Marginal Modeling of Correlated Ordinal Data Using a Multivariate Plackett Distribution

Geert MOLENBERGHS and Emmanuel LESAFFRE*

An extension of the bivariate model suggested by Dale is proposed for the analysis of dependent ordinal categorical data. The so-called multivariate Dale model is constructed by first generalizing the bivariate Plackett distribution to any dimensions. Because the approach is likelihood based, it satisfies properties that are not fulfilled by other popular methods, such as the generalized estimating equations approach. The proposed method models both the marginal and the association structure in a flexible way. The attractiveness of the multivariate Dale model is illustrated in three key examples, covering areas such as crossover trials, longitudinal studies with patients dropping out from the study, and discriminant analysis applications. The differences and similarities with the generalized estimating approach are highlighted.

KEY WORDS: Categorical data; Crossover trials; Cross-ratio; Dale model; Dropouts; Longitudinal studies; Multivariate density; Plackett distribution.

1. INTRODUCTION

During the last 20 years there has been an explosion of papers on repeated measurement problems, comprising a fascinating, yet complicated research area in statistics. Until recently, however, most attention was devoted to continuous response models, with emphasis on longitudinal studies that have various interesting aspects. Indeed, longitudinal studies typically have unbalanced designs, missing data, attrition, time-varying covariates, and other characteristics that make standard multivariate procedures inapplicable, as pointed out by Ware (1985). For continuous responses, Ware (1985) and Jennrich and Schluchter (1986) proposed a general approach for analysis using a linear model for the expected responses and structural models for the within-subject covariances. Central in this approach is the use of the Gaussian distribution, for which conditional and marginal models are of the same type. For categorical response models, such a flexible model is not yet available. Although there have been several proposals, none of them incorporated simultaneously a simple model for the conditional and marginal approach (for an extensive review, see Ashby et al. 1992).

The marginal approach has received much attention lately. Here, emphasis is on the efficient estimation of the effect of covariates on the marginal probabilities of a multivariate categorical response vector. From a practical viewpoint, two methods are currently in use: the empirical least squares method (EGLS), as implemented in the procedure CATMOD of SAS (Koch, Landis, Freeman, Freeman, and Lehnen 1977), and the more recent generalized estimating equations (GEE) approach of Liang and Zeger (1986). Neither of these two methods is likelihood based. In the discussion of the paper by Liang, Zeger, and Qaqish (1992), many discussants expressed their preference for likelihood methods. But up to now there has been no flexible, likelihood-based, regression model available for multivariate categorical responses. We are aware of only the multivariate probit model (Lesaffre and Molenberghs 1991); but although this model works well for a number of applications, it is not as flexible as needed.

We have developed a full likelihood method for the analysis of ordinal categorical responses allowing time-varying and subject-specific (continuous) covariates. The model is based on an extension of the two-dimensional Plackett distribution. Three key examples were chosen to illustrate the wide range of applications of the model. The first example is a crossover study with ordinal responses. The second example, a longitudinal clinical trial with an ordinal response and dropouts in time, illustrates the use of the model when some data are missing at random, ruling out the use of the GEE approach. The third example was chosen to illustrate the method's capabilities in discriminant analysis applications. The examples are presented in Section 2. Bivariate data models are presented in Section 3, and the multivariate extension is introduced in Section 4. Maximum likelihood estimation is briefly discussed in Section 5; and the examples are analyzed in Section 6.

2. EXAMPLES

2.1 Example 1: Primary Dysmenorrhea Data

The data are taken from a crossover trial that appeared in the paper of Kenward and Jones (1991). Eighty-six subjects were enrolled in a crossover study that compared placebo (A) with an analgesic at low and high doses (B and C) for the relief of pain in primary dysmenorrhea. The three treatments were administered in one of six possible orders: ABC, ACB, BAC, BCA, CAB, and CBA. The primary outcome score was the amount of relief, coded as none (1), moderate (2), and complete (3). There are 27 possible outcome combinations: (1, 1, 1), (1, 1, 2), . . . , (3, 3, 3), where (a_1, a_2, a_3) denotes outcomes a_i in period i. A table of the realized combinations can be found in the paper of Kenward and Jones (1991). For the analysis of the crossover data, these authors suggested a subject-specific approach based on the Rasch model. Here, too, it was of interest to estimate the treatment, period, and carryover effects.

* Geert Molenberghs is Research Assistant, Department of Mathematics and Computer Science, Universitaire Instelling Antwerpen, Universiteitsplein 1, B-2610 Wilrijk, Belgium. Emmanuel Lesaffre is Associate Professor, Biostatistical Centre, Department of Epidemiology, School of Public Health, Katholieke Universiteit Leuven, Capucijnenvoer 33, B-3000 Leuven, Belgium. The research of Geert Molenberghs was supported by Nationaal Fonds voor Wetenschappelijk Onderzoek. The authors greatly appreciate the various discussions they had with Mike Kenward, which improved the article considerably. They thank Ronald Brand, S. P. Verloove-Vanhorick, and the other POPS researchers for the kind permission to use their data.

