Sociedad de Estadística
e Investigación Operativa

Test

Volume 14, Number 1. June 2005

The Analysis of Ordered Categorical Data: An Overview and a Survey of Recent Developments

Ivy Liu
School of Mathematics, Statistics, and Computer Science
Victoria University of Wellington, New Zealand
Alan Agresti
Department of Statistics
University of Florida, USA

Sociedad de Estadística e Investigación Operativa
Test (2005) *Vol.* **14**, *No.* 1, pp. 1–73

Sociedad de Estadística e Investigación Operativa
Test (2005) *Vol.* **14**, *No.* 1, pp. 1–73

The Analysis of Ordered Categorical Data: An Overview and a Survey of Recent Developments

Ivy Liu
School of Mathematics, Statistics, and Computer Science
Victoria University of Wellington, New Zealand

Alan Agresti[*]
Department of Statistics
University of Florida, USA

Abstract

This article reviews methodologies used for analyzing ordered categorical (ordinal) response variables. We begin by surveying models for data with a single ordinal response variable. We also survey recently proposed strategies for modeling ordinal response variables when the data have some type of clustering or when repeated measurement occurs at various occasions for each subject, such as in longitudinal studies. Primary models in that case include *marginal models* and *cluster-specific (conditional) models* for which effects apply conditionally at the cluster level. Related discussion refers to multi-level and transitional models. The main emphasis is on maximum likelihood inference, although we indicate certain models (e.g., marginal models, multi-level models) for which this can be computationally difficult. The Bayesian approach has also received considerable attention for categorical data in the past decade, and we survey recent Bayesian approaches to modeling ordinal response variables. Alternative, non-model-based, approaches are also available for certain types of inference.

Key Words: Association model, Bayesian inference, cumulative logit, generalized estimating equations, generalized linear mixed model, inequality constraints, marginal model, multi-level model, ordinal data, proportional odds.

AMS subject classification: 62J99.

[*]Correspondence to: Alan Agresti. Department of Statistics, University of Florida. Gainesville, Florida, USA 32611-8545. E-mail: aa@stat.ufl.edu

This work was partially supported by a grant for A. Agresti from NSF and by a research study leave grant from Victoria University for I. Liu.

Received: January 2005; Accepted: January 2005

1 Introduction

This article reviews methodologies used for modeling ordered categorical (ordinal) response variables. Section 2 begins by reviewing the primary models used for data with a single ordinal response variable. The most popular models apply a link function to the cumulative probabilities, most commonly the logit or probit. Section 3 reviews recently proposed strategies for modeling ordinal response variables when the data have some type of clustering or when repeated measurement occurs at various occasions for each subject, such as in longitudinal studies. In this setting, *marginal models* analyze effects averaged over all clusters at particular levels of predictors. *Cluster-specific models* are conditional models, often using random effects, that describe effects at the cluster level. This section also considers multi-level models, which have a hierarchy of clustering levels, and transitional models, which model a response at a particular time in terms of previous responses.

The main emphasis in Sections 2 and 3 is on maximum likelihood (ML) model fitting. For certain models, however, ML fitting can still be computationally difficult. The main example is marginal modeling, for which it is awkward to express the likelihood function in terms of model parameters. In such cases, quasi-likelihood methods, such as the use of generalized estimating equations, are more commonly used. The Bayesian approach has also received considerable attention for categorical data in the past decade. Section 4 surveys Bayesian approaches to modeling ordinal data. Section 5 briefly describes alternative, non-model-based, approaches that are also available for certain types of inference. These include Mantel-Haenszel type methods, rank-based methods, and inequality-constrained methods. The article concludes by discussing other issues, such as dealing with missing data, power and sample size considerations, and available software.

2 Models for ordered categorical responses

Categorical data methods such as logistic regression and loglinear models were primarily developed in the 1960s and 1970s. Although models for ordinal data received some attention then (e.g. Snell, 1964; Bock and Jones, 1968), a stronger focus on the ordinal case was inspired by articles by McCullagh (1980) on logit modeling of cumulative probabilities and by Goodman (1979) on loglinear modeling relating to odds ratios that are natural for ordinal variables. This section overviews these and related modeling approaches.

2.1 Proportional odds model (A cumulative logit model)

Currently, the most popular model for ordinal responses uses logits of cumulative probabilities, often called *cumulative logits*. Early work with this approach includes Williams and Grizzle (1972) and Simon (1974). The model gained popularity primarily after the seminal article by McCullagh (1980) on regression modeling of ordinal responses. For a c-category ordinal response variable Y and a set of predictors $\mathbf{x}$ with corresponding effect parameters $\boldsymbol{\beta}$, the model has form

$$\text{logit}[P(Y \leq j \mid \mathbf{x})] = \alpha_j - \boldsymbol{\beta}'\mathbf{x}, \ \ j = 1, ..., c-1. \tag{2.1}$$

(The minus sign in the predictor term makes the sign of each component of $\boldsymbol{\beta}$ have the usual interpretation in terms of whether the effect is positive or negative, but it is not necessary to use this parameterization.) The parameters $\{\alpha_j\}$, called *cut points*, are usually nuisance parameters of little interest. This model applies simultaneously to all $c-1$ cumulative probabilities, and it assumes an identical effect of the predictors for each cumulative probability. Specifically, the model implies that odds ratios for describing effects of explanatory variables on the response variable are the same for each of the possible ways of collapsing a c-category response to a binary variable. This particular type of cumulative logit model, with effect $\boldsymbol{\beta}$ the same for all j, is often referred to as a *proportional odds model* (McCullagh, 1980).

To fit this model, it is unnecessary to assign scores to the response categories. One can motivate the model by assuming that the ordinal response Y has an underlying continuous response Y^* (Anderson and Philips, 1981). Such an unobserved variable is called a *latent variable*. Let Y^* have mean linearly related to $\mathbf{x}$, and have logistic conditional distribution with constant variance. Then for the categorical variable Y obtained by chopping Y^* into categories (with cutpoints $\{\alpha_j\}$), the proportional odds model holds for predictor $\mathbf{x}$, with effects proportional to those in the continuous model. If this latent variable model holds, the effects are invariant to the choices of the number of categories and their cutpoints. So, when the model fits well, different studies using different scales for the response variable should give similar conclusions. Farewell (1982) discussed the issue of the cutpoints themselves varying according to how different subjects perceive category boundaries. This is a problem that can now also be addressed with random intercepts in the model.

Some of the early literature with such models used weighted least squares for model fitting (e.g. Williams and Grizzle, 1972) but maximum likelihood is more versatile and these days is the preferred method. Walker and Duncan (1967) and McCullagh (1980) used Fisher scoring algorithms to do this. With the ML approach, one can base significance tests and confidence intervals for the model parameters $\boldsymbol{\beta}$ on likelihood-ratio, score, or Wald statistics.

When explanatory variables are categorical and data are not too sparse, one can also form chi-squared statistics to test the model fit by comparing observed frequencies to estimates of expected frequencies that satisfy the model. For sparse data or continuous predictors, such chi-squared fit statistics are inappropriate. The Hosmer – Lemeshow statistic for testing the fit of a logistic regression model for binary data has been generalized to ordinal responses by Lipsitz et al. (1996). This gives an alternative way to construct a goodness-of-fit test. It compares observed to fitted counts for a partition of the possible response (e.g., cumulative logit) values.

If the cumulative logit model fits poorly, one might include separate effects, replacing $\boldsymbol{\beta}$ in (2.1) by $\boldsymbol{\beta}_j$. This gives nonparallel curves for the $c-1$ logits. Such a model cannot hold over a broad range of explanatory variable values, because crossing curves for the logits implies that cumulative probabilities are out of order. However, such a model can be used as an alternative to the proportional odds model in a score test that the effects for different logits are identical (Brant, 1990; Peterson and Harrell, 1990).

When proportional odds structure may be inadequate, alternative possible strategies to improve the fit include (i) trying different link functions discussed in next section, such as the log-log, (ii) adding additional terms, such as interactions, to the linear predictor, (iii) generalizing the model by adding dispersion parameters (McCullagh, 1980; Cox, 1995), (iv) permitting separate effects for each logit for some but not all predictors (i.e., *partial proportional odds*; see Peterson and Harrell (1990), (v) using the ordinary model for a nominal response, which forms baseline-category logits by pairing each category with a baseline. In general, though, one should not use a different model based solely on an inadequate fit of the cumulative logit model according to a goodness-of-fit test. Especially when n is large, statistical significance need not imply an inadequate fit in a practical sense, and the decrease in bias obtained with a more complex model may be more than offset by the increased mean square error in estimating effects caused by the large increase in the number of model parameters. Rather than relying purely on testing, a sensible strategy is to fit models for the separate cumulative logits, and check whether the effects are different in a substantive sense, taking into account ordinary sampling variability. Alternatively, Kim (2003) proposed a graphical method for assessing the proportional odds assumption.

Applications of cumulative logit models have been numerous. An important one is for survival modeling of interval-censored data (e.g. Rossini and Tsiatis, 1996).

2.2 Cumulative link models

In addition to the logit link function for the cumulative probability, McCullagh (1980) applied other link functions that are commonly used for binary data, such as the probit, log-log and complementary log-log. A general model incorporating a variety of potential link functions is *cumulative link model*. It has form

$$G^{-1}[P(Y \leq j \mid \mathbf{x})] = \alpha_j - \boldsymbol{\beta}'\mathbf{x}, \;\; j = 1, ..., c-1, \tag{2.2}$$

where G^{-1} is a link function that is the inverse of a continuous cumulative distribution function (cdf) G.

Use of the standard normal cdf for G gives the *cumulative probit model*. It treats the underlying continuous variable Y^* as normal. If the model with logit link fits data well, so does the model with probit link, because the shapes of normal and logistic distributions are very similar. The choice of one over the other may depend on whether one wants to interpret effects in terms of odds ratios for the given categories (in which case the logit link is natural) or to interpret effects for an underlying normal latent variable for an ordinary regression model (in which case the probit link is more natural).

Use of the extreme value distribution, $G(\cdot) = 1 - \exp\{-\exp(\cdot)\}$ results in the model with complementary log-log link function, having form

$$\log\{-\log[1 - P(Y \leq j \mid \mathbf{x})]\} = \alpha_j + \boldsymbol{\beta}'\mathbf{x}.$$

McCullagh (1980) discussed its use and its connections with related proportional hazards models for survival data. The model using *log-log* link function $\log\{-\log[P(Y \leq j)]\}$ is appropriate when the complementary log-log link holds for the categories listed in reverse order. While the probit and logit link are most appropriate when an underlying continuous response is roughly bell-shaped, the complementary log-log link is suitable when the response curve is nonsymmetric. With such a link function, the cumulative probability approaches 0 at a different rate than it approaches 1 as the value of an explanatory variable increases. With small to moderate n it can be difficult to differentiate whether one cumulative link model fits better than another, so the choice of an appropriate link function should be based primarily on ease of interpretation. See Genter and Farewell (1985) for discussion of goodness-of-link testing.

2.3 Alternative multinomial logit models for ordinal responses

For ordinal responses, the *adjacent-categories logit* model (Simon, 1974; Goodman, 1983) is

$$\log[P(Y = j \mid \mathbf{x})/P(Y = j+1 \mid \mathbf{x})] = \alpha_j - \boldsymbol{\beta}'\mathbf{x}, \;\; j = 1, ..., c-1.$$

The model is a special case of the baseline-category logit model commonly used for nominal response variables (i.e., no natural ordering), with reduction in the number of parameters by utilizing the ordering to obtain a common effect. It utilizes single-category probabilities rather than cumulative probabilities, so it is more natural when one wants to describe effects in terms of odds relating to particular response categories. This model received considerable attention in the 1980s and 1990s, partly because of connections with certain ordinal loglinear models. See, for instance, Agresti (1992) who modeled paired preference data, extending the Bradley-Terry model to ordinal responses. For other work for such data, see Fahrmeir and Tutz (1994) and Böckenholt and Dillon (1997).

The cumulative logit and adjacent-categories logit model both imply stochastic orderings of the response distributions for different predictor values. Effects in adjacent-category logit models refer to the effect of a one-unit increase of a predictor on the log odds of response in the lower instead of the higher of any two adjacent categories, whereas the effect in (2.1) refers to the entire response scale. When the response variable has two categories, the cumulative logit and adjacent-categories logit models simplify to the ordinary logistic regression model.

An alternative logit model, called the *continuation-ratio* logit model, uses logits of form $\{\log[P(Y = j)/P(Y \geq j+1)]\}$ or $\{\log[P(Y = j+1)/P(Y \leq j)]\}$. Tutz (1990, 1991) referred to models using such logits as *sequential* type models, because the model form is useful when a sequential mechanism determines the response outcome. When the model includes separate effects $\{\boldsymbol{\beta}_j\}$, the multinomial likelihood factors into a product of the binomial likelihoods for the separate logits with respect to j. Then, separate fitting of models for different continuation-ratio logits gives the same results as simultaneous fitting.

McCullagh (1980) and Thompson and Baker (1981) treated the cumulative link model (2.2) as a special case of the multivariate generalized linear model

$$\mathbf{g}(\boldsymbol{\mu}_i) = \mathbf{X}_i\boldsymbol{\beta}, \tag{2.3}$$

where $\boldsymbol{\mu}_i$ is the mean of a vector of indicator responses $\mathbf{y}_i = (y_{i1}, y_{i2}, \ldots, y_{i,c-1})'$ for subject i (a 1 for the dummy variable value pertaining to the category in which the observation falls) and $\mathbf{g}$ is a vector of link functions. Similarly, the adjacent-category logit and continuation-ratio logit models can be considered within this class. Fahrmeir and Tutz (2001) gave details.

2.4 Other multinomial response models

Tutz (2003) extended the form of generalized additive models (Hastie and Tibshirani, 1990) by considering semiparametrically structured models for an ordinal response variable with various link functions. Tutz's model contains linear parts ($\mathbf{X}_i\boldsymbol{\beta}$), additive parts of covariates with an unspecified functional form, and interactions. This type of model is useful when there are several continuous covariates and some of them may have nonlinear relationships with $\mathbf{g}(\boldsymbol{\mu}_i)$. Tutz proposed a method of estimation based on an extension of penalized regression splines. For details of additive and semiparametric models for ordinal responses, see Hastie and Tibshirani (1987, 1990, 1993), Yee and Wild (1996), Kauermann (2000), and Kauermann and Tutz (2000, 2003).

Models considered thus far have recognized the categorical response scale by applying common link functions. In practice, such methods are still not well known to many researchers who commonly analyze ordinal data, such as social scientists who analyze questionnaire data with Likert-type scales. Probably the most common approach is to assign scores to the ordered categories and apply ordinary least squares regression modeling. See Heeren and D'Agostino (1987) for investigation of the robustness of such an approach.

A related approach attempts to recognize the categorical nature of the response variable in the regression model, by estimating the non-constant variance inherent to categorical measurement. This was first proposed, using weighted least squares for regression and ANOVA methods for categorical responses, by Grizzle et al. (1969). Let $\pi_j(\mathbf{x}) = P(Y = j \mid \mathbf{x})$. For an ordinal response, the *mean response model* has form

$$\sum_j v_j \pi_j(\mathbf{x}) = \alpha + \boldsymbol{\beta}'\mathbf{x}\,,$$

where v_j is the assigned score for response category j. With a categorical response scale, less variability tends to occur when the mean is near the high end or low end of the scale. Using such a model has the advantage of simplicity of interpretation, particularly when it is sufficient to summarize effects in terms of location rather than separate cell probabilities. It is simpler for many researchers to interpret effects expressed in terms of means rather than in terms of odds ratios. A structural problem, more common when c is small, is that the fit may give estimated means above the highest score or below the lowest score. Since the mean response model does not uniquely determine cell probabilities, unlike logit models, it does not imply stochastic orderings at different settings of predictors.

In principle, it is possible to use ML to fit mean response models using maximum likelihood, assuming a multinomial response, and this has received some attention (e.g. Lipsitz, 1992; Lang, 2004). Software is available for such fitting from Joseph Lang (Statistics Dept., Univ. of Iowa, e-mail: jblang@stat.uiowa.edu), using a method of maximizing the likelihood that treats the model formula as a constraint equation.

2.5 Modeling association with ordinal responses

The models discussed so far are like ordinary regression models in the sense that they distinguish between response and explanatory variables. *Association models*, on the other hand, are designed to describe association between variables, and they treat those variables symmetrically (Goodman, 1979, 1985). In an $r \times c$ contingency table, let X denote the row variable and Y denote the column variable. The cell counts $\{n_{ij}\}$ have expected values $\{\mu_{ij}\}$. Goodman (1985) proposed the association model with form

$$\log \mu_{ij} = \lambda + \lambda_i^X + \lambda_j^Y + \sum_{k=1}^{M} \beta_k u_{ik} v_{jk}, \tag{2.4}$$

where $M \leq \min(r-1, c-1)$. The saturated model (which gives a perfect fit) results when $M = \min(r-1, c-1)$.

