General Marginal Regression Models for the Joint Modeling of Event Frequency and Correlated Severities with Applications to Clinical Trials

Andrew S. Allen and Huiman X. Barnhart
Duke University

Abstract: In many clinical trials, information is collected on both the frequency of event occurrence and the severity of each event. For example, in evaluating a new anti-epileptic medication both the total number of seizures a patient has during the study period as well as the severity (e.g., mild, severe) of each seizure could be measured. In order to arrive at a full picture of drug or treatment performance, one needs to jointly model the number of events and their correlated ordinal severity measures. A separate analysis is not recommended as it is inefficient and can lead to what we define as "zero length bias" in estimates of treatment effect on severity. This paper proposes a general, likelihood based, marginal regression model for jointly modeling the number of events and their correlated ordinal severity measures. We describe parameter estimation issues and derive the Fisher information matrix for the joint model in order to obtain the asymptotic covariance matrix of the parameter estimates. A limited simulation study is conducted to examine the asymptotic properties of the maximum likelihood estimators. Using this joint model, we propose tests that incorporate information from both the number of events and their correlated ordinal severity measures. The methodology is illustrated with two examples from clinical trials: the first concerning a new drug treatment for epilepsy; the second evaluating the effect of a cholesterol lowering medication on coronary artery disease.

Key words: Joint modeling, marginal models, multiple endpoints, random length, shared parameters, zero length bias.

1. Introduction

In many clinical trials, data on both the frequency of a disease event as well as the severity of each event, often measured by an ordinal score, is recorded. Both of these endpoints (frequency and severity) can be important in evaluating treatment performance. Consider two examples. In a clinical trial of an anti-epileptic seizure medication both the frequency and the severity of seizures are important endpoints in evaluating treatment performance. In a clinical trial of

a cholesterol lowering drug both the number of arterial blockages as well as the degree of occlusion in each lesion are informative on drug performance. Data of this type are actually quite common. More examples include: depression may occur a random number of times with the severity of each depression episode being measured by an ordinal scale (e.g., 1=mildly depressed, 2=moderately depressed, 3=very depressed); patients suffering from migraine headache may have a random number of headaches with differing levels of pain; patients with cancer may have a random number of tumors of varying sizes, etc. Conceptually, we can think of these responses as a random ordinal vector of random length. A naive approach to such data would be to analyze the number of events separately from the multivariate severity measures. However, there are several reasons why this is not an ideal approach. First, even when one is primarily interested in each outcome separately, a separate analysis can result in a substantial decrease in efficiency (Albert *et al.*, 1997) as information contained in one endpoint can be used to estimate and make inference on parameters that describe the distribution of the other endpoint. Second, a separate analysis of severity can also give misleading results. Consider the clinical trial of a cholesterol lowering medication detailed in section 6.2. A separate analysis of severity is restricted to patients with severity measures, i.e., the group of patients with at least one blockage (event). In a randomized clinical trial, the patients with at least one event may consist of a biased sample of the test population. We refer to this phenomenon as "zero length bias". If the treatment under investigation is effective, patients in the treatment group are less likely to experience an event and therefore less likely to have a severity measure. An estimate of the marginal probability of observing a certain severity level conditional on observing an event, will not, in general, be equal to the unconditional marginal probability of observing this severity level. In the clinical setting, the unconditional marginal probabilities seem more relevant, i.e., the probability of having a severe blockage seems more relevant in assessing treatment than the probability of a severe blockage given a blockage has occurred. By jointly modeling the distribution of the number of events and their correlated severity measures, we propose a formal test for questions of this type (see details in section 5). In addition, the investigator may be interested in the relationship between the number of events and their severity. For example, one may want to test whether a drug changes the inhibition process among cancerous tumors (i.e., large tumors suppressing the ability of other tumors to grow near them) or estimate the probability of observing two or more severe events. Questions of this type can only be answered by modeling the joint distribution of the number of events and their correlated severity measures.

Barnhart (1992) gave the first probabilistic definition of random length elements. Barnhart and Sampson (1995) proposed a multiple population model to

jointly model the number of events and the correlated severities. Their methods are likelihood-based and assume that the event severities are distributed as a multivariate normal distribution. Recently, Barnhart has proposed a probit model for multivariate random length data (Barnhart, 1998). The probit model, however, assumes a discretization of a latent multivariate normal distribution for the multivariate ordinal responses and may not be flexible enough in many applications. Furthermore, these previous methods for multivariate random length data produce estimates of two treatment effects: one for event severity and one for the number of events. One needs to be cautious in interpreting the treatment effect for severity as it conditions on an event occuring. The test proposed in this manuscript provides a test of the overall treatment effect which includes both the number of events and the severity of these events.

This paper proposes a general, likelihood based, marginal regression model for jointly modeling the number of events and the vector of correlated ordinal severity measures. The proposed method is an extension to the joint modeling situation of the marginal models for multivariate categorical data proposed by Molenberghs and Lesaffre (1999). These models re-parameterize the multinomial distribution through a generalized linear model (GLIM) construction (McCullagh and Nelder, 1989), and allow great flexibility in the specification of the margins and associations.

The organization of this paper is as follows: in section 3 we present the proposed joint model; in section 4 we discuss estimation issues; in section 5 we present a limited simulation study; and in section 6 we present two examples from two clinical trials–one concerning an anti-seizure medication, the other a cholesterol lowering drug.

2. Data Examples

2.1 Example 1: Epilepsy clinical trial

This example involves data collected as part of a double-blind placebo controlled clinical trial designed to study the effect of felbamate (FBM), a new anti-epileptic drug, on epileptic seizure activity. A group of 40 epileptic patients were randomized to receive either placebo or felbamate. Patients were then followed for eighteen days (the first three being a titration period), and the number of partial seizures as well as a binary variable characterizing the severity of each seizure were recorded. The study was described in detail in Theodore et. al. (1995). For purposes of illustration, we collapsed the data in the following way. The study period (day 4 through day 18) was divided into four time periods. If an individual had any seizure activity in a given time period, we considered an event to have taken place. In addition, we assigned a severity score (mild or se-

vere) based on the severity of seizures in that time period. The number of events, therefore, varied between 0 and 4 while the severity of each event was either mild or severe. The data is given in table 3. A quick look at the data suggests that felbamate lowers both the average number and severity of seizures. Let treatment group be denoted by the indicator variable FBM_i with $FBM_i = 1$ if individual i is given felbamate and $FBM_i = 0$ if given placebo. Informative dropout was a problem in this dataset. Out of the original 40 individuals (20 randomized to felbamate and 20 randomized to placebo), 21 completed the protocol (13 in the felbamate group and 8 in the placebo group). Most of the dropouts in the placebo group dropped out because of increased seizure frequency and severity during the titration period. Those in the felbamate group tended to drop out due to drug induced side effects unrelated to seizure activity. To illustrate the current method, we ignore the problem of informative dropout and include only individuals that completed the protocol. Estimates based on this data set will tend to underestimate the effect of the drug therapy on seizure activity.