2.2 Example 2: A Multicenter Psychiatric Study

In a multicenter study, 315 patients were treated with fluvoxamine for psychiatric symptoms described as possibly resulting from a dysregulation of serotonin in the brain. Patients with one or more of the following diagnoses were included: depression, obsessive-compulsive disorder, and panic disorder. An anamnestic parameters, one has recorded sex; age; psychiatric antecedents (y/n); and duration of actual mental disease and initial severity, coded in 7 classes, from (1) = not ill through (7) = severely ill. The patients in the study were investigated at three subsequent visits. At each visit a detailed evaluation of the symptoms was made by scoring the presence and intensity of psychiatric symptoms, such as the depressive state, insomnia, anxiety, agression, and nausea. Furthermore, during each visit the therapeutic effect and the side effects were scored in an ordinal manner. One of the study's primary endpoints was the intensity of side effects at each visit. Four codes were given: (0) = no, (1) = not interfering with functionality of the patient, (2) = interfering significantly with functionality of the patient, and (3) = side effect surpassing the therapeutic effect. There was a considerable dropout rate at each visit. A logistic regression analysis shows that the probability of dropping out is highly significantly related to the occurrence of side effects at the previous visit. It was of interest to know which factors at baseline influenced this dropout rate significantly. The data set is available from the authors on request.

2.3 Example 3: The POPS Study

The POPS (Project On Preterm and Small-for-gestational age infants) is a multiclinic study that collected, in 1983, information about 1,338 infants with birthweight less than 1,500 g and/or gestational age less than 32 weeks born in the Netherlands in the same year (see Verloove-Vanhorick 1986 for more details). A total of 133 clinics were involved. The study population represents 94% of the births in that year with similar gestational age and birthweight characteristics. Prenatal, perinatal, and postnatal information, as well as 2-year follow-up data, were collected. The data base also contains information on the delivery and specific details of the infant. After 2 years, each child was reexamined. Lesaffre and Molenberghs (1991) studied the relationship between two ability scores at age 2 and related them to risk factors measured at delivery. Both ability scores were recorded in a dichotomous manner. They were available for 799 children. The first score ($ABIL_1$) checks whether the child can pile three bricks; $ABIL_1 = 1$ corresponds to "no," and $ABIL_1 = 2$ to "yes." The second score ($ABIL_2$) measures whether the child's physical movements are natural; $ABIL_2 = 1$ denotes "no" and $ABIL_2 = 2$ denotes "yes." $ABIL_2$ is a purely physical ability score, but $ABIL_1$ is a combination of physical and mental qualities. We also considered a third ability score, $ABIL_3$, which expresses whether or not the child is able to put a ball in a box when asked to do so. The problem was to determine the risk factors for low performance at the three tests. Further, it was of interest to compare the predicted probabilities, taking into account the relationship between the responses to those probabilities calculated under the assumption of independent responses.

3. A LIKELIHOOD METHOD FOR BIVARIATE DEPENDENT ORDINAL DATA

3.1 Two Binary Responses

Suppose that for each of K subjects in a study, a vector of two binary responses $\mathbf{Z}_k = (Z_{k1}, Z_{k2})^T$ is observed, together with a vector of covariates $\mathbf{x}$. The vector $\mathbf{x}$ can be different for each response, as in longitudinal studies with time-dependent covariates. Thus the study subjects are described by $(\mathbf{z}_k, \mathbf{x}_{kt})$, $(k = 1, \ldots, K; t = 1, 2)$. We want to establish the dependence of each of the two responses on the covariate vector(s), taking the dependence between the responses into account. This can be done using a marginal model that regresses each response on the covariate vector. But the marginal distributions do not fully determine the joint distribution. Thus an association parameter is needed to describe the dependence between the two responses. The bivariate probit model (Ashford and Sowden 1970; Lesaffre and Molenberghs 1991) describes the association via the correlation coefficient of an assumed underlying bivariate normal random variable $\mathbf{W}$.