The most commonly used models have $M = 1$ (Goodman, 1979; Haberman, 1974, 1981). The special cases then have *linear-by-linear association*, *row effects*, *column effects*, and *row and column effects (RC)*. The linear-by-linear association model treats the row scores $\{u_{i1}\}$ and the column scores $\{v_{j1}\}$ as fixed monotone constants. When $\beta_1 > 0$, Y tends to increase as X increases, whereas a negative sign of β shows a negative trend. Goodman (1979) proposed the special case $\{u_{i1} = i\}$ and $\{v_{j1} = j\}$, referring to it as the *uniform association* model because of the uniform value of all $(r-1)(c-1)$ *local odds ratios* constructed using pairs of adjacent rows and pairs of adjacent columns.

The row effects model (Goodman, 1979) fixes the column scores but treats row scores as parameters. It is also valid when X has nominal categories. The model with equally-spaced scores for Y relates to logit models for adjacent-category logits (Goodman, 1983). Likewise, the column effects model fixes the row scores but treats the column scores as parameters. The row-and-column-effects (RC) model treats both sets as parameters. In this case the model is not loglinear and ML estimation is more difficult (Haberman, 1981, 1995). In addition, anomalous behavior results for some inference, such as the likelihood-ratio statistic for testing that $\beta_1 = 0$ not having an asymptotic null chi-squared distribution (Haberman, 1981). Bartolucci and Forcina (2002) extended the RC model by allowing simultaneous modeling of marginal distributions with various ordinal logit models.

The general association model (2.4) for which $\beta_k = 0$ for $k > M^*$ is referred to as the $RC(M^*)$ model. See Becker (1990) for ML model fitting.

The goodness of fit of association models can be checked with ordinary chi-squared statistics, when cell counts are reasonably large. Association models have been generalized to include covariates (Becker, 1989; Becker and Clogg, 1989).

A related literature has developed for correspondence analysis models (Goodman, 1986, 1996; Gilula and Haberman, 1988; Gilula and Ritov, 1990) and equivalent canonical correlation models. The general canonical correlation model is

$$\pi_{ij} = \pi_{i+}\pi_{+j}\left(1 + \sum_{k=1}^{M} \lambda_k \mu_{ik} \nu_{jk}\right)$$

where $\pi_{i+} = \sum_j \pi_{ij}$ and $\pi_{+j} = \sum_i \pi_{ij}$. It has similar structure as the form in (2.4), but it uses an association term to model the difference between μ_{ij} and its independence value. When the association is weak, an approximate relation holds between parameter estimates in association models and canonical correlation models (Goodman, 1985), but otherwise the models refer to different types of ordinal association (Gilula et al., 1988). A major limitation of correspondence analysis and correlation models is non-trivial generalization to multiway tables. For overviews connecting the various models, as well as somewhat related latent class models, see Goodman (1986, 1996, 2004). For connections with graphical models, see Wermuth and Cox (1998) and Anderson and Böckenholt (2000). Association models and canonical correlation models do not seem to receive as much use in applications as the regression-type models (such as the cumulative logit models) that describe effects of explanatory variables on response variables.

3 Modeling clustered or repeated ordinal response data

We next review various strategies for modeling ordinal response variables when the data have some sort of clustering, such as with repeated measurement of subjects in a longitudinal study. The primary emphasis is on two classes of models. *Marginal models* describe so-called *population-averaged* effects which refer to an averaging over clusters at particular levels of predictors. *Subject-specific models* (also called *cluster-specific*) are *conditional models* that describe effects at the cluster level.

Denote T repeated responses in a cluster by $(Y_1, ...Y_T)$, thus regarding the response variable as multivariate. Although the notation here represents the same number of observations T in each cluster, the models and fitting algorithms apply also for the more general setting in which the number of observations can be different for each cluster (e.g., T_i in cluster i).

The marginal model with cumulative logit link has the form

$$\text{logit}[P(Y_t \le j \mid \mathbf{x}_t)] = \alpha_j - \boldsymbol{\beta}'\mathbf{x}_t,\ j = 1, ..., c-1,\ t = 1, ..., T, \qquad (3.1)$$

where $\mathbf{x}_t$ contains the values of the explanatory variables for observation t. This does not model the multivariate dependence among the T repeated responses, but focuses instead on the dependence of their T first-order marginal distributions on the explanatory variables.

3.1 Marginal models with a GEE approach

For ML fitting of this marginal model, it is awkward to maximize the log likelihood function, which results from a product of multinomial distributions from the various predictor levels, where each multinomial is defined for the c^T cells in the cross-classification of the T responses. The likelihood function refers to the complete joint distribution, so we cannot directly substitute the marginal model formula into the log likelihood. It is easier to apply a generalized estimating equations (GEE) method based on a multivariate generalization of quasi likelihood that specifies only the marginal regression models and a working guess for the correlation structure among the T responses, using the empirical dependence to adjust the standard errors accordingly.

The GEE methodology, originally specified for marginal models with univariate distributions such as the binomial and Poisson, extends to cumulative logit models (Lipsitz et al., 1994) and cumulative probit models (Toledano and Gatsonis, 1996) for repeated ordinal responses. Let $y_{it}(j) = 1$ if the response outcome for observation t in cluster i falls in category j. For instance, this may refer to the response at time t for subject i who makes T repeated observations in a longitudinal study, $(1 \le i \le N$, $1 \le t \le T)$. Let $\mathbf{y}_{it}$ be ($y_{it}(1)$, $y_{it}(2)$, …, $y_{it}(c-1)$). The covariance matrix for $\mathbf{y}_{it}$ is the one for a multinomial distribution. For each pair of categories (j_1, j_2), one selects a working correlation matrix for the pairs (t_1, t_2). Let $\boldsymbol{\beta}^* = (\boldsymbol{\beta}, \alpha_1, ..., \alpha_{c-1})$. The generalized estimating equations for estimating the model parameters take the form

$$u(\widehat{\boldsymbol{\beta}^*}) = \sum_{i=1}^{N} \widehat{\mathbf{D}}_i' \, \widehat{\mathbf{V}}_i^{-1} \, [\mathbf{y}_i - \widehat{\boldsymbol{\pi}}_i] = \mathbf{0}$$

where $\mathbf{y}_i = (\mathbf{y}_{i1}', \ldots, \mathbf{y}_{iT}')$ is the vector of observed responses for cluster i, $\boldsymbol{\pi}_i$ is the vector of probabilities associated with $\mathbf{Y}_i$, $\mathbf{D}_i' = \frac{\partial [\boldsymbol{\pi}_i]'}{\partial \boldsymbol{\beta}^*}$, $\mathbf{V}_i$ is the covariance matrix of $\mathbf{y}_i$, 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. See also Miller et al. (1993), Mark and Gail (1994), Heagerty and Zeger (1996), Williamson and Lee (1996), Huang et al. (2002) for marginal modeling of multiple ordinal measurements. In particular 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 weighted least squares approach developed for repeated categorical data by 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.

When marginal models are adopted, the association 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, such as a common local or cumulative odds ratio, rather than to expend much effort modeling it. When the association structure is itself of interest, a GEE2 approach is available for modeling associations using global odds ratios (Heagerty and Zeger, 1996).

3.2 Marginal models with a ML approach

As mentioned above, fitting marginal models is awkward with maximum likelihood. Algorithms have been proposed utilizing a multivariate logistic model that has a one-to-one correspondence between joint cell probabilities and parameters of marginal models as well as higher-order parameters of the joint distribution (Fitzmaurice and Laird, 1993; Glonek and McCullagh, 1995; Glonek, 1996). However, the correspondence is awkward for more than a few dimensions.

Another approach treats a marginal model as a set of constraint equations and uses methods of maximizing Poisson and multinomial likelihoods subject to constraints (Lang and Agresti, 1994; Lang, 1996, 2004). In these approaches, it is possible also to model simultaneously the joint distribution or higher-order marginal distributions. It is computationally intensive when T is large or when there are several predictors, especially if any of them is continuous. Recent theoretical and computational advances have made ML feasible for larger problems, both for constrained ML (Bergsma, 1997; Lang et al., 1997; Bergsma and Rudas, 2002; Lang, 2004) and for maximization with respect to joint probabilities expressed in terms of the marginal model parameters and an association model (Heumann, 1997).

3.3 Generalized linear mixed models

Instead of modeling the marginal distributions while treating the joint dependence structure as a nuisance, one can model the joint distribution also by using random effects for the clusters. The models have conditional interpretations with cluster-specific effects. When the response has distribution in the exponential family, generalized linear mixed models (GLMMs) add random effects to generalized linear models (GLMs). Similarly, the multivariate GLM defined in (2.3) can be extended to include random effects, giving a multivariate GLMM. In general, random effects in models can account for a variety of situations, including subject heterogeneity, unobserved covariates, and other forms of 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.

In contrast to the marginal cumulative logit model in (3.1), the model with random effects has form

$$\text{logit}[P(Y_{it} \leq j \mid \mathbf{x}_{it}, \mathbf{z}_{it})] = \alpha_j - \boldsymbol{\beta}'\mathbf{x}_{it} - \mathbf{u}_i'\mathbf{z}_{it}, \quad j = 1, ..., c-1,\ t = 1, ..., T,\ i = 1, ..., N, \tag{3.2}$$

where $\mathbf{z}_{it}$ refers to a vector of explanatory variables for the random effects and $\mathbf{u}_i$ are *iid* from a multivariate $N(\mathbf{0},\ \boldsymbol{\Sigma})$. See, for instance, Tutz and Hennevogl (1996). The simplest case takes $\mathbf{z}_{it}$ to be a vector of 1's and $\mathbf{u}_i$ to be a single random effect from a $N(0,\ \sigma^2)$ distribution. The form (3.2) can be extended to other ordinal models using different link functions (Hartzel et al., 2001a) as well as continuation-ratio logit models (Coull and Agresti, 2000).

When there more than a couple of terms in the vector of random effects, ML model fitting can be challenging. An early approach (Harville and Mee, 1984) used best linear unbiased prediction of parameters of an underlying continuous model. Since the random effects are unobserved, to obtain the likelihood function we construct the usual product of multinomials that would apply if they were known and then integrate out the random effects. Except in rare cases (such as the complementary log-log link with the log of a gamma or inverse Gaussian distribution for the random effects; see Crouchley (1995); Ten Have (1996)), this integral does not have closed form and it is necessary to use some approximation for the likelihood function. We can then maximize the approximated likelihood using a variety of standard methods.

Here, we briefly review algorithmic approaches for approximating the integral that determines the likelihood. For simple models such as random intercept models, straightforward Gauss-Hermite quadrature is usually adequate (Hedeker and Gibbons, 1994, 1996). An adaptive version of Gauss-Hermite quadrature (e.g. Liu and Pierce, 1994; Pinheiro and 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. For models with higher-dimensional integrals, more feasible methods use Monte Carlo methods, which use the randomly sampled nodes to approximate integrals. Booth and 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.

Alternatively, pseudo-likelihood methods avoid the intractable integral completely, making computation simpler (Breslow and Clayton, 1993). These methods are biased for highly non-normal cases, such as Bernoulli response data (Breslow and Clayton, 1993; Engel, 1998), with the bias increasing as variance components increase. We suspect that similar problems exist for the multinomial random effects models, both for estimating regression coefficients and variance components, when the multinomial sample sizes are small.

The form (3.2) assuming multivariate normality for the random effects has the possibility of a misspecified random effects distribution. Hartzel et al. (2001a) used a semi-parametric approach for ordinal models. The method applies an EM algorithm to obtain nonparametric ML estimates by treating the random effects as having a distribution with a set of mass points with unspecified locations and probabilities.

One application of the multivariate GLMM is to describe the degree of heterogeneity across stratified $r \times c$ tables. Hartzel et al. (2001b) used cumulative and adjacent-categories logit mixed models by treating the true stratum-specific ordinal log odds ratios as a sample with some unknown mean and standard deviation. It is natural to use random effects models when the levels of stratum variable are a sample, such as in many multi-center clinical trials. See Jaffrézic et al. (1999) for a model based on latent variables with heterogeneous variances.

3.4 Multi-level models

The models discussed so far in the section accommodate *two-level* problems such as repeated ordinal responses within subjects, for which subjects and clusters are the two levels. In practice, a study might involve more than one level of clustering, such as in many educational applications (e.g., students nested within schools which are nested within districts or a geographical region).

Possible approaches to handling multi-level cases include marginal modeling (typically implemented with the GEE method), multivariate GLMMs, and a Bayesian approach through specifying prior distributions at different levels of a hierarchical model. Like subject-specific two-level models, a GLMM can describe correlations using random effects for the subjects within clusters as well as for the clusters themselves. The ML approach maximizes a marginal likelihood by integrating the conditional likelihood over the distribution of the random effects. Again, such an approach requires numerical integration, which is typically difficult computationally for a high-dimensional random effect structure.

Qu et al. (1995) used a GEE approach to estimate a marginal model for ordinal responses. The correlation among the ordinal responses on the same subject was modelled through the correlation of assumed underlying continuous variables. For GLMM approaches, see Hedeker and Gibbons (1994) and Fielding (1999). For Bayesian approaches, see Tan et al. (1999), Chen and Dey (2000) and Qiu et al. (2002).

3.5 Other models (Transitional models and time series)

Another type of model to analyze repeated measurement data, called a *transitional model*, describes the distribution of a response conditional on past responses and explanatory variables. Many transitional models have Markov chain structure taking into account the time ordering, which is often useful in modeling time series data. This approach has received substantial attention for binary data (e.g. Bonney, 1986). Ekholm et al. (2003) proposed ordinal models in which the association between repeated responses is characterized by *dependence ratios* (Ekholm et al., 1995), which in a given cell equals the cell probability divided by its expected value under independence. Their model is a GLMM using random effects for the subjects when the association mechanism is purely exchangeable, and it is a transitional model with Markov chain structure when the association mechanism is purely Markov. See Kosorok and Chao (1996) for GEE and ML approaches with a transitional model in continuous time.

The choice among marginal (population-averaged), cluster-specific, and transitional models depends on whether one prefers interpretations to apply at the population or the subject level and on whether it is sensible to describe effects of explanatory variables conditional on previous responses. Cluster-specific models are especially useful for describing within-cluster effects, such as within-subject comparisons in a crossover study. Marginal models may be adequate if one is interested in summarizing responses for different groups (e.g., gender, race) without expending much effort on modeling the dependence structure. Different model types have different sizes for parameters for the effects. For instance, effects in a cluster-specific model are larger in magnitude than those in a population-averaged model. For a more detailed survey of methods for clustered ordered categorical data, see Agresti and Natarajan (2001).

4 Bayesian analyses for ordinal responses

In this section, we'll first discuss the Bayesian approach to estimating multinomial parameters that is relevant for univariate ordinal data, mentioning a couple of particular types of applications. Then we'll consider estimating probabilities in contingency tables when at least one variable is ordinal. Finally, we'll discuss more recent literature on the modeling of ordinal response variables, including multivariate ordinal responses.

4.1 Estimating multinomial parameters

With c categories in a single multinomial (ignoring explanatory variables, at first), suppose cell counts $(n_1, \ldots, n_c)$ have a multinomial distribution with $n = \sum n_i$ and parameters $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_c)'$. Let $\{p_i = n_i/n\}$ be the sample proportions. The conjugate density for the multinomial probability mass function is the Dirichlet, expressed in terms of gamma functions as

$$g(\boldsymbol{\pi}) = \frac{\Gamma\left(\sum \alpha_i\right)}{\left[\prod_i \Gamma(\alpha_i)\right]} \prod_{i=1}^{c} \pi_i^{\alpha_i - 1} \quad \text{for } 0 < \pi_i < 1 \text{ all } i, \quad \sum_i \pi_i = 1,$$

where $\{\alpha_i > 0\}$. Let $K = \sum \alpha_j$. The Dirichlet has $E(\pi_i) = \alpha_i/K$ and $\mathrm{Var}(\pi_i) = \alpha_i(K - \alpha_i)/[K^2(K+1)]$. The posterior density is also Dirichlet, with parameters $\{n_i + \alpha_i\}$, so the posterior mean is

$$E(\pi_i | n_1, \ldots, n_c) = (n_i + \alpha_i)/(n + K).$$

Let $\gamma_i = E(\pi_i) = \alpha_i/K$. This Bayesian estimator equals the weighted average

$$[n/(n+K)]p_i + [K/(n+K)]\gamma_i,$$

which is the sample proportion when the prior information corresponds to K trials with α_i outcomes of type i, $i = 1, \ldots, c$.