2.2 Example 2: Type II coronary intervention study

This example involves data collected during the National Heart, Lung and Blood Institute (NHLBI) Type II Coronary Intervention study (Brensike *et al.*, 1982, 1984). Patients with Type II hyperlipoproteinemia and coronary heart disease were randomly allocated to a daily dose of 24g of cholestyramine and diet (treatment group) or placebo and diet (control group). The goal of the study was to assess the effect of cholestyramine on coronary artery disease progression at the end of five years of study. The number and size of each lesion (blockage) in the lower arterial descending portion of the heart was determined by coronary angiogram at the fifth year of follow-up. We take the size of each lesion to be a severity score with small (=1), medium (=2), or large (=3) degrees of blockage. The observed number of blockages in this part of the heart varied from zero to four. A total of 143 patients were recruited into the study. At the end of five years of follow-up there were only 116 (57-placebo; 59-cholestyramine) patients who had their fifth year coronary angiogram done. For the purpose of illustration, this is the population we consider. Examination of the data did not suggest any problems with informative dropout. Let treatment group be denoted by the indicator variable TRT_i with $TRT_i = 1$ if individual i is given cholestyramine and $TRT_i = 0$ if given placebo. We also have two additional covariates. SBP_i is an indicator variable for whether systolic blood pressure is greater than 120, and RCA_i is an indicator variable of whether a regional contractile abnormality is present.

3. A Joint Model

Suppose we observe data $(k_i, \mathbf{y}_i, \mathbf{x}_i)$ for the ith individual, where the count k_i is a realization of a discrete random variable K_i (frequency of event in our context), $\mathbf{y}_i$ is a realization of a k_i-dimensional ordinal random variable $\mathbf{Y}_i = \left(Y_i^1, Y_i^2, ..., Y_i^{k_i}\right)'$ (severity measures in our context), and $\mathbf{x}_i$ is a vector of covariates. For simplicity, we will assume that each component of $\mathbf{Y}_i$ can take on r distinct values (i.e., r ordinal categories), though this assumption can be relaxed. Without loss of generality we will denote these values as $1, 2, ..., r$. We specify a joint distribution for K_i and $\mathbf{Y}_i$ by factoring it into two parts: a distribution for K_i and a distribution for $\mathbf{Y}_i|K_i$. Since the event severities and the counts both reflect the severity of the disease outcome, certain covariates that affect disease may affect both K_i and $\mathbf{Y}_i|K_i$. Therefore, we specify a dependence between them by allowing these two pieces to share information through common parameters. We summarize the proposed joint model here:

1. The distribution of the random length K_i, following Barnhart et al. (1999), is modeled as a slightly modified generalized linear model (GLIM) (Mc-Cullagh and Nelder, 1989) with three components:

    (a) The random length variable K_i has a discrete distribution:

    $$\Pr\left(K_i = k\right) = a\left(k\right) \exp\left(k\phi_i - b\left(\phi_i\right)\right), k = 0, 1, ..., J$$

    where the ϕ_i's are the natural parameters for the exponential family, $J > 0$ is assumed known (possibly infinite), and $a\left(\cdot\right)$ and $b\left(\cdot\right)$ are known positive functions;

    (b) The link function h: $\eta_i = h\left(E\left(K_i\right)\right)$, where we choose $h\left(\cdot\right)$ such that $h\left(b'\left(\cdot\right)\right)$ is either a strictly increasing or strictly decreasing function;

    (c) The systematic component: $\eta_i = \delta + \gamma \mathbf{x}_i' \beta$.

2. $\mathbf{Y}_i|K_i = k$ is modeled as a marginal model for multivariate categorical (MMMC) data (Molenberghs and Lesaffre, 1999). This model is also comprised of three components:

    (a) The distribution of $\mathbf{Y}_i|K_i = k$ is multinomial with cell probabilities π_i and corresponding orthant probabilities $\mathbf{p}_i$;

    (b) A vector of link functions $\mathbf{h}$:

    $$\begin{aligned}
    \eta_i^m &= \mathbf{h}^m\left(\mathbf{p}_i^m\right) \\
    \eta_i^a &= \mathbf{h}^a\left(\mathbf{p}_i\right)
    \end{aligned}$$

where $\mathbf{h}^m$ is a vector of marginal link functions, $\mathbf{h}^a$ is a vector of bivariate and higher order association link functions, and

$$\mathbf{p}_i^m = \left(\left(\mathbf{p}_i^1\right)', \left(\mathbf{p}_i^2\right)', \ldots, \left(\mathbf{p}_i^k\right)' \right)'$$

is a vector of univariate orthant probabilities. For example, a commonly used link function is the logit $(h(p) = log(p/1 - p))$. Additional examples can be found in section 3.2. For simplicity, we restrict the component functions of $\mathbf{h}^m$ to be the same scalar function, i.e., $\mathbf{h}^m = (h^m, ..., h^m)'$, where $h^m : [0, 1] \longmapsto \Re$ is either a strictly increasing or decreasing function. Therefore, the size of $\mathbf{h}^m$ is the same as $\mathbf{p}_i^m$ with each component operating on one and only one component of $\mathbf{p}_i^m$;

(c) The systematic component:

$$\begin{aligned} \eta_i^m &= \mathbf{X}_i^m \alpha + \left(\mathbf{x}_i' \beta\right) \mathbf{e} \\ \eta_i^a &= \mathbf{X}_i^a \nu \end{aligned}$$

where $\mathbf{X}_i^m$ and $\mathbf{X}_i^a$ are design matrices formed by linear combinations of covariates and intercept terms, and $\mathbf{e}$ is a vector of all 1's with its size being the same as η_i^m. Hence, different covariates may be used in the marginal (η_i^m) and association (η_i^a) models. Here we make the proportional odds like assumption that β is the same for all levels of the severity measure $\mathbf{Y}_i$. We discuss the MMMC framework in more detail later in this section.

3. $(K_i, \mathbf{Y}_i)$ is independent of $(K_j, \mathbf{Y}_j)$ for $i \neq j$.

