disease status is the only covariate in the location and scale pieces of the model, with  $x_{1i} = x_{2i} = 0$  for non-diseased,  $x_{1i} = x_{2i} = 1$  for diseased. Suppose that higher categories for the ordinal rating denote greater severity of disease. Then, when the cutpoint to distinguish disease status follows category  $k$  of the ordinal response, the false positive rate is  $1 - \Phi(\alpha_k)$  and the true positive rate is  $1 - \Phi[\exp(-\delta)(\alpha_k - \beta)]$ . The ROC curve is a plot of these rates for  $\alpha$  varying between  $(-\infty, \infty)$ . The area under this curve equals  $\Phi[\beta \exp(\delta)/\sqrt{(1 + \exp(-2\delta))}]$ , or simply  $\Phi(\beta/\sqrt{2})$  when there is no dispersion effect. Derivations similar to this yield ROC curves for the various covariate strata when (5) includes covariates in addition to the true disease status. For details, see Tosteson & Begg (1988).

In practice, study protocols typically require the re-reading of scans taken on a particular subject by multiple raters, thereby leading to potential correlations among the measurements taken on the same subject. Hence, the methods described in Sections 2 and 3 for clustered ordinal data have



found much use in this arena. The ordinal regression model in (5) can be extended to accommodate correlated observations. Toledano & Gatsonis (1996) considered a marginal model approach in which the cumulative probability of the  $k$ th category for the  $j$ th rating occasion on subject  $i$  depends on cut-points  $\alpha_{kj}$ , location parameters  $\beta_j$ , scale parameters  $\delta_j$  and covariates  $x_{1ij}$  and  $x_{2ij}$  through equation (5). The occasion-specific parameters determine a separate ROC curve at each rating occasion; for instance, when true disease status is the only covariate, the ROC curve at occasion  $j$  plots  $(1 - \Phi(\alpha), 1 - \Phi[\exp(-\delta_j)(\alpha - \beta_j)])$  for  $\alpha$  in  $(-\infty, \infty)$ . A comparison of the areas beneath pairs of curves is of primary interest in this context.

An estimate of the ROC curves and the areas beneath them is obtained by substituting estimates of  $\beta_j$  and  $\delta_j$ , which can be calculated by the GEE method. If the correlations between pairs of ratings on a subject are nuisance parameters, then one might implement the GEE method using a working assumption of independence. However, if these correlations are themselves of interest or if gains in efficiency are desired, then Toledano & Gatsonis (1996) suggested modeling the joint ratings with a bivariate normal distribution. To account for the various dependencies induced by a typical radiological study design, they posit three types of correlation: one to model correlations on the same case (or subject), a second to model correlations on a particular case that also share the same diagnostic modality, and a third to model the correlation on a particular case that also share the same reader. These correlation parameters may themselves depend on covariates  $x_{i3}$  using a link function such as the logit. The correlation parameters can also be estimated by the GEE method.

Toledano & Gatsonis (1999) and Ishwaran & Gatsonis (2000) have proposed modifications of the above methodology to accommodate missing data and cluster-specific formulations. One limitation of all these methods, and indeed cumulative link models in general, is that the regression parameters pertain to the distribution of the latent degree of suspicion variable rather than to the covariate effects on the observed data. Hence, Pepe (1997) suggested an alternative approach that directly models the covariate effects on the ROC curve. In a recent survey article on ROC methodology, Pepe (2000) also noted that future challenges include developing methods for situations where the true disease status may be imprecisely measured, or may be time-varying or is vector valued.

## 5.2 Missing Data in Longitudinal Studies

Missing data occurs in longitudinal studies when one or more of the responses on a subject are unavailable. If the missing responses are pre-planned by design, then the data are typically analyzed using standard methods for unbalanced data. The analysis of all the available data from unbalanced designs can be mathematically complicated, but is perfectly valid in that consistent estimates may be obtained without any regard to the missing data mechanism. However, this is no longer true when the missing responses are not pre-planned. In this situation, the mechanism underlying the missing responses becomes important and focusing on the observed data alone can result in biased estimates.

Little & Rubin (1987) distinguished among three possible missing data mechanisms. Suppose  $\mathbf{Y}^{(o)}$  denotes the observed responses,  $\mathbf{Y}^{(m)}$  the missing responses and  $D$  is a missing data indicator. Then, they classified a missing data process as *completely random* if  $D$  is statistically independent of both  $\mathbf{Y}^{(o)}$  and  $\mathbf{Y}^{(m)}$  and *random* if it is independent only of  $\mathbf{Y}^{(m)}$ . The process is termed *informative* otherwise. They demonstrated that a likelihood-based analysis using the observed responses  $\mathbf{Y}^{(o)}$  is valid only when the missing data process is either completely random or random, and when the distributions of  $\mathbf{Y}^{(o)}$  and  $D$  are separately parameterized. For this reason completely random and random missing processes are often termed as *ignorable*. However, random processes are not necessarily ignorable when non-likelihood-based methods, such as GEE are used (Mark & Gail, 1994). Kenward *et al.* (1994) provided an empirical illustration of the breakdown in the GEE estimates for data from a longitudinal psychiatric study in which the missingness pattern is not completely random. Hence, developing models for the joint distribution of the responses and the

missing data mechanism has been an active area of research in recent years.

The models for handling missingness differ primarily in the manner in which the joint distribution is factored. In *selection models*, the joint distribution of  $\mathbf{Y}^{(o)}$ ,  $\mathbf{Y}^{(m)}$  and  $D$  is factored as

$$f(\mathbf{Y}^{(o)}, \mathbf{Y}^{(m)}, D|\mathbf{x}) = f(D|\mathbf{Y}^{(o)}, \mathbf{Y}^{(m)}, \mathbf{x}) f(\mathbf{Y}^{(o)}, \mathbf{Y}^{(m)}|\mathbf{x}),$$

where  $f(\cdot)$  denotes a generic probability density (mass) function and  $\mathbf{x}$  is a vector of covariates. *Pattern mixture models* are based on the alternative factorization:

$$f(\mathbf{Y}^{(o)}, \mathbf{Y}^{(m)}, D|\mathbf{x}) = f(\mathbf{Y}^{(o)}, \mathbf{Y}^{(m)}|D, \mathbf{x}) f(D|\mathbf{x}).$$

When the missing data mechanism is ignorable, selection models seem preferable because they then do not require modeling the missing data mechanism (Little, Raghunathan & Tang, 2000). Both models accommodate non-ignorable missing data mechanisms, but in different ways. The selection model achieves this more directly by requiring fairly detailed specification of the drop-out distribution, while pattern-mixture models do so indirectly by placing restrictions on the distribution of the responses depending on the drop-out pattern. Typically little is known about the missing-data mechanism, so a sensitivity study is necessary to check how results depend on that specification.

