Published in final edited form as:

Biometrics. 2010 September ; 66(3): 966–974. doi:10.1111/j.1541-0420.2009.01359.x.

Assessing Toxicities in a Clinical Trial: Bayesian Inference for Ordinal Data Nested within Categories

L.G. Leon-Novelo,

University of Florida. Department of Statistics, 102 Griffin-Floyd Hall. P.O. Box 118545, Gainesville, FL 32611, USA

X. Zhou,

MS 442A, Genentech, Inc., South San Francisco, CA, USA

B. Nebiyu Bekele, and

Department of Biostatistics, University of Texas, M.D. Anderson Cancer Center, 1515 Holcombe Blvd., Box 447, Houston, TX 77584, USA

P. Müller

Department of Biostatistics, University of Texas, M.D. Anderson Cancer Center, 1515 Holcombe Blvd., Box 447, Houston, TX 77584, USA

L.G. Leon-Novelo: luis@stat.ufl.edu; X. Zhou: xzhou@gene.com; B. Nebiyu Bekele: bbekele@mdanderson.org; P. Müller: pmueller@mdanderson.org

Summary

This article addresses modeling and inference for ordinal outcomes nested within categorical responses. We propose a mixture of normal distributions for latent variables associated with the ordinal data. This mixture model allows us to fix without loss of generality the cutpoint parameters that link the latent variable with the observed ordinal outcome. Moreover, the mixture model is shown to be more flexible in estimating cell probabilities when compared to the traditional Bayesian ordinal probit regression model with random cutpoint parameters. We extend our model to take into account possible dependence among the outcomes in different categories. We apply the model to a randomized phase III study to compare treatments on the basis of toxicities recorded by type of toxicity and grade within type. The data include the different (categorical) toxicity types exhibited in each patient. Each type of toxicity has an (ordinal) grade associated to it. The dependence among the different types of toxicity exhibited by the same patient is modeled by introducing patient-specific random effects.

Keywords

Adverse events; Clinical trial; Gibbs sampling; Latent variable; Ordinal data nested within categories; Probit model

1. Introduction

We address modeling and inference for data that include ordinal outcomes nested within categories. The data format can alternatively be seen as multivariate ordinal data with each

7. Supplementary Materials

Web Appendices referenced in Sections 3 and 5 are available under the Paper Information link at the Biometrics website <http://www.biometrics.tibs.org>.

dimension of the multivariate outcome corresponding to one level of a categorical variable. The motivating application is to model adverse event (toxicity) data in clinical trials. Toxicity type and severity are usually recorded as categorical and ordinal outcome, respectively. In a randomized phase III study, in addition to the efficacy of the study agent, investigators and regulators are also interested in learning about the toxicity profile of the study agent. Traditionally, simple descriptive statistics such as cross-tabulations have been provided. However, this purely descriptive approach fails to offer an in-depth understanding of how the treatment affects both the toxicity type and the severity associated with a specific type of toxicity.

The multinomial probit (MNP) model (Aitchison and Bennett, 1970) and the multinomial logit model are popular model choices for implementing regression for categorical outcomes. However, the computational burden associated with implementing full posterior inference hinders the routine application of these models in applied work. In recent years, there have been some advances using classical and Bayesian approaches. In particular, the method of simulated moments by McFadden (1989), and Gibbs sampling with data augmentation as discussed in Albert and Chib (1993) and McCulloch and Rossi (1994), have made the required computations in the multinomial probit model more practical.

For inference with ordinal data, many authors have proposed methods in a classical (McCullagh, 1980) and Bayesian (Albert and Chib, 1993; Doss, 1994; Cowles, 1996) framework. A natural way to model ordinal data is to introduce an underlying continuous latent variable. The ordinal outcome is linked with the latent variable through a set of cutpoints. The probability of an ordinal outcome is represented by the probability that this latent continuous variable falls within a given interval defined by the cutpoints. The ordinal probit model is characterized by the assumption that this latent variable follows a normal distribution.

Albert and Chib (1993) proposed Bayesian inference for the ordinal probit regression parameters. The model includes a diffuse prior on the cutpoint parameters. Cowles (1996) proposed improved posterior simulation with a hybrid Gibbs/Metropolis-Hastings sampling scheme which updates the cutpoint parameters jointly with the other parameters. This approach reduces the high auto-correlation and achieves practical convergence within a reasonable number of iterations of the Markov chain Monte Carlo (MCMC) simulation.

Chib and Hamilton (2000) model an ordinal variable with a mixture of normal probit model with random cutpoints. The ordinal probit is part of a much larger encompassing model where the ordinal variable represents a random treatment assignment. Albert and Chib (2001) exploit a latent random variable representation for an ordinal sampling model to define a sequential ordinal model for discrete-time survival data. Their approach is based on continuation probabilities and requires J latent variables for J levels of the ordinal response. No inference on cutpoints is needed under this approach.

In this article we propose a mixture model for ordinal data based on a single latent variable and without the need to estimate cutpoint parameters. We show that in the proposed mixture model, the cutpoints can be fixed without loss of generality. While standard ordinal models assume that the regression lines which characterize the ordinal outcomes are parallel (thus leading to the proportional odds assumption when using a logistic link), our model is flexible in the sense that it is able to fit data when this parallel regressions assumption is violated. This is especially attractive when modeling multiple ordinal variables. We avoid the need to check that each variable meets the parallel regressions assumption. A similar model with a mixture of normals distribution for the latent probit score is introduced in Kottas *et al.* (2005). They use a non parametric mixture. Our model differs by using a finite

mixture, introducing a regression on covariates and using patient-specific random effects. Besides these innovations, the most important contribution of this paper is the application to inference for adverse event rates.

The rest of the paper is organized as follows. In Section 2, we describe a phase III clinical trial. In Section 3, we introduce a joint multinomial and ordinal probit model to estimate the cell probabilities of multiple categorical outcomes with different ordinal levels nested in each categorical outcome. The prior specifications and posterior inference are discussed in Section 4. We illustrate properties of the model by applying the model to a simulated dataset and data from a phase III clinical trial. Results are presented in Section 5. A summary and discussion of possible extensions are presented in Section 6. The web address of the web-based supplementary materials for this article is included in Section 7.