Note that the β's are common parameters shared in the specifications for both K_i and $\mathbf{Y}_i | K_i$. The scalar parameter γ can be thought of as scaling the marginal effect of the covariates on the multivariate severities to the random length scale. In addition, the sign and magnitude of γ control the association of the random length with the multivariate severities. If $\gamma = 0$, there is no relationship between the number of events and their severities. By the model assumptions, we have $E(K_i) = b'(\phi_i)$, thus $\eta_i = h(b'(\phi_i))$. Suppose $\gamma > 0$ and K_i and $K_{i'}$ are two random length variables such that $\mathbf{x}_i' \beta > \mathbf{x}_{i'}' \beta$, and therefore $\eta_i > \eta_{i'}$. Then if $h\left(b'(\cdot)\right)$ is a strictly increasing function, we have $\phi_i > \phi_{i'}$. It is easy to show that the distribution of K_i is stochastically increasing in ϕ_i. Therefore, $\mathbf{x}_i' \beta > \mathbf{x}_{i'}' \beta$ implies that K_i is stochastically larger than $K_{i'}$ when $\gamma > 0$. Similarly,

let $\mathbf{Y}_i | (K_i = k, \mathbf{x}_i)$ and $\mathbf{Y}_{i'} | (K_{i'} = k, \mathbf{x}_{i'})$ be conditional severity random variables such that $\mathbf{x}'_i \beta > \mathbf{x}'_{i'} \beta$, and therefore $\eta_i^m > \eta_{i'}^m$. If h^m is strictly increasing, then $p_i^t(j) > p_{i'}^t(j)$ for all t $(1 \leq t \leq k)$ and j $(1 \leq j \leq r)$, implying that $Y_i^t | (K_i = k, \mathbf{x}_i)$ is stochastically larger than $Y_{i'}^t | (K_{i'} = k, \mathbf{x}_i)$ for all t $(1 \leq t \leq k)$ when $\mathbf{x}'_i \beta > \mathbf{x}'_{i'} \beta$. Therefore, if both $h\left(b'(\cdot)\right)$ and h^m are strictly increasing functions, the above model describes the following: for $\gamma > 0$, the larger the value of $\mathbf{x}'_i \beta$ the more likely we are to observe a larger number of events as well as greater severity in any given event; for $\gamma < 0$, the larger the value of $\mathbf{x}'_i \beta$ the more likely we are to observe a smaller number of events as well as greater severity in any given event. Note that if h is a canonical link function, then $h\left(b'(\cdot)\right)$ is the identity function (McCullagh and Nelder, 1989), a strictly increasing function.

3.1 Marginal models for multivariate categorical data

The MMMC framework was proposed by Molenberghs and Lesaffre (1999). For simplicity, in this section "marginal" is used in the context of the conditional random variable $\mathbf{Y}_i | K_i$. Note that given $K_i = k$, all the information in $\mathbf{Y}_i$ is contained in the r^k contingency table obtained by cross-classifying $\mathbf{Y}_i$. Denote the jth entry in this table by $Z_i(\mathbf{j})$ $(Z_i(\mathbf{j}) \in \{0, 1\})$ where $\mathbf{j} = (j_1, j_2, ..., j_k)'$, $\mathbf{1} \leq \mathbf{j} \leq \mathbf{r}$ $(\mathbf{r} = (r, r, ..., r)'$, a k-dimensional vector). Assume this table is sampled from a multinomial distribution with $\mathbf{j}^{th}$ cell probability $\pi_i(\mathbf{j})$. Note that $E(Z_i(\mathbf{j})) = \pi_i(\mathbf{j})$. Define orthant probabilities $p_i(\mathbf{j}) = \sum_{\mathbf{l} \leq \mathbf{j}} \pi_i(\mathbf{l})$. This can be concisely written in matrix form as $\mathbf{p}_i = L\pi_i$ (Molenberghs and Lesaffre, 1999), where $\mathbf{p}_i$ is a vector of all orthant probabilities, π_i is a vector of all multinomial cell probabilities, and L is a $r^k \times r^k$ matrix of 0 and 1's. For example, in the bivariate binary case, the matrix L is given by

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix}.$$

Univariate (marginal) orthant probabilities are then given by setting all but one of their indices to their maximal value: $p_i^t(j) = p_i(r_1, ..., r_{t-1}, j, r_{t+1}, ..., r_k)$, where $r_1 = ... = r_{t-1} = r_{t+1} = ... = r_k = r$. The n-variate (marginal, $n < k$) orthant-probabilities are given in a similar way by setting all but n of the indices to their maximal value. For example, in the bivariate case $p_i^{t_1, t_2}(j_1, j_2) = p_i(r_1, ..., r_{t_1-1}, j_1, r_{t_1+1}, ..., r_{t_2-1}, j_2, r_{t_2+1}, ..., r_k)$. In the marginal models proposed by Molenberghs and Lesaffre, the cell probabilities in the above contingency table are expressed, via the link functions of orthant probabilities, as a function of covariates. This model is "marginal" since it explicitly models the univariate or-

thant probabilities, through link functions, as linear functions of covariates. The bivariate and higher-dimensional orthant probabilities are then used to model the association structure with link functions chosen so as to parameterize a given association structure.

3.2 Examples of marginal models for multivariate categorical data

The MMMC framework provides a class of models with the generality and flexibility necessary to model multivariate categorical data. Many previously proposed models are MMMC models. For example, assume $K_i = 2$ and $r = 2$, giving us two binary responses Y_{i1}, Y_{i2}. Let $\pi_i(j_1, j_2) = \Pr(Y_{i1} = j_1, Y_{i2} = j_2) = \Pr(Z_i(j_1, j_2) = 1)$. The bivariate logistic transform was defined by McCullagh and Nelder (1989) by the mapping $\pi \xrightarrow{L} \mathbf{p} \longmapsto \eta = (\eta_1, \eta_2, \eta_{12})$ where

$$
\begin{aligned}
\eta_{i1} &= \text{logit}\left[p_i^1(1)\right] = \text{logit}\left[\pi_i(1,1) + \pi_i(1,2)\right] \\
\eta_{i2} &= \text{logit}\left[p_i^2(1)\right] = \text{logit}\left[\pi_i(1,1) + \pi_i(2,1)\right] \\
\eta_{i12} &= \log\left[\frac{p_i(1,1)\left[1 - p_i^1(1) - p_i^2(1) + p_i(1,1)\right]}{\left[p_i^1(1) - p_i(1,1)\right]\left[p_i^2(1) - p_i(1,1)\right]}\right] = \log\left[\frac{\pi_i(1,1)\,\pi_i(2,2)}{\pi_i(1,2)\,\pi_i(2,1)}\right].
\end{aligned}
$$