Good (1965) used this approach of smoothing sample proportions in estimating multinomial probabilities. Such smoothing ignores any ordering of the categories. Sedransk et al. (1985) considered Bayesian estimation of a set of multinomial probabilities under the constraint $\pi_1 \leq \ldots \leq \pi_k \geq \pi_{k+1} \geq .. \geq \pi_c$ that is sometimes relevant for ordered categories. They used a truncated Dirichlet prior together with a prior on k if it is unknown.

The Dirichlet distribution is restricted by having relatively few parameters. For instance, one can specify the means through the choice of $\{\gamma_i\}$ and the variances through the choice of K, but then there is no freedom to alter the correlations. As an alternative to the Dirichlet, Leonard (1973) proposed using a multivariate normal prior distribution for multinomial logits. This induces a multivariate logistic-normal distribution for the multinomial parameters. Specifically, if $\mathbf{X} = (X_1, \ldots, X_c)$ has a multivariate normal distribution, then $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_c)$ with $\pi_i = \exp(X_i)/\sum_{j=1}^{c} \exp(X_j)$ has the logistic-normal distribution. This can provide extra flexibility. For instance, when the categories are ordered and one expects similarity of probabilities in adjacent categories, one might use an autoregressive form for the normal correlation matrix. Leonard (1973) suggested this approach in estimating a histogram.

An alternative way to obtain more flexibility with the prior specification is to use a hierarchical approach of specifying distributions for the Dirichlet parameters. This approach treats the $\{\alpha_i\}$ in the Dirichlet prior as unknown and specifies a second-stage prior for them. These approaches gain greater generality at the expense of giving up the simple conjugate Dirichlet form for the posterior. Once one departs from the conjugate case, there are advantages of computation and of ease of more general hierarchical structure by using a multivariate normal prior for logits, as in Leonard (1973).

4.2 Estimating probabilities in contingency tables

Good (1965) used Dirichlet priors and the corresponding hierarchical approach to estimate cell probabilities in contingency tables. In considering such data here, our notation will refer to two-way $r \times c$ tables with cell counts $\mathbf{n} = \{n_{ij}\}$ and probabilities $\boldsymbol{\pi} = \{\pi_{ij}\}$, but the ideas extend to any dimension. A Bayesian approach can compromise between sample proportions and model-based estimators. A Bayesian estimator can shrink the sample proportions $\mathbf{p} = \{p_{ij}\}$ toward a set of proportions satisfying a model.

Fienberg and Holland (1972, 1973) proposed estimates of $\{\pi_{ij}\}$ using data-dependent priors. For a particular choice of Dirichlet means $\{\gamma_{ij}\}$ for the Bayesian estimator

$$[n/(n+K)]p_{ij} + [K/(n+K)]\gamma_{ij},$$

they showed that the minimum total mean squared error occurs when

$$K = \left(1 - \sum \pi_{ij}^2\right) / \left[\sum (\gamma_{ij} - \pi_{ij})^2\right].$$

The optimal $K = K(\boldsymbol{\gamma}, \boldsymbol{\pi})$ depends on $\boldsymbol{\pi}$, and they used the estimate $K(\boldsymbol{\gamma}, \mathbf{p})$ plugging in the sample proportion $\mathbf{p}$. As $\mathbf{p}$ falls closer to the prior guess $\boldsymbol{\gamma}$, $K(\boldsymbol{\gamma}, \mathbf{p})$ increases and the prior guess receives more weight in the posterior estimate. They selected $\{\gamma_{ij}\}$ based on the fit of a simple model. For two-way tables, they used the independence fit $\{\gamma_{ij} = p_{i+}p_{+j}\}$ for the sample marginal proportions. When the categories are ordered, improved performance usually results from using the fit of an ordinal model. Agresti and Chuang (1989) considered this by adding a linear-by-linear association term to the independence model.

Rather than focusing directly on estimating probabilities, with prior distributions specified in terms of them, one could instead focus on association parameters. Evans et al. (1993) considered Goodman's RC model generalization of the independence model that has multiplicative row and column effects. This has linear-by-linear association structure, but with scores treated as parameters. Based on independent normal priors (with large variances) for the loglinear parameters for the saturated model, they used a posterior distribution for loglinear parameters to induce a marginal posterior distribution on the RC sub-model. As a by-product, this yields a statistic based on the posterior expected distance between the RC association structure and the general loglinear association structure in order to check how well the RC model fits the data. Computations use Monte Carlo with an adaptive importance sampling algorithm. For recent work on the more general $RC(M^*)$ model, see Kateri et al. (2005). They addressed the issue of determining the order of association M^*, and they checked the fit by evaluating the posterior distribution of the distance of the model from the full model.

4.3 Modeling ordinal responses

Johnson and Albert (1999) focused on Bayesian approaches to modeling ordinal response variables. Their main focus was on cumulative link models. Specification of priors is not simple, and they used an approach that specifies beta prior distributions for the cumulative probabilities at several values of the explanatory variables (e.g., see p. 133). They fitted the model using a hybrid Metropolis-Hastings/Gibbs sampler that recognizes an ordering constraint on the intercept $\{\alpha_j\}$ parameters. Among special cases, they considered an ordinal extension of the item response model.

Chipman and Hamada (1996) used the cumulative probit model but with a normal prior defined directly on $\boldsymbol{\beta}$ and a truncated ordered normal prior for the $\{\alpha_j\}$, implementing it with the Gibbs sampler. They illustrated with two industrial data sets. For binary and ordinal regression, Lang (1999) used a parametric link function based on smooth mixtures of two extreme value distributions and a logistic distribution. His model used a flat, non-informative prior for the regression parameters, and was designed for applications in which there is some prior information about the appropriate link function.

Bayesian ordinal models have been used for various applications. For instance, Johnson (1996) proposed a Bayesian model for agreement in which several judges provide ordinal ratings of items, a particular application being test grading. Like Albert and Chib (1993), Johnson assumed that for a given item, a normal latent variable underlies the categorical rating. For a given judge, cutpoints define boundary points for the categories. He suggested uniform priors over the real line for the cutpoints, truncated by their ordering constraints. The model is used to regress the latent variables for the items on covariates in order to compare the performance of raters. For other Bayesian analyses with ordinal data, see Cowles et al. (1996), Bradlow and Zaslavsky (1999), Tan et al. (1999), Ishwaran and Gatsonis (2000), Ishwaran (2000), Xie et al. (2000), Rossi et al. (2001), and Biswas and Das (2002).

Recent work has focused on multivariate response extensions, such as is useful for repeated measurement and other forms of clustered data. For modeling multivariate correlated ordinal responses, Chib and Greenberg (1998) considered a multivariate probit model. A multivariate normal latent random vector with cutpoints along the real line defines the categories of the observed discrete variables. The correlation among the categorical responses is induced through the covariance matrix for the underlying latent variables. See also Chib (2000). Webb and Forster (2004) parameterized the model in such a way that conditional posterior distributions are standard and easily simulated. They focused on model determination through comparing posterior marginal probabilities of the model given the data (integrating out the parameters). See also Chen and Shao (1999), who also briefly reviewed other Bayesian approaches to handling such data. They employed a scale mixture of multivariate normal links, a class of models that includes the multivariate probit, t-link, and logit. Chen and Shao offered both a noninformative and an informative prior and gave conditions which ensure that the posterior is proper. Tan et al. (1999) generalized the methodology in Qu and Tan (1998) to propose a Bayesian hierarchical proportional odds model with a complex random effect structure for three-level data, allowing each cluster to have a different number of observations. They used a Markov chain Monte Carlo (MCMC) fitting method that avoids numerical integration to estimate the fixed effects and the correlation parameters.

Finally, frequentist-based smoothing methods often can be given a Bayesian interpretation, through relating a penalty function to a prior distribution. For smoothing methods for ordinal data, see Titterington and Bowman (1985), Simonoff (1987), and Dong and Simonoff (1995). For a survey of Bayesian inference for categorical data, see Agresti and Hitchcock (2004).

5 Non-model based methods for ordinal responses

Over the years a variety of methods have been proposed that provide description and inference for ordinal data seemingly without any connection to modeling. These include large-sample and small-sample tests of independence and conditional independence, including ones based on randomization arguments, methods based on inequality constraints, and nonparametric-type methods based on ranks. In some cases, however, the methods have close connections with types of models summarized above.

5.1 CMH methods for stratified contingency tables

When the main focus is analyzing the association between two categorical variables X and Y while controlling for a third variable Z, it is common to display data using a three-way contingency table. Questions that occur for such data include: Are X and Y conditionally independent given Z? If they are conditionally dependent, how strong is the conditional association between X and Y? Does that conditional association vary across the levels of Z? Of course, such questions can be addressed with the models discussed in Section 2, but here we briefly discuss other approaches.

The most popular non-model-based approach of testing conditional independence for stratified data are generalized Cochran–Mantel–Haenszel (CMH) tests. These are generalizations of the test originally proposed for two groups and a binary response (Mantel and Haenszel, 1959). By treating X and Y as either nominal or ordinal, Landis et al. (1978) proposed generalized CMH statistics for stratified $r \times c$ tables that combine information from the strata. For instance, when both X and Y are ordinal, a chi-squared statistic having $df = 1$ summarizes correlation information between two variables, combined over strata. It detects a linear trend in the effect, for fixed or rank-based scores assigned to the rows and columns. The test works well when the X and Y associations are similar in each stratum. Also, when one expects the true associations to be similar, taking this into account results in a more powerful test. There is a close connection between the generalized CMH tests and different multinomial logit models, in which the generalized CMH tests occur as score tests. See Agresti (2002, Chapter 7) for details.

In addition to the test statistics, related estimators are available for various ordinal odds ratios discussed in the next subsection, under the assumption that they are constant across the strata (Liu and Agresti, 1996; Liu, 2003). When each stratum has a large sample, such estimators are similar to ML estimators based on multinomial logit models. However, when the data are very sparse, such as when the number of strata grows with the sample size, these estimators (like the classic Mantel-Haenszel estimator for stratified 2×2 tables) are superior to ML estimators. Landis et al. (1998) and Stokes et al. (2000) reviewed CMH methods.

5.2 Ordinal odds ratios

For two-way tables, various types of odds ratios apply for ordinal responses. *Local odds ratios* apply to 2×2 tables consisting of pairs of adjacent rows and adjacent columns. *Global odds ratios* apply when both variables are collapsed to a dichotomy. *Local-global odds ratios* apply for a mixture, using an adjacent pair of variables for one variable and a collapsed dichotomy for the other one. *Continuation odds ratios* use pairs of adjacent categories for one variable, and a fixed category contrasted with all the higher categories (or all the lower categories) for the other variable. For a survey of types of odds ratios for ordinal data, see Douglas et al. (1991).

Local odds ratios arise naturally in loglinear models and in adjacent-category logit models (Goodman, 1979, 1983). Local-global odds ratios arise naturally in cumulative logit models, using the global dimension for the ordinal response (McCullagh, 1980). The global odds ratio does not relate naturally to models discussed in Section 2, but it is a sensible odds ratio for a bivariate ordinal response. For instance, in a stratified contingency table, let Y_k and X_k denote the column and row category respectively for a bivariate ordinal response for each subject in stratum k. The form of the global odds ratio assuming a constant association in each stratum is

$$\theta = \frac{P(Y_k \le j \mid X_k \le i)/P(Y_k > j \mid X_k \le i)}{P(Y_k \le j \mid X_k > i)/P(Y_k > j \mid X_k > i)}.$$

Pearson and Heron (1913), Plackett (1965), Wahrendorf (1980), Dale (1984, 1986), Molenberghs and Lesaffre (1994), Williamson et al. (1995), Glonek and McCullagh (1995), Glonek (1996), Williamson and Kim (1996), and Liu (2003) have discussed modeling using this type of odds ratio.

5.3 Rank-based approaches

We mentioned above how a generalized CMH statistic for testing conditional can use rank scores in summarizing an association. A related strategy, but not requiring any scores, bases the association and inference about it on a measure that strictly uses ordinal information. Examples are generalizations of *Kendall's tau* for contingency tables. These utilize the numbers of concordant and discordant pairs in summarizing information about an ordinal trend. A pair of observations is concordant if the member that ranks higher on X also ranks higher on Y and it is discordant if the member that ranks higher on X ranks lower on Y.

For instance, Goodman and Kruskal's gamma (Goodman and Kruskal, 1954) equals the difference between the proportion of concordant pairs and the proportion of discordant pairs, out of the untied pairs. The standard measures fall between -1 and +1, with value of 0 implied by independence. Such measures are also useful both for independent samples and for association and comparing distributions for matched pairs (e.g. Agresti, 1980, 1983). Formulas for standard errors of the extensions of Kendall's tau, derived using the delta method, are quite complex. However, they are available in standard software, such as SAS (PROC FREQ).

Other rank-based methods such as a version of the Jonckheere-Terpstra test for ordered categories are also available for testing independence for contingency table with ordinal responses. Chuang-Stein and Agresti (1997) summarized the use of them in detecting a monotone dose-response relationship. See also Cohen and Sackrowitz (1992). Agresti (1981) considered association between a nominal and ordinal variable. There is scope for exploring generalizations to ordered categorical data or recent work done in extending rank-based methods to repeated measurement and other forms of clustered data (e.g. Brunner et al., 1999).

5.4 Using inequality constraints

Another way to utilize ordered categories is to assume inequality constraints on parameters for those categories that describe association structure. For a two-way contingency table, for instance, we could use cumulative, local, global or continuation odds ratios to describe the association, imposing constraints on their values such as having all $(r-1)(c-1)$ log odds ratios be nonnegative. For instance, we could obtain ML estimates of cell probabilities subject to such a condition, and construct tests of independence, such as a likelihood-ratio (LR) test against this alternative. Bartolucci et al. (2001) proposed a general framework for fitting and testing models incorporating the constraint on different type of ordinal odds ratios. Among the various ordinal odds ratios, such a constraint on the local odds ratio is the most restrictive. For instance, uniformly nonnegative local log odds ratios imply uniformly nonnegative global log odds ratios.

When $r = 2$, Oh (1995) discussed estimation of cell probabilities under the nonnegative log odds ratio condition for different types of ordinal odds ratios. The asymptotic distributions for most of the LR tests are *chi-bar-squared*, based on a mixture of independent chi-squared random variables of form $\sum_{d=1}^{r} \rho_d \chi^2_{d-1}$, where χ^2_d is a chi-squared variable with d degrees of freedom (with $\chi^2_0 \equiv 0$), and $\{\rho_d\}$ is a set of probabilities. However, the chi-bar-squared approximations may not hold well for the general $r \times c$ case, especially for small samples and unbalanced data sets (Wang, 1996). Agresti and Coull (1998) suggested Monte Carlo simulation of exact conditional tests based on the LR test statistic. Bartolucci and Scaccia (2004) discussed the conditional approach of conditioning on the observed margins. They estimated the parameters of the models under constraints on various ordinal odds ratio by maximizing an estimate of the likelihood ratio based on a Monte Carlo approach. A P-value for the Pearson chi-squared statistics can be computed on the basis of Monte Carlo simulations when the observed table is sparse. See Agresti and Coull (2002) for a survey of literature on inequality-constrained approaches.

An alternative approach is to place ordering constraints on parameters in models. For instance, for a factor with ordered levels, one could replace unordered parameters $\{\beta_i\}$ by a monotonicity constraint on the effects such as $\beta_1 \leq \beta_2 \leq \ldots \leq \beta_r$. See, for instance, Agresti et al. (1987) for the row effects loglinear model and Ritov and Gilula (1991) for the multiplicative row and column effects (RC) model.

5.5 Measuring agreement

A substantial literature has evolved over the years on measuring agreement between ratings on ordinal scales. The weighted kappa measure is a generalization of kappa that gives weights to different types of disagreements (Spitzer et al., 1967). See Gonin et al. (2000) for this measure in a modeling context using GEE. For a variety of other measures, some using basic ideas of concordance and discordance useful for measuring ordinal association, see Svensson (1998a, 2000a,b). Other work on diagnostic agreement involves the use of ROC curves (e.g. Tosteson and Begg, 1988; Toledano and Gatsonis, 1996; Ishwaran and Gatsonis, 2000; Lui et al., 2004). A considerable literature also takes a modeling approach, such as using loglinear models (Agresti, 1988; Becker and Agresti, 1992; Rogel et al., 1998) or latent variable models (Uebersax and Grove, 1993; Uebersax, 1999) or Bayesian approaches (Johnson, 1996; Ishwaran, 2000).

6 Other issues

This final section briefly visits a variety of areas having some literature for ordinal response data.

6.1 Exact inference

With the various approaches described in previous sections, standard inference is based on large-sample asymptotics. For small-sample or sparse data, researchers have proposed alternative methods. Most of these use a conditional approach that eliminates nuisance parameters by conditioning on their sufficient statistics.

