Random Effects in Censored Ordinal Regression: Latent Structure and Bayesian Approach

Minge Xie,1,* Douglas G. Simpson,2 and Raymond J. Carroll3
1Department of Statistics, Rutgers University, Piscataway, New Jersey 08855, U.S.A.
2Department of Statistics, University of Illinois, Champaign, Illinois 61820, U.S.A.
3Department of Statistics, Texas A&M University, College Station, Texas 77843-3143, U.S.A.
* email: mxie@stat.rutgers.edu

SUMMARY. This paper discusses random effects in censored ordinal regression and presents a Gibbs sampling approach to fit the regression model. A latent structure and its corresponding Bayesian formulation are introduced to effectively deal with heterogeneous and censored ordinal observations. This work is motivated by the need to analyze interval-censored ordinal data from multiple studies in toxicological risk assessment. Application of our methodology to the data offers further support to the conclusions developed earlier using GEE methods yet provides additional insight into the uncertainty levels of the risk estimates.

KEY WORDS: Censored response; Gibbs sampler/Metropolis algorithm; Hierarchical Model; Risk assessment

1. Introduction

Ordered categorical responses arise naturally in epidemiology, toxicological studies, and industrial experiments. Frequently, they are associated with some explanatory variables, in which case ordinal regression analysis is appropriate. Sometimes, due to insufficient information or knowledge, observations are censored. A few authors (e.g., Dickey, Jiang, and Kadane, 1987; Simpson et al., 1996) have investigated censored categorical data. Approaches to random effects in censored ordinal regression are presented in this paper.

Our study of random effects in ordinal regression is motivated by the need to combine information from multiple experiments in toxicological risk assessment. Here we focus on the analysis of existing data on acute inhalation exposure to a widely used industrial solvent, perchloroethylene (PERC), also known as tetrachloroethylene. It is known that exposure to PERC can cause varied responses, including central nervous system (CNS) effects such as dizziness, headache, nausea, unconsciousness, and even death. PERC has been the subject of a number of published inhalation exposure–response studies, but many of these studies are small and narrowly focused. For example, according to Guth and Raymond (1996), very few experiments on PERC vary both the exposure concentration and duration. The data available to us are combined from multiple studies. The information includes test species (rats, mice, humans), response effects, target endpoints (CNS effects, liver damage, distress of the respiratory tract), protocols, etc. Table 1 gives a profile of the data. Combining data from multiple experiments is one way to obtain information on both concentration and exposure. Analysis of data from different experiments also provides information on important interspecies differences and between-experiment variation.

The multiplicity of endpoints considered in toxicology experiments raises a challenging issue in combining information: How should we put different toxicological outcomes on a common scale for analysis? One theoretical approach is to view the collection of possible endpoints as a highly multivariate response. Although theoretically appealing, such an approach may lead to serious missing data problems in which the fraction of missing responses is considerably higher than the fraction of observed responses. Following Simpson et al. (1996), we instead employ ordinal severity scoring to provide comparable outcomes for analysis. The idea is to view each endpoint as an indirect measure of toxicological severity. A toxicologist scores each observed response on an ordinal toxicological severity scale. The ordinal responses are then analyzed across experiments. Guth et al. (1997) provided detailed information on how the responses in the PERC database were scored. Because of the multiple data sources, we need to handle correlation between observations. Simpson et al. (1996) considered a marginal analysis approach based on generalized estimating equations. Here we consider ordinal mixed model analysis, building the correlation structure into the model. Our purpose in performing this complementary analysis of the PERC data is to examine the robustness of the conclusions to alternative analysis. In addition, the mixed model analysis has the ability to model separate sources of variation due to different random effects, and it could potentially provide a more efficient analysis.

Random effects models have been the subject of a large body of statistical research. For categorical response data, one major difficulty lies in evaluation of the high-dimensional integrals that appear in the likelihood function. This difficulty has stimulated development of various approximations (cf.
Random Effects in Censored Ordinal Regression

Table 1
Profile information on perchloroethylene (PERC)

<table>
<thead>
<tr>
<th>Species</th>
<th>Number of experiments</th>
<th>Total number of observations</th>
<th>Number of censored observations</th>
<th>Median concentration (g/m³)</th>
<th>Median duration (hours)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>3</td>
<td>9</td>
<td>0</td>
<td>0.72</td>
<td>2.00</td>
</tr>
<tr>
<td>Mouse</td>
<td>13</td>
<td>287</td>
<td>63</td>
<td>25.12</td>
<td>4.00</td>
</tr>
<tr>
<td>Rat</td>
<td>7</td>
<td>84</td>
<td>12</td>
<td>25.70</td>
<td>4.00</td>
</tr>
</tbody>
</table>

Stiratelli, Laird, and Ware, 1984; Schall, 1991; Breslow and Clayton, 1993) and simulation-based approaches (cf., McCulloch [1994, 1997] and Bayesian Gibbs sampling approach by Zeger and Karim [1991]). We propose a new Gibbs sampling approach that is particularly convenient for handling random effects for censored ordinal observations. Our method is closely related to the approach of Albert and Chib (1993) for Bayesian analysis of binary and categorical regressions with probit or t-distribution links. Albert and Chib (1993) also extended their approach to a binary hierarchical model different from the models we discuss in this paper. We consider general link functions and censored observations. Also, we compare Bayesian confidence intervals with likelihood ratio-type intervals.