The η_i's are usually taken to be linear functions of the covariate vector $\mathbf{x}_i$. Note that the association is modeled with an odds ratio, given as $\exp(\eta_{i12})$. A bivariate probit model can be defined for these data by the mapping $\pi \xrightarrow{L} \mathbf{p} \longmapsto \eta = (\eta_1, \eta_2, \eta_{12})$ with

$$
\begin{aligned}
\eta_{i1} &= \Phi^{-1}\left[p_i^1(1)\right] = \Phi^{-1}\left[\pi_i(1,1) + \pi_i(1,2)\right] \\
\eta_{i2} &= \Phi^{-1}\left[p_i^2(1)\right] = \Phi^{-1}\left[\pi_i(1,1) + \pi_i(2,1)\right] \\
\eta_{i12} &= \log\left\{1 + r_t\left(L^{-1}\mathbf{p}_i\right)\right\} - \log\left\{1 - r_t\left(L^{-1}\mathbf{p}_i\right)\right\} \\
&= \log\left\{1 + r_t(\pi_i)\right\} - \log\left\{1 - r_t(\pi_i)\right\},
\end{aligned}
$$

where $r_t(\cdot)$ is the tetrachoric correlation. If we define $\Phi(x, y, \rho)$ to be the cumulative distribution function of a standard bivariate normal distribution with correlation ρ, the tetrachoric correlation is the value of ρ such that $\pi_i(2,2) = 1 - \Phi(\eta_{i1}, \eta_{i2}, \rho)$. Thus two popular models for univariate categorical data, the logit and the probit model, are unified in a multivariate setting with the marginal parameters having the same interpretation as in the univariate setting. Many more models are possible, and the bivariate logit and probit models can be generalized to higher dimensions. Molenberghs and Lesaffre (1999) discuss possible association structures for different marginal models. An important class of marginal models are the so called multivariate logistic models of Glonek and McCullagh (1995). The multivariate logistic transform is defined by the mapping

$\pi \overset{\mathbf{L}}{\longmapsto} \mathbf{p} \longmapsto \eta$ where

$$\begin{aligned} \eta &= C' \log\left(AL^{-1}\mathbf{p}_i\right) \\ &= C' \log\left(A\pi_i\right), \end{aligned}$$

and involves contrasts of log probabilities–given by linear combinations of multinomial cell probabilities. The explicit construction of C and A is given in Glonek and McCullagh (1994). This class of models extends the logit models for categorical data given in Agresti (1990, pp. 306-338) to the multivariate setting. Global and local odds ratios are common measures of association that are included in the multivariate logistic model class. The bivariate Dale model (Dale, 1986) and its n-variate extensions (multivariate Dale model: Molenberghs and Lesaffre, 1994) are also members of the MMMC family. These Dale models specify dependences between the ordinal outcomes in terms of (generalized) global odds ratios while there are no restrictions on the types of marginal distributions that can be used. The above bivariate logistic transform is actually a binary version of the bivariate Dale model.

An important feature of marginal models is that if a set of outcomes satisfy a marginal model, then so does every subvector of outcomes. As a result, given that we observe a severity outcome, the interpretation of the parameters does not depend on the actual number of outcomes. This property is important in the random length data context where the length of the outcome is random. The parameters will have the same interpretation regardless of size of the outcome vectors observed.

4. Maximum Likelihood Estimation

Collectively denote the model parameters as $\theta = \left(\alpha', \nu', \beta', \delta, \gamma\right)'$. The log-likelihood function for the model proposed in the above section is

$$l(\theta) = \log\left\{\prod_{i=1}^{N} P(K_i) f\left(\mathbf{Y}_i|K_i\right)^{\delta(K_i)}\right\}$$

where N is the total number of individuals and $\delta(\cdot)$ is an indicator function such that $\delta(K_i) = 1$ if $K_i > 0$, and $\delta(K_i) = 0$ if $K_i = 0$. Let $I(\theta)$ be the total Fisher information matrix for θ contained in the N independent multivariate random length observations $\mathbf{Y}_i$ with random length K_i, $i = 1, ..., N$. Then by Theorem 3.3.1 of Barnhart (1992), we have

$$I(\theta) = I^*(\theta) + \sum_{i=1}^{N}\sum_{k=1}^{J} P(K_i = k) I(\theta|k)$$

where $I^*(\theta)$ is the total Fisher information matrix for θ contained in the random length data $\{K_i, i = 1, ..., N\}$, and $I(\theta|k)$ is the total Fisher information matrix for θ contained in the conditional severities $\mathbf{Y}_i|K_i = k$. The appendix gives the form of $I^*(\theta)$ and $I(\theta|k)$ for the proposed model. Intuitively, the information is the sum of the information from the random length and the weighted average of information from severity over all possible lengths.

The maximum likelihood estimator (MLE) for θ can be obtained through a Fisher scoring algorithm or by numerical maximization of the likelihood through a standard optimization algorithm. Both techniques require the inversion of the link functions at each step of the algorithm to obtain the multinomial cell probabilities. A number of techniques are possible: Newton-Raphson (Glonek and McCullagh, 1995), iterative proportional fitting (Molenberghs and Lesaffre, 1999), Lagrange multipliers (Haber and Brown, 1986), etc. In the simulation and examples that follow, we use a multivariate Dale type approach where the orthant probabilities are found sequentially as solutions to polynomial equations (Molenberghs and Lesaffre, 1994). Barnhart (1992) has shown that, under the usual regularity conditions, the MLE is consistent and asymptotically normal. The asymptotic covariance matrix of the MLE, $\widehat{\theta}$, is obtained as the inverse of the above information matrix and is estimated by $I^{-1}\left(\widehat{\theta}\right)$.

5. Simulation

We performed a limited simulation study to evaluate the asymptotic behavior of the MLE for the simple case where the number of events is either one or two. For three different sets of parameters, we generate 1000 data sets where each data set consists of ordinal responses (3 levels) from two different treatment groups with 500 subjects in each. As before we specify the model in two parts. The first part, the distribution of $\mathbf{Y}_i|K_i = k$, is taken to be multinomial with cell probabilities related to the covariates, via the orthant probabilities, by

$$
\begin{aligned}
\alpha_1 - \beta x_i &= \mathrm{logit}\left[p_i^1(1)\right] \\
\alpha_2 - \beta x_i &= \mathrm{logit}\left[p_i^1(2)\right] \\
\alpha_1 - \beta x_i &= \mathrm{logit}\left[p_i^2(1)\right] \\
\alpha_2 - \beta x_i &= \mathrm{logit}\left[p_i^2(2)\right] \\
\nu &= \log\left[\frac{p_i(1,1)\left[1 - p_i^1(1) - p_i^2(1) + p_i(1,1)\right]}{\left[p_i^1(1) - p_i(1,1)\right]\left[p_i^2(1) - p_i(1,1)\right]}\right] \\
\nu &= \log\left[\frac{p_i(1,2)\left[1 - p_i^1(1) - p_i^2(2) + p_i(1,2)\right]}{\left[p_i^1(1) - p_i(1,2)\right]\left[p_i^2(2) - p_i(1,2)\right]}\right]
\end{aligned}
\tag{5.1}
$$