For instance, Fisher's exact test for 2×2 tables has been extended to $r \times c$ tables with ordered categories using an exact conditional distribution for a correlation-type summary based on row and column scores (Agresti et al., 1990). When it is computationally difficult to enumerate the complete conditional distribution, one can use Monte Carlo methods to closely approximate the exact P-value. Kim and Agresti (1997) generalized this for exact testing of conditional independence in stratified $r \times c$ tables. For ordinal variables a related approach uses test statistics obtained by expressing the alternative in terms of various types of monotone trends, such as uniformly nonnegative values of ordinal odds ratios of various types (Agresti and Coull, 1998). Bartolucci and Scaccia (2004) gave an alternative approach based on such a constraint.

Instead of testing independence or conditional independence, one could test the fit of a model. An exact goodness-of-fit test has been proposed for various loglinear and logit models using the MCMC method (Forster et al., 1996). Booth and Butler (1999) discussed alternative computational approaches using a general simulation method for exact tests of goodness of fit for various loglinear models. They noted that importance sampling breaks down for large values of *df*; in that case, MCMC methods seem to be the method of choice.

If the nuisance parameters of a model do not have reduced sufficient statistics (such as the cumulative link models), the conditional exact approach fails. In such a case, an unconditional approach that eliminates nuisance parameters using a "worst-case" scenario is applicable. The P-value is a tail probability of a test statistic maximized over all possible values for the nuisance parameters. However, it is a computational challenge for the model with several nuisance parameters. In general, "exact" methods lead to conservative inferences for hypothesis tests and confidence intervals, due to discreteness.

6.2 Missing data

Missing data are an all-too-common problem, especially in longitudinal studies. Little and Rubin (1987) distinguished among three possible missing data mechanisms. They classified *missing completely at random* (MCAR) to be a process in which the probability of missingness is completely independent of both unobserved and observed data. The process is *missing at random* (MAR) if conditional on the observed data, the missingness is independent of the unobserved measurements. They demonstrated that a likelihood-based analysis using the observed responses is valid only when the missing data process is either MCAR or MAR and when the distributions of observed data and missingness are separately parameterized.

A useful feature of multivariate GLMMs is their treatment of missing data when the missingness is either MCAR or MAR, because subjects who are missing at a given time point are not excluded from the analysis. For example, partial proportional odds models with cluster-specific random effects used in Hedeker and Mermelstein (2000) have this feature. They also suggested the use of pattern-mixture modeling (Little, 1995) for dealing with non-random missingness. However, random processes are not necessarily ignorable when non-likelihood-based methods, such as GEE are used (Mark and Gail, 1994). For instance, Kenward et al. (1994) provided an empirical illustration of the breakdown in the GEE method when the process is not MCAR.

With reference to ordinal data, the literature on missing data includes a score test of independence in two-way tables with extensions for stratified data with ordinal models (Lipsitz and Fitzmaurice, 1996), modelling nonrandom drop-out in a longitudinal study with ordinal responses (Molenberghs et al., 1997; Ten Have et al., 2000) and Bayesian tobit modeling in studies with longitudinal ordinal data (Cowles et al., 1996). In some cases, information may be missing in a key covariate rather than the responses. Toledano and Gatsonis (1999) encountered this in a study comparing two modalities for the staging of lung cancer. For transitional models, Miller et al. (2001) considered fitting of models with continuation ratio and cumulative logit links when the data have nonresponse.

6.3 Sample size and power

For the comparison of two groups on an ordinal response, Whitehead (1993) gave sample size formulas for the proportional odds models to achieve a particular power. This requires anticipating the c marginal response proportions as well as the size of the effect, based on an asymptotic approach to the hypothesis test. With equal marginal probabilities, the ratio of the sample size $N(c)$ needed when the response variable has c categories relative to the sample size $N(2)$ needed when it is binary is approximately

$$N(c)/N(2) = .75/[1 - 1/c^2].$$

Relative to a continuous response ($c = \infty$), using c categories provides efficiency $(1 - 1/c^2)$. The loss of information from collapsing to a binary response is substantial, but little gain results from using more than 4 or 5 categories.

A special case of comparing two-sample ordinal data with the Wilcoxon rank-sum statistic, Hilton and Mehta (1993) proposed a different approach of sample size determination by evaluating the exact conditional distribution using a network algorithm with simulation. Lee et al. (2002) compared the performance of the above methods based on asymptotic and exact approaches and provided guidelines.

Ohman-Strickland and Lu (2003) provided sample size calculations based on subject-specific models comparing two treatments, where the subjects are measured before and after receiving a treatment. They considered both cumulative and adjacent-categories logit models that contain a subject random effect, when the effect of interest is the treatment-by-time interaction term.

6.4 Software for analyzing ordinal data

For univariate ordinal response models, several software packages can fit cumulative link models via ML estimation. Examples are PROC LOGISTIC and PROC GENMOD in SAS (see Stokes et al., 2000). Bender and Benner (2000) showed the steps of fitting continuation-ratio logit models using PROC LOGISTIC, where the original data set is restructured by repeatedly including the data subset and two new variables. This allows model fitting with a common effect. Although S-Plus and R are widely used by statisticians, ordinal models are not part of standard versions. The package *VGAM* at `http://www.stat.auckland.ac.nz/~yee/VGAM/` can fit both cumulative logit and continuation-ratio logit models using ML methods. The package also fits *vector generalized linear* and *additive models* described in Yee and Wild (1996). We have not used it, but apparently STATA has capability of fitting ordinal regression models and related models with random effects, as well as marginal models.

For clustered or repeated ordinal responses data, a 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 including multi-level models for normal or other random effects distributions, all for a variety of link functions. Currently, SAS offers the greatest scope of methods for repeated categorical data (see the *SAS/STAT User's Guide* online manual at their website). The procedure GENMOD can perform GEE analyses (with independence working correlation) for marginal models using that family of links, including cumulative logit and probit (Stokes et al., 2000, Chapter 15). There is, however, no capability of ML fitting of marginal models. The approach of ML fitting of marginal models using constrained methods of maximization is available in S-plus and R functions available from Prof. J. B. Lang (Statistics Dept., Univ. of Iowa).

For cumulative link models containing random effects, one can use PROC NLMIXED in SAS. 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. See Hartzel et al. (2001b) and the SAS website mentioned above for examples. A FORTRAN program (MIXOR) is also available for cumulative logit models with random effects (Hedeker and Gibbons, 1996). It uses Gauss-Hermite numerical integration, but standard errors are based on expected information whereas NLMIXED uses observed information.

StatXact and LogXact, distributed by Cytel Software (Cambridge, Massachusetts), provide several exact methods to handle a variety of problems, including tests of independence in two-way tables with ordered or unordered categories, tests of conditional independence in stratified tables, and inferences for parameters in logistic regression. Exact inferences for multinomial logit models to handle ordinal responses are apparently available in LogXact, version 6. The models include adjacent-category and cumulative logit models.

6.5 Final comments

As this article has shown, the past quarter century has seen substantial developments in specialized methods for ordinal data. In the next quarter century, perhaps the main challenge is to make these methods better known to methodologists who commonly encounter ordinal data. Our hunch is that most methodologists still analyze such data by either assigning scores and using ordinary normal-theory methods or ignore the ordering and use standard methods for nominal variables (e.g., Pearson chi-squared test of independence for two-way contingency tables) despite the loss of power and parsimony that such an approach entails. Indeed, the only treatment of contingency tables in most introductory statistics books is of the Pearson chi-squared test.

So, has ordinal data analysis entered a state of maturity, or are there important problems yet to be addressed? We invite the discussants of this article to state their own beliefs on what is needed in the future, either for additional research or in order to make these methods more commonly used in practice. We also invite them to point out areas or key works that we neglected to mention.

DISCUSSION

Gerhard Tutz
Institut für Statistik
Ludwig-Maximilians-Universität München, Germany

I would like to congratulate the authors for this careful and clear overview on developments in the analysis of ordered categorical data, developments to which the authors themselves have contributed much. It certainly may be used by statisticians and practioners to find appropriate methods for their own analysis of ordinal data and inspire future research.

My commentary refers less to the excellent overview than to the use of ordinal regression models in practice. Since McCullagh (1980) seminal paper ordinal regression has become known and used in applied work in many areas. However, what is widely known as "the ordinal model" is the cumulative type model

$$P(Y \leq j|x) = F(\eta_j(x)) \quad \text{with the linear predictor} \quad \eta_j(x) = \alpha_j - x'\beta.$$

Also extensions, e.g. to mixed models are mostly based on this type of model while alternative link functions have been widely ignored outside the inner circle of statistician interested in ordinal regression models. I think that in particular the sequential type model has many advantages over the cumulative model and deserves to play a much stronger role in data analysis. The sequential type model may be written in the form

$$P(Y = j|Y \geq j, x) = F(\eta_j(x)), \eta_j(x) = \alpha_j - x'\beta$$

where F is a strictly monotone distribution function. If F is chosen as the logistic distribution function one obtains the continuation-ratio logit model. Since the model may be seen as modelling the (binary) transition from category r to category $r+1$, given category r has been reached, any model which is used in binary regression may be used. One of the main advantages of the sequential model is its potential for extensions if the basic model is not appropriate. If the cumulative model turns out to be too crude an approximation one might consider the more general predictor

$$\eta_j(x) = \alpha_j - x'\beta_j$$

which is more flexible by allowing for category-specific effects β_j. For the logistic cumulative link one obtains the non-proportional or partial proportional odds model. But, for the cumulative type models the restriction $\eta_1(x) \leq \cdots \leq \eta_{c-1}(x)$ implies severe restrictions on the parameters and the range of $x-$values where the models can hold. Even if estimates are found, for a new value x the plug-in of $\hat{\beta}_j$ might yield improper probabilities. Often, numerical problems occur when fitting the general cumulative type model. Even step-halving in iterative estimation procedures may fail to find estimates $\hat{\beta}_j$ which at least yield proper probabilities within the given sample. If estimates do not exist one has to rely upon the simpler model with global effect β. The problem is that inferences drawn from the proportional odds model may be misleading if the model is inappropriate (for an example, see Bender and Grouven, 1998).

The sequential model does not suffer from these drawbacks since no ordering of the predictor values is required. In addition, estimation for the simple and the more general model with category-specific effects may be performed by using fitting procedures for binary variables which are widely available. The widespread focus on the cumulative model is the more regrettable because problems with the restriction on parameters become more severe if one considers more complex models like marginal or mixed models.

A second commentary concerns the structuring of the predictor term in more complex models. The structuring should be flexible in order to fit the data well but should also be simple and interpretable. In the simplest case for repeated measurements in marginal and mixed models the jth predictor of observation t within cluster i has the basic form

$$\eta_{tj}(x_{it}) = \alpha_j - x'_{it}\beta$$

assuming that α_j and β are constant across observations $t = 1, \ldots, T$ (for the mixed model cluster-specific effects have to be included). In particular if the number of response categories c and/or T is not too small the assumption is rather restrictive. Although simple and interpretable models certainly should be preferred over flexible models, one may only find structures in data which are reflected by the model. Thus using models as devices to detect structures future research might consider flexible predictors like

$$\eta_{tj}(x_{it}) = \alpha_{tj} - x'_{it}\beta_j \quad \text{or} \quad \eta_{tj}(x_{it}) = \alpha_{tj} - x'_{it}\beta_t$$

which contain category-specific and observation-specific parameters. While the latter extension yields a varying-coefficient model, which is familiar at least for unidimensional response, the first extension is specific for ordinal responses. In both cases the number of parameters increases dramatically if c or T is large. In order to avoid the fitting of noise and obtain interpretable structures parameters usually have to be restricted. For example the variation of β_j can be restricted by using penalized maximum likelihood techniques when the usual log-likelihood l is replaced by a penalized likelihood

$$l_P = l - \sum_{j,s} \lambda_s(\beta_{j+1,s} - \beta_{j,s})^2.$$

with smoothing parameters $\lambda_s,\ s = 1,...,c-1$. For the cumulative type model where additional restrictions have to be taken into account the penalized likelihood yields models between the proportional and non-proportional model, these models resulting as extreme cases. If λ_s is very large a proportional odds model is fitted for the sth component of the vector x_{it}. Thus existence of estimates may be reached by modelling closer to the proportional odds model. Of course the threshold parameters α_{tj} have to be restricted across $t = 1, ..., T$ and $j = 1, ..., c-1$. Extensions might also refer to smooth structuring of the predictor. Additive structures are an interesting alternative to the linear predictor and have been used in the additive ordinal regression model for cross-sectional data. In the repeated measurements case it is certainly challenging to find smooth structures which vary across t (or j) and still are simple and interpretable.

A further remark concerns the mean response model. Although it has to be mentioned in an overview because it may be attractive for practioners in my opinion it should not be considered an ordinal model. By assigning scores to categories one obtains a discrete response model which treats the responses as metrically scaled. If effects are summarized in terms of location these summaries refer to the assigned scores. If these are artificial for ordered categories, so will be the results since different scores will yield different effects.

An issue which is related to the analysis of ordinal data but seems to have been neglected in research is prediction of ordinal outcomes. There is a huge body of papers on prediction of binary or categorical outcomes, mostly within the framework of classification. It is surprising that the ordering of classes has hardly been considered in discriminant analysis or machine learning. The approach to use simple regression if the number of response categories is large and classification methods if it is small is common but certainly inappropriate. Since models that work well when used to identify and quantify the influence of explanatory variables are not necessarily the best models in prediction it seems worthwhile to develop instruments for the prediction of ordered outcomes which make use of the order information by using modern instruments of statistical learning theory as given for example in Hastie et al. (2001).

Jeffrey S. Simonoff
Leonard N. Stern School of Business
New York University, USA

It is my pleasure to comment on this article. As Liu and Agresti note in Sections 2.4 and 6.5 of the paper, it is probably true that most methodologists still analyze ordinal categorical data by either assuming numerical scores and using Gaussian-based methods (ignoring the categorical nature of the data), or using categorical data methods that ignore the ordering in the data. The latter strategy obviously throws away important information, while the former (among other weaknesses) doesn't answer what seems to me to be the fundamental regression question: what is $P(Y = j)$ for given predictor values? Hopefully this excellent article will help encourage data analysts to use methods that are both informative and appropriate for ordinal categorical data.

Liu and Agresti focus on parametric models. I would like to say a little more about a subject I've been interested in for more than 25 years, nonparametric and semiparametric (smoothing-based) approaches to ordinal data, which are only briefly mentioned in Sections 2.4 and 4.3. More details and further references can be found in Simonoff (1996, Chapter 6), Simonoff (1998), and Simonoff and Tutz (2001), including discussion of computational issues.

I will focus here on local likelihood approaches. Nonparametric categorical data smoothing is based on weakening global parametric assumptions by hypothesizing simple (polynomial) relationships locally. To keep things simple, consider the situation with a single predictor. Let $x_1 < \cdots < x_g$ be the distinct values of the predictor variable; at measurement point x_i there are m_i objects distributed multinomially $\mathbf{n}_i \sim M(m_i, \mathbf{p}_i)$. The log-likelihood function is then

$$L = \sum_{i=1}^{g} \left[\sum_{j=1}^{c-1} n_{ij} \log \left(\frac{p_j(x_i)}{1 - \sum_{\ell=1}^{c-1} p_\ell} \right) + m_i \log \left(1 - \sum_{\ell=1}^{c-1} p_\ell(x_i) \right) \right] \tag{1}$$

$$= \sum_{i=1}^{g} \left\{ \sum_{j=1}^{c-1} n_{ij} \theta_j(x_i) - m_i \log \left[1 + \sum_{\ell=1}^{c-1} \exp \theta_\ell(x_i) \right] \right\}, \tag{2}$$

where

$$p_j = \frac{\exp \theta_j}{1 + \sum_{\ell=1}^{c-1} \exp \theta_\ell} \tag{3a}$$

for $j = 1, \ldots, c-1$, and

$$p_c = \frac{1}{1 + \sum_{\ell=1}^{c-1} \exp \theta_\ell}. \tag{3b}$$

The local (linear) likelihood estimate of p_s at x maximizes the local log-likelihood, which replaces θ_j in (2) with a linear function

$$\beta_{0j} + \beta_{xj}(x - x_i) + \beta_{yj} \left(\frac{s}{c} - \frac{j}{c} \right), \tag{4}$$

and weights each term in the outer brackets by a product kernel function $K_{1,h_1}(s/c, j/c) K_{2,h_2}(x, x_i)$, where $K_{q,h}(a,b) = K_q[(a-b)/h]$ (for $q = 1, 2$), K_1 and K_2 are continuous, symmetric unimodal density functions, and h_1 and h_2 are smoothing parameters that control the amount of smoothing over the response and predictor variables, respectively. The estimate $\hat{p}_s(x)$ then substitutes $\hat{\beta}_{0s}$ for θ_s and $\hat{\beta}_{0\ell}$ for θ_ℓ into (3a) or (3b). This scheme is based on the idea that the probability vector $\mathbf{p}(x)$ is smooth, in the sense that nearby values of x imply similar probabilities, and nearby categories of the response are similar.