Dale (1986) proposed a family of bivariate response models that also meets the requirements of preceding description and avoids the (restrictive) assumption of a single underlying density. The model arises from the decomposition of the joint probabilities $p_{j_1 j_2}(\mathbf{x}) = P(Z_1 = j_1, Z_2 = j_2 | \mathbf{x})$, $(j_1, j_2 = 1, 2)$ into "main effects" and "interactions." The marginal probabilities describe the main effect, and the log cross-ratio is the interaction term. Formally, this decomposition is given by

$$h_1(p_{1+}(\mathbf{x})) = \beta_1^T \mathbf{x}, \tag{3.1}$$

$$h_2(p_{+1}(\mathbf{x})) = \beta_2^T \mathbf{x}, \tag{3.2}$$

and

$$h_3\left(\frac{p_{11}(\mathbf{x})p_{22}(\mathbf{x})}{p_{12}(\mathbf{x})p_{21}(\mathbf{x})}\right) = \beta_3^T \mathbf{x}, \tag{3.3}$$

where h_1, h_2, and h_3 are link functions in the generalized linear model terminology and $p_{1+}(\mathbf{x})$ and $p_{+1}(\mathbf{x})$ are the marginal probabilities for observing $Z_1 = 1$ and $Z_2 = 1$. The most popular choice for $h_1 \equiv h_2$ is the logit function, whereas for h_3 the natural logarithmic function is commonly used. In that case, one has two marginal logistic regression models, and the logarithm of the cross-ratio

$$\ln \psi(\mathbf{x}) = \ln\left(\frac{p_{11}(\mathbf{x})p_{22}(\mathbf{x})}{p_{12}(\mathbf{x})p_{21}(\mathbf{x})}\right)$$

is linear in the covariates. But other links are equally possible while the links h_1 and h_2 may be different.

The joint probabilities follow from the marginal probabilities in the following way (omitting the dependence of the different terms on $\mathbf{x}$ for ease of notation):

$$p_{11} = \frac{1 + (p_{1+} + p_{+1})(\psi - 1) - S(p_{1+}, p_{+1}, \psi)}{2(\psi - 1)} \quad \text{if } \psi \neq 1,$$

$$= p_{1+}p_{+1} \quad \text{if } \psi = 1, \tag{3.4}$$

and $p_{12} = p_{1+} - p_{11}$, $p_{21} = p_{+1} - p_{11}$, and $p_{22} = 1 - p_{12} - p_{21} - p_{11}$, with

$$S(q_1, q_2, \psi) = \sqrt{[1 + (q_1 + q_2)(\psi - 1)]^2 + 4\psi(1 - \psi)q_1q_2}.$$

The preceding description also arises as the discrete realization of a continuous bivariate Plackett distribution (Plackett 1965). This is seen as follows. Suppose that the bivariate random vector $\mathbf{W} = (W_1, W_2)^T$ has joint distribution function $F(w_1, w_2)$, with marginal distributions $F(w_t)$ $(t = 1, 2)$. Define the (global) cross-ratio function $\psi(w_1, w_2)$ by

$$\psi(w_1, w_2) = \frac{p_{11}p_{22}}{p_{12}p_{21}} = \frac{F(1 - F_1 - F_2 + F)}{(F_1 - F)(F_2 - F)}, \quad (3.5)$$

with $F_t \equiv F_t(w_t)$, $(t = 1, 2)$, and $F \equiv F(w_1, w_2)$. It is clear that $\psi(w_1, w_2)$ satisfies $0 \le \psi \le \infty$. The components $p_{j_1j_2}$ in (3.5) are the quadrant probabilities in $\mathbb{R}^2$ with vertex at (w_1, w_2). For a Plackett distribution, the global cross-ratio $\psi(w_1, w_2) \equiv \psi$ is constant. Equation (3.5) can be seen as a defining equation for F, once F_1, F_2, and ψ are known. The Plackett distribution then gives rise to the preceding bivariate response model if its mean vector $\mu = (\mu_1, \mu_2)^T$ depends linearly on the covariate vector and if it is assumed that $\mathbf{Z}$ is a discretized version of the continuous vector $\mathbf{W}$, in the sense that $Z_t = 1 \Leftrightarrow \theta_t \le W_t$ for $t = 1, 2$. Here θ_1 and θ_2 are two a priori defined thresholds. In other words, Dale's bivariate response model is obtained if the bivariate response vector $\mathbf{Z}$ is a discretized version of $\mathbf{W}$ using the threshold vector θ, and if the covariate vector shifts the mean vector of the distribution of $\mathbf{W}$ over the plane, thereby possibly changing also the association parameter ψ as a function of $\mathbf{x}$.

3.2 The Bivariate Global Cross-ratio Model

Dale (1986) generalized the preceding approach to model $r_1 \times r_2$ contingency tables (representing pairs of ordered categorical variables with r_1 and r_2 levels) in the presence of explanatory variables $\mathbf{x}$. This is called the global cross-ratio model; we will refer to it here as the bivariate Dale model (BDM).