$$\nu = \log \left[\frac{p_i(2,1) \left[1 - p_i^1(2) - p_i^2(1) + p_i(2,1)\right]}{\left[p_i^1(2) - p_i(2,1)\right] \left[p_i^2(1) - p_i(2,1)\right]} \right]$$

$$\nu = \log \left[\frac{p_i(2,2) \left[1 - p_i^1(2) - p_i^2(2) + p_i(2,2)\right]}{\left[p_i^1(2) - p_i(2,2)\right] \left[p_i^2(2) - p_i(2,2)\right]} \right]$$

in the case when $k = 2$ and by

$$\alpha_1 - \beta x_i = \text{logit}\left[p_i(1)\right] \tag{5.2}$$
$$\alpha_2 - \beta x_i = \text{logit}\left[p_i(2)\right]$$

when $k = 1$. Here x_i is an indicator variable of treatment group. Hence, a proportional odds cumulative-logit model is being assumed for the marginal distributions of $\mathbf{Y}_i | K_i$ while the association is being described by global odds-ratios. The distribution of K_i is taken to be Bernoulli with

$$\text{logit}(\Pr(K_i = 1 | x_i)) = \delta - \gamma \beta x_i. \tag{5.3}$$

For each of 1000 data sets, data for 500 subjects in each of two treatment groups $(x_i = 1, 0)$ were simulated in the following way: first, the length of the severity vector was determined by drawing from a Bernoulli distribution with probability given by equation (5.3); then given the value of k, the severity vector was obtained by drawing from a multinomial distribution with cell probabilities $\pi_i = L^{-1}\mathbf{p}_i$, where $\mathbf{p}_i$ is determined $(k = 2)$ by equation (5.1) or $(k = 1)$ equation (5.2) . The parameter vectors used for the simulations were $\theta = (\alpha_1, \alpha_2, \nu, \beta, \delta, \gamma)' = (0.5, 1.25, -1.0, \nu, 0.8, 1.2)'$ where $\nu = 0.5, 1.0, 2.0$. For each data set, the above model was fit by numerically maximizing the log-likelihood. Table 1 displays the means, standard errors (s.e.) and ranges of the parameter estimates from the 1000 simulations.

We found the empirical distribution of the MLE to be symmetric and approximately normal as evidenced by high degree of linearity found in normal probability plots (not shown). In table 2, we compare the standard errors of the MLE computed from the inverse of the Fisher total information matrix, $I^{-1}(\theta)$, with the standard errors from the simulation. These results indicate that the two standard errors are quite close.

6. Examples

In this section, we illustrate the proposed joint model with two examples from clinical trials: the first concerning a new drug treatment for epilepsy; the second evaluating the effect of a cholesterol lowering medication on coronary artery disease.

A. S. Allen and H. X. Barnhart

Table 1: MLE estimates from simulated data

	$\widehat{\alpha}_1$	$\widehat{\alpha}_2$	$\widehat{\nu}$	$\widehat{\beta}$	$\widehat{\delta}$	$\widehat{\gamma}$
True Value	0.5	1.25		-1	0.8	1.2
$\nu = 0.5$						
mean	0.500	1.252	0.493	-0.998	0.799	1.219
s.e.	0.080	0.086	0.173	0.107	0.096	0.189
min	0.274	1.003	-0.120	-1.325	0.558	0.616
max	0.739	1.527	1.126	-0.629	1.109	1.982
$\nu = 1.0$						
mean	0.501	1.253	0.995	-1.000	0.807	1.222
s.e.	0.082	0.088	0.171	0.112	0.094	0.192
min	0.233	0.954	0.493	-1.470	0.549	0.695
max	0.754	1.588	1.545	-0.643	1.109	2.093
$\nu = 2.0$						
mean	0.501	1.254	2.000	-1.001	0.807	1.223
s.e.	0.085	0.091	0.177	0.117	0.094	0.199
min	0.238	0.956	1.518	-1.495	0.549	0.700
max	0.765	1.558	-0.656	2.566	1.109	2.143

Table 2: Standard errors from simulation and from information matrix

	$\widehat{\alpha}_1$	$\widehat{\alpha}_2$	$\widehat{\nu}$	$\widehat{\beta}$	$\widehat{\delta}$	$\widehat{\gamma}$
True Value	0.5	1.25		-1	0.8	1.2
$\nu = 0.5$						
s.e. from simulation	0.080	0.086	0.173	0.107	0.096	0.189
s.e. from $I^{-1}(\theta)$	0.081	0.087	0.171	0.107	0.097	0.185
$\nu = 1.0$						
s.e. from simulation	0.082	0.088	0.171	0.112	0.094	0.192
s.e. from $I^{-1}(\theta)$.	0.083	0.089	0.173	0.110	0.097	0.188
$\nu = 2.0$						
s.e. from simulation	0.085	0.091	0.177	0.117	0.094	0.199
s.e. from $I^{-1}(\theta)$	0.086	0.092	0.181	0.115	0.097	0.192

Table 3: Epilepsy seizure data

Placebo		Felbamate	
Subject ID	Event Sequence	Subject ID	Event Sequence
3	1,2,1	4	*
6	2,2	7	*
15	1,2,1	9	1
16	*	12	*
17	1,2,1	13	1,1
24	2,2	19	1
25	1,1,1	20	2
39	2,1,2	21	*
		30	1,1
		33	2,1
		36	1
		40	*

*-indicates no seizure; 1-indicates mild seizure; 2-indicates severe seizure

6.1 Example 1: Epilepsy clinical trial

This example involves data collected as part of a double-blind placebo controlled clinical trial designed to study the effect of felbamate (FBM), a new antiepileptic drug, on epileptic seizure activity. A group of 40 epileptic patients were randomized to receive either placebo or felbamate. Patients were then followed for eighteen days (the first three being a titration period), and the number of partial seizures as well as a binary variable characterizing the severity of each seizure were recorded. The study was described in detail in Theodore et. al. (1995). For purposes of illustration, we collapsed the data in the following way. The study period (day 4 through day 18) was divided into four time periods. If an individual had any seizure activity in a given time period, we considered an event to have taken place. In addition, we assigned a severity score (mild or severe) based on the severity of seizures in that time period. The number of events, therefore, varied between 0 and 4 while the severity of each event was either mild or severe. The data is given in table 3. A quick look at the data suggests that felbamate lowers both the average number and severity of seizures. Let treatment group be denoted by the indicator variable FBM_i with $\text{FBM}_i = 1$ if individual i is given felbamate and $\text{FBM}_i = 0$ if given placebo. Informative dropout was a problem in this dataset. Out of the original 40 individuals (20 randomized to felbamate and 20 randomized to placebo), 21 completed the protocol (13 in

Table 4: MLE estimates of epilepsy data from joint model

Parameter	Estimate	S.E.	Z-score
α	-0.274	0.593	-0.462
ν	2.310	0.728	3.173
β	-1.536	1.249	-1.230
δ	1.112	0.170	6.541
γ	0.880	0.751	1.172
$\gamma\beta$	-1.352	0.352	-3.841

the felbamate group and 8 in the placebo group). Most of the dropouts in the placebo group dropped out because of increased seizure frequency and severity during the titration period. Those in the felbamate group tended to drop out due to drug induced side effects unrelated to seizure activity. To illustrate the current method, we ignore the problem of informative dropout and include only individuals that completed the protocol. Estimates based on this data set will tend to underestimate the effect of the drug therapy on seizure activity.