We will now review some of the modeling approaches that have been suggested for longitudinal ordinal data with missing outcomes. Molenberghs, Kenward & Lesaffre (1997) considered the problem of incorporating informative drop-outs in a longitudinal study with  $T$  common times of measurement for each subject on an ordinal response. At each time, a vector of covariates  $\mathbf{x}_i$  and a dropout indicator  $D (= 2, \dots, T + 1)$  are also measured, where  $D = T + 1$  indicates no drop-out. They used a selection model factorization, modeling the marginal distribution of the hypothetical complete data by a multivariate model based on describing patterns of association for pairs of variables using global odds ratios, as in Molenberghs & Lesaffre (1994). The conditional distribution of the drop-out variable is assumed to have the form:

$$\Pr(D = d|\mathbf{y}^{(o)}, \mathbf{y}^{(m)}, \mathbf{x}) = p_d(H_d, y_d^{(m)}, \phi),$$

where  $H_d$  denotes the observed history through time  $t_{d-1}$ ,  $y_d^{(m)}$  is the unobserved value at time  $t_d$ , and  $\phi$  is a set of unknown parameters. They suggest modeling the logit of  $p_d$  as a linear function of  $H_d$ ,  $y_d^{(m)}$ , and covariates chosen from  $\mathbf{x}_i$ . They maximized the resulting likelihood function using the EM algorithm.

An alternate approach was proposed by Cowles, Carlin & Connett (1996) for modeling longitudinal compliance data. Their model evolved from a study of the effectiveness of an inhaled bronchodilator in smoking cessation. The observed data at each visit are  $(y_{1ij}, y_{2ij}, y_{3ij})$ , where  $y_{1ij}$  is an ordinal variable with three categories signifying the participant's level of compliance with the study protocol and  $y_{2ij}$  and  $y_{3ij}$  are additional continuous measures of compliance. The values of  $y_{2ij}$  and  $y_{3ij}$  may be unobservable, depending on the value of  $y_{1ij}$ , and hence the problem becomes one of making inferences in the presence of intermittent missingness. The variable  $y_{1ij}$  plays the same role as the drop-out indicator  $D$  in the previous paragraph. Cowles *et al.* (1996) proposed a pattern mixture model that factors the joint distribution of the data into the marginal distribution of  $y_{1ij}$  and the distribution of  $\{y_{2ij}, y_{3ij}\}$  conditional on  $y_{1ij}$ . They used a cumulative probit model for the first distribution and a tobit model for the second, while incorporating the correlations among  $(y_{1ij}, y_{2ij}, y_{3ij})$  through patient-specific random effects. They estimated parameters within a Bayesian framework using diffuse normal prior distributions for the regression coefficients and inverse Wishart distributions for the variance components.

So far, we have discussed missingness in the responses. In some cases, however, information may be missing in a key covariate rather than the responses. Toledano & Gatsonis (1999) encountered this in a study comparing two modalities for the staging of lung cancer. The true disease status (presence or absence of bronchial invasion) was missing for several subjects for reasons potentially related to

their outcome measurements. To avoid biases that could result from a non-ignorable missing data mechanism, Toledano & Gatsonis (1999) suggested modifying the GEE equations by weighting the contribution of the  $i$ th subject by  $D_i/\pi_i$ , where  $D_i$  denotes whether the covariate is observed ( $D_i = 1$ ) or missing ( $D_i = 0$ ) and  $\pi_i = \Pr(D_i = 1)$ . Hence, only the complete cases contribute to the GEE equations, but subjects who are less likely to have an observed covariate value are accorded more weight. The probability  $\pi_i$  is then linked with covariates and the observed outcomes through a link function. The parameters of this relationship are also estimated using a set of generalized estimating equations.

The literature on the behavior of test statistics in the presence of incomplete data is relatively sparse. Lipsitz & Fitzmaurice (1996) derived the score test of independence for two-way contingency tables and showed that the test statistic continues to have an asymptotic chi-squared null distribution. Methods for inference in the presence of missingness is an area that is likely to see considerable research in future years. However, as pointed out by Molenberghs *et al.* (1997), such methods should be used with caution because the assumptions regarding the missing data mechanism are unexaminable and hence the inferences may not be robust. They suggested regarding any such modeling as merely a component of a sensitivity analysis.

### 5.3 Modeling Agreement

When the repeated measurement takes the form of ratings by several observers, agreement between pairs of observers or between each observer and a gold standard is of interest. For observations on a sample of cases by a pair of observers, one can summarize their ratings in a square contingency table with the same categories in each dimension. Traditionally, it has been popular to describe agreement on an ordinal scale using weighted kappa (Fleiss, Cohen & Everitt, 1969). This is a summary measure that compares the frequency of cases of agreement to the frequency expected by chance if the table satisfied the model of independence, where a distance metric between pairs of categories quantifies the extent of disagreement.

Recent work has focused on the modeling of interrater agreement and handling more than two observers. A variety of modeling strategies have been suggested. These include latent trait models (Johnson, 1996; Uebersax & Grove, 1993), quasi-symmetric loglinear models and related association models and latent class models (Agresti, 1988; Becker, 1989; Becker, 1990; Becker & Agresti, 1992; Agresti & Lang, 1993b), random effects (Williamson & Manatunga, 1997), and the area under a ROC curve (Toledano & Gatsonis, 1996).