This formulation can be generalized to allow for higher order local polynomials and multiple predictors by adding the appropriate polynomial terms to (4); multiple predictors also can be handled using separate smooth terms through generalized additive modeling (Hastie and Tibshirani, 1990). It also can be adapted to $r \times c$ contingency tables with ordered categories (based on the Poisson likelihood) by taking x to be a predictor that takes on the values $\{1, \ldots, r\}$, providing an alternative to the models described in Liu and Agresti's Section 4.2. Such estimates are particularly effective for large sparse tables, where standard methods can fail. There is also a close connection here to nonparametric density estimation. If the table is viewed as a binning of a (possibly multidimensional) continuous random variable, the local likelihood probability estimates (suitably standardized) become indistinguishable from the local likelihood density estimates of Hjort and Jones (1996) and Loader (1996) as the number of categories becomes larger and the bins narrow.

These methods are completely nonparametric, in that the only assumption made is that the underlying probabilities are smooth. If it is believed that a parametric model provides a useful representation of the data, at least locally, a semiparametric approach of model-based smoothing can be more effective. Consider, for example, the proportional odds model, once again (to keep things simple) with a single predictor. Local likelihood estimation at x is again based on weighting the terms in the outer brackets in (1) or (2) using a kernel function (note that now the likelihood would incorporate the proportional odds restriction on $\mathbf{p}$), now substituting a polynomial for $\text{logit}[P(Y \leq j|x)]$,

$$\text{logit}[P(Y \leq j|x)] = \gamma_{0j} + \gamma_1(x - x_i) + \cdots + \gamma_t(x - x_i)^t.$$

The estimated probability $\hat{p}_j(x)$ then satisfies

$$\hat{p}_j(x) = \frac{\exp(\hat{\gamma}_{0j})}{1 + \exp(\hat{\gamma}_{0j})} - \frac{\exp(\hat{\gamma}_{0,j-1})}{1 + \exp(\hat{\gamma}_{0,j-1})}$$

(since it is the cumulative logit that is being modeled here). This generalizes to multiple predictors by including the appropriate polynomial terms or by using separate additive smooth terms. Thus, the model-based estimate smooths over predictors in the usual (distance-related) way, but smooths over the response based on a nonparametric smooth curve for the cumulative logit relationship. This provides a simple way of assessing the appropriateness of the linear assumption in the proportional odds model, as the nonparametric curve(s) can be plotted and compared to a linear relationship.

Finally, I would like to echo Liu and Agresti's hope for increased use of ordinal categorical data models in the future, meaning parametric, but also nonparametric and semiparametric, approaches. The latter methodologies are ideally suited to the easy availability of high quality graphics in packages like R and S-PLUS, providing useful diagnostic checks for parametric model assumptions. It has been my experience that both students and researchers appreciate the advantages of ordinal categorical data models over Gaussian-based analyses when they are exposed to them, so perhaps merely "getting the word out" can go a long way to increasing their use. I congratulate the editors for inviting this paper and discussion, helping to do just that.

Maria Kateri
Department of Statistics and Insurance Science
University of Piraeus, Greece

First of all I would like to congratulate Ivy Liu and Alan Agresti for their interesting overview and exhaustive survey. In a comprehensive paper they managed to cover the developments of ordered categorical data analysis to various directions.

Before expressing my thoughts for possible future directions of research and ways to familiarize greater audience with methods for ordinal data, as asked by the authors, I will shortly refer to the analysis of square ordinal tables (not mentioned in the review) and comment on the modelling of association in contingency tables, which has not shared the popularity of regression type models, as also noted by Liu and Agresti (Section 2.5).

Models for square ordinal contingency tables

The well known models of symmetry (S), quasi symmetry (QS) and marginal homogeneity (MH) are applicable to square tables with commensurable classification variables. In case the classification variables are ordinal, the class of appropriate models is enlarged with most representative the conditional symmetry or triangular asymmetry model (cf. McCullagh, 1978) and the model of diagonal asymmetry (D). For a detailed review see Goodman (1985). The Bayesian analysis of some of these models is provided by Forster (2001).

These models are of special interest with nice interpretational aspects. They are connected with measuring agreement and change (Von Eye and Spiel, 1996) as well as with mobility tables (cf. Lawal, 2004). When the contingency table of interest is the transition probability matrix of a Markov chain, then the models of QS, MH and D asymmetry are related to "reversibility", "equilibrium state" and "random walk" respectively (cf. Lindsey, 1999). In this context, the mover-stayer model is also worth mentioning.

Modelling association in contingency tables

The main qualitative difference between association and correlation models is that although both of them are models of dependence, the first are (under certain conditions) the closest to independence in terms of the Kullback-Leibler distance, while the later in terms of the Pearsonian distance (Gilula et al., 1988). Rom and Sarkar (1992), Kateri and Papaioannou (1994) and Goodman (1996) introduced general classes of dependence models which express the departure from independence in terms of generalized measures and include the association and correlation models as special cases. Kateri and Papaioannou (1997) proved a similar property for the QS model. It is (under certain conditions) the closest model to complete symmetry in terms of the Kullback-Leibler distance. Using the Pearsonian distance, the Pearsonian QS model is introduced and further a general class of quasi symmetric models is created, through the f-divergence, which includes the standard QS as special case.

An interesting issue in the analysis of contingency tables, related to ordinality of the classification variables, is that of collapsing rows or/and columns of the table. The predominant criteria for collapsing (or not) are those of homogeneity (cf. Gilula, 1986; Wermuth and Cox, 1998) and structure (Goodman, 1981, 1985). Kateri and Iliopoulos (2003) proved that these two criteria, for which it was believed that they can sometimes be contradictory (Goodman, 1981; Gilula, 1986), are always in agreement.

The future of ordinal data analysis

According to my opinion, the future of research on ordinal categorical data lies mainly on the analysis of repeated measures or more general correlated data. Although the developments in the area the last two decades are impressing, as it is also evidenced by the review of Liu and Agresti, the available methodology is inferior compared to that of uncorrelated data and there are still directions and topics open to further research. Some of them are quoted briefly next.

Relatively little work has been done on the design of studies with ordinal repeated data, particularly for sample size and power calculations (Rabbee et al., 2003). The analysis of multilevel models could also be further developed. Undoubtfully, models for ordinal longitudinal responses can be complicated and the corresponding estimation procedures not straighforward. In the context of marginal models, the regression parameters are estimated mainly by the generalized estimation equations (GEE), based on a "working" correlation matrix. For a presentation of the existing GEE approaches and an outline of their features and drawbacks see also Sutradhar (2003). The estimation can be hard for random effect models as well, especially as the number of random components increases. Hence, the estimation procedures for the parameters of such complex models could be improved and the properties of the estimators further investigated. An alternative is the use of a non-parametric random effects approach based on latent class or finite mixture modelling (see Vermunt and Hagenaars, 2004, and the references therein). I expect non-parametric techniques to see growth, especially for high-dimensional problems. Non-parametric procedures have then to be compared to the corresponding parametric ones in terms of power and robustness.

I belong to those who are convinced that the future of Statistics is Bayesian, thus categorical data analysis will certainly be further developed towards this direction. Bayesian procedures are also tractable because they can derive relatively easy small sample inference.

Finally, I believe that methods for ordinal categorical data analysis are not popular and rarely used in practice mainly due to two reasons. First, these methods are not taught in most of the undergraduate programs on Statistics and secondly there are not user-friendly procedures available for most of them. It is thus crucial the development of a unified user-friendly software for the analysis of ordinal categorical data. This would also facilitate the teaching of these methods at a more introductory level of studies. The setting up of an association or a forum for categorical data analysis with regular meetings and an on-line newsletter would help further towards this direction.

Emmanuel Lesaffre
Biostatistical Centre
Catholic University of Leuven, Belgium

The authors are to be congratulated for an excellent overview of methods to analyze ordered categorical data, covering early and more recent approaches. I will use the last paragraph of the authors' paper to structure my comments. Hence I will reflect on (1) the use of methods for ordinal data in practice, (2) neglected research areas and (3) future areas of research.

The use of methods for ordinal data in practice

It strikes me that in my personal statistical consultancy that for ordinal responses I often end up using an ordinary binary logistic regression in stead of an ordinal logistic regression. The reason is that my co-workers (most often medical doctors) understand somewhat better the output of a binary logistic regression than that of an ordinal logistic regression and that very often the results of an ordinal logistic regression are only marginally more significant than those of a binary logistic regression. The latter seems to be often the case. Clearly, this is a purely personal and most likely very incomplete observation, but perhaps this is also the observation of other applied statisticians. I am curious to hear what the experience of the authors is in this respect and whether they can pinpoint simulation and/or practical studies which clearly show the benefit of ordinal logistic over binary logistic regression. Further, the problem with all new methods is the absence of software. What can we do about this? Statistical journals could motivate the authors to submit along with their paper also their software. But, I am afraid this is not enough to get statistical methods used in practice, the large software houses (e.g. SAS) needs to be convinced for their usefulness.

Neglected research areas

With such an extensive overview it is hard to find areas that were not covered. But it surprised me somewhat that the stereotype model of Anderson (1984) has not been mentioned. This model has been suggested for the analysis of "assessed" ordered categorical variables, such as the "extent of pain relief" scored by medical doctors as "worse" to "complete relief". The model allows checking whether the relationship of the (ordered) response with the regression vector is indeed ordered. Anderson concludes in his introduction that "There is no merit in fitting an ordered relationship as a routine, simply because the response is ordered." Therefore, Anderson suggested modeling the relationship of the "assessed" ordinal response y (ranging from 1 to k) as follows:

$$P(Y = s) = \frac{\exp\left(\beta_{0s} - \phi_s \boldsymbol{\beta}^T \mathbf{x}\right)}{\sum_{t=1}^{k} \exp\left(\beta_{0t} - \phi_t \boldsymbol{\beta}^T \mathbf{x}\right)} \tag{1}$$

where $1 = \phi_1 > \phi_2 > \ldots > \phi_k = 0$. Model (1) is a special case of a multinomial logistic regression model. The model is uni-dimensional with respect to its relationship to the regressors and is in this sense quite different from the classical approaches for ordinal data. Model (1) can also be extended to express a two- and higher (maximally $(k-1)$-dimensional) relationship with the regressors. McCullagh concludes in his discussion of the paper by stating that "This paper is a substantial contribution to an important and lively area of Statistics". Yet, I failed to see many applications of this approach in the literature, nor it seems that this approach to have inspired others. I wonder how the authors feel about this approach and whether they have an explanation why this approach never got into practice.

As Anderson noticed, many ordinal scores are of the "assessed" type. This implies that scorers have to classify subjects on a discrete scale. Hence, misclassification (of the ordinal score) is a relevant topic in this area but not covered in this overview. It is known that the effects of misclassification can be important causing a distortion of the estimated relation between the response and the regressors. For a general overview I refer to Gustafson (2003). We have described in a recent paper on a geographical dental analysis (Mwalili et al., 2005) how correction for misclassification can be done in an ordinal (logistic) regression with a possibly corrupted ordinal response. I believe this is an important research area and wonder how the authors feel about this.

Further, I did not find much on goodness-of-fit tests in the overview, but perhaps this is an area of neglected research and not necessarily that it has been neglected in the overview.

Finally, I just want to mention that methods for ordinal response models have been studied specific statistical applications and were not mentioned (inevitably) in this overview. As an example we mention the use of group sequential methods for ordinal responses (Spiessens et al., 2002).

Future research

As already clear from the previous section, I would value very much (a) research showing the gain by exploiting the ordinal nature of the data (in stead of looking at the binary version) and (b) efforts to minimize the misclassification of the ordinal score and (c) research into optimal ways to correct for misclassification the ordinal score.

Thomas M. Loughin
Department of Statistics
Kansas State University, USA

First, I would like to thank Professors Liu and Agresti for writing such a broad and accessible review of ordinal data analysis. As someone who is both a researcher and an active statistical consultant, I especially appreciate the clear, uncluttered presentation and the extensive bibliography. I would also like to thank the editors for providing me the opportunity to read and comment on this work.

This paper demonstrates that there may be several models that could be used for the analysis of any particular ordered categorical data set, as well as some non-model-based procedures. Ultimately, though, any statistical analysis we conduct leads to inferences of some kind. Whether a particular analysis approach can provide inferences that are "good" must be one of the primary considerations in selecting an analysis method. In the frequentist context, "good" inference means tests with approximately the right type I error rates and good power against interesting alternatives, and confidence intervals that are short, yet maintain approximately nominal coverage. Practitioners must know something about the relative strengths and weaknesses of inferences resulting from candidate procedures in order to make informed choices for effective statistical analysis.

We must bear in mind, to paraphrase Box, that all models are approximations. Even in a very good model, parameters must be approximated based on available data. For many models, different computational approaches can be employed which can yield different parameter estimates. Furthermore, most inferences are based on asymptotic arguments, which by definition provide approximate inference for finite sample sizes, and may additionally rely on further assumptions and approximations.

Thus, there are many potential sources for errors to creep into inferences made from any statistical model. What are the relative sizes of these errors? How valid are inferences based on models for categorical data, in general, and for ordered categorical data, in particular?

Unfortunately, the news is not always good for nominal data modeling procedures. Simulations of overdispersed count data in relatively simple structures have been conducted by Campbell et al. (1999) and Young et al. (1999). They conclude that tests based on an ordinary linear model analysis are at least as good as those from various generalized linear models often recommended for overdispersed data, which are further subject to convergence problems that do not affect the ordinary linear models. Bilder et al. (2000) investigate a variety of model-based and non-model-based approaches to the analysis of multiple-response categorical data (such as vectors of responses to "mark-all-that-apply" survey questions), including numerous approaches suggested by Agresti and Liu (1999, 2001). They find that most of the model-based approaches do not provide reliable tests for a special case of independence that is relevant in multiple-response problems, either because of difficulty or instability in fitting the models or because of instability in the subsequent Wald statistics. On the other hand, a comparatively simple sum of Pearson Statistics (also proposed by Agresti and Liu, 1999) combined with bootstrap inference provides tests with error rates that are consistently near the nominal level and with good power. Finally, ongoing simulations with my own graduate students on count and proportion data in split-plot designs demonstrate clearly that these problems cannot be handled adequately by fixed-effect generalized linear models. We find further that, *even when we use the exact same generalized linear mixed model structure for the analysis as was used to generate the data*, the generalized linear mixed model analysis is generally somewhat conservative, and gets more so when the expected counts are small. An analysis by ordinary mixed models maintains proper size and shows better power when the number of binomial trials is constant across all experimental units. There is some preliminary indication that the ordinary mixed model analysis may become liberal when numbers of binomial trials vary substantially across treatment combinations, and transformation only partially alleviates the problem. Therefore, there may be room in the practitioner's toolkit for both types of analysis.

These results are all for counts and proportions in nominal problems. We can anticipate that the smoothing that results from assuming some particular ordinal structure can help to reduce instability that might plague a comparable nominal analysis, but at what cost in terms of robustness? How much is known, based on simulation or other assessment, about how well these specialized procedures perform for standard inferences that might be of typical interest? I hope that Professors Liu and Agresti can offer a few words on this issue, or perhaps point out some additional references where these assessments are made for ordinal methods.

I suspect that Professors Liu and Agresti are quite correct in their suspicion in Section 6.5 regarding the choice of analysis for many "methodologists" (by which I assume they mean "non-statisticians analyzing data"). They ask us to comment on what we believe is required to raise awareness of these more recently-developed analysis methods. Certainly, little progress will be made on that front until software is available in popular statistical packages. Even well-trained statisticians, faced with time constraints, may be likely to apply simple and available approximations when the alternatives are complex methods with little organized software.

But better software, by itself, will not compel people to switch to newer methods. I liken this to the development of a new surgical procedure. Surgeons around the world will take the time to learn new techniques only if either the new techniques are easier (or less expensive or less time-consuming) than the old and provide at least comparable results, or if the results produced by the new techniques are demonstrably better than they can currently achieve. Without such evidence, why should they change? What reason have we given methodologists to change their statistical analysis methods for ordinal data? The structural elegance of a model is irrelevant to them if it can't generate a useful p-value or confidence interval. Many of the new methods already have a clear *disadvantage* by being considerably more complex or difficult to work with than traditional methods of analysis. Unless new methods are demonstrated to have a clear advantage *in performance* over simpler analysis methods, methodologists and many statisticians will continue to make use of the same tools that their professional ancestors used.