Let $\mathbf{Z} = (Z_1, Z_2)^T$ be a random vector that takes on values (j_1, j_2), where $1 \le j_t \le r_t$, $(t = 1, 2)$. The outcomes, corresponding to a given covariate vector $\mathbf{x}$, can be arranged as an $r_1 \times r_2$ contingency table $(y_{j_1j_2})(j_t = 1, \ldots, r_t; t = 1, 2)$:

$$\left[\begin{array}{ccc|c|ccc} y_{11} & \cdots & y_{1j_2} & y_{1,j_2+1} & \cdots & y_{1r_2} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ y_{j_11} & \cdots & y_{j_1j_2} & y_{j_1,j_2+1} & \cdots & y_{j_1r_2} \\ \hline y_{j_1+1,1} & \cdots & y_{j_1+1,j_2} & y_{j_1+1,j_2+1} & \cdots & y_{j_1+1,r_2} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ y_{r_11} & \cdots & y_{r_1j_2} & y_{r_1,j_2+1} & \cdots & y_{r_1r_2} \end{array}\right]. \quad (3.6)$$

Similarly, the probabilities can be represented as an $r_1 \times r_2$ table:

$$\left[\begin{array}{ccc|c|ccc} p_{11} & \cdots & p_{1j_2} & p_{1,j_2+1} & \cdots & p_{1r_2} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ p_{j_11} & \cdots & p_{j_1j_2} & p_{j_1,j_2+1} & \cdots & p_{j_1r_2} \\ \hline p_{j_1+1,1} & \cdots & p_{j_1+1,j_2} & p_{j_1+1,j_2+1} & \cdots & p_{j_1+1,r_2} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ p_{r_11} & \cdots & p_{r_1j_2} & p_{r_1,j_2+1} & \cdots & p_{r_1r_2} \end{array}\right]. \quad (3.7)$$

Dichotomizing contingency table (3.6) at (j_1, j_2) (double lines) leads to a 2×2 contingency table,

$$\begin{array}{|c|c|} \hline \{Z_1 \le j_1, Z_2 \le j_2\} & \{Z_1 \le j_1, Z_2 > j_2\} \\ \hline \{Z_1 > j_1, Z_2 \le j_2\} & \{Z_1 > j_1, Z_2 > j_2\} \\ \hline \end{array}, \quad (3.8)$$

of which the probabilities are given by

$$p_{11}(j_1, j_2, \mathbf{x}) = P(Z_1 \le j_1, Z_2 \le j_2 | \mathbf{x}),$$

$$P_{12}(j_1, j_2, \mathbf{x}) = P(Z_1 \le j_1, Z_2 > j_2 | \mathbf{x}),$$

$$P_{21}(j_1, j_2, \mathbf{x}) = P(Z_1 > j_1, Z_2 \le j_2 | \mathbf{x}),$$

and

$$P_{22}(j_1, j_2, \mathbf{x}) = P(Z_1 > j_1, Z_2 > j_2 | \mathbf{x}).$$

Marginal probabilities are obtained by summing over subscripts: $P_{1+}(j_1, \mathbf{x}) = P(Z_1 \le j_1 | \mathbf{x})$ and $P_{+1}(j_2, \mathbf{x}) = P(Z_2 \le j_2 | \mathbf{x})$.

In analogy with (3.1)–(3.3), the link functions are described by:

$$h_1(P_{1+}(j_1, \mathbf{x})) = \alpha_{1j_1} + \beta_1^T \mathbf{x}, \quad (j_1 = 1, \ldots, r_1 - 1), \quad (3.9)$$

$$h_2(P_{+1}(j_2, \mathbf{x})) = \alpha_{2j_2} + \beta_2^T \mathbf{x}, \quad (j_2 = 1, \ldots, r_2 - 1), \quad (3.10)$$

and

$$h_3(\psi(j_1, j_2, \mathbf{x})) = \beta_3^T \mathbf{x}, \quad (j_t = 1, \ldots, r_t - 1; t = 1, 2), \quad (3.11)$$

where the global cross-ratio $\psi(j_1, j_2, \mathbf{x})$ is given by

$$\psi(j_1, j_2, \mathbf{x}) = \frac{P_{11}(j_1, j_2, \mathbf{x})P_{22}(j_1, j_2, \mathbf{x})}{P_{12}(j_1, j_2, \mathbf{x})P_{21}(j_1, j_2, \mathbf{x})}.$$

Note that for every contingency table (3.6)—or, equivalently, table of probabilities (3.7)—a set of $(r_1 - 1) \times (r_2 - 1)$ global cross-ratios is obtained:

$$\left[\begin{array}{ccc|c|ccc} \psi_{11} & \cdots & \psi_{1j_2} & \psi_{1,j_2+1} & \cdots & \psi_{1,r_2-1} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \psi_{j_11} & \cdots & \psi_{j_1j_2} & \psi_{j_1,j_2+1} & \cdots & \psi_{j_1,r_2-1} \\ \hline \psi_{j_1+1,1} & \cdots & \psi_{j_1+1,j_2} & \psi_{j_1+1,j_2+1} & \cdots & \psi_{j_1+1,r_2-1} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \psi_{r_1-1,1} & \cdots & \psi_{r_1-1,j_2} & \psi_{r_1-1,j_2+1} & \cdots & \psi_{r_1-1,r_2-1} \end{array}\right].$$

More complex choices for the linear predictors on the right side of (3.9)–(3.11) are possible. For instance, h_3 can incorporate terms depending on j_1 and j_2, representing row, column, and cell effects.