The rest of the article is organized as follows. Section 2 introduces a class of ordinal mixed models and the associated underlying latent structures. Section 3 presents the Bayesian formulation and proposes a Gibbs algorithm to fit the model (algorithm details provided in the Appendix). It also shows that the improper priors for variance components will result in improper joint posteriors, a result similar to that obtained under different model settings (see Natarajan and McCulloch, 1995; Hobert and Casella, 1996). Section 4 compares the Gibbs method with two existing non-Bayesian approaches: mixed (random intercept) model analysis with numerical integration and marginal (generalized estimating equation) analysis. Confidence intervals suitable for parameters with severe nonlinearity are discussed. Section 5 applies our methodology to PERC data. Section 6 provides further comments and discussions.

2. Models and Latent Structures
Let \( Y_{ij} \) be the \( j \)th response in the \( i \)th cluster, \( j = 1, \ldots, m_i \), \( i = 1, \ldots, n \). \( Y_{ij} \) takes one of the ordinal values \( \{0, 1, \ldots, S\} \). Let \( x_{ij} \) be a \( p \times 1 \) covariate vector of regression variables corresponding to fixed effects, \( \beta \), of interest. Also let \( z_{ij} \) denote a \( q \times 1 \) covariate vector of random effects \( b_i \). The observations between clusters are independent. Given \( b_i \), the responses \( Y_{ij} \) are assumed to be conditionally independent and satisfy the following mixed model:

\[
\Pr\{Y_{ij} \geq s \mid b_i\} = \begin{cases} \left( x_{ij}^T \beta + z_{ij}^T b_i - \alpha_s \right), & \text{if } s = 1, 2, \ldots, S, \\ 1, & \text{if } s = 0, \end{cases} \tag{1}
\]

where \( \alpha_1 < \cdots < \alpha_S \) and the inverse link function \( H \) is a cumulative distribution function (c.d.f.) of a continuous random variable. The random effects \( b_i \) are assumed to be samples from a multivariate normal, say \( N(0, D) \). If there is no random effect, i.e., \( b_1 = \cdots = b_n = 0 \), and the link function \( H^{-1}(p) = \log\{p/(1 - p)\} \), then (1) reduces to the proportional odds model (McCullagh and Nelder, 1989).

Sometimes the response is censored, i.e., the distinction between the categories for an outcome is missing, and what is reported is merely a set of categories (Dickey et al., 1987). We assume here that the censoring mechanism is defined intrinsically and hence is ignorable in the sense of Little and Rubin (1987). Since any reported set of ordinal categories can be expressed as a union of finite sets of neighboring categories, we focus, without loss of generality, on the case of interval-censored observations with one set of neighboring categories. For an interval-censored observation \( y_{ij} \), the exact value that \( y_{ij} \) takes is unknown, but the boundaries where \( y_{ij} \) lies are reported, say, \( u_{ij} \) and \( v_{ij} \) (\( u_{ij} \leq v_{ij} \)). Without loss of generality, \( u_{ij} \) and \( v_{ij} \) can take values from \( \{0, 1, \ldots, S\} \). If an observation is uncensored, \( u_{ij} = v_{ij} = y_{ij} \).

From (1), the censored \( Y_{ij} \) can be modeled by an interval probability,

\[
\Pr\{u_{ij} \leq Y_{ij} \leq v_{ij} \mid b_i\} = H\left(x_{ij}^T \beta + z_{ij}^T b_i - \alpha_{u_{ij}}\right) - H\left(x_{ij}^T \beta + z_{ij}^T b_i - \alpha_{v_{ij} + 1}\right), \tag{2}
\]

where we let \( \alpha_0 = -\infty \) and \( \alpha_{S+1} = \infty \). Let \( \theta \) denote all the unknown parameters. The likelihood function for observations from (1) or (2) is

\[
L(\theta \mid (u_{ij}, v_{ij}), i, j) = \prod_{i=1}^{n} \prod_{j=1}^{m_i} \Pr\{u_{ij} \leq Y_{ij} \leq v_{ij} \mid b_i\} f(b_i) db_i, \tag{3}
\]

where \( f(b_i) \) is the density function of the \( q \times 1 \) random vector \( b_i \). If \( q > 1 \), the primary difficulty in implementing likelihood inference is to evaluate the \( q \)-dimensional integral in the likelihood function and its first and second partial derivatives.