We applied the proposed joint model to the data in table 3. The distribution of the number of events K_i (time periods with seizures) was assumed to be Poisson,

$$\Pr\left(K_i = k\right) = a\left(k\right)\exp\left(k\phi_i - b\left(\phi_i\right)\right)$$

where $a\left(k\right) = 1/k!$, $\phi_i = \eta_i = \delta + \gamma\beta\text{FBM}_i$, and $b\left(\phi_i\right) = \exp\left(\delta + \gamma\beta\text{FBM}_i\right)$. The distribution of $\mathbf{Y}_i|K_i = k$ was taken to be multinomial with cell probabilities related to the covariates via the orthant probabilities as before. Here we assumed the univariate orthant probabilities were related to the covariates through the logit link, i.e.,

$$\alpha - \beta\text{FBM}_i = \text{logit}\left[p_i^t\left(1\right)\right],\ t = 1, 2, ..., k,$$

and the bivariate associations were described with global odds ratios by,

$$\nu = \log\left[\frac{p_i^{t_1,t_2}\left(1,1\right)\left[1 - p_i^{t_1}\left(1\right) - p_i^{t_2}\left(1\right) + p_i^{t_1,t_2}\left(1,1\right)\right]}{\left[p_i^{t_1}\left(1\right) - p_i^{t_1,t_2}\left(1,1\right)\right]\left[p_i^{t_2}\left(1\right) - p_i^{t_1,t_2}\left(1,1\right)\right]}\right],$$

$t_1 < t_2$ $\left(t_1, t_2 \in \{1, 2, ..., k\}\right)$. The trivariate and higher order associations were also described by global odds ratios and were assumed to be zero on the log scale. The parameters for this model were $\alpha, \nu, \beta, \delta$, and γ. Table 4 gives the maximum likelihood estimates and the corresponding standard errors computed from the inverse of the Fisher total information matrix. The estimate for the last

term, $\gamma\beta$, was computed using the MLE invariance principle while the estimate's standard error was calculated using a standard delta-method approach. The negative values for $\widehat{\gamma\beta}$ and $\widehat{\beta}$ suggest that felbamate decreases the number and severity (conditional on k) of seizures, while the positive value for $\widehat{\gamma}$ confirms that the effect of felbamate is in the same direction for both the number and severity of seizures.

We also ran two separate analyses, modeling the length and severity individually (not shown). The model for the random length was assumed Poisson as above, while the multivariate severities were modeled using GEE with a logit link and an exchangeable working covariance matrix. Point estimates and standard errors from this approach were similar to corresponding parameters from the above joint model. However, the joint model is likelihood based and directly models the relationship between the random length and severities allowing more flexibility in the types of questions that can be addressed. For example it would be impossible, via the separate analysis, to address the question: "Does treatment significantly decrease my chances of having more than two severe events?" The joint model also makes explicit what is normally implicit in a separate analysis of severity: that you need to observe an event in order to observe a severity, i.e., that the severity analysis is conditional on the occurrence of at least one event. Ostensibly, a relevant question (perhaps **the** relevant question) regarding treatment effect with respect to seizure activity is whether the marginal probabilities of no seizure, mild seizure, or severe seizure are different between the two treatment groups. When some subjects have no events this cannot be answered by a separate analysis of severity since the marginal probability of having a mild or severe event is the product of both pieces from the joint model, i.e.,

$$\begin{aligned} \Pr\left(Y = \mathrm{MILD} \,|\, \mathbf{x}\right) &= \Pr\left(Y = \mathrm{MILD} \,|\, k > 0, \mathbf{x}\right) \Pr\left(K > 0 \,|\, \mathbf{x}\right) \\ \Pr\left(Y = \mathrm{SEVERE} \,|\, \mathbf{x}\right) &= \Pr\left(Y = \mathrm{SEVERE} \,|\, k > 0, \mathbf{x}\right) \Pr\left(K > 0 \,|\, \mathbf{x}\right). \end{aligned}$$

This follows from the fact that

$$\Pr\left(Y = \mathrm{SEVERE}|k = 0, \mathbf{x}\right) = \Pr(Y = \mathrm{MILD}|k = 0, \mathbf{x}) = 0$$

and that the marginal model specification for $\mathbf{Y}_i|K_i$ implies

$$\begin{aligned} \Pr\left(Y = \mathrm{SEVERE}|k = i, \mathbf{x}\right) &= \Pr(Y = \mathrm{SEVERE}|k = j, \mathbf{x}) \quad \text{and} \\ \Pr\left(Y = \mathrm{MILD}|k = i, \mathbf{x}\right) &= \Pr(Y = MILD|k = j, \mathbf{x}) \end{aligned}$$

for all $i, j > 0$. In order to illustrate how the joint model can be used to generate a test for a treatment effect, we compare the two multinomial distributions, $\pi_i = (\Pr\left(K_i = 0|\mathrm{FBM}_i\right), \ \Pr\left(Y_i = \mathrm{MILD}|\mathrm{FBM}_i\right), \Pr\left(Y = \mathrm{SEVERE}|\mathrm{FBM}_i\right))'$, for the

two treatment groups respectively using contrasts of baseline category logits. Specifically we test

$$H_0 : \log \left(\frac{\Pr \left(Y = k \,|\, \mathrm{FBM}_i = 1 \right)}{\Pr \left(K_i = 0 \,|\, \mathrm{FBM}_i = 1 \right)} \right) = \log \left(\frac{\Pr \left(Y = k \,|\, \mathrm{FBM}_i = 0 \right)}{\Pr \left(K_i = 0 \,|\, \mathrm{FBM}_i = 0 \right)} \right),$$

for $k =$ SEVERE, MILD. using a Wald chi-square test (details in appendix). The resulting chi-square test statistic was 16.2. Comparing this to χ_2^2 gives a P-value of 0.0003. This indicates that, when compared to placebo, felbamate significantly increases the chance of no seizure while lowering the chances of having a severe seizure. The test effectively uses information from both endpoints being investigated, seizure frequency and severity, and is made possible by the joint-modeling framework.