2. A Phase III Clinical Trial

Studies have suggested that retinoid chemoprevention may help control second primary tumors, disease recurrence, and mortality for stage I non-small-cell lung cancer (NSCLC) patients. A National Cancer Institute (NCI) intergroup phase III trial of 1166 patients with pathologic stage I NSCLC was conducted to validate the efficacy of isotretinoin, a retinoid hypothesized to have chemopreventive properties. Patients were randomly assigned to receive either placebo or isotretinoin (30 mg/day) for 3 years in a double-blinded study. Patients were stratified at randomization by tumor stage, histology, and smoking status. A total of 589 patients received isotretinoin while the remaining patients received placebo.

One of the objectives of the phase III study is to assess the treatment effect on different types of toxicity and the grade associated with each of them. In this paper we focus on inference related to this objective only. The treatment-related toxicities include: cheilitis (inflammation of the lips), conjunctivitis (inflammation of the membrane lining the eyelids), arthralgia (pain in the joints), hypertriglyceridemia (excess of triglycerides in the blood), headache, and abnormal vision. With the exception of hypertriglyceridemia, toxicity was graded using the Common Toxicity Criteria, a toxicity scale used by the NCI for Adverse Events. Triglyceride toxicity was recorded on a scale of 1, 2 or 3 depending on the triglyceride level relative to normal, with 1 indicating between 2.5 and 5 times the normal level, 2 for between 5 and 10 times, and 3 for beyond 10 times or if a patient experienced complications (e.g., pancreatitis) at any grade of triglyceride toxicity.

If patients experienced multiple incidents of the same toxicity, only one incident at the highest grade was counted. When multiple different toxicities are reported for the same patient, the corresponding observed toxicity levels are expected to correlate. We will introduce the desired correlation with patient-specific random effects in the model. Reported toxicity rates are per patient, i.e., toxicity rates are probabilities that a patient reports a certain type of toxicity and grade. A summary of the data is shown in Table 1.

3. A Hierarchical Model for Ordinal Data Nested within Categories

For each recorded adverse event, the data report one variable. This variable is an ordinal outcome, z_{ij} , which reports the grade at which the j^{th} categorical outcome was observed on the i^{th} individual, $i \in \{1, \dots, n\}$ and $j \in \{1, \dots, J\}$, where, respectively, n and J denote the total number of patients and the total number of different toxicity types recorded in the study. The variable z_{ij} takes values $k = 0, 1, \dots, K_j$. The observation $z_{ij} = k$ indicates that the i^{th} patient exhibited the toxicity of type j at grade k . The additional grade $k = 0$ is used to indicate that toxicity j was not recorded for patient i .

Let X be a $(n \times H)$ matrix of possible regressors, with the i^{th} row, x_i , recording H covariates for the i^{th} patient, $i = 1, \dots, n$. For inference on a dose effect one could use $x_{i1} = 0$ when the drug is not present, $x_{i1} = 1$ for the lowest dose of the drug, $x_{i1} = 2$ for the second lowest dose and so on. In our specific example, we have a dichotomous covariate. We use $x_{i1} = 1$ when patient i is treated with isotretinoin, and $x_{i1} = -1$ for placebo. Considering just the treatment effect we have $H = 1$ and the i^{th} row of the covariate matrix, X , is just x_{i1} . In general, the covariates could be occasion-specific and indexed by patient i and toxicity j . We only use patient-specific covariates in the application and proceed therefore for simplicity with patient-specific covariates only.

We set up an ordinal probit regression for z_{ij} on covariates x_i . The cell probability $Pr(z_{ij} = k)$ is represented as the probability that a continuous latent variable v_{ij} falls into the interval $(\theta_{kj}, \theta_{k+1,j})$. A patient specific random effect r_i induces correlation across all toxicity observations for the same patient. Multiple cutpoints are required for the K_j ordinal outcomes:

$$z_{ij} = k \quad \text{if} \quad \theta_{kj} \leq v_{ij} < \theta_{k+1,j} \quad \text{for} \quad k = 0, 1, \dots, K_j. \quad (1)$$

$$v_{ij} = x_i^T \beta_j + r_i + \xi_{ij}, \quad (2)$$

where,

$$r_i \sim N(0, \sigma_r^2), \quad \xi_{ij} \sim \sum_{g=1}^G p_{jg} N(\mu_{jg}, \sigma_\xi^2) \quad (3)$$

and

$$(p_{j1}, p_{j2}, \dots, p_{jG}) \sim \text{Dirichlet}(\alpha_{j1}, \alpha_{j2}, \dots, \alpha_{jG}), \quad j = 1, \dots, J. \quad (4)$$

Here, β_j parameterize the ordinal probit model for z_{ij} . Notice that β_j does not include an intercept parameter. An intercept is already implicitly included in μ_{jg} . Consider the implied model for each v_{ij} after marginalizing with respect to r_i . The marginal distribution for each

$v_{ij}, j = 1, \dots, J$, is a mixture of normal distributions sharing the scale parameter $(\sqrt{\sigma_r^2 + \sigma_\xi^2})$, with distinct location parameters $(x_i^T \beta_j + \mu_{jg})$, fixed number of components (G) and weights (p_{jg}). It can be shown that without loss of generality we can fix the cutpoints θ_{jk} when working with the mixture of normal model in (3) instead of a single normal. See also the discussion below. While we assume that G is fixed, for $j = 1, \dots, J$, in (3), using different G_j for each toxicity is possible without additional complications.