For every table (3.8), we assume that (3.5) holds with ψ replaced by $\psi(j_1, j_2, \mathbf{x})$, indicating that ψ is allowed to depend on the cutpoints and on the covariates. Further, $F(\cdot\,|\mathbf{x}) \equiv F_{j_1 j_2}(\cdot\,|\mathbf{x}) = P_{11}(j_1, j_2, \mathbf{x})$, and $F(\cdot\,|\mathbf{x})$ can also be expressed in terms of the assumed underlying Plackett distribution: $F(\cdot\,|\mathbf{x}) = P(W_1 \le \theta_{1j_1}, W_2 \le \theta_{2j_2}|\mathbf{x})$. Observe that for each double dichotomy of the $r_1 \times r_2$ table, a different underlying Plackett distribution is assumed. When it can be assumed that $\psi(j_1, j_2, \mathbf{x}) \equiv \psi(\mathbf{x})$, for $j_t = 1, \ldots, r_t - 1$ $(t = 1, 2)$, there is a single underlying Plackett distribution, exactly as for the binary response model.

3.3 Some Properties of Dale's Model

Dale's model has appealing properties. First, there is the flexibility with which the marginal structure is modeled; that is, the cumulative marginal probabilities can be fitted in the generalized linear models framework. Second, the marginal parameters are orthogonal onto the association parameters in the sense that the corresponding elements in the expected covariance matrix are identically 0 (Palmgren 1989). Further, the associations can be modeled in a flexible way, including covariate-, row-, and column-, and cell-specific terms (see Dale 1986).

The BDM does not require marginal scores for the responses and is essentially invariant under any monotonic transformation of the marginal response variables. Further, if adjacent marginal categories are combined, then the model for the new table has fewer parameters, but these parameters have the same interpretation as in the model for the original expanded table, because the parameters pertain to cutpoints between categories. This is in contrast to models based on local association (Goodman 1981).

But despite these advantages, the model has not been generalized to three or more dimensions. McCullagh and Nelder (1989) described the three-way decomposition into "main" and "interaction" parameters, but they did not indicate how to compute the probabilities. Recently, Liang et al. (1992) independently suggested marginal models that are very similar to the general Dale model we propose.

4. A LIKELIHOOD METHOD FOR MULTIVARIATE DEPENDENT ORDINAL DATA

The computational basis of the BDM is the Plackett distribution. Therefore, we first generalize the bivariate Plackett distribution to n dimensions. In this section we present a general description and mention some properties without proof. An extensive mathematical description of the multivariate Plackett distribution will be the subject of a separate publication. The multivariate Plackett distribution will be used to construct the multivariate Dale model, which is the basis for our full-likelihood approach.

4.1 Definition of the Multivariate Plackett Distribution

Given the marginal distributions $F_1(w_1)$ and $F_2(w_2)$ and the cross-ratio ψ, the Plackett distribution is the solution of the second-degree polynomial equation

$$\psi(F - a_1)(F - a_2) - (F - b_1)(F - b_2) = 0, \quad (4.1)$$

where $a_1 = F_1$, $a_2 = F_2$, $b_1 = 0$, and $b_2 = F_1 + F_2 - 1$. The solution of this equation is given by (3.4). To yield a genuine distribution function, the solution F of (4.1) should satisfy the Fréchet inequalities (Fréchet 1951): $\max(b_1, b_2) \le F \le \min(a_1, a_2)$. This approach can be generalized to n dimensions. To define the multivariate Plackett distribution, consider the set of $2^n - 1$ generalized cross-ratios with values in $[0, +\infty]$: ψ_t, $(1 \le t \le n)$; $\psi_{t_1 t_2}$, $(1 \le t_1 \le t_2 \le n)$; $\ldots$; $\psi_{t_1 \ldots t_k}$, $(1 \le t_1 < \cdots < t_k \le n)$; $\ldots$; $\psi_{1 \ldots n}$. The one-dimensional ψ_t's are precisely the odds of the univariate probabilities; that is,

$$\psi_t = \frac{p_1^{(t)}}{p_2^{(t)}} = \frac{F_t}{1 - F_t}, \quad (4.2)$$

$(1 \le t \le n)$. The bivariate associations $\psi_{t_1 t_2}$ are defined as in (3.5):

$$\psi_{t_1 t_2} = \frac{p_{11}^{t_1 t_2} p_{22}^{t_1 t_2}}{p_{12}^{t_1 t_2} p_{21}^{t_1 t_2}} = \frac{F_{t_1 t_2}(1 - F_{t_1} - F_{t_2} + F_{t_1 t_2})}{(F_{t_1} - F_{t_1 t_2})(F_{t_2} - F_{t_1 t_2})}, \quad (4.3)$$

$(1 \le t_1 < t_2 \le n)$. As soon as ψ_{t_1}, ψ_{t_2}, and $\psi_{t_1 t_2}$ are known, $F_{t_1 t_2}$ can be calculated. The cross-ratio $\psi_{t_1 t_2}$ can also be viewed as the odds ratio of $\psi_{t_1(1)}$ and $\psi_{t_2(2)}$, computed as in (4.2), within the first and second level of dimension t_2.