We illustrate a simple loglinear approach for the case of two observers,  $A$  and  $B$ . For a sample of cases, let  $y_{ij}$  denote the number of cases rated in category  $i$  by  $A$  and in category  $j$  by  $B$ , and let  $\mu_{ij} = E(y_{ij})$ . Let  $u_1 < u_2 < \dots < u_c$  be scores assigned to the ordered response categories. A useful loglinear model for ordinal contingency tables is the model of *linear-by-linear association*,

$$\log \mu_{ij} = \lambda + \lambda_i^A + \lambda_j^B + \beta u_i u_j.$$

The model of independence is the special case with  $\beta = 0$ . This model often fits cross classifications of ordinal variables well, but when both dimensions have the same categories (as in observer agreement studies) typically more cases occur on the main diagonal than this model predicts. A model that allows for such extra agreement is

$$\log \mu_{ij} = \lambda + \lambda_i^A + \lambda_j^B + \beta u_i u_j + \delta I(i = j),$$

where  $I()$  is an indicator that equals 1 when  $i = j$  and 0 otherwise. The agreement may be described by odds ratios such as

$$\frac{\mu_{ii}\mu_{jj}}{\mu_{ij}\mu_{ji}} = \exp \{ \beta(u_i - u_j)^2 + 2\delta \}.$$

Hence, the level of agreement depends on a general association component (with size depending on  $\beta$ ), the distance between the categories, and the parameter  $\delta$  describing extra agreement on the main diagonal.

A paper (Cook & Farewell, 1995) for binary responses uses a conditioning argument for a two-stage agreement analysis in which the first stage focuses on cluster-specific agreement and the second stage on marginal agreement. This interesting analysis could be extended to ordinal responses. Also, current approaches typically treat the observers as a fixed set, whereas in most applications it is more natural to treat them as a sample. A cumulative logit model with random effects both for subjects rated and for the observers may be promising for many applications.

#### 5.4 Comparing Marginal Distributions of Contingency Tables

In repeated measurement studies, effects of interest may be either between-subject or within-subject. For categorical responses, methods for the within-subject comparison of marginal distributions of contingency tables have a long history (Bishop *et al.*, 1975). Two model-based methods have traditionally been used. One is the likelihood-ratio test for the marginal model, based on comparing the maximized log likelihood under the assumption of marginal homogeneity and without that assumption. The other is based on comparing loglinear models of symmetry and quasi symmetry; assuming quasi symmetry, marginal homogeneity is equivalent to symmetry. Examples presented in the literature tend to be simple, with few dimensions and usually even without incorporating covariates, since both approaches have computational difficulties as the table that cross classifies the responses gets more sparse. Yet another approach to testing, computationally simpler but not model-based so not yielding estimated effects under the alternative, can be based on a randomization argument (Landis *et al.*, 1988) or an extension of the Cochran–Mantel–Haenszel test to ordered response categories (Landis, Heyman & Koch, 1978).

Landis & Koch (1977) gave an example in which seven pathologists made diagnoses about carcinoma of the uterine cervix, with ratings on a five-point ordinal scale ranging from negative to invasive carcinoma, for a sample of 118 cases. Consider testing the hypothesis of marginal homogeneity, meaning that the seven pathologists had the same distribution for the response scale. Although such a problem seems simple, computational problems result from having a contingency table with  $5^7 = 78,125$  cells; that is, the model of marginal homogeneity applies to the one-way margins but maximization of the likelihood requires using all 78,125 cells.

The more recent literature suggests two approaches that are computationally simpler than the likelihood-based approach, with or without loglinear models: The GEE marginal model approach described in Section 2 and the random effects cluster-specific model approach described in Section 3. We illustrate for the Landis & Koch (1977) example, using a cumulative logit model. Let  $\gamma_k(j; i)$  denote the cumulative probability in category  $k$  for the rating of subject  $i$  by pathologist  $j$ . The marginal model allowing location shifts among the margins for the seven pathologists is

$$\text{logit}[\gamma_k(j; i)] = \alpha_k + \beta_j.$$

For identifiability, we use the constraint  $\beta_7 = 0$ . The corresponding random intercept model for the cluster-specific approach is

$$\text{logit}[\gamma_k(j; i)] = u_i + \alpha_k + \beta_j,$$

where  $\{u_i\}$  are *iid* from a  $N(0, \sigma)$  distribution. ML is less awkward with this model than with the marginal one, as it applies to a  $7 \times 5 \times 118$  table, namely a  $7 \times 5$  table giving the response for each of the seven pathologists (one observation for each row) for each of the 118 cases. Assuming either model, marginal homogeneity is the special case  $\beta_1 = \beta_2 = \dots = \beta_7$ .

Table 3 shows estimates and standard errors of the estimated  $\{\beta_j\}$  for the two models. The estimates

using GEE for the marginal model use the independence working correlation, so they are the same as the ML estimates one would obtain by treating the 7 responses on each subject as independent. The estimates using the random effects model tend to be much larger in absolute value. This reflects a strong within-case correlation, summarized by  $\hat{\sigma} = 3.85$  for the normal random effects distribution.

The GEE Wald statistic for testing marginal homogeneity ( $\beta_1 = \beta_2 = \dots = \beta_7$ ) equals 114.2, based on  $df = 6$ . This is similar to results obtained in a related approach by Agresti, Lipsitz & Lang (1992), who used the “pseudo ML”  $\{\hat{\beta}_j\}$  estimates based on a working independence structure and obtained their covariance matrix using the jackknife; they reported a Wald statistic value of 113.6, with  $df = 6$ . The random effects approach is likelihood-based, so one can construct a likelihood-ratio test of marginal homogeneity. The test statistic equals 253.6, with  $df = 6$ .

**Table 3**

*Results of fitting cumulative logit models (with standard errors in parentheses) to 5<sup>7</sup> table from Landis & Koch (1977) cross classifying seven observers on a 5-category ordinal scale.*

Observer Effect	GEE	Random Effects
A	-0.451 (.108)	-1.201 (.300)
B	-0.391 (.093)	-0.919 (.299)
C	0.319 (.118)	0.558 (.301)
D	0.632 (.105)	1.545 (.313)
E	-0.491 (.098)	-1.379 (.300)
F	1.252 (.161)	2.907 (.344)

Note: Estimate set equal to 0 for observer G. Random effects fit used PROC NL MIXED in SAS, based on adaptive Gauss–Hermite quadrature with 20 quadrature points.

### 5.5 Survival Modeling of Interval-censored Data

An application that is tangentially related to the analysis of clustered ordinal response data is survival analysis with interval censoring. This is a type of censoring for time-to-event data in which one observes only a time interval in which a response occurred rather than the exact time. For instance, in a series of clinical examinations, the time of an event is known only to fall between the examination times preceding and subsequent to the event. Thus, the ordered categories refer to a discretized version of an inherently continuous response.