To this end, it does not suffice to publish results of comparative studies in a top statistical journal. To truly "get the word out" about modern methods, there is a need for easy-to-swallow statistical papers published in the literature of the people who need to use them, and for application papers demonstrating their use. A brief review of Professor Agresti's website shows that he has been active in this regard. More of us (myself included) need to take a larger role in publishing work of practical utility to non-statistical readers.

Elisabeth Svensson
Department of Statistics
Örebro University, Sweden

This paper by Liu and Agresti contains a comprehensive review mainly concerning the development of models for ordinal data. It was a pleasure for me reading this paper as it reflects the methodological and computational progress of the models presented by Agresti (1989a,b) and McCullagh (1980), and these papers were important sources of inspiration in my development of methods for paired ordinal data. The authors have raised interesting problems to discuss. In some sense modelling ordinal data has matured, but there is still more to consider. I would like to add an approach to analysis of ordinal data, supplementary to the measures based on concordance/discordance mentioned in the paper (Section 5.3). I agree that an important task for the future is to make the models and methods known and used, both by statisticians and others. It is a matter of knowledge but also to bring about an awareness regarding the important link between the properties of data and the choice of suitable methods of analysis.

As Liu and Agresti mention in Section 5, besides the methodological developments in modelling ordinal data there is an ongoing research regarding non-model methods of analysis. My research concerns development of methods for evaluation of paired ordinal data irrespective of the number of possible response categories, including continuous data from visual analogue scales (VAS). The paired data could come from inter- or intra-rater agreement studies, but my methods are also applicable to analysis of change in paired ordinal responses. The basic idea behind my approach is the *augmented ranking* that takes account of the information obtained by the dependence between the paired assessments; therefore the pairs of ordinal data are assigned ranks that are tied to the pairs of observations, in contrast to the classical marginal methods, where the ranks are tied to each marginal distribution. This ranking approach makes it possible to evaluate and measure the marginal determined systematic change in responses (bias or group-specific change) separately from the additional individual variability that is unexplained by the marginal distribution (subject-specific change), (Svensson, 1993; Svensson and Holm, 1994). The approach takes account of the rank-invariant properties of data and allows for small data sets and zero cell frequencies of the contingency table as well.

A presence of systematic change/disagreement in scale position and/or concentration between the two assessments is expressed by the measures of *relative position* and *relative concentration*; the latter is useful in case of nonlinear change in ordinal responses. The additional individual variability in the pairs of assessments is related to the so-called *rank-transformable pattern*, which is the expected pattern of paired classifications, conditional on the two sets of marginal distribution, when the internal ordering of all individuals is unchanged even though the individuals could have changed the ordinal response value between the assessments. Two measures of subject-specific dispersion from the rank-transformable pattern are suggested. The *measure of Disorder* is based on indicators of discordant pairs (Svensson, 2000a,b), and the *Relative Rank Variance* is defined by the sum of squares of the augmented mean rank differences (Svensson, 1993, 1997). Alternatively the subject-specific dispersion from the rank-transformable pattern can be expressed in term of closeness or homogeneity. A simple measure is the *coefficient of monotonic agreement*, which is a measure of association of indicators of ordered and disordered pairs in relation to the rank-transformable pattern. The *augmented rank-order agreement coefficient* expresses the correlation of augmented ranks (Svensson, 1997). Both measures express the dispersion of paired ordinal data from the rank-transformable pattern, which only in the case of equal marginal distributions coincides with the main diagonal of a square contingency table. The approach is applied to agreement studies (Svensson and Holm, 1994) but also to inter-scale comparisons (e.g. Gosman-Hedström and Svensson, 2000; Svensson, 2000a,b) and evaluation of change (e.g. Sonn and Svensson, 1997; Svensson, 1998a; Svensson and Starmark, 2002). Hopefully in the future, the augmented ranking approach could be involved in statistical models for ordinal responses as well.

I agree with Liu and Agresti that in applied research, little or no attention is paid to the relationship between measurement properties of data, although this is fundamental to the choice of statistical methods; a problem highlighted by Hand (1996). With the increasing use of questionnaires and other types of instruments for qualitative variables together with the complexity in research questions, it is important for the quality of research that specialized methods and models for ordinal data are used.

According to a survey among doctoral students in medicine, tradition and the preference for established statistical methods often have a delaying effect on the acceptance of new statistical approaches (Svensson, 2001). The major reasons mentioned why novel statistical methods for ordinal data were not used were lack of knowledge, the tradition within research groups and that traditional normal-theory methods were demanded in order to compare results with those of previous studies. My way of making novel methods known and more used in practice is to give courses to applied research groups also involving the supervisors (Svensson, 1998b). The information delay concerning development of statistical methods can also be bridged by means of joint courses for statisticians and users (Svensson, 1998b, 2001).

This overview by Liu and Agresti is a valuable tool in making good models and methods for analysis of ordinal data better known and would inspire statisticians to apply these methods to research problems in practice.

Ana M. Aguilera
Department of Statistics
University of Granada, Spain

First, I would like to congratulate Professors Liu and Agresti for their excellent review of the literature on methods for analyzing ordinal categorical data. I have enjoyed very much reading this article and I'm sure that it will contribute to make ordinal methods more popular among statistical users. The paper is so complete that it is really hard to add something new. However I would like to mention a recent improvement of the ordinal logit models that can be very interesting in practice.

Bastien et al. (2005) have developed, as a particular case of partial least squares generalized linear regression (PLS-GLR), a PLS ordinal logistic regression algorithm that solves the problem that appears when the number of explicative variables is bigger than the number of sample observations or there is a high dependence framework among predictors of a logistic regression model (multicollinearity). It is known that in order to overcome these problems PLS uses as covariates latent uncorrelated variables given by linear spans of the original predictors that take into account the relationship among the original covariates and the response. PLS-GLR is an ad hoc adaptation of PLS algorithm where each one of the linear models that involves the response variable is changed by the corresponding logit model meanwhile the remaining linear fits are kept. An alternative solution to these problems is the principal component logistic regression (PCLR) model developed by Aguilera et al. (2005) for the case of a binary response that could be easily generalized for an ordinal multiple response by using, for example, cumulative logits. PCLR is based on using as covariates of the logit model a reduced set of optimum principal components selected according to their ability for explaining the response.

With respect to the interesting question, formulated by the authors in the last section of the paper, about the future of ordinal data analysis, I think that in addition to write popularizing works on the subject, similar to this paper and the pioneer book by Agresti (1984), it is essential that standard versions of the software packages widely used by statisticians include in their menus the most usual methodologies for ordinal data analysis. In fact, some packages very used in education (SPSS, among others) do not offer basic tools as the ordinal-nominal association measures studied in Agresti (1981), ordinal log-linear models or stepwise selection procedures for the proportional odds models.

On the other hand, I think that research on important non-solved problems must also continue. In the context of functional data analysis, for example, functional generalized linear models (FGLM) have been recently introduced by James (2002) with the aim of explaining a response variable in terms of a functional covariate whose observations are functions instead of vectors as in the classic multivariate analysis. The particular case of the logistic link (functional logit models) has been deeply studied in Escabias et al. (2004b) where functional principal component analysis of the functional predictor is used for solving the multicollinearity problem and getting an accurate estimation of the parameter function. This functional principal component logit model has been successfully applied for predicting the risk of drought in certain area in terms of the continuous-time evolution of its temperature (Escabias et al., 2004a). In this line of study, I think that it would be interesting in the future to adapt FGLM to the case of an ordinal response by using different link functions as the ones mentioned in this paper, and to study their practical performance.

Finally, I would like to thank the editors of TEST journal for giving me the opportunity of discussing this paper.

Rejoinder by I. Liu and A. Agresti

We thank the discussants for reading our article and taking the time to prepare thoughtful and interesting comments. In particular, we're pleased to see that this discussion has highlighted some interesting topics and approaches that we did not discuss in our survey article. In the following reply, Dr. Agresti apologizes for what must seem like an over-emphasis on his own publications, but it's easiest to reply in terms of work with which one is most familiar!

Gerhard Tutz makes a strong argument that more attention should be paid to sequential models. An important point here is that it generalizes more easily than the cumulative logit model when the simple proportional odds form of model fits poorly. One reason the sequential model may not be more commonly used is that results depend on whether categories are ordered from low to high or from high to low, and in most applications with ordinal data it is natural for results to be invariant to this choice.

We agree with Dr. Tutz that the mean response model is less desirable for modeling ordinal response data. However, we like to keep it in our tool bag for a couple of reasons: This type of model is easier for non-statisticians to understand than logistic models, and as the number of response categories increases it is nice to have a model that interfaces with ordinary regression without recourse to assuming an underlying latent structure. Although there is the disadvantage of choosing a metric, in practice we face the same issue in dealing with ordinal explanatory variables in any of the standard models.

Dr. Tutz makes an interesting observation about the lack of literature on prediction for ordinal responses. Also, using more general structure for the predictor term as suggested by Dr. Tutz is useful for improving the flexibility of model fitting. As he notes, the challenge is to find smooth structures that are simple and interpretable. Such methods will probably be more commonly used once methodologists become more familiar with actual applications where this provides useful additional information.

Along with Gerhard Tutz, Jeffrey Simonoff is one of the world's experts on smoothing methods for categorical data. We very much appreciated his discussion, which leads those of us who are not as familiar with this area as we should be through the basic ideas. His discussion about this together with Tutz's show a couple of ways that smoothing can recognize inherent ordinality. These methods are also likely to be increasingly relevant as more applications have large, sparse data sets.

Maria Kateri mentions that our review did not focus on the analysis of square contingency tables with ordered categories. She mentioned some models that apply to such tables. The quasi-symmetry model is a beautiful one that has impressive scope, and the properties she mentions about minimizing distance to symmetry give us further appreciation for it. We'd like to mention a couple of other models for square ordinal tables that we feel are especially useful. One is an ordinal version of quasi symmetry that assigns scores to the categories and treats the main effect terms in the model as quantitative rather than qualitative. With this simplification, the sufficient statistics for the main effects are the row mean and the column mean rather than the entire marginal distributions. See Agresti (1993) and Agresti (2002, Section 10.4.6). Another useful model is a cumulative logit model with a location shift for the margins. This can be done conditionally (with a subject-specific effect, as in Agresti and Lang (1993) and Agresti (2002, Exercise 12.35)) or marginally (Agresti, 2002, Section 10.3.1). The GEE approach with the marginal version of the latter model is a special case of the methodology discussed in Section 3.1 of this paper, but for such simple tables our preference is for maximum likelihood fitting.

Thanks to Dr. Kateri for mentioning other nonparametric approaches to random effects modeling beyond what we mentioned in Section 3.3. Dr. Kateri may well be correct that our future is a Bayesian one. However, specifying priors in a sensible fashion for ordinal models seems challenging for practical implementation, especially for the more complex models that she notes will be increasingly important.

An advantage of most ordinal models, compared to those for a nominal response, is having a single parameter for each effect. So, it is disappointing, but not surprising, to hear from Emmanuel Lesaffre that many are still more comfortable using ordinary logistic regression. Dr. Lesaffre mentions that in his experience the results of an ordinal logistic regression are only marginally more significant than those of a binary logistic regression. For the most part, this has not been our experience. The results in the article by Whitehead (1993) that we quote in Section 6.3 suggest that although there's usually not much to be gained with more than 4 or 5 categories, using more than two categories can result in more powerful inferences than collapsing to a binary response. Whitehead's results apply when the outcome probabilities are similar, and if most observations occur in one category then collapsing to that category versus the others will not make such a noticeable difference.

It is indeed an oversight that we failed to mention the interesting article by Anderson (1984). With fixed scores $\{\phi_t\}$ that model relates to the loglinear association models of Goodman (1979), and when those scores are equally-spaced it is probably most easily understood when expressed as an adjacent-categories logit model. This type of model has received a fair amount of attention, albeit not nearly as much as the cumulative logit model. For parameter scores, it relates to other association models of Goodman's, and the possible reduction of power from the increase in parameters needed to describe an association as well as the inferential complications that can occur may be a reason it has not received more attention. Perhaps the main reason, though, was the untimely death of John Anderson, so the model lacked an advocate. The articles by Agresti et al. (1987) and Ritov and Gilula (1991) did attempt to deal with forms of the model for two-way tables.

We agree wholeheartedly with Dr. Lesaffre that misclassification is an area that deserves much more attention. Probably a survey paper by someone (Dr. Lesaffre being a prime candidate!) on this topic would help to familiarize the rest of us with the possible ways of handling this important problem. Regarding his query about goodness of fit, we are aware of the article we cited by Lipsitz et al. (1996) but nothing else, so this may well be a neglected area, particularly with regard to marginal and mixed models for multivariate responses.

Dr. Loughin makes the excellent point that a new method is not truly "ready for prime time" unless we can show it performs well according to the criteria he mentions, and it is unlikely to be adopted much in practice unless practitioners see clear advantages to doing so. We believe the suggestion he makes is correct that the smoothing inherent in ordinal methods can provide improvements relative to models for nominal responses. We all know the benefits of parsimony, including reduced mean square error in estimating measures of interest, even if the simpler model fits a bit worse than a more complex model. That's one reason we rarely would use the generalization of the proportional odds model that has separate parameters for each cumulative logit. In terms of a reference to where assessments were made that show the advantage of ordinal methodology, Agresti and Yang (1987) showed that sparseness of data is much more of a problem for nominal-scale models than ordinal models. Finally, the remarks Dr. Loughin makes about generalized linear mixed models are worrisome but very interesting and highlight the need for further study of how well these methods work in practical situations.

Like Maria Kateri, Elisabeth Svensson focuses on paired ordinal data, highlighting an approach she's used that is non-model-based. We welcome the initiative she suggests of connecting this approach to statistical models for ordinal responses. This would seem to be helpful for generalizing her approach to make summary comparisons while adjusting for covariates.

Dr. Svensson has worked admirably to bridge the gap between statisticians and practitioners with joint courses. Those of us involved in research need to remember the fundamental importance of this, alluded to also by Dr. Loughin, and try to participate in such activities. At the University of Florida, it's worked well to have an applied course in categorical data analysis that can be taken both by undergraduate statistics majors and by graduate students in other areas who have already studied basic statistical methods – this course enrolls about 50 students every year, about 75% of them being graduate students from other disciplines. In the context of applications, we should remember that many important advances in categorical data analysis made by Leo Goodman became commonly used because he would accompany an article in a statistics journal with a companion article in a social science journal that would show both the usefulness of the method and how to use it.

Dr. Aguilera has focused on a couple of interesting problems on which there seems to be much that can be done. The work cited by Aguilera et al. (2005) and Bastien et al. (2005) addresses issues arising with large numbers of explanatory variables relative to observations. Undoubtedly interesting applications await for such methodology, such as perhaps large, sparse tables arising in genomics. The same can certainly be said for functional data analysis methods.

In summary, based on the remarks of these discussants, there still seems to be considerable scope for statisticians to become pioneers in further developing methods for ordinal data. Particular attention can be paid to smoothing methods and prediction, mixed models, misclassification and missing data, Bayesian modeling, dealing with data sets with large numbers of variables, and finding ways to highlight to methodologists that it is worth their time to learn methods specially designed for ordinal data. We look forward to seeing how this area evolves further over the next decade.

References

Agresti, A. (1980). Generalized odds ratios for ordinal data. *Biometrics*, 36:59–67.

Agresti, A. (1981). Measures of nominal–ordinal association. *Journal of the American Statistical Association*, 76:524–529.

AGRESTI, A. (1983). Testing marginal homogeneity for ordinal categorical variables. *Biometrics*, 39:505–510.

AGRESTI, A. (1984). *Analysis of Ordinal Categorical Data*. Wiley, New York.

AGRESTI, A. (1988). A model for agreement between ratings on an ordinal scale. *Biometrics*, 44:539–548.

AGRESTI, A. (1989a). A survey of models for repeated ordered categorical response data. *Statistics in Medicine*, 8:1209–1224.

AGRESTI, A. (1989b). Tutorial on modelling ordered categorical response data. *Psychological Bulletin*, 105:290–301.

AGRESTI, A. (1992). Analysis of ordinal paired comparison data. *Applied Statistics*, 41:287–297.

AGRESTI, A. (1993). 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. (2002). *Categorical Data Analysis*. Wiley, New Jersey, 2nd ed.

AGRESTI, A. and CHUANG, C. (1989). Model–based Bayesian methods for estimating cell proportions in cross–classification tables having ordered categories. *Computational Statistics and Data Analysis*, 7:245–258.

AGRESTI, A., CHUANG, C., and KEZOUH, A. (1987). Order–restricted score parameters in association models for contingency tables. *Journal of the American Statistical Association*, 82:619–623.