The mixture model can alternatively be written as a hierarchical model by introducing a latent indicator variable w_{ij} . Specifically, conditional on $w_{ij} = g$, β_j and r_i , the latent variable v_{ij} follows a normal distribution: $v_{ij} | w_{ij} = g, \beta_j, r_i, \mu_{jg} \sim N(x_i^T \beta_j + r_i + \mu_{jg}, \sigma_\xi^2)$. The prior probability for $w_{ij} = g$ is $Pr(w_{ij} = g) = p_{jg}$. Let $\Phi(\cdot)$ denote the standard normal cdf. Marginalizing with respect to both r_i and the latent variable v_{ij} , we have:

$$Pr(z_{ij}=k|w_{ij}=g, \beta_j, \mu_{jg}) = \Phi\left(\frac{\theta_{k+1,j} - x_i^T \beta_j - \mu_{jg}}{\sqrt{\sigma_r^2 + \sigma_\xi^2}}\right) - \Phi\left(\frac{\theta_{k,j} - x_i^T \beta_j - \mu_{jg}}{\sqrt{\sigma_r^2 + \sigma_\xi^2}}\right) \quad (5)$$

$$Pr(w_{ij}=g) = p_{jg}. \quad (6)$$

For each category j the probability of a response at level k is

$$\pi_{jk} \equiv Pr(z_{ij}=k) = \sum_{g=1}^G Pr(z_{ij}=k|w_{ij}=g) p_{jg} \quad (7)$$

For reasons of identifiability we fix the variance of the normal kernels in (3). We recommend $\sigma_\xi^2=1$ and $\sigma_r^2=4$. See the argument below.

For later reference we state the joint probability model. Let $\mathbf{z} = (z_{ij}: i = 1, \dots, n, j = 1, \dots, J)$ denote the data. Let f_{β_j} denote the prior distribution for β_j . Assume that μ_{jg} for $j = 1, \dots, J$ and $g = 1, \dots, G$ are *a priori* independent, that for each j , given the imputed hyperparameter $\varphi_j, \mu_{j1}, \dots, \mu_{jG}$ is a random sample from $f_\mu(\cdot|\varphi_j)$ and that $\varphi_1, \dots, \varphi_J \stackrel{\text{iid}}{\sim} f_\varphi$. Let $\boldsymbol{\beta} = (\beta_1, \dots, \beta_J)$ denote all probit regression coefficients, and similarly for $\boldsymbol{\mu}, \mathbf{r}, \mathbf{v}, \mathbf{w}$ and $\boldsymbol{\varphi}$. Let $N(x|m, s)$ indicate a normal pdf with mean m and variance s evaluated at x . The joint distribution of $\mathbf{z}$ and all the model parameters is

$$\begin{aligned} p(\mathbf{z}, \mathbf{v}, \mathbf{r}, \boldsymbol{\beta}, \boldsymbol{\mu}, \mathbf{w}, \boldsymbol{\varphi}) = & \prod_{i=1}^n \prod_{j=1}^J 1[\theta_{z_{ij},j} < v_{ij} < \theta_{z_{ij}+1,j}] \prod_{i=1}^n \prod_{j=1}^J N(v_{ij}|x_i^T \beta_j + r_i + \mu_{j,w_{ij}}, \sigma_\xi^2) \\ & \times \prod_{j=1}^J f_{\beta_j}(\beta_j) \prod_{i=1}^n N(r_i|0, \sigma_r^2) \\ & \times \prod_{j=1}^J \left\{ \prod_{g=1}^G p_{jg}^{\#\{i:w_{ij}=g\}} \prod_{g=1}^G p_{jg}^{\alpha_{jg}-1} \right\} \prod_{j=1}^J \left\{ f_\varphi(\varphi_j) \prod_{g=1}^G f_\mu(\mu_{jg}|\varphi_j) \right\} \end{aligned} \quad (8)$$

To show that the cutpoints θ_{jk} in (1) can be fixed, consider the following simplified version

of the right side in (3). Let $v \sim \sum_{g=1}^G p_g N(\mu_g, \sigma^2)$ and define $\pi_k \equiv Pr(z = k) = Pr(\theta_k \leq v < \theta_{k+1})$, $k = 0, 1, \dots, K$. We show by a constructive argument that an appropriate choice of $(G, p_g, \mu_g, g = 1, \dots, G)$ can approximate an arbitrary set of desired cell probabilities $(\pi_0^*, \pi_1^*, \dots, \pi_K^*)$. In particular, the parallel regression assumptions of the probit model is not required. A similar argument was used in Kottas *et al.* (2005) for infinite Dirichlet process mixtures of normal distributions. Consider a mixture of normal distributions with $G \geq K$ components. Place one component of the mixture into each interval $[\theta_k, \theta_{k+1})$ by choosing $\mu_k = \frac{1}{2}(\theta_k + \theta_{k+1})$, and set $p_k = \pi_k^*$. Specify σ such that $1 - \varepsilon$ of the probabilities of each kernel is between the adjacent cutpoints. This trivially achieves $|\pi_k - \pi_k^*| < \varepsilon$ for $k = 0, 1, \dots, K$. Therefore, the cutpoints θ_{jk} in (1) can be fixed without loss of generality.

We recommend as a default choice of cutpoint parameters θ_k : $\theta_0 = -\infty$, $\theta_{K+1} = \infty$, $\{\theta_1, \dots, \theta_K\} = \{0, \pm 4, \pm 8, \dots\}$. For reasons of identifiability, we suggest fixing $\sigma = 1$. This choice

implicitly restricts cell probabilities $\pi_2, \dots, \pi_{K-1}$ to be at most 0.95. The first and last cell probabilities, π_1 and π_K , are unrestricted. This is important in the context of the later application to adverse event rates, when the first level of the ordinal outcome corresponds to no toxicity, which is often greater than 0.95. If larger cell probabilities are desired for intermediate outcomes, the widths between θ_k and θ_{k+1} can be increased or decreased accordingly. Although the proposed model could represent any configuration of cell probabilities with $G \geq K$, we recommend $G < K$ to avoid identifiability issues. Also, $G < K$ suffices to represent the cell probabilities for most applications. Figure 1 illustrates the model for one toxicity with three possible outcomes, i.e., $z_{ij} \in \{0, 1, 2\}$ and $K_j = 2$. The figure shows how the proposed model with $G = 2$ fits the cell probabilities π_{1k} . The figure shows the mixture of normal distribution of the latent random variable, v_{i1} , under two values of the covariate $x_i = -1, 1$. In both mixtures, the darkly shaded, lightly shaded and white areas correspond to the probability of the ordinal outcome taking the values 0, 1 and 2, respectively. Notice that this particular set of cell probabilities does not satisfy the parallel regression assumption. Therefore, these probabilities cannot be represented by a model with a unimodal distributed latent random variable and random cutpoints. For more detail see point 1 in Web Appendix A. Finally, we fix σ_r^2 at $\sigma_r^2=4$, implying non-negligible prior probability for random effects r_i to be in a range that covers several cutpoints θ_k .