Here, repeated measurement may occur on each subject, but ultimately the response is simply the interval in which the event occurred. So, the cluster of observations for a subject reduces to a univariate measure that is the interval censoring of the time-to-event variable. This application leads naturally to a proportional odds type of model. Let  $T_i$  be the unobserved event time for subject  $i$ . Then the model has the form

$$\text{logit}[P(T_i \leq t)] = \alpha(t) - \mathbf{x}_i' \boldsymbol{\beta},$$

where  $\alpha(\cdot)$  is an unspecified baseline function that is an infinite-dimensional nuisance parameter and  $\boldsymbol{\beta}$  describes the effects of interest. For details, see Rossini & Tsiatis (1996), Rabinowitz, Betensky & Tsiatis (2000), and Xie, Simpson & Carroll (2000). For use of a proportional hazards sort of model with an ordinal response, for which the link function is the complementary log-log link, see McCullagh (1980) and Farewell (1982).

### 5.6 Paired Preference Modeling

The Bradley–Terry model describes outcomes of pairwise competitions of a set of items (e.g. pro players in tennis matches, brands of a type of wine in taste comparisons). Each item has a parameter, and the logit of the probability of preference of item  $i$  over  $j$  equals the difference between the parameters for those items. This model extends to response scales on which there is an ordering for the evaluation. For instance, one wine might be rated as much better, slightly better, equal, slightly worse, or much worse than another wine. Direct applications apply of ordinal models using cumulative logits or adjacent-categories logits (Agresti, 1992; Fahrmeir & Tutz, 1994; Böckenholt & Dillon, 1997).

In this type of application, clustering and subsequent dependence can arise from repeated comparisons by the same subjects. For instance, a given rater may compare several pairs of wines. For future work, a random effects approach seems natural here to help account for the various types of dependence that can occur in such studies and to account for the raters typically being a sample of the possible ones.

### 5.7 Interface with Rank-based Methods

Traditional nonparametric methods deal with ranks of observations. Methods for ordered categorical data may be regarded as dealing with a crude type of ranking in which the categorical response corresponds to a large number of ties. As the number of categories increase and the data are “more continuous”, there is a natural connection of rank-based methods and methods for ordered categories. There is scope for exploring these connections with some of the recent work done in extending rank-based methods to repeated measurement and other forms of clustered data. So far the main emphasis of the rank-based literature has been on hypothesis testing. See, for instance, Akritas & Brunner (1997) and Brunner, Munzel & Puri (1999).

### 5.8 Other Applications

The literature on methods for clustered ordinal data continues to grow at a rapid rate. For a variety of other methods and applications, see Catalano (1997), Gange *et al.* (1995), Gansky *et al.* (1994), Gibbons *et al.* (1994), Glynn & Rosner (1994), Kenward & Jones (1994), Lesaffre *et al.* (1994, 1996), Lindsey *et al.* (1997), Qu & Tan (1998), Qu *et al.* (1995), Wang (1996), Williams *et al.* (1996), Williamson & Kim (1996), and Williamson & Lee (1996).

## 6 Software

New methods are rarely used in practice unless accompanied by user-friendly software. Of particular benefit would be a program that can handle a variety of strategies for multivariate ordinal logit models, including ML fitting of marginal models, GEE methods, and mixed models, all for a variety of link functions. Even with binary data, such goals currently require a variety of software, and more basic software needs exist that are not nearly as ambitious. For instance, some major statistical packages (e.g., SPSS) do not yet contain procedures for fitting univariate cumulative logit models or other models for multinomial responses.

Currently, SAS offers the greatest scope of methods for repeated categorical data (see the *SAS/STAT User's Guide* online manual at their website). Starting in Version 7, PROC GENMOD can fit cumulative link models and can perform GEE analyses (with independence working correlation) for marginal models using that family of links, including cumulative logit and probit. Table 4 illustrates this with the cumulative logit link for the marginal model fitted to Table 1. There is, however, no capability of ML fitting of marginal models. For cumulative link models containing random effects,

one can use PROC NLMIXED (starting in Version 7). This uses adaptive Gauss–Hermite quadrature for integration with respect to the random effects distribution to determine the likelihood function. NLMIXED is not naturally designed for multinomial responses, but one can use it for such models by specifying the form of the likelihood. Table 4 illustrates for the cumulative logit link. See Hartzel *et al.* (2001), Agresti, Booth, Hobert & Caffo (2000), and the SAS website mentioned above for other examples.

**Table 4**

*Example of SAS code (Version 8) for using PROC GENMOD and PROC NLMIXED to fit cumulative logit marginal and random effects models allowing interaction to Table 1.*

```

data example;
input case treat occas outcome ;
datalines; ** two lines for each subject, as illustrated ;
    1    1    0    1
    1    1    1    1
    2    1    0    1
    2    1    1    1
    3    1    0    1
    3    1    1    1
.....
    239  0    0    4
    239  0    1    4
;
proc genmod data=example;
class case;
model outcome = treat*occas treat*occas /
  dist=multinomial link=cumlogit ;
repeated subject=case / type=indep corrw;
run;
proc nlmixed data=example qpoints=40;
  bounds i2>0, i3>0; ** ensures intercepts have proper ordering ** ;
  eta1 = i1 + treat*beta1 + occas*beta2 + treat*occas*beta3 + u;
  eta2 = i1+i2 + treat*beta1 + occas*beta2 + treat*occas*beta3 + u;
  eta3 = i1+i2+i3 + treat*beta1 + occas*beta2 + treat*occas*beta3 + u;
  p1 = 1/(1 + exp(-eta1)); ** probability in category 1 ** ;
  p2 = 1/(1 + exp(-eta2)) - 1/(1 + exp(-eta1));
  p3 = 1/(1 + exp(-eta3)) - 1/(1 + exp(-eta2));
  p4 = 1 - 1/(1 + exp(-eta3));
  if (outcome=1) then y1=1; else y1=0;
  if (outcome=2) then y2=1; else y2=0;
  if (outcome=3) then y3=1; else y3=0;
  if (outcome=4) then y4=1; else y4=0;
  ll = y1*log(p1) + y2*log(p2) + y3*log(p3) + y4*log(p4);
  model outcome ~ general(ll); ** Define general log likelihood ** ;
  estimate 'alpha2' i1+i2;
  estimate 'alpha3' i1+i2+i3;
  random u ~ normal(0, su*su) subject=case;
run;

```

A FORTRAN program (MIXOR) is also available for cumulative logit models with random effects (Hedeker & Gibbons, 1996). It uses Gauss–Hermite numerical integration, but standard errors are based on expected information whereas NLMIXED uses observed information.