For models (1) and (2), there is an associated latent structure. Given random effects \( b_i \), define a continuous latent random variable \( W_{ij} \) that satisfies

\[
\Pr\left\{W_{ij} - x_{ij}^T \beta - z_{ij}^T b_i < t \mid b_i\right\} = 1 - H(-t). \tag{4}
\]

Let \( Y_{ij} = s \) if and only if \( W_{ij} \) lies in the interval of \( \{\alpha_s, \alpha_{s+1}\} \), where \( \alpha_1 < \cdots < \alpha_S \) is a set of ordered thresholds. \( Y_{ij} \) specified in this way satisfies (1). Further, \( W_{ij} \) censored within...
In a special case with deterministic scale parameter $Q_{ij}$ in the link function. Under this extension, $(4)$ can be expressed as, given $b_i$ and scalar $Q_{ij}$,

$$\Pr \left\{ \left( W_{ij} - x_{ij}^T \beta - z_{ij}^T b_i \right) / Q_{ij} \leq t \mid b_i, Q_{ij} \right\} = 1 - \tilde{H}(-t),$$

where the function $\tilde{H}$ is defined in such a way that $\tilde{H}(t)$ is a c.d.f. of a standard $t$-distribution (see Albert and Chib, 1997). If the $Q$'s are i.i.d.

$$\text{distributed (i.i.d.) random variables with c.d.f. } F(Q_{ij}).$$

If the $Q$'s are equal, they can be set to one, leading to $(4)$. If the $Q$'s are known but are not equal, this model belongs to the class of scaled link models discussed in Xie, Simpson, and Carroll (1993).

In our context, the class of conditionally independent hierarchical models, which may be thought of as the Bayesian analog of $(5)$, is the following. For $i = 1, \ldots, n$, $j = 1, \ldots, m_i$, and $s = 1, \ldots, S$,

$$Y_{ij} \mid W_{ij}, \alpha_1, \ldots, \alpha_S \sim \delta \left( \sum_{s=1}^{S} I(W_{ij} \geq \alpha_s) \right),$$

$$W_{ij} - x_{ij}^T \beta - z_{ij}^T b_i / Q_{ij} \mid \beta, b_i, Q_{ij} \sim 1 - \tilde{H}(-t), \quad \alpha_s \propto \pi(\alpha_s),$$

$$\beta \propto \pi(\beta), b_i \sim N_0(0, D), \quad Q_{ij} \sim F(Q_{ij}), \quad D^{-1} \sim \pi(D^{-1}),$$

where $\pi(\cdot)$ denotes a prior distribution of the corresponding random variable, $\delta(c)$ is the c.d.f. for a point mass (random) variable with mass at value $c$, and $I(c)$ is an indicator function that takes value one if set $C$ is true and value zero otherwise. If $Q_{ij}$ is deterministic, then $Q_{ij} \sim F(Q_{ij})$ does not need to appear in the model $(6)$ can be viewed as the Bayesian analog of $(4)$.

We present a simulation-based Gibbs sampling method to fit the Bayesian hierarchical model. As in Albert and Chib (1993), our algorithm cycles through $(\beta, b, D, W, \alpha)$ or $(\beta, b, D, W, \alpha, Q)$ if the $Q$'s are random variables, where $\alpha = (\alpha_1, \ldots, \alpha_S), b = (b_1^T, \ldots, b_S^T)^T$, and $W$ and $Q$ are defined in the Appendix. Let $y$ denote all responses. Due to conditional independence of certain related variables, the fully conditional posteriors, which are obtained by updating each Gibbs cycle, can be simplified to $f(\beta \mid b, W, Q), f(b \mid \beta, D, W, Q), f(D \mid b), f(W \mid \beta, b, \alpha, Q, y), f(\alpha \mid W, y), f(\alpha \mid W, y), \text{and } f(Q \mid \beta, b, W)$. Simulations from $f(D \mid b), f(W \mid \beta, b, \alpha, Q, y), f(\alpha \mid W, y), f(Q \mid \beta, b, W)$ are explicit. If $H(t) = \tilde{H}(t)$, the same goes for $f(\beta \mid b, W, Q)$ and $f(b \mid \beta, D, W, Q)$. In general, explicit forms of $f(\beta \mid b, W, Q)$ and $f(b \mid \beta, D, W, Q)$ may not be available. We propose a Metropolis algorithm to simulate the deviates. See the Appendix for detailed specification of the Gibbs formulas.

The Gibbs sampler developed is very efficient for models with probit, $t$-distribution, or logit links. Compared with the Gibbs sampling approach by Zeger and Karim (1991), our approach avoids the low efficiency of rejection sampling, but it needs some effort to simulate deviates from $f(W \mid \beta, b, \alpha, Q, y)$, which is explicit. One distinct advantage of our Gibbs approach is that it is particularly suitable for handling censored observations.

We adopt a one-run Gibbs sampling scheme suggested by Zeger and Karim (1991), and the one-run Gibbs Markov chain is viewed as having converged when the sample series become stable. The sample means and standard errors of the Gibbs samples after the convergence point are used to estimate the posterior expectations and associated standard deviations for $\alpha, \beta$, and $D$ and their functional.

As a general approach for dealing with the mixed effects models using a Bayesian paradigm, many authors have suggested using noninformative priors, either due to lack of prior information or because they are convenient (cf., Zeger and Karim, 1991). However, the use of improper priors for variance components may lead to nonexistence of the posteriors (Box and Tiao, 1973). Under a Gibbs sampling approach, it is possible that the fully conditional distributions are all proper but the joint posterior distribution does not exist (Natarajan and McCulloch, 1995; Hobert and Casella, 1996). Although model $(6)$ is more complicated than the models considered by Natarajan and McCulloch (1995) and Hobert and Casella (1996), the same argument applies here. For instance, in a special case of $(6)$ with $D = \sigma^2\tilde{D}$, where $\tilde{D}$ is known and positive definite matrix, we have Theorem 1.