4. Priors, Posterior and Simulation Scheme

We use conjugate priors for the probit regression parameters, centering the prior to represent the prior judgment about the marginal prevalence of the outcomes and the effects of the covariates. We use

$$f_{\beta_j}(\beta_j) \equiv N(m_{\beta_j}, \sigma_{\beta}). \quad (9)$$

As default choice for G , the size of the mixture model in (3), we suggest $G = K - 1$. Our recommendation is based on empirical evidence. On one hand, small values of G create faster mixing Markov chains, but may not be sufficient to fit the data. On the other, large values of G may overparameterize the model leading to poorly mixing Markov chains. Alternatively, using reversible jump MCMC (Green, 1995), G could be included in the parameter vector and estimated as part of the inference. But since the parameters of interest are the cell probabilities π_{jk} , and inference on mixture-specific parameters is not of interest we prefer the approach with fixed large values of G . Formally implied inference on the parameters of the mixture model, including μ_{jg} and p_{jg} , should not be interpreted. Problems related to label switching (arbitrary permutation of the terms in the mixture) and node duplication (replicating essentially identical terms) make the posterior distribution on μ_{jg} and p_{jg} meaningless to interpret.

We assume that w_{ij} takes on discrete values 1, 2, ..., G with prior probability $p_{j1}, \dots, p_{jG}$, respectively. For the location parameter (μ_{jg}) in the components of the mixture of normal model (3), we use independent normal priors $f_{\mu}(\mu_{jg}|\varphi_j) \equiv N(\varphi_j, \sigma_{\mu}^2)$ with a conjugate hyperprior $f_{\varphi_j}(\varphi) \equiv N(0, \sigma_{\varphi}^2)$.

Keeping in mind the default suggestion for the cutpoints θ_k we recommend $\sigma_{\mu} \approx 1/\sqrt{J}\Sigma_j K_j$, i.e., half the span from the first to last cutpoint, averaging across all toxicities.

An investigator might be interested in assessing how different dose levels affect toxicity grade. Our model may be used in this context. For a cytotoxic agent, it is usually assumed that a higher dose incurs worse toxicity. The parameter β_j , the dose effect on the toxicity

grade, may be restricted to be positive when there are only two dose levels and the lower dose group is the reference group. When there are M dose groups, β_j becomes an $(M - 1)$ -dimensional vector. In this case one could enforce monotonicity of toxicity with increasing dose by introducing an order constraint on β_j as follows: assuming that the lowest dose is the reference group and the highest dose group is group M , monotonicity can be represented as $\beta_{(M-1)j} > \beta_{(M-2)j} > \dots > \beta_{1j}$. This assumption guarantees *a priori* that a higher dose incurs worse toxicity.

All full conditional posterior distributions are derived from the joint probability model (8). Since we adopt conjugate priors for all parameters, all full conditional posterior distributions have tractable closed forms, allowing straightforward implementation of Markov Chain Monte Carlo (MCMC) posterior simulation using a Gibbs sampler. We start with initial values for the latent variables $\mathbf{v}$ and $\mathbf{w}$. The values for $\mathbf{v}$ need to comply with (1). One iteration of the Gibbs sampler is described by the following transition probabilities:

- a. For each $j = 1, \dots, J$, we first marginalize with respect to φ_j (recall that φ_j is the prior mean for μ_{jg}). Then, sample $(\beta_j, \mu_{j1}, \dots, \mu_{jG})$ from the joint complete conditional posterior distribution. That is, from the posterior distribution for the linear regression:

$$v_{ij} - r_i = x_i^T \beta_j + \sum_{g=1}^G 1(w_{ij}=g) \mu_{jg} + \varepsilon_i, \quad \text{for } i=1, \dots, n,$$

with response $y_i^* = v_{ij} - r_i$. The residuals are $\varepsilon_1, \dots, \varepsilon_n \stackrel{\text{iid}}{\sim} N(0, \sigma_\varepsilon^2)$ and the prior is the multivariate normal distribution resulting from the product of (9) and the prior distribution of $(\mu_{j1}, \dots, \mu_{jG})$ when marginalizing with respect to φ . The latter is a multivariate normal distribution of dimension G with mean zero and covariance matrix $\sigma_\mu^2 I_G + 1_G 1_G^T \sigma_\varphi^2$, where I_G is the identity matrix of dimension G and 1_G is the column vector of length G with all entries equal to 1.

- b. For each $i = 1, \dots, n$, we sample r_i from the complete conditional posterior distribution, or, equivalently, from the posterior distribution for a linear regression with response $y_j^{**} = v_{ij} - (x_i^T \beta_j + \mu_{t, w_{ij}})$,

$$y_j^{**} = r_i + \varepsilon_j, \quad \text{for } j=1, \dots, J,$$

with $\varepsilon_1, \dots, \varepsilon_J \stackrel{\text{iid}}{\sim} N(0, \sigma_\varepsilon^2)$ and prior $r_i \sim N(0, \sigma_r^2)$.