The approach of ML fitting of marginal models using constrained methods of maximization is available in a S-plus function available from Prof. J. B. Lang (Statistics Dept., Univ. of Iowa). The ML marginal fitting approach based on maximization with respect to joint probabilities expressed in terms of the marginal model parameters and an association model is available in software called MAREG. It is designed for either ML or GEE fitting of marginal models for ordinal responses (Heumann, 1997). For details, see <http://www.stat.uni-muenchen.de/~andreas/mareg/winmareg.html>.

The software package STATA has some random effects and GEE capability for modeling clustered ordinal categorical data, but we do not have any experience with these procedures. Perhaps surprisingly, there do not yet seem to be any generally available Splus or R functions for generalized linear mixed modeling or marginal modeling of ordinal response data. Splus does have GEE capability for binary data and other univariate distributions, but we are unaware of any functions for multinomial responses.

## Acknowledgments

Prof. Agresti's research was partially supported by grants from the National Institutes of Health and the National Science Foundation.

## References

- Agresti, A. (1988). A model for agreement between ratings on an ordinal scale. *Biometrics*, **44**, 539–548.
- Agresti, A. (1989). A survey of models for repeated ordered categorical response data. *Statistics in Medicine*, **8**, 1209–1224.
- Agresti, A. (1992). Analysis of ordinal paired comparison data. *Applied Statistics*, **41**, 287–297.
- Agresti, A. (1993a). Computing conditional maximum likelihood estimates for generalized Rasch models using simple loglinear models with diagonal parameters. *Scandinavian Journal of Statistics*, **20**, 63–71.
- Agresti, A. (1993b). Distribution-free fitting of logit models with random effects of repeated categorical responses. *Statistics in Medicine*, **12**, 1969–1987.
- Agresti, A. (1999). Modelling ordered categorical data: Recent advances and future challenges. *Statistics in Medicine*, **18**, 2191–2207.
- Agresti, A. & Lang, J.B. (1993a). A proportional odds model with subject-specific effects for repeated ordered categorical responses. *Biometrika*, **80**, 527–534.
- Agresti, A. & Lang, J.B. (1993b). Quasi-symmetric latent class models, with application to rater agreement. *Biometrics*, **49**, 131–139.
- Agresti, A., Booth, J., Hobert, J. & Caffo, B. (2000). Random effects modeling of categorical response data. *Sociological Methodology*, **30**, 27–81.
- Agresti, A., Lipsitz, S. & Lang, J.B. (1992). Comparing marginal distributions of large, sparse contingency tables. *Computational Statistics and Data Analysis*, **14**, 55–73.
- Aitkin, M. (1999). A general maximum likelihood analysis of variance components in generalized linear models. *Biometrics*, **55**, 117–128.
- Akritis, M.G. & Brunner, E. (1997). A unified approach to rank tests for mixed models. *Journal of Statistical Planning and Inference*, **61**, 249–277.
- Albert, J.H. & Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, **88**, 669–679.
- Albert, P.S. (1994). A Markov model for sequences of ordinal data from a relapsing-remitting disease. *Biometrics*, **50**, 51–60.
- Albert, P.S., Hunsberger, S.A. & Biro, F.M. (1997). Modeling repeated measures with monotonic ordinal responses and misclassification, with application to studying maturation. *Journal of the American Statistical Association*, **92**, 1304–1311.
- Anderson, J.A. & Philips, P.R. (1981). Regression, discrimination and measurement models for ordered categorical variables. *Applied Statistics*, **30**, 22–31.
- Becker, M.P. (1989). Using association models to analyse agreement data: Two examples. *Statistics in Medicine*, **8**, 1199–1207.
- Becker, M.P. (1990). Quasisymmetric models for the analysis of square contingency tables. *Journal of the Royal Statistical Society, Series B, Methodological*, **52**, 369–378.