6.2 Example 2: Type II coronary intervention study

This example involves data collected during the National Heart, Lung and Blood Institute (NHLBI) Type II Coronary Intervention study (Brensike *et al.*, 1982, 1984). Patients with Type II hyperlipoproteinemia and coronary heart disease were randomly allocated to a daily dose of 24g of cholestyramine and diet (treatment group) or placebo and diet (control group). The goal of the study was to assess the effect of cholestyramine on coronary artery disease progression at the end of five years of study. The number and size of each lesion (blockage) in the lower arterial descending portion of the heart was determined by coronary angiogram at the fifth year of follow-up. We take the size of each lesion to be a severity score with small (=1), medium (=2), or large (=3) degrees of blockage. The observed number of blockages in this part of the heart varied from zero to four. A total of 143 patients were recruited into the study. At the end of five years of follow-up there were only 116 (57-placebo; 59-cholestyramine) patients who had their fifth year coronary angiogram done. For the purpose of illustration, this is the population we consider. Examination of the data did not suggest any problems with informative dropout. Let treatment group be denoted by the indicator variable TRT_i with $\mathrm{TRT}_i = 1$ if individual i is given cholestyramine and $\mathrm{TRT}_i = 0$ if given placebo. We also have two additional covariates. SBP_i is an indicator variable for whether systolic blood pressure is greater than 120, and RCA_i is an indicator variable of whether a regional contractile abnormality is present.

We applied the above joint model to the data. Again, the distribution of K_i was assumed to be Poisson,

$$\Pr \left(K_i = k \right) = a \left(k \right) \exp \left(k \phi_i - b \left(\phi_i \right) \right)$$

where $a(k) = 1/k!$, $\phi_i = \eta_i = \delta + \gamma(\beta_1 \mathrm{TRT}_i + \beta_2 \mathrm{RCA}_i + \beta_3 \mathrm{SBP}_i)$, and $b(\phi_i) = \exp[\delta + \gamma(\beta_1 \mathrm{TRT}_i + \beta_2 \mathrm{RCA}_i + \beta_3 \mathrm{SBP}_i)]$. The distribution of $\mathbf{Y}_i | K_i = k$ is taken to be multinomial with cell probabilities related to the covariates via the orthant probabilities as before. Here we assume the univariate orthant probabilities are related to the covariates through a cumulative logit link, i.e.,

$$
\begin{aligned}
\alpha_1 - (\beta_1 \mathrm{TRT}_i + \beta_2 \mathrm{RCA}_i + \beta_3 \mathrm{SBP}_i) &= \mathrm{logit}\left[p_i^t(1) \right] \\
\alpha_2 - (\beta_1 \mathrm{TRT}_i + \beta_2 \mathrm{RCA}_i + \beta_3 \mathrm{SBP}_i) &= \mathrm{logit}\left[p_i^t(2) \right], \ t = 1, 2, ..., k,
\end{aligned}
$$

and the bivariate associations are described with global odds ratios by,

$$
\begin{aligned}
\nu &= \log\left[\frac{p_i^{t_1,t_2}(1,1)\left[1 - p_i^{t_1}(1) - p_i^{t_2}(1) + p_i^{t_1,t_2}(1,1)\right]}{\left[p_i^{t_1}(1) - p_i^{t_1,t_2}(1,1)\right]\left[p_i^{t_2}(1) - p_i^{t_1,t_2}(1,1)\right]} \right] \\
\nu &= \log\left[\frac{p_i^{t_1,t_2}(1,2)\left[1 - p_i^{t_1}(1) - p_i^{t_2}(2) + p_i^{t_1,t_2}(1,2)\right]}{\left[p_i^{t_1}(1) - p_i^{t_1,t_2}(1,2)\right]\left[p_i^{t_2}(2) - p_i^{t_1,t_2}(1,2)\right]} \right] \\
\nu &= \log\left[\frac{p_i^{t_1,t_2}(2,1)\left[1 - p_i^{t_1}(2) - p_i^{t_2}(1) + p_i^{t_1,t_2}(2,1)\right]}{\left[p_i^{t_1}(2) - p_i^{t_1,t_2}(2,1)\right]\left[p_i^{t_2}(1) - p_i^{t_1,t_2}(2,1)\right]} \right] \\
\nu &= \log\left[\frac{p_i^{t_1,t_2}(2,2)\left[1 - p_i^{t_1}(2) - p_i^{t_2}(2) + p_i^{t_1,t_2}(2,2)\right]}{\left[p_i^{t_1}(2) - p_i^{t_1,t_2}(2,2)\right]\left[p_i^{t_2}(2) - p_i^{t_1,t_2}(2,2)\right]} \right],
\end{aligned}
$$

$t_1 < t_2$ $(t_1, t_2 \in \{1, 2, ..., k\})$. The trivariate and higher order associations are also described by global odds ratios and are assumed zero on the log scale. The parameters for this model are $\alpha_1, \alpha_2, \nu, \beta_1, \beta_2, \beta_3, \delta$, and γ. Table 5 gives the maximum likelihood estimates and the corresponding standard errors computed from the inverse of the Fisher total information matrix. The estimates and standard errors for $\gamma\beta_1, \gamma\beta_2$, and $\gamma\beta_3$ are computed as described before. The negative values for $\widehat{\gamma\beta_1}$ and $\widehat{\beta_1}$ suggest that cholestyramine decreases the number and severity (conditional on k) of seizures. The positive value for $\widehat{\gamma}$ confirms that, on average, the effect of cholestyramine and the other two covariates are in the same direction for both the number and severity of seizures. Again, we ran two separate analyses modeling the length and severity individually (not shown). The model for the random length was again assumed Poisson, while the multivariate severities were modeled using GEE with a cumulative logit link and an exchangeable working covariance matrix. Point estimates and standard errors from this approach differed from the corresponding parameters in the above joint model, though the overall pattern and directions of association were preserved. In particular, both

Table 5: MLE estimates of heart data from joint model

Parameter	Estimate	S.E.	Z-score
α_1	-0.795	0.269	-2.953
α_2	1.769	0.321	5.518
ν	0.873	0.309	2.824
β_1	-0.302	0.263	-1.146
β_2	0.621	0.405	1.535
β_3	-0.091	0.210	-0.433
δ	0.461	0.129	3.583
γ	0.677	0.498	1.360
$\gamma\beta_1$	-0.204	0.116	-1.763
$\gamma\beta_2$	0.420	0.175	2.407
$\gamma\beta_3$	-0.062	0.095	-0.647

β_1 is the parameter for TRT_i; β_2 is the parameter
for RCA_i; β_3 is the parameter for SBP_i.

approaches failed to find significant effects of cholestyramine on either of the disease outcomes (number or degree of occlusion of blockages).