AGRESTI, A. and COULL, B. A. (1998). Order–restricted inference for monotone trend alternatives in contingency tables. *Computational Statistics and Data Analysis*, 28:139–155.

AGRESTI, A. and COULL, B. A. (2002). The analysis of contingency tables under inequality constraints. *Journal of Statistical Planning and Inference*, 107:45–73.

AGRESTI, A. and HITCHCOCK, D. (2004). Bayesian inference for categorical data analysis: A survey. Technical report, Department of Statistics, University of Florida.

Agresti, A. and Lang, J. (1993). A proportional odds model with subject-specific effects for repeated categorical responses. *Biometrika*, 80:527–534.

Agresti, A. and Liu, I. M. (1999). Modeling a categorical variable allowing arbitrarily many category choices. *Biometrics*, 55:936–943.

Agresti, A. and Liu, I. M. (2001). Strategies for modeling a categorical variable allowing multiple category choices. *Sociological Methods & Research*, 29:403–434.

Agresti, A., Mehta, C. R., and Patel, N. R. (1990). Exact inference for contingency tables with ordered categories. *Journal of the American Statistical Association*, 85:453–458.

Agresti, A. and Natarajan, R. (2001). Modeling clustered ordered categorical data: A survey. *International Statistical Review*, 69:345–371.

Agresti, A. and Yang, M. (1987). An empirical investigation of some effects of sparseness in contingency tables. *Computational Statistics and Data Analysis*, 5:9–21.

Aguilera, A. M., Escabias, M., and Valderrama, M. J. (2005). Using principal components for estimating logistic regression with high dimensional multicollinear data. *Computational Statistics and Data Analysis*. To appear.

Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88:669–679.

Anderson, C. J. and Böckenholt, U. (2000). Graphical regression models for polytomous variables. *Psychometrika*, 65:497–509.

Anderson, J. A. (1984). Regression and ordered categorical variables (with discussion). *Journal of the Royal Statistical Society. Series B*, 46(1):1–30.

Anderson, J. A. and Philips, P. R. (1981). Regression, discrimination and measurement models for ordered categorical variables. *Applied Statistics*, 30:22–31.

Bartolucci, F. and Forcina, A. (2002). Extended RC association models allowing for order restrictions and marginal modeling. *Journal of the American Statistical Association*, 97:1192–1199.

Bartolucci, F., Forcina, A., and Dardanoni, V. (2001). Positive quadrant dependence and marginal modelling in two–way tables with ordered margins. *Journal of the American Statistical Association*, 96:1497–1505.

Bartolucci, F. and Scaccia, L. (2004). Testing for positive association in contingency tables with fixed margins. *Computational Statistics and Data Analysis*, 47:195–210.

Bastien, P., Esposito-Vinzi, V., and Tenenhaus, M. (2005). PLS generalised linear regression. *Computational Statistics and Data Analysis*, 48:17–46.

Becker, M. P. (1989). Models for the analysis of association in multivariate contingency tables. *Journal of the American Statistical Association*, 84:1014–1019.

Becker, M. P. (1990). Maximum likelihood estimation of the RC(M) association model. *Applied Statistics*, 39:152–167.

Becker, M. P. and Agresti, A. (1992). Loglinear modeling of pairwise interobserver agreement on a categorical scale. *Statistics in Medicine*, 11:101–114.

Becker, M. P. and Clogg, C. C. (1989). Analysis of sets of two–way contingency tables using association models. *Journal of the American Statistical Association*, 84:142–151.

Bender, R. and Benner, A. (2000). Calculating ordinal regression models in SAS and S–PLUS. *Biometrical Journal*, 42:677–699.

Bender, R. and Grouven, U. (1998). Using binary logistic regression models for ordinal data with non-proportional odds. *Journal of Clinical Epidemiology*, 51:809–816.

Bergsma, W. P. (1997). *Marginal Models for Categorical Data*. Ph.D. thesis, Tilburg Univ, Netherlands.

Bergsma, W. P. and Rudas, T. (2002). Marginal models for categorical data. *Annals of Statistics*, 30:140–159.

Bilder, C. R., Loughin, T. M., and Nettleton, D. (2000). Multiple marginal independence testing for pick any/c variables. *Communications in Statistics. Simulation and Computation*, 29:1285–1316.

Biswas, A. and Das, K. (2002). A Bayesian analysis of bivariate ordinal data: Wisconsin epidemiologic study of diabetic retinopathy revisited. *Statistics in Medicine*, 21:549–559.

Bock, R. D. and Jones, L. V. (1968). *The Measurement and Prediction of Judgement and Choice*. Holden–Day, San Francisco.

Böckenholt, U. and 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. and Butler, R. W. (1999). An importance sampling algorithm for exact conditional tests in log–linear models. *Biometrika*, 86:321–332.

Booth, J. G. and 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*, 61:265–285.

Bradlow, E. T. and Zaslavsky, A. M. (1999). A hierarchical latent variable model for ordinal data from a customer satisfaction survey with "no answer" responses. *Journal of the American Statistical Association*, 94:43–52.

Brant, R. (1990). Assessing proportionality in the proportional odds model for ordinal logistic regression. *Biometrics*, 46:1171–1178.

Breslow, N. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88:9–25.

Brunner, E., Munzel, U., and Puri, M. L. (1999). Rank–score tests in factorial designs with repeated measures. *Journal of Multivariate Analysis*, 70:286–317.

Campbell, N. L., Young, L. J., and Capuano, G. A. (1999). Analyzing over-dispersed count data in two-way cross-classification problems using generalized linear models. *Journal of Statistical Computation and Simulation*, 63:263–291.

Chen, M.-H. and Dey, D. K. (2000). A unified Bayesian approach for analyzing correlated ordinal response data. *Revista Brasileira de Probabilidade e Estatistica*, 14:87–111.

Chen, M.-H. and Shao, Q.-M. (1999). Properties of prior and posterior distributions for multivariate categorical response data models. *Journal of Multivariate Analysis*, 71:277–296.

Chib, S. (2000). *Generalized Linear Models: A Bayesian Perspective*. Marcel Dekker, New York.

Chib, S. and Greenberg, E. (1998). Analysis of multivariate probit models. *Biometrika*, 85:347–361.

Chipman, H. and Hamada, M. (1996). Bayesian analysis of ordered categorical data from industrial experiments. *Technometrics*, 38:1–10.

Chuang-Stein, C. and Agresti, A. (1997). A review of tests for detecting a monotone dose–response relationship with ordinal responses data. *Statistics in Medicine*, 16:2599–2618.

Cohen, A. and Sackrowitz, H. B. (1992). Improved tests for comparing treatments against a control and other one–sided problems. *Journal of the American Statistical Association*, 87:1137–1144.

Coull, B. A. and 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., and 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.

Crouchley, R. (1995). A random–effects model for ordered categorical data. *Journal of the American Statistical Association*, 90:489–498.

Dale, J. R. (1984). Local versus global association for bivariate ordered responses. *Biometrika*, 71:507–514.

Dale, J. R. (1986). Global cross–ratio models for bivariate, discrete, ordered responses. *Biometrics*, 42:909–917.

Dong, J. and Simonoff, J. S. (1995). A geometric combination estimator for *d*–dimensional ordinal sparse contingency tables. *Annals of Statistics*, 23:1143–1159.

Douglas, R., Fienberg, S. E., Lee, M.-L. T., Sampson, A. R., and Whitaker, L. R. (1991). Positive dependence concepts for ordinal contingency tables. In H. W. Block, A. R. Sampson, and T. H. Savits, eds., *Topics in Statistical Dependence*, pp. 189–202. Institute of Mathematical Statistics (Hayward, CA).

Ekholm, A., Jokinen, J., McDonald, J. W., and Smith, P. W. F. (2003). Joint regression and association modeling of longitudinal ordinal data. *Biometrics*, 59:795–803.

Ekholm, A., Smith, P. W. F., and McDonald, J. W. (1995). Marginal regression analysis of a multivariate binary response. *Biometrika*, 82:847–854.

Engel, B. (1998). A simple illustration of the failure of PQL, IRREML and APHL as approximate ML methods for mixed models for binary data. *Biometrical Journal*, 40:141–154.

Escabias, M., Aguilera, A. M., and Valderrama, M. J. (2004a). Modelling environmental data by functional principal component logistic regression. *Environmetrics*, 16:95–107.

Escabias, M., Aguilera, A. M., and Valderrama, M. J. (2004b). Principal component estimation of functional logistic regression: Discussion of two different approaches. *Journal of Nonparametric Statistics*, 16:365–384.

EVANS, M., GILULA, Z., and GUTTMAN, I. (1993). Computational issues in the Bayesian analysis of categorical data: Log–linear and Goodman's RC Model. *Statistica Sinica*, 3:391–406.

FAHRMEIR, L. and TUTZ, G. (1994). Dynamic stochastic models for time–dependent ordered paired comparison systems. *Journal of the American Statistical Association*, 89:1438–1449.

FAHRMEIR, L. and TUTZ, G. (2001). *Multivariate Statistical Modelling Based on Generalized Linear Models*. Springer, New York, 2nd ed.

FAREWELL, V. T. (1982). A note on regression analysis of ordinal data with variability of classification. *Biometrika*, 69:533–538.

FIELDING, A. (1999). Why use arbitrary points scores?: Ordered categories in models of educational progress. *Journal of the Royal Statistical Society. Series A*, 162:303–328.

FIENBERG, S. E. and HOLLAND, P. W. (1972). On the choice of flattening constants for estimating multinomial probabilities. *Journal of Multivariate Analysis*, 2:127–134.

FIENBERG, S. E. and HOLLAND, P. W. (1973). Simultaneous estimation of multinomial cell probabilities. *Journal of the American Statistical Association*, 68:683–691.

FITZMAURICE, G. M. and LAIRD, N. M. (1993). A likelihood–based method for analysing longitudinal binary responses. *Biometrika*, 80:141–151.

FORSTER, J. J. (2001). Bayesian inference for square contingency tables. Unpublished manuscript.

FORSTER, J. J., MCDONALD, J. W., and SMITH, P. W. F. (1996). Monte Carlo exact conditional tests for log–linear and logistic models. *Journal of the Royal Statistical Society. Series B*, 58:445–453.

GENTER, F. C. and FAREWELL, V. T. (1985). Goodness–of–link testing in ordinal regression models. *Canadian Journal of Statistics*, 13:37–44.

GILULA, Z. (1986). Grouping and association in contingency tables: An exploratory canonical correlation approach. *Journal of the American Statistical Association*, 81:773–779.

Gilula, Z. and Haberman, S. J. (1988). The analysis of multivariate contingency tables by restricted canonical and restricted association models. *Journal of the American Statistical Association*, 83:760–771.

Gilula, Z., Krieger, A. M., and Ritov, Y. (1988). Ordinal association in contingency tables: Some interpretive aspects. *Journal of the American Statistical Association*, 83:540–545.

Gilula, Z. and Ritov, Y. (1990). Inferential ordinal correspondence analysis: Motivation, derivation and limitations. *International Statistical Review*, 58:99–108.

Glonek, G. F. V. (1996). A class of regression models for multivariate categorical responses. *Biometrika*, 83:15–28.

Glonek, G. F. V. and McCullagh, P. (1995). Multivariate logistic models. *Journal of the Royal Statistical Society. Series B*, 57:533–546.

Gonin, R., Lipsitz, S. R., Fitzmaurice, G. M., and Molenberghs, G. (2000). Regression modelling of weighted κ by using generalized estimating equations. *Journal of the Royal Statistical Society. Series C: Applied Statistics*, 49:1–18.

Good, I. J. (1965). *The Estimation of Probabilities: An Essay on Modern Bayesian Methods*. MIT Press, Cambridge, MA.

Goodman, L. A. (1979). Simple models for the analysis of association in cross–classifications having ordered categories. *Journal of the American Statistical Association*, 74:537–552.

Goodman, L. A. (1981). Criteria for determining wether certain categories in a cross classification table should be combined, with special reference to occupational categories in an occupational mobility table. *American Journal of Sociology*, 87:612–650.

Goodman, L. A. (1983). The analysis of dependence in cross–classifications having ordered categories, using log–linear models for frequencies and log–linear models for odds. *Biometrics*, 39:149–160.

Goodman, L. A. (1985). The analysis of cross–classified data having ordered and/or unordered categories: Association models, correlation models, and asymmetry models for contingency tables with or without missing entries. *Annals of Statistics*, 13:10–69.

GOODMAN, L. A. (1986). Some useful extensions of the usual correspondence analysis approach and the usual log–linear models approach in the analysis of contingency tables. *International Statistical Review*, 54:243–309.

GOODMAN, L. A. (1996). A single general method for the analysis of cross–classified data: reconciliation and synthesis of some methods of pearson, yule, and fisher, and also some methods of correspondence analysis and association analysis. *Journal of the American Statistical Association*, 91:408–428.

GOODMAN, L. A. (2004). Three different ways to view cross–classified categorical data: Rasch–type models, log–linear models, and latent–class models. Technical report.

GOODMAN, L. A. and KRUSKAL, W. H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association*, 49:732–764.

GOSMAN-HEDSTRÖM, G. and SVENSSON, E. (2000). Parallel reliability of the Functional Independence Measure and the Barthel ADL index. *Disability and Rehabilitation*, 22:702–715.

GRIZZLE, J. E., STARMER, C. F., and KOCH, G. G. (1969). Analysis of categorical data by linear models. *Biometrics*, 25:489–504.

GUSTAFSON, P. (2003). *Measurement Error and Misclassification in Statistics and Epidemiology: Impacts and Bayesian Adjustments*, vol. 13 of *Interdisciplinary Statistics*. Chapman and Hall/CRC Press, Boca Raton, Florida.

HABERMAN, S. J. (1974). Log–linear models for frequency tables with ordered classifications. *Biometrics*, 36:589–600.

HABERMAN, S. J. (1981). Tests for independence in two–way contingency tables based on canonical correlation and on linear–by–linear interaction. *Annals of Statistics*, 9:1178–1186.

HABERMAN, S. J. (1995). Computation of maximum likelihood estimates in association models. *Journal of the American Statistical Association*, 90:1438–1446.

Hand, D. J. (1996). Statistics and the theory of measurement. *Journal of the Royal Statistical Society. Series A*, 159:445–492.

Hartzel, J., Agresti, A., and Caffo, B. (2001a). Multinomial logit random effects models. *Statistical Modelling*, 1:81–102.

Hartzel, J., Liu, I.-M., and Agresti, A. (2001b). 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. and Mee, R. W. (1984). A mixed–model procedure for analyzing ordered categorical data. *Biometrics*, 40:393–408.

Hastie, T. and Tibshirani, R. (1987). Non–parametric logistic and proportional odds regression. *Applied Statistics*, 36:260–276.

Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models*. Chapman and Hall, London.

Hastie, T. and Tibshirani, R. (1993). Varying–coefficient models. *Journal of the Royal Statistical Society. Series B*, 55:757–796.

Hastie, T., Tibshirani, R., and Friedman, J. H. (2001). *The Elements of Statistical Learning*. Springer–Verlag, New York.

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

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

Hedeker, D. and Gibbons, R. D. (1996). MIXOR: A computer program for mixed–effects ordinal regression analysis. *Computer Methods and Programs in Biomedicine*, 49:157–176.

Hedeker, D. and Mermelstein, R. J. (2000). Analysis of longitudinal substance use outcomes using ordinal random–effects regression models. *Addiction*, 95:S381–S394.

Heeren, T. and D'Agostino, R. (1987). Robustness of the two independent samples t–test when applied to ordinal scaled data. *Statistics in Medicine*, 6:79–90.

Heumann, C. (1997). *Likelihoodbasierte Marginale Regressionsmodelle fur Korrelierte Kategoriale Daten*. Ph.D. thesis, Ludwig–Maximilians–Universitat, Munchen.

Hilton, J. F. and Mehta, C. R. (1993). Power and sample size calculations for exact conditional tests with ordered categorical data. *Biometrics*, 49:609–616.

Hjort, N. L. and Jones, M. C. (1996). Locally parametric nonparametric density estimation. *Annals of Statistics*, 24:1619–1647.

Huang, G.-H., Bandeen-Roche, K., and Rubin, G. S. (2002). Building marginal models for multiple ordinal measurements. *Journal of the Royal Statistical Society. Series C: Applied Statistics*, 51:37–57.

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

Ishwaran, H. and Gatsonis, C. A. (2000). A general class of hierarchical ordinal regression models with applications to correlated ROC analysis. *The Canadian Journal of Statistics*, 28:731–750.

Jaffrézic, F., Robert-Granié, C. R., and Foulley, J.-L. (1999). A quasi–score approach to the analysis of ordered categorical data via a mixed heteroskedastic threshold model. *Genetics Selection Evolution*, 31:301–318.