- Becker, M.P. & Agresti, A. (1992). Log-linear modelling of pairwise interobserver agreement on a categorical scale. *Statistics in Medicine*, **11**, 101–114.
- Bedrick, E.J., Christensen, R. & Johnson, W. (1996). A new perspective on priors for generalized linear models. *Journal of the American Statistical Association*, **91**, 1450–1460.
- Bergsma, W.P. (1997). *Marginal models for categorical data*. Ph.D. thesis, Tilburg Univ., The Netherlands.
- Bishop, Y.M.M., Fienberg, S.E. & Holland, P.W. (1975). *Discrete multivariate analyses: Theory and practice*. MIT Press.
- Böckenholt, U. & Dillon, W.R. (1997). Modeling within-subject dependencies in ordinal paired comparison data. *Psychometrika*, **62**, 411–434.
- Bonney, G.E. (1986). Regressive logistic models for familial disease and other binary traits. *Biometrics*, **42**, 611–625.
- Booth, J.G. & Hobert, J.P. (1999). Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo Em algorithm. *Journal of the Royal Statistical Society, Series B, Methodological*, **61**, 265–285.
- Breslow, N.E. & Day, N.E. (1980). *Statistical methods in cancer research Volume I: The analysis of case-control studies*. IARC Scientific Publications.
- Brunner, E., Munzel, U. & Puri, M.L. (1999). Rank-score tests in factorial designs with repeated measures. *Journal of Multivariate Analysis*, **70**, 286–317.
- Carr, G.J., Hafner, K.B. & Koch, G.G. (1989). Analysis of rank measures of association for ordinal data from longitudinal studies. *Journal of the American Statistical Association*, **84**, 797–804.
- Catalano, P.J. (1997). Bivariate modelling of clustered continuous and ordered categorical outcomes. *Statistics in Medicine*, **16**, 883–900.
- Chen, M-H. (1998). Bayesian analysis of correlated mixed categorical data by incorporating historical prior information. *Communications in Statistics–Theory Meth.*, **27**, 1341–1361.
- Chen, M-H. & Shao, Q-M. (1999). Properties of prior and posterior distributions for multivariate categorical response data models. *Journal of Multivariate Analysis*, **71**, 277–296.
- Chuang, C. & Francom, S.F. (1986). A structural model for the Markov chain of change in vote intention. *Communications in Statistics, Part A–Theory and Methods*, **15**, 3475–3487.
- Conaway, M.R. (1989). Analysis of repeated categorical measurements with conditional likelihood methods. *Journal of the American Statistical Association*, **84**, 53–62.
- Cook, R.J. & Farewell, V.T. (1995). Conditional inference for subject-specific and marginal agreement: Two families of agreement measures. *The Canadian Journal of Statistics*, **23**, 333–344.
- Coull, B.A. & Agresti, A. (2000). Random effects modeling of multiple binomial responses using the multivariate binomial logit-normal distribution. *Biometrics*, **56**, 73–80.
- Cowles, M.K., Carlin, B.P. & Connett, J.E. (1996). Bayesian tobit modeling of longitudinal ordinal clinical trial compliance data with nonignorable missingness. *Journal of the American Statistical Association*, **91**, 86–98.
- Cox, C. (1995). Location-scale cumulative odds models for ordinal data: A generalized non-linear model approach. *Statistics in Medicine*, **14**, 1191–1203.
- Croon, M. (1990). Latent class analysis with ordered latent classes. *The British Journal of Mathematical and Statistical Psychology*, **43**, 171–192.
- Croon, M.A. (1993). Latent class models with ordered latent classes. An approach to nonparametric latent trait modeling. *Statistical Modelling and Latent Variables*, pp. 81–94.
- Crouchley, R. (1995). A random-effects model for ordered categorical data. *Journal of the American Statistical Association*, **90**, 489–498.
- Dale, J.R. (1986). Global cross-ratio models for bivariate, discrete, ordered responses. *Biometrics*, **42**, 909–917.
- Diggle, P.J., Liang, K.-Y. & Zeger, S.L. (1996). *Analysis of Longitudinal Data*. Clarendon Press.
- Ezzet, F. & Whitehead, J. (1991). A random effects model for ordinal responses from a crossover trial (disc: P906-907). *Statistics in Medicine*, **10**, 901–906.
- Fahrmeir, L. & Tutz, G. (1994). Dynamic stochastic models for time-dependent ordered paired comparison systems. *Journal of the American Statistical Association*, **89**, 1438–1449.
- Farewell, V.T. (1982). A note on regression analysis of ordinal data with variability of classification. *Biometrika*, **69**, 533–538.
- Fitzmaurice, G.M. & Laird, N.M. (1993). A likelihood-based method for analysing longitudinal binary responses. *Biometrika*, **80**, 141–151.
- Fleiss, J.L., Cohen, J. & Everitt, B.S. (1969). Large-sample standard errors of kappa and weighted kappa. *Psychological Bulletin*, **72**, 322–327.
- Francom, S.F., Chuang-Stein, C. & Landis, J.R. (1989). A log-linear model for ordinal data to characterize differential change among treatments. *Statistics in Medicine*, **8**, 571–582.
- 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.
- Gansky, S.A., Koch, G.G. & Wilson, J. (1994). Statistical evaluation of relationships between analgesic dose and ordered ratings of pain relief over an eight-hour period. *Journal of Biopharmaceutical Statistics*, **4**, 233–265.
- Gibbons, R.D., Hedeker, D., Charles, S.C. & Frisch, P. (1994). A random-effects probit model for predicting medical malpractice claims. *Journal of the American Statistical Association*, **89**, 760–767.
- Glonek, G.F.V. (1996). A class of regression models for multivariate categorical responses. *Biometrika*, **83**, 15–28.
- Glonek, G.F.V. & McCullagh, P. (1995). Multivariate logistic models. *Journal of the Royal Statistical Society, Series B, Methodological*, **57**, 533–546.
- Glynn, R.J. & Rosner, B. (1994). Comparison of alternative regression models for paired binary data. *Statistics in Medicine*, **13**, 1023–1036.
- Goodman, L.A. (1979). Multiplicative models for square contingency tables with ordered categories. *Biometrika*, **66**, 413–418.