**Theorem 1:** For a model $(6)$ with $D = \sigma^2\tilde{D}$ and $\pi(\sigma^2) \propto (\sigma^2)^{-\alpha+1}$, if $a \geq 0$ or $a < -(1/2)q$, then the posterior distribution $f(\beta, \sigma^2, W \mid y)$ does not exist.

**Proof.** We only need to determine when the denominator

$$\pi(\beta)\pi(\alpha)\pi(\sigma^2)\pi(\tilde{D}) \propto \sigma^2$$

can not converge (without of loss of generality, $Q$'s are ignored). Note that (i) $A_1 = f(y_i \mid W_i, \alpha)f(W_i \mid \beta, b_i)dW_i$, is bounded, and (ii) with $b_i$ reparameterized to $b_i = \sigma^{-1}b_i$ and when $\sigma$ is small, $A_1$ is also bounded away from zero (we explicitly assume that all $x_{ij}^T\beta$'s fall in a compact set). The rest of the proof follows the appendix of Natarajan and McCulloch (1995, p. 642).

**Theorem 1** implies that, if $a = 0$ and $\tilde{D} = I$, i.e., the priors of variance components are noninformative, the posterior distribution $f(\beta, \sigma^2, W \mid y)$ does not exist. In this case, the Gibbs sample series do not converge to a proper distribution. To avoid the problem, we use a proper prior for the variance components in our Gibbs sampler. Often the prior proper is selected such that it carries very little information related to the observed data. For example, in the model that was considered in Theorem 1, we may assume that the prior $\pi(\sigma^{-2})$ is a proper gamma($d, \lambda$) distribution with a large scale parameter.
1/\lambda. When proper priors are selected for the variance components, if \( H(t) \) is either the logistic or the probit function, it can be shown that \( f(\beta, \sigma^2, W \mid y) \) exists even when both \( \pi(\alpha_s) \) and \( \pi(\beta) \) are noninformative flat priors. In the example considered in Section 5, we select noninformative flat priors for both \( \alpha \) and \( \beta \) and a proper prior for the variance component.

4. Non-Bayesian Approaches and Likelihood Ratio-Type Confidence Intervals

The integrals in the likelihood function (3) and its derivatives can sometimes be evaluated through numerical integration. For instance, a special case of (1) is the random intercept model, where \( z_{ij} = 1 \) and \( b_i \) is a random variable,

\[
\Pr\{Y_{ij} \geq s \mid b_i\} = H\left(z_{ij}^T \beta + b_i - \alpha_s\right). \tag{7}
\]

In this case, the maximum likelihood estimates can be obtained through numerical integration, and we are able to compare the estimates from the Gibbs sampler with the numerical maximum likelihood estimates.

In a random effects model, the observations within the same cluster are correlated. An alternative, known as the marginal modeling approach, can be used in dealing with this type of data. Typically, (1) or (7) will be replaced by the fixed effects marginal probabilities,

\[
\Pr\{Y_{ij} \geq s\} = H\left(z_{ij}^T \beta - \tilde{\alpha}_s\right), \tag{8}
\]

and the generalized estimating equations (GEE) technique is often adopted in the model fitting (cf., Diggle, Zeger, and Liang, 1994). Let \( \hat{\theta} \) represent all the parameters that appear in (8). We propose a variation of the GEE approach to fit the marginal model (8): A generalized log-likelihood function is defined as a weighted sum of the individual log-marginal density functions,

\[
l(\hat{\theta} \mid y) = \sum_{i=1}^{n} \sum_{j=1}^{m_i} W_{ij} \log \left\{ H\left(z_{ij}^T \beta - \tilde{\alpha}_{uj}\right) - H\left(z_{ij}^T \beta - \tilde{\alpha}_{uj+1}\right) \right\},
\]

and the parameter estimates \( \hat{\theta} \) are obtained by maximizing \( l(\hat{\theta} \mid y) \). The \( W \)'s in \( l(\hat{\theta} \mid y) \) are the weights used to improve the efficiency of the estimators (see Simpson et al. [1996] for more details). As in the standard GEE approach, a sandwich formula \( H^{-1} J H^{-1} \) should be used to estimate the covariance matrix of the estimator, where

\[
H = -\mathbb{E}\left\{ \frac{\partial^2}{\partial \hat{\theta}^2} l(\hat{\theta} \mid y) \right\} \bigg|_{\hat{\theta} = \hat{\theta}}
\]

and

\[
J = \text{var}\left\{ \frac{\partial}{\partial \hat{\theta}} l(\hat{\theta} \mid y) \right\} \bigg|_{\hat{\theta} = \hat{\theta}}.
\]

Assume the random intercept model (7) is the correct model. If the probit link function \( \Phi^{-1}(t) \) are used in both models (7) and (8), it can be shown that the fixed effects in (8) are \((1+\sigma^2)^{-1/2} \) times the corresponding fixed effects in (7) (see Simpson et al., 1996). Because \( |e^1/(1+e^1) - \Phi(t/1.702)| < 0.001 \) (Baker, 1992, p. 92), if the logit link is used in both models (7) and (8), the multiplier between the fixed effects will be approximately \((1+\sigma^2)^{-1/2} \).