The three-dimensional cross-ratios can be defined similarly to the three-factor interactions in log-linear models (see Agresti 1990) and are analogous to the aforementioned extension. Thus the cross-ratio $\psi_{t_1 t_2 t_2}$ is defined as the ratio of two conditional cross-ratios $\psi_{t_1 t_2(1)}$ and $\psi_{t_1 t_2(2)}$, the two-dimensional cross-ratios defined within the first and second level of dimension t_3. The numerator of $\psi_{t_1 t_2 t_3}$ contains $F_{t_1 t_2 t_3}$ with a positive sign, and the denominator contains $F_{t_1 t_2 t_3}$ with a negative sign. Again, the knowledge of the cross-ratios enables one to determine $F_{t_1 t_2 t_3}$.

But care must be taken when specifying the cross-ratios, because not every combination leads to a valid solution. This is not surprising, as the correlation matrix in a multivariate probit model is similarly constrained to be positive definite. An example is given in Section 4.2.

The n-dimensional probabilities can be computed if all lower-dimensional probabilities together with the global cross-ratio of dimension n are known. Let $p_{j_1 \ldots j_k}^{t_1 \ldots t_k}$ be the $(j_1, \ldots, j_k)$-orthant probability of the k-dimensional marginal table, formed by dimensions $(t_1, \ldots, t_k)$. We present the defining equation for $F_{t_1 \ldots t_k}$:

$$\psi_{t_1 \ldots t_k} = \frac{\prod_{(j_1, \ldots, j_k) \in A_k^+} p_{j_1 \ldots j_k}^{t_1 \ldots t_k}}{\prod_{(j_1, \ldots, j_k) \in A_k^-} p_{j_1 \ldots j_k}^{t_1 \ldots t_k}}, \quad (4.4)$$

where $A_k^+ = \{(j_1, \ldots, j_k) \in \{1, 2\}^k | 2 \text{ divides } \sum_{l=1}^k j_l - k\}$ and $A_k^- = \{1, 2\}^k \backslash A_k^+$. In particular, for $F_{1 \ldots n}$,

$$\psi_{1 \ldots n} = \frac{\prod_{(j_1, \ldots, j_n) \in A_n^+} p_{j_1 \ldots j_n}}{\prod_{(j_1, \ldots, j_n) \in A_n^-} p_{j_1 \ldots j_n}}. \quad (4.5)$$

For example, for $n = 3$, $A_1^+ = \{1\}$, $A_2^+ = \{(1, 1), (2, 2)\}$, and $A_3^+ = \{(1, 1, 1), (1, 2, 2), (2, 1, 2), (2, 2, 1)\}$. Based on these expressions, (4.4) yields (4.2), (4.3), and the three-dimensional odds ratio

$$\psi_{123} = \frac{p_{111} p_{122} p_{212} p_{221}}{p_{112} p_{121} p_{211} p_{222}}.$$

The orthant probabilities $p_{j_1 \ldots j_n}$ are determined by the distribution F. A general expression can be derived that will be useful for the automated computation of the orthant probabilities. Some notation is needed. Let $\beta(\mathbf{j}) \equiv \beta(j_1, \ldots, j_n)$ be the set of places for which j_t is equal to 1 (e.g., $\beta(1, 2, 1, 1) = \{1, 3, 4\}$); then

$$p_{j_1 \ldots j_n} = \sum_{\mathbf{s} \supset \beta(\mathbf{j})} \operatorname{sgn}(\mathbf{s}) F_{\mathbf{s}}, \tag{4.6}$$

where

$$\begin{aligned} \operatorname{sgn}(\mathbf{s}) &= 1 \quad \text{if } \#\mathbf{s} - \#\beta(\mathbf{j}) \text{ is even,} \\ &= -1 \quad \text{otherwise,} \end{aligned}$$

and $F_{\mathbf{s}} = F_{s_1 \ldots s_m}$, with $s_1 \leq \cdots \leq s_m$. In the three-dimensional case, the octant probabilities are

$$\begin{aligned} p_{111} &= F_{123}, \\ p_{112} &= F_{12} - F_{123}, \\ p_{121} &= F_{13} - F_{123}, \\ p_{211} &= F_{23} - F_{123}, \\ p_{122} &= F_1 - F_{12} - F_{13} + F_{123}, \\ p_{212} &= F_2 - F_{12} - F_{23} + F_{123}, \\ p_{221} &= F_3 - F_{13} - F_{23} + F_{123}, \end{aligned}$$

and

$$p_{222} = 1 - F_1 - F_2 - F_3 + F_{12} + F_{13} + F_{23} - F_{123}. \tag{4.7}$$

As an example, consider p_{212}. In this case, $\beta(2, 1, 2) = \{2\}$, and there are four possible vectors $\mathbf{s}$: (2), (1, 2), (2, 3), and (1, 2, 3). Therefore, (4.6) yields the expression for p_{212} in (4.7).