James, G. M. (2002). Generalized linear models with functional predictors. *Journal of Royal Statistical Society. Series B*, 64:411–432.

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.

Johnson, V. E. and Albert, J. H. (1999). *Ordinal Data Modeling*. Springer–Verlag, New York.

Kateri, M. and Iliopoulos, G. (2003). On collapsing categories in two-way contingency tables. *Statistics*, 37:443–455.

Kateri, M., Nicolaou, A., and Ntzoufras, I. (2005). Bayesian inference for the RC(m) association model. *Journal of Computational and Graphical Statistics*, 14:116–138.

Kateri, M. and Papaioannou, T. (1994). f-divergence association models. *International Journal of Mathematical and Statistical Science*, 3:179–203.

Kateri, M. and Papaioannou, T. (1997). Asymmetry models for contingency tables. *Journal of the American Statistical Association*, 92:1124–1131.

Kauermann, G. (2000). Modeling longitudinal data with ordinal response by varying coefficients. *Biometrics*, 56:692–698.

Kauermann, G. and Tutz, G. (2000). Local likelihood estimation in varying coefficient models including additive bias correction. *Journal of Nonparametric Statistics*, 12:343–371.

Kauermann, G. and Tutz, G. (2003). Semi– and nonparametric modeling of ordinal data. *Journal of Computational and Graphical Statistics*, 12:176–196.

Kenward, M. G., Lesaffre, E., and 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, D. and Agresti, A. (1997). Nearly exact tests of conditional independence and marginal homogeneity for sparse contingency tables. *Computational Statistics and Data Analysis*, 24:89–104.

Kim, J.-H. (2003). Assessing practical significance of the proportional odds assumption. *Statistics & Probability Letters*, 65:233–239.

Koch, G. G., Landis, J. R., Freeman, J. L., Freeman, J., D. H., and Lehnen, R. G. (1977). A general methodology for the analysis of experiments with repeated measurement of categorical data. *Biometrics*, 33:133–158.

Kosorok, M. R. and Chao, W.-H. (1996). The analysis of longitudinal ordinal response data in continuous time. *Journal of the American Statistical Association*, 91:807–817.

Landis, J. R., Heyman, E. R., and 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., Sharp, T. J., Kuritz, S. J., and Koch, G. G. (1998). Mantel–haenszel methods. In *Encyclopedia of Biostatistics*, pp. 2378–2691. Wiley, Chichester, UK.

Lang, J. B. (1996). Maximum likelihood methods for a general class of log–linear models. *Annals of Statistics*, 24:726–752.

Lang, J. B. (1999). Bayesian ordinal and binary regression models with a parametric family of mixture links. *Computational Statistics and Data Analysis*, 31:59–87.

Lang, J. B. (2004). Multinomial–Poisson homogeneous models for contingency tables. *Annals of Statistics*, 32:340–383.

Lang, J. B. and 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., and Smith, P. W. F. (1997). Association marginal modeling of multivariate polytomous responses: A maximum likelihood approach for large, sparse tables. *Journal of the American Statistical Association*, 94:1161–1171.

Lawal, H. B. (2004). Review of non-independence, asymmetry, skew-symmetry and point-symmetry models in the analysis of social mobility data. *Quality & Quantity*, 38:259–289.

Lee, M.-K., Song, H.-H., Kang, S.-H., and Ahn, C. W. (2002). The determination of sample sizes in the comparison of two multinomial proportions from ordered categories. *Biometrical Journal*, 44:395–409.

Leonard, T. (1973). A Bayesian method for histograms. *Biometrika*, 60:297–308.

Lindsey, J. K. (1999). *Models for Repeated Measurements*. Oxford University Press, New York, 2nd ed.

Lipsitz, S. R. (1992). Methods for estimating the parameters of a linear model for ordered categorical data. *Biometrics*, 48:271–281.

Lipsitz, S. R. and 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., Fitzmaurice, G. M., and Molenberghs, G. (1996). Goodness–of–fit tests for ordinal response regression models. *Applied Statistics*, 45:175–190.

Lipsitz, S. R., Kim, K., and Zhao, L. (1994). Analysis of repeated categorical data using generalized estimating equations. *Statistics in Medicine*, 13:1149–1163.

Little, R. J. A. (1995). Modeling the drop–out mechanism in repeated–measures studies. *Journal of the American Statistical Association*, 90:1112–1121.

Little, R. J. A. and Rubin, D. B. (1987). *Statistical Analysis with Missing Data*. Wiley, New York.

Liu, I. (2003). Describing ordinal odds ratios for stratified $r \times c$ tables. *Biometrical Journal*, 45:730–750.

Liu, I.-M. and Agresti, A. (1996). Mantel–Haenszel–type inference for cumulative odds ratio. *Biometrics*, 52:1222–1234.

Liu, Q. and Pierce, D. A. (1994). A note on Gauss–Hermite quadrature. *Biometrika*, 81:624–629.

Loader, C. R. (1996). Local likelihood density estimation. *Annals of Statistics*, 24:1602–1618.

Lui, K.-J., Zhou, X.-H., and Lin, C.-D. (2004). Testing equality between two diagnostic procedures in paired–sample ordinal data. *Biometrical Journal*, 46:642–652.

Mantel, N. and Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, 22:719–748.

Mark, S. D. and Gail, M. H. (1994). A comparison of likelihood–based and marginal estimating equation methods for analysing repeated ordered categorical responses with missing data. *Statistics in Medicine*, 13:479–493.

McCullagh, P. (1978). A class of parametric models for the analysis of square contingency tables with ordered categories. *Biometrika*, 65:413–418.

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

Miller, M. E., Davis, C. S., and Landis, J. R. (1993). The analysis of longitudinal polytomous data: Generalized estimating equations and connections with weighted least square. *Biometrics*, 49:1033–1044.

Miller, M. E., Ten Have, T. R., Reboussin, B. A., Lohman, K. K., and Rejeski, W. J. (2001). A marginal model for analyzing discrete outcomes from longitudinal survey with outcomes subject to multiple–cause nonresponse. *Journal of the American Statistical Association*, 96:844–857.

Molenberghs, G., Kenward, M. G., and Lesaffre, E. (1997). The analysis of longitudinal ordinal data with nonrandom drop–out. *Biometrika*, 84:33–44.

Molenberghs, G. and Lesaffre, E. (1994). Marginal modeling of correlated ordinal data using a multivariate Plackett distribution. *Journal of the American Statistical Association*, 89:633–644.

Mwalili, S. M., Lesaffre, E., and Declerck, D. (2005). A Bayesian ordinal logistic regression model to correct for interobserver measurement error in a geographical oral health study. *Applied Statistics*, 54(1):1–17.

Oh, M. (1995). On maximum likelihood estimation of cell probabilities in $2 \times k$ contingency tables under negative dependence restrictions with various sampling schemes. *Communications in Statistics. Theory and Methods*, 24:2127–2143.

Ohman-Strickland, P. A. and Lu, S.-E. (2003). Estimates, power and sample size calculations for two–sample ordinal outcomes under before–after study designs. *Statistics in Medicine*, 22:1807–1818.

Pearson, K. and Heron, D. (1913). On theories of association. *Biometrika*, 9:159–315.

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

Pinheiro, J. C. and Bates, D. M. (1995). Approximations to the log–likelihood function in the non–linear mixed–effects model. *Journal of Computational and Graphical Statistics*, 4:12–35.

Plackett, R. L. (1965). A class of bivariate distributions. *Journal of the American Statistical Association*, 60:516–522.

Qiu, Z., Song, P. X.-K., and Tan, M. (2002). Bayesian hierarchical models for multi–level repeated ordinal data using WinBUGS. *Journal of Biopharmaceutical Statistics*, 12:121–135.

Qu, Y. S., Piedmonte, M. R., and Medendorp, S. V. (1995). Latent variable models for clustered ordinal data. *Biometrics*, 51:268–275.

Qu, Y. S. and Tan, M. (1998). Analysis of clustered ordinal data with subclusters via a Bayesian hierarchical model. *Communications in Statistics. Theory and Methods*, 27:1461–1476.

Rabbee, N., Coull, B. A., Mehta, C., Patel, N., and Senchaudhuri, P. (2003). Power and sample size for ordered categorical data. *Statistical Methods in Medical Research*, 12:73–84.

Ritov, Y. and Gilula, Z. (1991). The order–restricted RC model for ordered contingency tables: Estimation and testing for fit. *Annals of Statistics*, 19:2090–2101.

Rogel, A., Boalle, P. Y., and Mary, J. Y. (1998). Global and partial agreement among several observers. *Statistics in Medicine*, 17:489–501.

Rom, D. and Sarkar, S. K. (1992). Grouping and association in contingency tables: An exploratory canonical correlation approach. *Journal of Statistical Planning and Inference*, 33:205–212.

Rossi, P. E., Gilula, Z., and Allenby, G. M. (2001). Overcoming scale usage heterogeneity: A Bayesian hierarchical approach. *Journal of the American Statistical Association*, 96:20–31.

Rossini, A. J. and 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.

Sedransk, J., Monahan, J., and Chiu, H. Y. (1985). Bayesian estimation of finite population parameters in categorical data models incorporating order. *Journal of the Royal Statistical Society. Series B*, 47:519–527.

Simon, G. (1974). Alternative analyses for the singly–ordered contingency table. *Journal of the American Statistical Association*, 69:971–976.

Simonoff, J. S. (1987). Probability estimation via smoothing in sparse contingency tables with ordered categories. *Statistics & Probability Letters*, 5:55–63.

Simonoff, J. S. (1996). *Smoothing Methods in Statistics*. Springer-Verlag, New York.

Simonoff, J. S. (1998). Three sides of smoothing: Categorical data smoothing, nonparametric regression, and density estimation. *International Statistical Review*, 66:137–156.

Simonoff, J. S. and Tutz, G. (2001). Smoothing methods for discrete data. In M. G. Schimek, ed., *Smoothing and Regression. Approaches, Computation and Application*, pp. 193–228. John Wiley and Sons, New York.

Snell, E. J. (1964). A scaling procedure for ordered categorical data. *Biometrics*, 20:592–607.

Sonn, U. and Svensson, E. (1997). Measures of individual and group changes in ordered categorical data: Application to the ADL Staircase. *Scandinavian Journal of Rehabilitation Medicine*, 29:233–242.

Spiessens, B., Lesaffre, E., Verbeke, G., and Kim, K. M. (2002). Group sequential methods for an ordinal logistic random-effects model under misspecification. *Biometrics*, 58:569–575.

Spitzer, R. L., Cohen, J., Fleiss, J. L., and Endicott, J. (1967). Quantification of agreement in psychiatric diagnosis. *Archives of General Psychiatry*, 17:83–87.

Stokes, M. E., Davis, C. S., and Koch, G. G. (2000). *Categorical Data Analysis Using the SAS system*. SAS Institute, Cary, NC, 2nd ed.

SUTRADHAR, B. C. (2003). An overview on regression models for discrete longitudinal responses. *Statistical Science*, 18:377–393.

SVENSSON, E. (1993). *Analysis of systematic and random differences between paired ordinal categorical data*. Ph.D. thesis, Department of Statistics, Göteborg University. Publications 21.

SVENSSON, E. (1997). A coefficient of agreement adjusted for bias in paired ordered categorical data. *Biometrical Journal*, 39:643–657.

SVENSSON, E. (1998a). Ordinal invariant measures for individual and group changes in ordered categorical data. *Statistics in Medicine*, 17:2923–2936.

SVENSSON, E. (1998b). Teaching biostatistics to clinical research groups. In S. K. L. W. K. T. Pereira-Mendoza, L. and W. K. Wong, eds., *Proceedings of the Fifth International Conference on Teaching Statistics*, pp. 289–294. International Statistical Institute, Singapore.

SVENSSON, E. (2000a). Comparison of the quality of assessments using continuous and discrete ordinal rating scales. *Biometrical Journal*, 42:417–434.

SVENSSON, E. (2000b). Concordance between ratings using different scales for the same variable. *Statistics in Medicine*, 19:3483–3496.

SVENSSON, E. (2001). Important considerations for optimal communication between statisticians and medical researchers in consulting, teaching and collaborative research -with a focus on the analysis of ordered categorical data. In C. Batanero, ed., *Training Researchers in the Use of Statistics*, pp. 23–35. International Association for Statistical Education, Granada.

SVENSSON, E. and HOLM, S. (1994). Separation of systematic and random differences in ordinal rating scales. *Statistics in Medicine*, 13:2437–2453.

SVENSSON, E. and STARMARK, J. E. (2002). Evaluation of individual and group changes in social outcome after aneurysmal subarachnoid haemorrhage: A long-term follow-up study. *Journal of Rehabilitation Medicine*, 34:251–259.

TAN, M., QU, Y. S., MASCHA, E., and SCHUBERT, A. (1999). A Bayesian hierarchical model for multi–level repeated ordinal data: analysis of oral practice examinations in a large anaesthesiology training program. *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., Miller, M. E., Reboussin, B. A., and James, M. M. (2000). Mixed effects logistic regression models for longitudinal ordinal functional response data with multiple cause drop–out from the longitudinal study of aging. *Biometrics*, 56:279–287.

Thompson, R. and Baker, R. J. (1981). Composite link functions in generalized linear models. *Applied Statistics*, 30:125–131.

Titterington, D. M. and Bowman, A. W. (1985). A comparative study of smoothing procedures for ordered categorical data. *Journal of Statistical Computation and Simulation*, 21:291–312.

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

Toledano, A. Y. and Gatsonis, C. (1999). Generalized estimating equations for ordinal categorical data: arbitrary patterns of missing responses and missingness in a key covariate. *Biometrics*, 55:488–496.

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

Tutz, G. (1990). Sequential item response models with an ordered response. *British Journal of Mathematical and Statistical Psychology*, 43:39–55.

Tutz, G. (1991). Sequential models in categorical regression. *Computational Statistics and Data Analysis*, 11:275–295.

Tutz, G. (2003). Generalized semiparametrically structured ordinal models. *Biometrics*, 59:263–273.

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

Uebersax, J. S. (1999). Probit latent class analysis with dichotomous or ordered category measures: Conditional independence/dependence models. *Applied Psychological Measurement*, 23:283–297.

Uebersax, J. S. and Grove, W. M. (1993). A latent trait finite mixture model for the analysis of rating agreement. *Biometrics*, 49:823–835.

Vermunt, J. and Hagenaars, J. A. (2004). Ordinal longitudinal data analysis. In R. C. Hauspie, N. Cameron, and L. Molinari, eds., *Methods in Human Growth Research*, pp. 374–393. Cambridge University Press, Cambridge.

Von Eye, A. and Spiel, C. (1996). Standard and non-standard log-linear symmetry models for measuring change in categorical variables. *American Statistician*, 50:300–305.

Wahrendorf, J. (1980). Inference in contingency tables with ordered categories using Plackett's coefficient of association for bivariate distributions. *Biometrika*, 67:15–21.

Walker, S. H. and Duncan, D. B. (1967). Estimation of the probability of an event as a function of several independent variables. *Biometrika*, 54:167–179.

Wang, Y. (1996). A likelihood ratio test against stochastic ordering in several populations. *Journal of the American Statistical Association*, 91:1676–1683.

Webb, E. L. and Forster, J. J. (2004). Bayesian model determination for multivariate ordinal and binary data. Technical report.

Wermuth, N. and Cox, D. R. (1998). On the application of conditional independence to ordinal data. *International Statistical Review*, 66:181–199.

Whitehead, J. (1993). Sample size calculations for ordered categorical data. *Statistics in Medicine*, 12:2257–2271.

Williams, O. D. and Grizzle, J. E. (1972). Analysis of contingency tables having ordered response categories. *Journal of the American Statistical Association*, 67:55–63.

Williamson, J. M. and Kim, K. (1996). A global odds ratio regression model for bivariate ordered categorical data from opthalmologic studies. *Statistics in Medicine*, 15:1507–1518.

Williamson, J. M., Kim, K., and Lipsitz, S. R. (1995). Analyzing bivariate ordinal data using a global odds ratio. *Journal of the American Statistical Association*, 90:1432–1437.

Williamson, J. M. and Lee, M.-L. T. (1996). A GEE regression model for the association between an ordinal and a nominal variable. *Communications in Statistics. Theory and Methods*, 25:1887–1901.

Xie, M., Simpson, D. G., and Carroll, R. J. (2000). Random effects in censored ordinal regression: Latent structure and Bayesian approach. *Biometrics*, 56:376–383.

Yee, T. W. and Wild, C. J. (1996). Vector generalized additive models. *Journal of the Royal Statistical Society. Series B*, 58:481–493.

Young, L. J., Campbell, N. L., and Capuano, G. A. (1999). Analysis of overdispersed count data from single-factor experiments: A comparative study. *Journal of Agricultural, Biological, and Environmental Statistics*, 4:258–275.