Often we need to provide confidence intervals for parameters that are nonlinear functions of the regression parameters, e.g., the effective dose, ED_{100}, defined in Section 5. In the Bayesian Gibbs sampling approach, the confidence limits at 100(1 - \alpha)% confidence level can be estimated by 100(\alpha/2)% and 100(1 - \alpha/2)% Gibbs sample percentiles. This is comparable to the likelihood ratio-type confidence intervals in non-Bayesian approaches, which are often used to make inferences about nonlinear parameters for which asymmetric confidence intervals may be appropriate.

In the random intercept model (7), the likelihood ratio-type confidence interval can be evaluated by numerical integration. However, under the marginal model (8), we propose an approach to compute generalized likelihood ratio-type intervals. Partition \( \hat{\theta} \) into \( \hat{\theta}_1 \) and \( \hat{\theta}_2 \) and let \( \tilde{\theta}_2(\hat{\theta}_1) = \arg\max_{\hat{\theta}_2} l(\hat{\theta}_1, \hat{\theta}_2) \).

We define \( D_n(\hat{\theta}_1) = -2\{l(\hat{\theta}_1, \tilde{\theta}_2(\hat{\theta}_1)) - l(\hat{\theta})\} \) as a generalization of the log-likelihood ratio statistic. Let \( q(\alpha, \hat{\theta}) \) be the 100(1 - \alpha)% quantile of the weighted sum of \( x^T_j \) distribution described in Theorem 2. An approximate 100(1 - \alpha)% generalized likelihood confidence interval (region) can be constructed as \( \{\hat{\theta}_1 : D_n(\hat{\theta}_1) \leq q(\alpha, \hat{\theta})\} \).

Adapting a result by Kent (1982) to the current context, we have Theorem 2.

**THEOREM 2:** Suppose \( \mathcal{H} \) and \( \mathcal{H}^{-1} J \mathcal{H}^{-1} \) are partitioned accordingly as the parameter \( \hat{\theta} \). Denote

\[
\mathcal{H}_{1112} = \mathcal{H}_{11} - \mathcal{H}_{12} \mathcal{H}_{22} \mathcal{H}_{11}.
\]

If (8) is true, then \( D_n(\hat{\theta}_1) \) is asymptotically distributed as \( \chi^2_q \), where \( W_1, \ldots, W_q \) are independent \( \chi^2_q \) distributions and \( \lambda_1, \ldots, \lambda_q \) are eigenvalues of \( \mathcal{H}_{11}^{-1} (\mathcal{H}^{-1} J \mathcal{H}^{-1})_{11} \).

**Proof.** For the basic idea of the proof, see Kent (1982).

5. Application to Perchloroethylene Data

In a risk assessment study on CNS effects of acute (single exposure up to 12 hours) inhalation exposure to PERC, a group of scientists from the U.S. Environmental Protection Agency conducted a literature search. They assembled a PERC database from all available data from published sources, proceedings, and technical reports (Guth and Raymond, 1996). These files were carefully screened, and poorly documented studies were removed. Since the original response measurements from the literature were quite different, in order to put these very different quantitative measurements on a common scale, the response measurements were reduced to ordinal severity categories of no effect (NE), adverse effect (AE), and severe effect (SE). The scoring was based on some previous work, including Guidelines for Developing of Emergency Exposure Levels for Hazard Substances (National Research Council, 1993) and on biological considerations. It was performed by a toxicologist.

Because there is often insufficient information to determine the biological significance in a particular response, substantial censoring occurs in severity scoring. For example, if the level of toxicity associated with a continuous response (such as percentage change in duration immobility) is in fact not
known or not agreed upon, severity score then is assigned to
two or more adjacent severity categories. Also, a few lethality
studies targeting fatal exposures provided little information
on further scoring on NE or AE. For such studies, censored
observations are assigned.

Simpson et al. (1996) analyzed the same data set, where
they adopted the marginal modeling approach to deal with
the correlation induced from the study effects. They specified
the probability of response \( Y_j \) for species \( j \) as

\[
\Pr(Y_j \geq s) = H(\bar{\gamma}_j + x_1\beta_{1j} + x_2\beta_2 - \bar{\alpha}_s), \quad (9)
\]

where \((x_1, x_2) = (\log_{10}\text{concentration}, \log_{10}\text{duration})\), \(H(t) = e^{t}/(1 + e^{t})\), and different species have their own regression
intercepts \( \bar{\gamma}_j \) and dose slope parameters \( \beta_{1j} \).

From our perspective, we are interested in exposure effects
of PERC on human beings and animals as well as any poten-
tial random study effects. Population average of individual
studies in the sense of Diggle et al. (1994, p. 133) shall not
be the focus. In our inference, the random effects approach
seems more appropriate. Indeed, our modeling procedure shall
start with a conditional model specific to each study since, as
pointed out by the associate editor, presumably each separate
investigator (of the disparate studies) would only model
data set in his or her own study. The studies in the PERC
database can be viewed as random selections from the large
pool of all potential studies, and variation induced by the
nonsystemic study effects should be modeled by random ef-
fects. In this paper, the results obtained in random effects
analysis are compared with the earlier marginal analysis to
investigate the robustness of the conclusion and to contrast
the approaches. The approach presented in this paper allows
Bayesian inference. It might also form the basis of higher level
inference across different chemicals in our future study.