Again we generated a test for a treatment effect, comparing the two multinomial distributions,

$$
\begin{aligned}
\pi_i = \ & \left(\Pr\left(K_i = 0|\text{TRT}_i,\text{RCA}_i, \text{SBP}_i\right), \Pr\left(Y_i = 1|\text{TRT}_i,\text{RCA}_i, \text{SBP}_i\right),\right. \\
& \left. \Pr(Y = 2|TRT_i, \text{RCA}_i, SBP_i), \Pr\left(Y = 3|\text{TRT}_i,\text{RCA}_i, \text{SBP}_i\right)\right)',
\end{aligned}
$$

for the two treatment groups (holding RCA_i and SBP_i at their population averages) using contrasts of baseline category logits, i.e.,

$$
\begin{aligned}
H_0: \quad & \log\left(\frac{\Pr\left(Y = k \mid \text{TRT}_i = 1,\text{RCA}_i, \text{SBP}_i\right)}{\Pr\left(K_i = 0 \mid \text{TRT}_i = 1,\text{RCA}_i, \text{SBP}_i\right)}\right) \\
& = \log\left(\frac{\Pr\left(Y = k \mid \text{TRT}_i = 0,\text{RCA}_i, \text{SBP}_i\right)}{\Pr\left(K_i = 0 \mid \text{TRT}_i = 0,\text{RCA}_i, \text{SBP}_i\right)}\right), \quad k = 1, 2 \text{ and } 3.
\end{aligned}
$$

The Wald chi-square test statistic for H_0 was 10.2. Comparing this to χ_3^2 gives a P-value of 0.017. This indicates that cholestyramine is effective in reducing the chance of developing a lesion as well as the chance of having a larger lesion size.

7. Discussion

We have proposed a general, likelihood based, joint model for multivariate ordinal random length data where both the number of events and the severity of events are outcome measures. A limited simulation showed that the MLE is asymptotically normal. We also showed how one can construct a meaningful test for treatment effect based on both the number and severity of events within the joint modeling framework. We noted that informative dropout was a problem for the epilepsy data example. This was beyond the scope of this paper, though we are currently developing methodology for modeling the informative dropout process within the joint modeling framework.

Acknowledgments

Appendix

Form of $I(\theta)^*$ and $I(\theta|k)$:

Using similar computation methods as in GLIM (McCullagh and Nelder, 1989), one can show that

$$I(\theta)^* = \sum_{i=1}^{N} \left(\frac{\partial \eta_i}{\partial \theta}\right)' \left((h'(E(K_i)))^2 \, Var(K_i)\right)^{-1} \left(\frac{\partial \eta_i}{\partial \theta}\right)$$

where $\eta_i = \delta + \gamma(\mathbf{x}_i' \beta)$ as above. Using these same GLIM techniques, it is easy to show that

$$I_i(\theta|k) = \mathrm{X}_i' \left[\left(\frac{\partial \eta_i}{\partial \pi_i}\right)'\right]^{-1} Cov(Z_i)^{-1} \left(\frac{\partial \eta_i}{\partial \pi_i}\right)^{-1} \mathrm{X}_i$$

where $\eta_i = \left((\eta_i^m)', (\eta_i^a)'\right)'$ and X_i is the design matrix obtained by combining $\mathbf{x}_i$, X_i^m, and X_i^a.

Chi-square test:

Let C be the contrast of multinomial cell estimates that are to be tested. Then a Wald chi-square test statistic,

$$C \left[\frac{\partial C}{\partial \theta} I\left(\hat{\theta}\right)^{-1} \left(\frac{\partial C}{\partial \theta}\right)'\right]^{-1} C'$$

s asymptotically distributed as $\chi^2_{rank(C)}$.

References

Agresti, A. (1990). *Categorical Data Analysis.* New York: Wiley.

Albert, P. S., Follmann, D. A., and Barnhart, H. X. (1997). A generalized estimating equation approach for modeling random length binary data. *Biometrics* **53**, 116-1124.

Barnhart, H. X. (1992). Models for multivariate random length data with applications in clinical trials. Ph.D. Dissertation, University of Pittsburgh.

Barnhart, H. X. and Sampson, A. R. (1995). Multiple population models for multivariate random length data–With applications in clinical trials. *Biometrics* **51**, 195-204.

Barnhart, H. X. (1998). A probit model for multivariate random length ordinal data. *Communications in Statistics* **27** 1693-1713.

Barnhart, H. X., Kosinski, A. S., and Sampson, A. R. (1999). A regression model for multivariate random length data. *Statistics in Medicine* **18**,119-211.

Brensike, J. F., Kelsey, S. F., Passamani, E. R., *et al.* (1982). National Heart, Lung, and Blood Institute Type II coronary intervention study: Design, methods, and baseline characteristics. *Controlled Clinical Trials* **3**, 91-111.

Brensike, J. F., Levy, R. I., Kelsey, S. F., *et al.* (1984). Effects of therapy with cholestyramine on progression of coronary arteriosclerosis: Results of the NHLBI TYPE II coronary intervention study. *Circulation* **69**, 313-324.

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

Glonek, G. F. V. and McCullagh, P. (1994). Multivariate logistic models. *Technical Report 94-31.* School of Information Science and Technology, Flinders University of South Australia, Adelaide.

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

Haber, M. and Brown, M. B. (1986). Maximum likelihood methods for log-linear models when expected frequencies are subject to linear constraints. *Journal of the American Statistical Association* **81,** 477-482.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models,* 2nd edition. London: Chapman and Hall.

Molenberghs, G. and Lesaffre, E. (1999). Marginal modelling of multivariate categorical data. *Statistics in Medicine* **18**, 2237-2255.

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.

Theodore, W. H., Albert, P. S., Stertz, B., Malow, B., Ko, D., White, S., Flamini, R., And Ketter, T. (1995). Felbamate monotherapy: Implications for antiepileptic drug development. *Epilepsia* **36,** 1105-1110.

Andrew S. Allen
Department of Biostatistics and Bioinformatics
and Duke Clinical Research Institute
Duke University
7057 North Pavilion
Box 3850 Med Ctr
Durham, NC 27710, USA
andrew.s.allen@duke.edu

Huiman X. Barnhart
Department of Biostatistics and Bioinformatics
and Duke Clinical Research Institute
Duke University
7057 North Pavilion
Box 3850 Med Ctr
Durham, NC 27710, USA
huiman.barnhart@duke.edu