- Goodman, L.A. (1994). On quasi-independence and quasi-dependence in contingency tables, with special reference to ordinal triangular contingency tables. *Journal of the American Statistical Association*, **89**, 1059–1063.
- Grizzle, J.E., Starmer, C.F. & Koch, G.G. (1969). Analysis of categorical data by linear models (corr: V26 p860). *Biometrics*, **25**, 489–504.
- Haber, M. (1985). Maximum likelihood methods for linear and log-linear models in categorical data. *Computational Statistics and Data Analysis*, **3**, 1–10.
- Hartzel, J., Liu, I-Ming & Agresti, A. (2001). Describing heterogeneous effects in stratified ordinal contingency tables, with application to multi-center clinical trials. *Computational Statistics and Data Analysis*, **35**, 429–449.
- 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.
- Heinen, T. (1993). *Discrete latent variable models*. Tilburg University Press.
- Heumann, C. (1997). *Likelihoodbasierte marginale regressionsmodelle für korrelierte kategoriale daten*. Ph.D. thesis, Ludwig-Maximilians-Universität, München.
- Ishwaran, H. & Gatsonis, C. (2000). A general class of hierarchical ordinal regression models with applications to correlated ROC analysis. *Canadian Journal of Statistics*, **28**, 731–750.
- Jaffrézic, F., Robert-Granié, C.R. & Foulley, J.-L. (1999). A quasi-score approach to the analysis of ordered categorical data via a mixed heteroskedastic threshold model. *Genet. Sel. Evol.*, **31**, 301–318.
- Jansen, J. (1990). On the statistical analysis of ordinal data when extravariation is present. *Applied Statistics*, **39**, 75–84.
- Johnson, V.E. (1996). On Bayesian analysis of multirater ordinal data: An application to automated essay grading. *Journal of the American Statistical Association*, **91**, 42–51.
- Kenward, M.G. & Jones, B. (1991). The analysis of categorical data from cross-over trials using a latent variable model. *Statistics in Medicine*, **10**, 1607–1619.
- Kenward, M.G. & Jones, B. (1994). The analysis of binary and categorical data from crossover trials. *Statistical Methods in Medical Research*, **3**, 325–344.
- 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.
- Kim, K. (1995). A bivariate cumulative probit regression model for ordered categorical data. *Statistics in Medicine*, **14**, 1341–1352.
- Koch, G.G., Landis, J.R., Freeman, J.L., Freeman, D.H., Jr. & Lehnen, R.G. (1977). A general methodology for the analysis of experiments with repeated measurement of categorical data. *Biometrics*, **33**, 133–158.
- Landis, J.R. & Koch, G.G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*, **33**, 363–374.
- Landis, J.R., Heyman, E.R. & Koch, G.G. (1978). Average partial association in three-way contingency tables: A review and discussion of alternative tests. *International Statistical Review*, **46**, 237–254.
- Landis, J.R., Miller, M.E., Davis, C.S. & Koch, G.G. (1988). Some general methods for the analysis of categorical data in longitudinal studies. *Statistics in Medicine*, **7**, 109–137.
- Lang, J.B. (1996). Maximum likelihood methods for a generalized class of log-linear models. *The Annals of Statistics*, **24**, 726–752.
- Lang, J.B. & Agresti, A. (1994). Simultaneously modeling joint and marginal distributions of multivariate categorical responses. *Journal of the American Statistical Association*, **89**, 625–632.
- Lang, J.B., McDonald, J.W. & Smith, P.W.F. (1999). Association-marginal modeling of multivariate categorical responses: A maximum likelihood approach. *Journal of the American Statistical Association*, **94**, 1161–1171.
- Lesaffre, E., Molenberghs, G. & Dewulf, L. (1996). Effect of dropouts in a longitudinal study: An application of a repeated ordinal model. *Statistics in Medicine*, **15**, 1123–1141.
- Lesaffre, E., Verbeke, G. & Molenberghs, G. (1994). A sensitivity analysis of two multivariate response models. *Computational Statistics and Data Analysis*, **17**, 363–391.
- Li, Y. & Lin, X. (2000). Testing random effects in uncensored/censored clustered data with categorical responses. *Unpublished manuscript*.
- 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 (disc: P24-40). *Journal of the Royal Statistical Society, Series B, Methodological*, **54**, 3–24.
- Lindsey, J.K. (1999). *Models for repeated measurements, second edition*. Oxford: Oxford University Press.
- Lindsey, J.K., Jones, B. & Ebbutt, A.F. (1997). Simple models for repeated ordinal responses with an application to a seasonal rhinitis clinical trial. *Statistics in Medicine*, **16**, 2873–2882.
- Lipsitz, S.R. & Fitzmaurice, G.M. (1996). The score test for independence in  $R \times C$  contingency tables with missing data. *Biometrics*, **52**, 751–762.
- Lipsitz, S.R., Kim, K. & Zhao, L. (1994). Analysis of repeated categorical data using generalized estimating equations. *Statistics in Medicine*, **13**, 1149–1163.
- Little, R.J.A. & Rubin, D.B. (1987). *Statistical analysis with missing data*. New York: Wiley.

- Little, R.J., Raghunathan, T. & Tang, G. (2000). Pattern and pattern-set mixture models for nonignorable nonresponse. *Proceedings, 20th International Biometric Conference*.
- Lumley, T. (1996). Generalized estimating equations for ordinal data: A note on working correlation structures. *Biometrics*, **52**, 354–361.
- Mark, S.D. & Gail, M.H. (1994). A comparison of likelihood-based and marginal estimating equation methods for analysing repeated ordered categorical responses with missing data (Disc: p495-498). *Statistics in Medicine*, **13**, 479–493.
- Masters, G.N. (1985). A comparison of latent trait and latent class analyses of Likert-type data. *Psychometrika*, **50**, 69–82.
- McCullagh, P. (1977). A logistic model for paired comparisons with ordered categorical data. *Biometrika*, **64**, 449–454.
- McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society, Series B, Methodological*, **42**, 109–142.
- McCullagh, P. & Nelder, J.A. (1989). *Generalized linear models (second edition)*. Chapman & Hall.
- Miller, M.E., Davis, C.S. & Landis, J.R. (1993). The analysis of longitudinal polytomous data: 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.
- Molenberghs, G., Kenward, M.G. & Lesaffre, E. (1997). The analysis of longitudinal ordinal data with nonrandom drop-out. *Biometrika*, **84**, 33–44.
- Natarajan, R. (1995). Estimation in models for multinomial response data: Bayesian and frequentist approaches. *Unpublished Ph.D. dissertation*, Cornell University.
- Natarajan, R. & Kass, R.E. (2000). Reference Bayesian methods for generalized linear mixed models. *Journal of the American Statistical Association*, **95**, 227–237.
- 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.
- Pendergast, J.F., Gange, S.J., Newton, M.A., Lindstrom, M.J., Palta, M. & Fisher, M.R. (1996). A survey of methods for analyzing clustered binary response data. *International Statistical Review*, **64**, 89–118.
- Pepe, M. (1997). A regression modelling framework for receiver operating characteristic curves in medical diagnostic testing. *Biometrika*, **84**, 595–608.
- Pepe, M. S. (2000). Receiver operating characteristic methodology. *Journal of the American Statistical Association*, **95**, 308–311.
- Peterson, B. & Harrell, F.E., Jr. (1990). Partial proportional odds models for ordinal response variables. *Applied Statistics*, **39**, 205–217.
- Pinheiro, J.C. & Bates, D.M. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics*, **4**, 12–35.
- Qu, Y. & Tan, M. (1998). Analysis of clustered ordinal data with subclusters via a Bayesian hierarchical model. *Communications in Statistics—Theory Meth.*, **27**, 1461–1475.
- Qu, Y., Piedmonte, M.R. & Medendorp, S.V. (1995). Latent variable models for clustered ordinal data. *Biometrics*, **51**, 268–275.
- Rabinowitz, D., Betensky, R.A. & Tsiatis, A.A. (2000). Using conditional logistic regression to fit proportional odds models to interval censored data. *Biometrics*, **56**, 511–518.
- Rasch, G. (1961). On general laws and the meaning of measurement in psychology. *Proc. 4th Berkeley Symp. Math. Statist. Probab.*, **4**, 321–333.
- Rossini, A.J. & Tsiatis, A.A. (1996). A semiparametric proportional odds regression model for the analysis of current status data. *Journal of the American Statistical Association*, **91**, 713–721.
- Rost, J. (1985). A latent class model for rating data. *Psychometrika*, **50**, 37–49.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika, Monograph Suppl.*, **17**.
- Sheiner, L.B., Beal, S.L. & Dunne, A. (1997). Analysis of nonrandomly censored ordered categorical longitudinal data from analgesic trials. *Journal of the American Statistical Association*, **92**, 1235–1244.
- Spiegelhalter, D.J., Thomas, A., Best, N.G. & Gilks, W.R. (1996). BUGS: Bayesian inference using Gibbs sampling, Version .050. *MRC Biostatistics Unit, Cambridge: UK*.
- Stram, D.O., Wei, L.J. & Ware, J.H. (1988). Analysis of repeated ordered categorical outcomes with possibly missing observations and time-dependent covariates. *Journal of the American Statistical Association*, **83**, 631–637.
- Tan, M., Qu, Y., Mascha, E. & Schubert, A. (1999). A Bayesian hierarchical model for multi-level repeated ordinal data: analysis of oral practice examinations in a large anaesthesiology training programme. *Statistics in Medicine*, **18**, 1983–1992.
- 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.
- Ten Have, T.R., Landis, J.R. & Hartzel, J. (1996). Population-averaged and cluster-specific models for clustered ordinal response data. *Statistics in Medicine*, **15**, 2573–2588.
- Tjur, T. (1982). A connection between Rasch's item analysis model and a multiplicative Poisson model. *Scandinavian Journal of Statistics*, **9**, 23–30.
- Toledano, A.Y. & Gatsonis, C. (1995). Regression analysis of correlated receiver operating characteristic data. *Academic Radiology*, **2**, S30–S36.
- Toledano, A.Y. & Gatsonis, C. (1996). Ordinal regression methodology for ROC curves derived from correlated data. *Statistics in Medicine*, **15**, 1807–1826.
- Toledano, A.Y. & Gatsonis, C. (1999). Generalized estimating equations for ordinal categorical data: arbitrary patterns of