We assume the response \( Y_{ij} \) in the \( i \)th study for the \( j \)th
species follows

\[
\Pr(Y_{ij} \geq s | b_i) = H(b_i + \gamma_j + x_1\beta_{1j} + x_2\beta_2 - \alpha_s), \quad (10)
\]

where \((x_1, x_2)\) varies for different observations, \(H(t) = e^{t}/(1+ e^{t})\), and the study random intercept \( b_i \sim N(0, \sigma^2)\). Note that
the variation induced by species differences is considered sys-
temic and is modeled by fixed effects.

We fit this random intercept model to the PERC data using
the Gibbs sampler proposed in Section 3. Noninforma-
tive flat priors are used for the fixed effects and a proper
\((\sigma^2, \sigma^2)\) with \( \sigma^2 = 30 \) is used for \((\sigma^2)^{-1}\). The
starting value for \( \sigma^2 \) is two and the starting values for other
fixed effects are \((1 + 2.91^2/1.702^2)^{-1/2} \) times the estimates ob-
tained by Simpson et al. (1996). The Gibbs chain becomes
stable after about a couple thousand iterations (time series
figures not shown in the paper), and the last 5000 samples
out of a total of 9000 Gibbs iterations are used to estimate
the parameters and their standard deviations.

Table 2 lists the numerical results. It also reports the results
from fitting marginal model (9) and the maximum likelihood
estimates from fitting the random intercept model (10) using
numerical integration. The estimates from the marginal model
are the same as those in Table 4 of Simpson et al. (1996) ex-
cept for a corrected value of \( \beta_{\text{conc, mouse}} \). From Table 2,
the Bayesian estimates are very close to the maximum likeli-
hood estimates, and the estimates from the marginal model
approach are about \((1 + 2.91^2/1.702^2)^{-1/2} = 1.98 \) times
less than the estimates obtained from the random effect analy-
sis. This agreement suggests that reasonable estimates of the
mixed model can be obtained from the marginal analysis.

One important concept often seen in the risk assessment
literature is \( 100q\% \) effective dose, \( ED_{100q} \), which is defined
as the hypothetical dose level that would lead to a \( 100q\% \)
rate of adverse or severe effect at a given duration and for a
given species. Under the random intercept model (10), \( ED_{100q}^{(j)} \)

### Table 2

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Marginal/GEE (standard errors)</th>
<th>Random effect/numerical integration</th>
<th>Random effect/Gibbs(^a) (standard errors)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha_1 )</td>
<td>17.89 (6.08)</td>
<td>32.71 (12.08)</td>
<td>35.17 (12.45)</td>
</tr>
<tr>
<td>( \alpha_2 )</td>
<td>20.95 (6.54)</td>
<td>37.70 (12.40)</td>
<td>40.33 (12.66)</td>
</tr>
<tr>
<td>( \gamma_{\text{mouse}} )</td>
<td>8.00 (3.78)</td>
<td>22.30 (12.75)</td>
<td>24.84 (13.30)</td>
</tr>
<tr>
<td>( \gamma_{\text{rat}} )</td>
<td>-11.73 (7.65)</td>
<td>-8.10 (12.18)</td>
<td>-5.60 (12.09)</td>
</tr>
<tr>
<td>( \beta_{\text{conc, human}} )</td>
<td>5.95 (1.93)</td>
<td>10.78 (3.82)</td>
<td>11.53 (4.12)</td>
</tr>
<tr>
<td>( \beta_{\text{conc, mouse}} )</td>
<td>2.68 (1.22)</td>
<td>2.88 (1.21)</td>
<td>2.79 (1.09)</td>
</tr>
<tr>
<td>( \beta_{\text{conc, rat}} )</td>
<td>6.98 (2.21)</td>
<td>9.52 (2.41)</td>
<td>9.57 (2.16)</td>
</tr>
<tr>
<td>( \beta_{\text{dur}} )</td>
<td>2.52 (1.02)</td>
<td>3.71 (1.05)</td>
<td>3.78 (1.03)</td>
</tr>
<tr>
<td>( \sigma )</td>
<td>—</td>
<td>2.91 (0.82)</td>
<td>3.20 (0.86)</td>
</tr>
</tbody>
</table>

\(^a\) Parameters estimates = posterior means. Standard error = square root of posterior variance.
Random Effects in Censored Ordinal Regression

Table 3
Comparison of log10(ED10) estimates (duration = 1 hour)
(95% confidence interval limits are in parentheses)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Marginal/GEE</th>
<th>Random effect/numerical integration</th>
<th>Random effect/Gibbsa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human, sev1</td>
<td>2.64</td>
<td>2.63</td>
<td>2.84</td>
</tr>
<tr>
<td></td>
<td>(1.95, 4.36)</td>
<td>(2.22, 3.23)</td>
<td>(2.24, 3.32)</td>
</tr>
<tr>
<td>Mouse, sev1</td>
<td>2.87</td>
<td>2.10</td>
<td>2.61</td>
</tr>
<tr>
<td></td>
<td>(2.32, 3.78)</td>
<td>(1.70, 2.73)</td>
<td>(0.04, 3.82)</td>
</tr>
<tr>
<td>Rat, sev1</td>
<td>3.93</td>
<td>3.83</td>
<td>4.00</td>
</tr>
<tr>
<td></td>
<td>(3.42, 4.62)</td>
<td>(3.54, 4.17)</td>
<td>(3.61, 4.31)</td>
</tr>
</tbody>
</table>

a Confidence limits are the 2.5 and 97.5% Gibbs sample percentiles.