- c. We update β_j with a random walk Metropolis-Hasting transition probability. We generate $\tilde{\beta} \sim N(\beta_j, c^2)$ where $c > 0$ and compute the posterior distribution of β_j (marginalizing with respect to $\mathbf{w}$):

$$p(\beta_j | \mu, \mathbf{r}, \mathbf{p}, \mathbf{z}) \propto f_{\beta_j}(\beta_j) \prod_{i=1}^n \left[\Phi_{x_i^T \beta_j \sigma_\varepsilon^2} \left(\theta_{z_{ij}+1, j} - r_i - \sum_{g=1}^G p_{jg} \mu_{jg} \right) - \Phi_{x_i^T \beta_j \sigma_\varepsilon^2} \left(\theta_{z_{ij}, j} - r_i - \sum_{g=1}^G p_{jg} \mu_{jg} \right) \right]$$

where $\Phi_{\mu \sigma^2}(x)$ represents the normal cdf with mean μ and variance σ^2 evaluated at x . Let $A = p(\tilde{\beta} | \mu, \mathbf{r}, \mathbf{p}, \mathbf{z}) / p(\beta_j | \mu, \mathbf{r}, \mathbf{p}, \mathbf{z})$. With probability $\min\{1, A\}$ we replace β_j by $\tilde{\beta}$.

- d. The latent indicator variables w_{ij} in equation (6) are sampled from the complete conditional posterior distribution (marginalized with respect to z_{ij}):

$$Pr(w_{ij}=g|\beta_j, r_i, \mu_{jg}, p_{jg}) \propto p_{jg} [\Phi_{\mu_{jg}, \sigma_\xi^2}(\theta_{z_{ij}+1, j} - x_i^T \beta_j - r_i) - \Phi_{\mu_{jg}, \sigma_\xi^2}(\theta_{z_{ij}, j} - x_i^T \beta_j - r_i)].$$

- e. For each j , we update $(p_{j1}, \dots, p_{jG})$ from

$$f(p_{j1}, \dots, p_{jG} | w_{1j}, \dots, w_{nj}) = \text{Dirichlet}(\alpha'_{j1}, \dots, \alpha'_{jG}),$$

$$\text{with } \alpha'_{jg} = \alpha_{jg} + \#\{i: w_{ij}=g\}.$$

- f. The latent variables v_{ij} are updated by draws from the truncated normal distribution

$$f(v_{ij}|\beta_j, r_i, w_{ij}=g, \mu_{jg}, z_{ij}=k) \propto N(v_{ij}|x_i^T \beta_j + r_i + \mu_{jg}, \sigma_\xi^2) I(\theta_{kj} \leq z_{ij} < \theta_{k+1, j}).$$

Finally, for each toxicity type and grade, we evaluate the posterior probability of toxicity for a future patient. For each type of toxicity, using the equations (5) and (7), the posterior probability $\pi_{jk} \equiv Pr(z_{fj} = k | \mathbf{z})$, $f = n + 1$, that a future patient with covariate vector x_f exhibits the toxicity j at level k is estimated from the Gibbs sampler outputs. Let η^m denote the imputed value of the generic parameter η after m iterations of the Gibbs sampler. We report

$$\pi_{jk} = \sum_{g=1}^G Pr(z_{fj}=k | w_{fj}=g, \mathbf{z}) Pr(w_{fj}=g | \mathbf{z}) \\ \approx \frac{1}{M} \sum_{m=1}^M \sum_{g=1}^G \left\{ \Phi \left(\frac{\theta_{k+1, j} - x_f^T \beta_j^m - \mu_{jg}^m}{\sqrt{\sigma_r^2 + \sigma_\xi^2}} \right) - \Phi \left(\frac{\theta_{kj} - x_f^T \beta_j^m - \mu_{jg}^m}{\sqrt{\sigma_r^2 + \sigma_\xi^2}} \right) \right\} p_{jg}^m,$$

where M is the total number of MCMC iterations retained after an initial burn-in.

5. Applications

5.1 A Simulated Dataset

We use a simulated dataset to validate the model. A total of $n = 1000$ subjects were assigned into two groups, A and B , of equal size. For each subject i , there were four ($J = 4$) ordinal outcomes labeled $z_i = (z_{i1}, \dots, z_{i4})$ with four ($K = 3$) possible values, i.e., $z_{ij} \in \{0, 1, 2, 3\}$. The observations z_i were generated according to the model defined by (1)–(3) but with r_i generated from a mixture of normal distributions, $r_i \sim 0.75 N(0, 1) + 0.25 N(4, 1)$. In (1) we fixed the values of the cutpoints, θ_{jk} , as $(\theta_{0k}, \theta_{1k}, \dots, \theta_{4k}) = (-\infty, -3, 1, 2, \infty)$ for all j . We deliberately chose cutpoints different from the default cutpoints that are used in the analysis model. In (2) we set $x_i = -1$ for group A and $x_i = 1$ for group B . The slope parameters in the probit regression were set to $(\beta_1, \dots, \beta_4) = (0, 0.5, 1, 1.5)^T$. In (3) we set the variance $\sigma_\xi^2 = 1$, the number of components in the normal mixture $G = 3$, the normal means $(\mu_{j1}, \mu_{j2}, \mu_{j3}) = (-5, 0, 3)$ and the mixture weights $(p_{j1}, p_{j2}, p_{j3}) = (0.9, 0.05, 0.05)$ for all j . The true cell probabilities are reported in Table 2.

We fit model (1)–(4) with priors for the parameters as described in Section 4. More specifically, we assume that the probit parameters in (2), β_j , follow the normal distribution specified in (9) with $m_{\beta_j} = 0$ and $\sigma_{\beta_j}^2 = 1$. In (3) we assume G , and weights distributed according to (4) with $\alpha_{j1} = \alpha_{j2} = 1$ for all j . The normal means are assigned conjugate priors,

$\mu_{jg} \sim N(\varphi, \sigma_\mu^2=16)$ and $\varphi \sim N(0, \sigma_\varphi^2=10^4)$. We use the default choices $\sigma_\xi^2=1, \sigma_r^2=4$ and $\sigma_\mu^2=16$. The values of the cutpoints in (1), θ_{jk} , were set to $(\theta_{j0}, \dots, \theta_{j4}) = (-\infty, -4, 0, 4, \infty)$ for all j .