The set of $2^n - 1$ generalized cross-ratios fully specifies the n-dimensional Plackett distribution. But from the preceding reasoning it is not clear whether such a distribution always exists. Further, if existence and uniqueness are guaranteed, just how to calculate the distribution is not yet clear, because it is only implicitly specified by (4.4). These matters are discussed in the next section.

4.2 Computational Aspects of the Distribution

Note that the probabilities in the numerator (denominator) of (4.5) involve $+F_{12 \ldots n}(-F_{12 \ldots n})$, and that both numerator and denominator contain an even number of factors. Thus (4.5) may be abbreviated as

$$\psi = \frac{\prod_{i=1}^{2^{n-1}} (F - b_i)}{\prod_{i=1}^{2^{n-1}} (F - a_i)}, \tag{4.8}$$

where $\psi \equiv \psi_{1 \ldots n}$ and $F \equiv F_{1 \ldots n}$. The a_i and b_i are functions of the $(n - 1)$- and lower-dimensional probabilities (or, equivalently, cross-ratios). A valid solution must satisfy

$$\max_i b_i \leq F \leq \min_i a_i. \tag{4.9}$$

But this condition is not satisfied for all choices of a_i and b_i. To see this, take the three-way Plackett distribution. Then, according to (4.9), the one- and two-dimensional marginal distributions must satisfy the following inequalities: $F_{ij} + F_{ik} \leq F_i + F_{jk}$, $(i \neq j \neq k \neq i)$, and $F_1 + F_2 + F_3 \leq 1 + F_{12} + F_{13} + F_{23}$. If $F_1 = F_2 = F_3 = \frac{1}{2}$, $\psi_{12} = .05$, $\psi_{13} = 1$, and $\psi_{23} = 20$, then $F_{13} + F_{23} > F_3 + F_{12}$, and (4.9) cannot be satisfied. Such constraints are not exceptional and can be found in several multivariate distributions; for instance, for the multivariate normal distribution, the correlation matrix must be positive definite.

If (4.9) is satisfied, then existence and uniqueness of a solution is guaranteed by the following lemma. The verification of (4.9) is straightforward, as the functions b_i and a_i are linear functions of the lower-order marginal probabilities:

Lemma 1. Let $P(C) = \psi \prod_{i=1}^{m} (C - a_i) - \prod_{i=1}^{m} (C - b_i)$, where m is even, $0 < \psi < +\infty$ and $b_1 = \max_{1 \leq i \leq m} b_i < \min_{1 \leq i \leq m} a_i = a_1$, then the interval $]b_1, a_1[$ contains exactly one real root of $P(C)$.

Proof. The inequalities $P(a_1) = -\prod_{i=1}^{m} (a_1 - b_i) < 0$ and $P(b_1) = \psi \prod_{i=1}^{m} (b_1 - a_i) > 0$, together with the continuity of $P(C)$, establish the existence. Now, $\partial P / \partial C = \psi \sum_{i=1}^{m} \prod_{j \neq i} (C - a_j) - \sum_{i=1}^{m} \prod_{j \neq i} (C - b_j) = \psi \sum_i T_i - \sum_i S_i$. T_i is a product of $(m - 1)$ negative factors, whence T_i is negative. S_i is positive, so $P(C)$ is strictly decreasing in $]b_1, a_1[$. The result is shown.

It follows from the proof that the regula falsi method with starting points a_1 and b_1 always leads to the solution. Though in general a_1 and b_1 are close to each other and convergence is quickly reached, it is desirable to look for even faster methods. It is our experience that a Newton iteration with starting point say, $\frac{1}{2}(a_1 + b_1)$, converges to the root, generally in three or four steps (with a convergence criterion $|c_{k+1} - c_k| < 10^{-8}$).

An algebraic solution to the two-dimensional problem has been given by Mardia (1970) and by Dale (1986). The three-way Plackett distribution can also be solved algebraically using Ferrari's method for solving fourth-degree polynomials (Chemical Rubber Co. 1972, p. 106). But the solution cannot be written in a mathematically elegant way. From the four-way Plackett distribution on, one must rely on numerical techniques. It is a fundamental result of algebra that a polynomial of degree higher than 5 has no algebraic solution. We recognize that this may seem as a major disadvantage, but a similar problem occurs in other, classical statistical models, such as the probit model, where the probability is known only after evaluating the standard normal integral, usually needing six or seven iterations.