\[ \beta_{ij}^{-1}\{H_{ij}(q) + \alpha_{s} - \gamma_{j} - \beta_{2} x_{2}\} \]

where \( H_{ij}(t) = \int H(t + \sigma u) d\Phi(u) \); under the marginal model (9),

\[ \text{ED}_{10}^{j}(H_{ij}(q) + \alpha_{s} - \gamma_{j} - \beta_{2} x_{2}) \]

If the random intercept model (10) is true, it can be shown mathematically that \( \text{ED}_{10}^{j} \) and \( \text{ED}_{10}^{j} \) are the same (see also Simpson et al., 1996).

For each of the last 5000 Gibbs samples, the ED10’s are computed at the unit duration (1 hour) for human, mouse, and rat species, and the corresponding 95% confidence interval limits are obtained using the 2.5 and 97.5% Gibbs sample percentiles. Table 3 compares the three unit ED10 estimates and confidence intervals with those obtained from the two non-Bayesian approaches discussed in Section 4. Again, the results seem consistent with each other. The confidence intervals used here exploit the invariance of the estimation criterion to the parameterization.

6. Conclusions

We studied random effects in censored ordinal regression and developed a Bayesian Gibbs method to fit the model. Our methodologies were applied to the PERC data, and the results were compared with those obtained from other analyses. The agreement between the maximum likelihood estimates and the Bayesian estimates from the Gibbs algorithm offers support, to some extent, for our random intercept model analysis. Compared with the delta-method standard errors commonly used in the method of generalized estimating equations, the confidence intervals used here exploit the invariance of the estimation criterion to the parameterization.

ACKNOWLEDGEMENTS

Xie’s research was supported by NSF grant DMS9803273. Simpson’s research was supported by NSF grant DMS9505290. Carroll’s research was supported by a grant from the National Institute of Environmental Health Sciences (P30-ES09106). The authors wish to thank James Calvin, the associate editor, and two referees for their helpful comments and suggestions.

RÉSUMÉ

Cet article discute de l’introduction d’effets aléatoires dans une régression ordinaire censurée et présente une approche d’échantillonnage de Gibbs pour ajuster le modèle de régression. Une structure latente et sa formulation bayésienne correspondante sont introduites pour traiter efficacement les observations ordinaires censurées et hétérogènes. Ce travail est motivé par le besoin d’analyser des données ordinaires censurées par intervalles, provenant de multiples études mesurant un risque toxicologique. Une application de notre méthodologie aux données permet de soutenir les conclusions déjà développées utilisant les méthodes GEE (Simpson et al., 1996) et permet d’avoir une idée plus claire sur les niveaux d’incertitude des estimations des risques.
REFERENCES


APPENDIX

Gibbs Sampler

Let \( \text{diag}(c_1, \ldots, c_k) \) represent a diagonal matrix with diagonal elements (or submatrices) equal to \( c_1, \ldots, c_k \). Define

\[
W_i = (W_{i1}, \ldots, W_{imi})^T,
\]

\[
X_i = (x_{i1}, \ldots, x_{imi})^T,
\]

\[
Z_i = (z_{i1}, \ldots, z_{imi})^T,
\]

and

\[
Q_i = \text{diag}(Q_{i1}, \ldots, Q_{imi}).
\]

for \( i = 1, \ldots, n \). Write

\[
W = \left( W_1^T, \ldots, W_n^T \right)^T,
\]

\[
X = \left( X_1^T, \ldots, X_n^T \right)^T,
\]

\[
Z = \text{diag} (Z_1^T, \ldots, Z_n^T),
\]

and

\[
Q = \text{diag} (Q_1, \ldots, Q_n).
\]

Simulating Deviates from \( f(\beta \mid b, W, Q) \)

In a special case with \( \bar{H}(t) = \Phi(t) \), if \( \pi(\beta) \) has a normal prior \( N(\beta, \Sigma) \),

\[
f(\beta \mid W, b, Q) \sim N \left( \beta, \left( X^TQX + \Sigma^{-1} \right)^{-1} \right),
\]

(A.1)

where \( \beta = \left( X^TQX + \Sigma^{-1} \right)^{-1} \left( \Sigma^{-1}\beta + X^TQ(W - Z)b \right) \). If \( \pi(\beta) \propto \text{constant} \), (A.1) still holds; however, \( \Sigma^{-1} \) should be replaced by zero.

In other cases, \( f(\beta \mid W, b, Q) \) may not have an explicit form. A Metropolis algorithm (cf., Tierney, 1994) is proposed to produce the random draws: We need to specify a candidate distribution \( h(\beta) \), from which potential deviates \( \beta^* \) are simulated, and an acceptance function,

\[
A(\beta, \beta^*) = \min \left\{ 1, \frac{f(\beta^* \mid W, b, Q)h(\beta)}{f(\beta \mid W, b, Q)h(\beta^*)} \right\},
\]

which provides the probability of accepting a new value \( \beta^* \) versus keeping the previous value \( \beta \).