- missing responses in a key covariate. *Biometrics*, **55**, 488–496.
- Tomizawa, S. (1995). A generalization of the marginal homogeneity model for square contingency tables with ordered categories. *Journal of Educational and Behavioral Statistics*, **20**, 349–360.
- Tosteson, A.N. & Begg, C.B. (1988). A general regression methodology for ROC curve estimation. *Medical Decision Making*, **8**, 204–215.
- Tutz, G. (1991). Sequential models in categorical regression. *Computational Statistics and Data Analysis*, **11**, 275–295.
- Tutz, G. & Hennevogl, W. (1996). Random effects in ordinal regression models. *Computational Statistics and Data Analysis*, **22**, 537–557.
- Uebersax, J.S. & Grove, W.M. (1993). A latent trait finite mixture model for the analysis of rating agreement. *Biometrics*, **49**, 823–835.
- Wahrendorf, J. (1980). Inference in contingency tables with ordered categories using Plackett's coefficient of association for bivariate distributions. *Biometrika*, **67**, 15–21.
- Wang, Y.-G. (1996). A quasi-likelihood approach for ordered categorical data with overdispersion. *Biometrics*, **52**, 1252–1258.
- Wedderburn, R.W.M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss–Newton method. *Biometrika*, **61**, 439–447.
- Williams, P.L., Molenberghs, G. & Lipsitz, S.R. (1996). Analysis of multiple ordinal outcomes in developmental toxicity studies. *Journal of Agricultural, Biological, and Environmental Statistics*, **1**, 250–274.
- Williamson, J. & Kim, K. (1996). A global odds ratio regression model for bivariate ordered categorical data from ophthalmologic studies. *Statistics in Medicine*, **15**, 1507–1518.
- Williamson, J. & Lee, M.-L.T. (1996). A GEE regression model for the association between an ordinal and a nominal variable. *Communications in Statistics, Part A—Theory and Methods*, **25**, 1887–1901.
- Williamson, J.M. & Manatunga, A.K. (1997). Assessing interrater agreement from dependent data. *Biometrics*, **53**, 707–714.
- Williamson, J.M., Kim, K. & Lipsitz, S.R. (1995). Analyzing bivariate ordinal data using a global odds ratio. *Journal of the American Statistical Association*, **90**, 1432–1437.
- Xie, M., Simpson, D.G. & Carroll, R.J. (2000). Random effects in censored ordinal regression: Latent structure and Bayesian approach. *Biometrics*, **56**, 376–383.
- Zeger, S.L., Liang, K.-Y. & Albert, P.S. (1988). Models for longitudinal data: A generalized estimating equation approach (corr: V45 P347). *Biometrics*, **44**, 1049–1060.

## Résumé

Cet article passe en revue diverses stratégies pour la modélisation de variables de réponse qualitatives et ordinales, quand les données présentent des phénomènes de grappes. Il prolonge une étude similaire réalisée pour les variables dichotomiques par Pendergast *et al.* (1996). Les mesures répétées à plusieurs reprises pour chaque individu, par exemple dans les études longitudinales, constituent un cas particulier important. Il existe à présent une variété beaucoup plus grande de modèles et de méthodes d'ajustement, que lors de la réalisation d'une étude similaire pour les variables ordinales et répétées il y a une dizaine d'années (Agresti, 1989). Il est mis particulièrement l'accent sur deux types de modèles, les modèles marginaux pour lesquels on prend la moyenne des effets sur toutes les grappes pour des niveaux donnés des variables explicatives, et les modèles par grappes pour lesquels les effets sont pris en compte au niveau de chaque grappe. Nous présentons les deux types de modèles dans le cas de variables ordinales, brosons un panorama de la littérature pour chacun des deux types, et discutons les relations entre eux. Puis nous résumons d'autres approches pour la modélisation et l'estimation de paramètres, notamment une approche bayésienne. Nous discutons aussi des applications et des domaines qui devraient susciter de nouvelles recherches, par exemple les méthodes de traitement des données manquantes ou d'évaluation de l'exactitude de tests. Enfin, nous considérons la disponibilité actuelle de logiciels pour mettre en oeuvre les méthodes discutées dans cet article.

*Mots clés:* Inférence bayésienne; Modèles logit cumulatifs; Équations d'estimation généralisées; Modèles logit; Modèles marginaux; Paires appariées; Données manquantes; Données ordinales; Risques proportionnels; Effets aléatoires; Mesures répétées; Tableaux de contingence.

[Received July 2000, accepted February 2001]