We simulated a total of 110,000 iterations of the posterior MCMC scheme. After an initial burn-in of 10,000 iteration, the imputed parameters were saved after each 10th iteration, yielding to a posterior Monte Carlo sample size of 10,000. The marginal posterior probabilities for each combination of toxicity type and grade were estimated and compared with the true cell probabilities in Table 2. The model reports reasonable estimates of the cell probabilities; in 26 out of 32 cells the true cell probability is within the reported 95% central credible interval.

For comparison we also implemented an ordinal probit regression with random cutpoints, but a (single) normal distribution for the latent probit scores (Albert and Chib, 1993). For a fair comparison we included patient-specific random effects as in the proposed model. Thus the alternative model is (1)–(4) with $G = 1$ and random cutpoints θ_2 through θ_{K_j} ($\theta_1 = 0$ is fixed). We used the posterior MCMC implementation of Cowles (1996). We summarized the marginal posterior distributions for the cell probabilities applying this model in the same way as in Table 2 (not shown). We find that for only 10 out of the 32 cell probabilities the central 95% posterior credible intervals contain the simulation truth. To assess the efficiency of the posterior MCMC for the proposed model versus the ordinal probit regression we recorded serial autocorrelations of the Markov chain simulations. We find comparable values and briefly discuss them in point 2 of Web Appendix A. In summary, we conclude that the proposed model and the conventional ordinal probit regression lead to comparable results with slightly more flexibility of the proposed model. We caution against over-interpreting the comparison. The simulated data set is relatively large with $n = 1000$, and the reported comparison is based on only one simulated data set.

Finally we investigated robustness of posterior inference with respect to choices of the prior hyperparameters. To explore prior sensitivity we considered several alternative choices. The value of $G = 1$ was not enough to fit the data. Most of the central 95% credible intervals for the probabilities of toxicity, π_{jk} , did not contain π_{jk} . The inference when using $G = 3$ is essentially the same as when using $G = 2$. Shifting the values of all α_{jg} to either 1/3 or 3 did not change inference appreciably. Similarly, using a non informative prior, $p(\beta) \propto 1$, for the regression parameter did not substantially affect the inference. Overall, we found that the posterior estimates are quite robust with respect to prior specification.

5.2 A Phase III Clinical Trial of Retinoid Isotretinoin

We applied the proposed model for inference in the phase III clinical trial introduced in Section 2. In contrast to the ordinal probit model (Albert and Chib, 1993), the proposed model does not require the parallel regression assumption. In Web Appendix B is shown that the data does not support this assumption.

As in the simulation study, we chose $N(0, 1)$ priors for the ordinal probit parameters. The size of the mixture was fixed at $G = 2$ with equal *a priori* weights by setting $\alpha_{jg} = 1$. A vague hyperprior centered at 0 with the variance of $\sigma_\varphi^2=10^4$ was imposed on φ . The cutpoints were chosen following the default choices. The variances were set to the default choices $\sigma_\xi^2=1, \sigma_r^2=4$ and $\sigma_\mu^2=16$.

Saving every 10th iteration after a 10,000 iteration burn-in, a Monte Carlo posterior sample of size 10,000 was saved to estimate cell probabilities. Table 3 displays the estimated cell

probabilities together with central 95% credible intervals. Note the near zero probabilities for some of the higher toxicity grades. The proposed model is appropriate to handle sparse tables. The estimates formally confirm and quantify what is expected from inspection of the data. There were more incidences of cheilitis and conjunctivitis observed in the isotretinoin group than in the placebo group. Elevated triglyceride levels were found more frequently in the isotretinoin group than in the placebo group. In contrast, more patients experienced headache in the placebo group. Posterior inference confirms that adverse event rates under treatment and placebo differ significantly. Figure 2 summarizes posterior inference for the probit regression parameters. It is 99% certain that the treatment (isotretinoin) had an undesired (i.e., $\beta_j > 0$) effect on cheilitis, conjunctivitis and hyper-triglyceride.

The posterior distribution allows us to report coherent probabilities for any event of interest. In particular, we can report inference on joint and conditional probabilities of adverse events across different toxicities. For example, Table 4 reports conditional probabilities for each adverse event (at any grade) given an adverse event in another toxicity for the same patient. For comparison the table also reports the marginal probabilities (in the diagonal). The considerable variation of probabilities in each row confirms that the toxicities exhibited by the same patient are not independent. The inclusion of patient-specific random effects r_i was critical in fitting this data. For example, the first row reports that the probability that a patient exhibits abnormal vision is low marginally, but considerably increased when the patient has experienced fatigue or headache. For more discussion about the importance of the patient-specific random effect see Web Appendix C.

Figure 3 shows the estimated distribution for the underlying latent variable, v_{ij} , in equation (2) conditioned on $r_i = 0$. Let $\bar{\beta}_j = E(\beta_j | \mathbf{z})$, $\bar{p}_{jg} = E(p_{jg} | \mathbf{z})$ and $\bar{\mu}_{jg} = E(\mu_{jg} | \mathbf{z})$. The figure shows $x_i^T \bar{\beta}_j + \sum_g \bar{p}_{jg} N(\bar{\mu}_{jg}, 1)$, for $x = -1, 1$ and $j = 1 \dots, 7$.

6. Discussion

We have proposed a Bayesian hierarchical model to analyze ordinal data nested within categories. Our model characterizes the ordinal/categorical data structure by a variation of the ordinal probit model. We provide posterior summaries to assess treatment effects. In the phase III clinical trial example, traditional analysis might simply group the toxicities levels into two: no toxicity (0) and some toxicity (1+2+3+4), and then, apply a Chi-squared test or Fisher's exact test to compare the two treatment groups for a particular toxicity type instead. The proposed approach provides alternative model-based posterior inference. In accordance with the natural structure of the data, our model treats toxicity grade as an ordinal data. The proposed model accounts for the (high) dependence across different toxicities within the same patient. The proposed model allows for extensions to more complicated designs by appropriate changes in the linear model (2). For example, one could accommodate repeated observations of adverse events by replacing the latent variable v_{ij} in (2) by v_{ijh} for the h^{th} repeated observation of type of toxicity j for patient i , and defining a new set of random effects R_{ij} .