In the logistic or t-distribution link (when the degrees of freedom \( k \) is large) case, \( f(\beta \mid W, b, Q) \) can be well approximated by a normal density function,

\[
g(W \mid \beta, b, Q) = \prod_{i,j} \frac{1}{2\pi c} e^{-\left( \frac{1}{2} \right) c^2 (w_{ij} - x_{ij}^T \beta - z_{ij}^T b_i)^2}
\]
Random Effects in Censored Ordinal Regression

where $c = 1.702$ if $\tilde{H}(t)$ is the inverse logit function and $c = k/(k - 2)$ if $\tilde{H}(t)$ is a $t$-distribution c.d.f. Take $h(\beta) \propto g(W | \beta, b, Q)\pi(\beta)$. $A(\beta, \beta^*) \approx 1$, and the Metropolis algorithm is highly efficient (Tierney, 1994). If $\pi(\beta)$ is $N(\hat{\beta}, \Sigma)$ or a flat prior, $h(\beta)$ is a normal density with form (A.1); however, $Q$ should be replaced by $c^2Q$.

**Simulating Deviates from $f(b | \beta, D, W, Q)$**

Simulating deviates from $f(b_i | \beta, D, W_i, Q_i)$ for $i = 1, \ldots, n$ is analogous to the $\beta$ case discussed above.

**Simulating Deviates from $f(D | b)$**

As discussed in Section 3, $\pi(D^{-1})$ will be a proper prior. In our context, it is a Wishart distribution $Wishart_q(P, \gamma)$, with precision matrix $P^{-1}$ and degrees of freedom $\gamma$ ($\gamma > 0$). Therefore,

$$D^{-1} | b \sim \text{Wishart}_q \left\{ \left( P^{-1} + \sum_i b_i b_i^T \right)^{-1}, \gamma + n \right\},$$

with precision matrix $P^{-1} + \sum_i b_i b_i^T$ and degrees of freedom $\gamma + n$.

**Simulating Deviates from $f(\alpha | W, y)$**

We in fact simulate deviates from $f(\alpha_s | W, \{\alpha_{s'}, s' \neq s\}, y)$ for $1 \leq s \leq S$. Note that $f(\alpha_s | W, \{\alpha_{s'}, s' \neq s\}, y) \propto \pi(\alpha_s) I(a_1 < a_s < a_2)$, where

$$a_1 = \max \left\{ \max_{i, j: v_{ij} = s-1} \{W_{ij}, \alpha_{s-1}\} \right\},$$

$$a_2 = \min \left\{ \min_{i, j: u_{ij} = s} \{W_{ij}, \alpha_{s+1}\} \right\}.$$

$f(\alpha_s | W, \{\alpha_{s'}, s' \neq s\}, y)$ is a truncated distribution of the prior $\pi(\alpha_s)$.

**Simulating Deviates from $f(W | \beta, b, \alpha, Q, y)$**

Let $\mu_{ij} = x_{ij}^T \beta + z_{ij}^T b_i$ and $\tilde{W}_{ij} = (W_{ij} - \mu_{ij})/Q_{ij}$ for $t \in [a_{u_{ij}}, a_{v_{ij}+1})$. Then

$$\Pr\{\tilde{W}_{ij} \leq t | \beta, b_i, \alpha, Q_{ij}, (u_{ij}, v_{ij})\} = \frac{\tilde{H}(\mu_{ij} - a_{u_{ij}}) - \tilde{H}(\mu_{ij} - t)}{\tilde{H}(\mu_{ij} - a_{u_{ij}}) - \tilde{H}(\mu_{ij} - a_{v_{ij}+1})}.$$

If $e$ is a deviate from Uniform$(0, 1)$, then

$$e^* = -\tilde{H}^{-1}\left( (1 - e)\tilde{H}(\mu_{ij} - a_{u_{ij}}) + e\tilde{H}(\mu_{ij} - a_{v_{ij}+1}) \right)$$

will be a deviate from $f(\tilde{W}_{ij} | \beta, b_i, \alpha, Q_{ij}, y_{ij})$. Therefore, $\tilde{W}_{ij} + Q_{ij}e^*$ is a deviate from $f(W_{ij} | \beta, b_i, \alpha, Q_{ij}, y_{ij})$.

If the scales $Q$ are random, then we have an additional step.

**Simulating Deviates from $f(Q | \beta, b, W)$**

Sometimes, treating the link function as a mixture of distributions may simplify the algorithm. For instance, for models with student $t$-distribution link functions (when the degrees of freedom $k$ is small), Albert and Chib (1993) suggested selecting $H(t) = \Phi(t)$ and treating $Q$'s as independent $\chi^2_k$ distributions. In this case, $f(Q | \beta, b, W)$ has an explicit form,

$$Q_{ij} | W_{ij}, \beta, b_i \sim \text{Gamma}\left( \frac{k + 1}{2}, \frac{1}{2k} \left( W_{ij} - x_{ij}^T \beta - z_{ij}^T b_i \right)^2 \right).$$