This model also has interesting applications in other areas such as health outcomes research and clinical trial design. For example, some studies have shown that even when treatments are known to be effective, many patients who could benefit from them are not getting these treatments. Beta blocker medication, given after heart attacks, can reduce mortality; blood-thinning medication can prevent stroke; and thrombolytic therapy given immediately after a heart attack can reduce the damage from the attack. The outcome instrument has focused on assessing the overall level of functioning after receiving the treatment conditional on patients' prognostic characteristics. The overall level of functioning is a quantified variable

on an ordinal scale. Therefore, by assessing the ordinal outcomes within each category, health outcome researchers will be able to identify and address the barriers to better care and, eventually, translate these findings into practical strategies to improve care.

Some limitations remain. One critical choice for the proposed model is the size of the mixture in modeling ordinal outcomes. We suggested as a rule of thumb to set the size of mixture, G , equal to the number ordinal levels minus two ($K - 1$). Alternatively, one could treat G as an unknown parameter and use reversible jump MCMC.

Also, the use of fixed cutpoints naturally precludes any extensions that rely on random cutpoints. For example, Albert and Chib (2001) propose a model to characterize discrete time survival outcomes. The model is based on the continuation-ratio ordinal regression model's of Agresti (2002). Our proposed model can not naturally be used for this application.

In summary, we have introduced an approach for flexible, model-based inference for the adverse events reported in a Phase III clinical trial. The model includes dependence across adverse events for the same patient. The computational effort of implementation is comparable to a traditional ordinal probit regression.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The toxicity dataset from the NCI intergroup lung cancer study was generously provided by Dr. J. Jack Lee from the Department of Biostatistics and Applied Mathematics and Dr. Scott Lippman from the Department of Clinical Cancer Prevention at the University of Texas, M.D. Anderson Cancer Center.

References

- Agresti, A. Categorical Data Analysis (Wiley Series in Probability and Statistics). Wiley-Interscience; 2002.
- Aitchison J, Bennett JA. Polychotomous quantal response by maximum indicant. *Biometrika*. 1970; 57:253–262.
- Albert J, Chib S. Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*. 1993; 88:669–679.
- Albert J, Chib S. Sequential ordinal modeling with applications to survival data. *Biometrics*. 2001; 57:829–836. [PubMed: 11550934]
- Chib S, Hamilton BH. Bayesian analysis of cross-section and clustered data treatment models. *Journal of Econometrics*. 2000; 97:25–50.
- Cowles MK. Accelerating Monte Carlo Markov chain convergence for cumulative-link generalized linear models. *Statistics and Computing*. 1996; 6:101–111.
- Doss H. Bayesian nonparametric estimation for incomplete data via successive substitution sampling. *The Annals of Statistics*. 1994; 22:1763–1786.
- Green PJ. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*. 1995; 82:711–732.
- Kottas A, Müller P, Quintana F. Nonparametric Bayesian modeling for multivariate ordinal data. *Journal of Computational and Graphical Statistics*. 2005; 14:610–625.
- McCullagh P. Regression models for ordinal data. *Journal of the Royal Statistical Society, Series B: Methodological*. 1980; 42:109–142.
- McCulloch R, Rossi PE. An exact likelihood analysis of the multinomial probit model. *Journal of Econometrics*. 1994; 64:207–240.

McFadden D. A method of simulated moments for estimation of discrete response models without numerical integration (STMA V31 2344). *Econometrica*. 1989; 57:995–1026.

Figure 1.

Illustration of the distribution of the latent variable v_{i1} in the model described in (1)–(3) when the covariate takes values -1 and 1 . Here we consider: $J = 1$ type of toxicity, no patient-specific random effect r_i , $G = 2$ components in the mixture of normals and three possible ordinal outcomes, i.e., $K = 2$. In both mixtures, the darkly shaded, lightly shaded and white areas correspond to the probabilities π_{1k} of the ordinal outcome taking the values 0 , 1 and 2 , respectively.

Figure 2. Boxplots of the simulated posterior samples of the ordinal probit model parameters (β_j 's). Boxes corresponding to samples of β_j 's with 0.01-quantile greater than zero are shaded.

Figure 3.

Estimated mixture of normal distribution of the latent variable v_{ij} conditioned on $r_i = 0$ for $j = 1, \dots, 7$. Shaded curve corresponds to placebo ($x = -1$) and dashed to isotretinoin ($x = 1$). Darkly and lightly shaded areas represent, respectively, the probability of no toxicity (π_{j0}) and toxicity at grade 1 (π_{j1}) under placebo.

Toxicity frequency for randomized eligible patients by study arms. In the placebo (Isotretinoin) group, 171 (427) out of 577 (589) patients exhibited some type of toxicity. In bold the proportion of patients in the study arm belonging to the cell.

Table 1

Toxic effect	No Tox	Placebo			
		G1	G2	G3	G4
Abnormal vision	565(0.979)	9(0.016)	0(0)	2(0.003)	1(0.002)
Arthralgia	548(0.95)	19(0.033)	10(0.017)	0(0)	-
Cheilitis	493(0.854)	76(0.132)	8(0.014)	0(0)	-
Conjunctivitis	530(0.919)	43(0.075)	3(0.005)	1(0.002)	-
Fatigue	558(0.967)	12(0.021)	5(0.009)	2(0.003)	-
Headache	554(0.96)	16(0.028)	3(0.005)	4(0.007)	-
Hyper-triglyceride	551(0.955)	22(0.038)	4(0.007)	0(0)	-
Isotretinoin					
Abnormal vision	579(0.983)	8(0.014)	1(0.002)	1(0.002)	0(0)
Arthralgia	544(0.924)	30(0.051)	10(0.017)	5(0.008)	-
Cheilitis	212(0.36)	245(0.416)	122(0.207)	10(0.017)	-
Conjunctivitis	449(0.762)	98(0.166)	31(0.053)	11(0.019)	-
Fatigue	572(0.971)	14(0.024)	3(0.005)	0(0)	-
Headache	580(0.985)	9(0.015)	0(0)	0(0)	-
Hyper-triglyceride	514(0.873)	64(0.109)	10(0.017)	1(0.002)	-

Table 2

Simulated data set, marginal posterior cell probabilities with central 95% credible intervals. In bold are the true cell probabilities.

Group A				
Ordinal level	0	1	2	3
T1	0.619(0.64) (0.576, 0.658)	0.299(0.274) (0.264, 0.337)	0.023(0.027) (0.017, 0.031)	0.059(0.059) (0.045, 0.074)
T2	0.65(0.683) (0.609, 0.689)	0.272(0.246) (0.237, 0.309)	0.028(0.021) (0.02, 0.036)	0.05(0.05) (0.037, 0.064)
T3	0.662(0.72) (0.62, 0.703)	0.253(0.22) (0.216, 0.29)	0.029(0.017) (0.019, 0.041)	0.056(0.042) (0.042, 0.072)
T4	0.736(0.757) (0.697, 0.774)	0.183(0.193) (0.15, 0.218)	0.036(0.016) (0.024, 0.049)	0.045(0.034) (0.032, 0.061)
Group B				
T1	0.646(0.64) (0.604, 0.685)	0.274(0.274) (0.239, 0.311)	0.022(0.027) (0.016, 0.03)	0.057(0.059) (0.044, 0.073)
T2	0.581(0.586) (0.538, 0.623)	0.335(0.306) (0.297, 0.372)	0.03(0.036) (0.023, 0.038)	0.054(0.071) (0.041, 0.069)
T3	0.52(0.517) (0.475, 0.564)	0.379(0.348) (0.341, 0.417)	0.036(0.049) (0.027, 0.045)	0.065(0.086) (0.05, 0.082)
T4	0.405(0.432) (0.361, 0.45)	0.469(0.4) (0.434, 0.502)	0.056(0.061) (0.045, 0.069)	0.069(0.107) (0.055, 0.086)

Table 3

Marginal posterior cell probabilities (central 95% credible intervals) of toxicity. The table only reports grades up to $k = 3$. The marginal probabilities for grade G4 abnormal vision are 0 (with 95% C.I. (0, 0.002)) under placebo, and 0.001 (with 95% C.I., (0, 0.002)) under isotretinoin.

Toxic effect	No Tox	G1	G2	G3
Placebo				
Abn. vision	0.974 (0.962, 0.985)	0.02 (0.011, 0.033)	0.002 (0.001, 0.005)	0.003 (0.001, 0.006)
Arthralgia	0.937 (0.918, 0.954)	0.042 (0.029, 0.057)	0.019 (0.013, 0.026)	0.002 (0.001, 0.005)
Cheilitis	0.808 (0.774, 0.84)	0.172 (0.145, 0.201)	0.02 (0.012, 0.03)	0 (0, 0)
Conjunctivitis	0.896 (0.872, 0.918)	0.077 (0.063, 0.093)	0.025 (0.015, 0.036)	0.001 (0, 0.002)
Fatigue	0.967 (0.952, 0.978)	0.024 (0.015, 0.035)	0.009 (0.005, 0.015)	0.001 (0, 0.001)
Headache	0.962 (0.945, 0.976)	0.029 (0.017, 0.044)	0.008 (0.004, 0.013)	0.001 (0, 0.003)
H.-triglyceride	0.971 (0.955, 0.983)	0.028 (0.017, 0.042)	0.001 (0, 0.003)	0 (0, 0)
Isotretinoin				
Abn. vision	0.972 (0.958, 0.983)	0.023 (0.013, 0.035)	0.002 (0, 0.004)	0.003 (0.001, 0.006)
Arthralgia	0.922 (0.9, 0.941)	0.054 (0.039, 0.073)	0.021 (0.014, 0.028)	0.003 (0.002, 0.006)
Cheilitis	0.398 (0.355, 0.443)	0.403 (0.365, 0.439)	0.177 (0.153, 0.203)	0.022 (0.015, 0.031)
Conjunctivitis	0.777 (0.739, 0.814)	0.163 (0.133, 0.196)	0.052 (0.04, 0.066)	0.007 (0.004, 0.011)
Fatigue	0.959 (0.943, 0.973)	0.029 (0.019, 0.042)	0.011 (0.006, 0.017)	0.001 (0, 0.002)
Headache	0.968 (0.953, 0.98)	0.024 (0.015, 0.036)	0.007 (0.004, 0.012)	0.001 (0, 0.002)
H.-triglyceride	0.851 (0.82, 0.879)	0.086 (0.064, 0.11)	0.056 (0.038, 0.075)	0.007 (0.002, 0.013)

Table 4

Probability of different toxicities (at any grade; rows) conditional on the same patient having experienced other toxicities (at any grade; columns). The two numbers x/y in each cell report probabilities under placebo/isotretinoin. For comparison the diagonal reports in bold the marginal probabilities of exhibiting each type of toxicity.

	Abnormal vision (AV)	Arthralgia (AR)	Cheilitis (CH)	Conjunctivitis (CO)	Fatigue (FA)	Headache (HE)	Hyper- triglyceride (HT)
AV	0.026/0.028	0.175/0.186	0.071/0.043	0.108/0.099	0.171/0.209	0.249/0.266	0.078/0.11
AR	0.428/0.518	0.063/0.078	0.12/0.107	0.162/0.199	0.227/0.316	0.326/0.391	0.134/0.201
CH	0.533/0.923	0.366/0.821	0.192/0.602	0.302/0.824	0.348/0.774	0.442/0.823	0.334/0.729
CO	0.438/0.78	0.267/0.566	0.163/0.305	0.104/0.223	0.256/0.511	0.344/0.603	0.176/0.411
FA	0.223/0.306	0.121/0.166	0.061/0.053	0.082/0.095	0.033/0.041	0.168/0.238	0.061/0.105
HE	0.366/0.304	0.195/0.161	0.087/0.044	0.125/0.087	0.189/0.186	0.038/0.032	0.095/0.098
HT	0.089/0.581	0.062/0.382	0.051/0.181	0.049/0.275	0.053/0.381	0.074/0.456	0.029/0.149