4.3 The Multivariate Dale Model

Given the multivariate Plackett distribution, the multivariate Dale model (MDM) is a straightforward extension of the BDM. Let $\mathbf{W} = (W_1, \ldots, W_n)^T$ have a multivariate Plackett distribution with univariate marginals $F_t(W_t)$, $(t = 1, \ldots, n)$ and a particular set of generalized global cross-ratios. Further, let $\mathbf{Z} = (Z_1, \ldots, Z_n)^T$ be a vector of ordered categorical variables with Z_t, assuming values $j_t = 1, \ldots, r_t$, $(t = 1, \ldots, n)$. Thus, in analogy with the bivariate case, $\mathbf{Z}$ is a discrete realization of $\mathbf{W}$. Assume that we have a sample

of K individuals. For the kth subject, a vector of responses $\mathbf{z}_k$, together with a covariate vector $\mathbf{x}_k$, is observed. Both the marginal distributions and the cross-ratios can depend on the covariates.

For each multi-index $\mathbf{j} = (j_1, \ldots, j_n)$ with $1 \le j_t < r_t$, $(t = 1, \ldots, n)$, define a *2^n-dichotomization table (multiple dichotomy)*: $T_{\mathbf{j}} = \{\mathcal{O}_{\mathbf{s}}(\mathbf{j}) | \mathbf{s} \in \{-1, 1\}^n\}$, where

$$\mathcal{O}_{\mathbf{s}}(\mathbf{j}) = \{\mathbf{Z} | Z_t \le j_t \text{ if } s_t = -1 \quad \text{and} \quad Z_t > j_t \text{ if } s_t = 1\}.$$

This means that at every n-dimensional cutpoint, the data table is collapsed into a $2 \times 2 \times \cdots \times 2$ table. Observe the analogy with the bivariate case. For $n = 2$, $T_{\mathbf{j}}$ contains the four corners of the $r_1 \times r_2$ contingency table, split up at $\mathbf{j} = (j_1, j_2)$.

Every table is assumed to arise as a discretization of a multivariate Plackett distribution. The n marginal distributions are modeled, together with all pairs of two-way cross-ratios. In addition, three-way up to n-way interactions (i.e., generalized cross-ratios) are included to fully specify the joint distribution. Formally, we assume that for each $T_{\mathbf{j}}$, (4.5) holds with a cross-ratio possibly depending on $\mathbf{j}$ and $\mathbf{x}$; that is, $\psi_{1\ldots n}$ is replaced by $\psi(\mathbf{j}; \mathbf{x})$. Further,

$$\begin{aligned} F \equiv F_{\mathbf{j}}(\cdot | \mathbf{x}) &= P(Z_1 \le j_1, \ldots, Z_n \le j_n | \mathbf{x}) \\ &= P(W_1 \le \theta_{1j_1}, \ldots, W_n \le \theta_{nj_n} | \mathbf{x}). \end{aligned}$$

The model description is completed by specifying link functions and linear predictors for both the univariate marginals and the association parameters. If we assume a marginal proportional odds model, then the marginal links can be written as

$$\eta_{tj}(\mathbf{x}) = h_t(P(Z_t \le j | \mathbf{x})) = \alpha_{tj} + \boldsymbol{\beta}_t^T \mathbf{x}, \qquad (1 \le t \le n, 1 \le j < r_t). \quad (4.10)$$

Expression (4.10) can be represented in terms of the latent variables: $h_t(P(W_t \le \theta_{tj} | \mathbf{x})) = \alpha_{tj} + \boldsymbol{\beta}_t^T \mathbf{x}$, $(1 \le t \le n, 1 \le j < r_t)$. As in the bivariate case, common choices for the link functions h_t are the logit and the probit links.

The cross-ratios are usually log-linearly modeled. Covariate terms may be included, together with row-, column-, and cell-specific terms. A possible choice consists of complex models for the bivariate associations and simple ones for the higher-order associations. For a fixed pair of variables (t_1, t_2), where $1 \le t_1 < t_2 \le n$, one can model the log cross-ratio as

$$\begin{aligned} \gamma_{t_1 t_2}^{j_1 j_2}(\mathbf{x}) &= \ln \psi_{t_1 t_2}(j_1, j_2, \mathbf{x}) \\ &= \nu + \rho_{j_1} + \kappa_{j_2} + \tau_{j_1 j_2} + \mathbf{x}^T \boldsymbol{\beta}_{t_1 t_2}. \end{aligned} \quad (4.11)$$

Here ν is an intercept parameter, ρ_{j_1} $(j_1 = 1, \ldots, r_1 - 1)$ are row-specific parameters, κ_{j_2} $(j_2 = 1, \ldots, r_2 - 1)$ are column-specific parameters, and $\tau_{j_1 j_2}$ $(j_1 = 1, \ldots, r_1 - 1; j_2 = 1, \ldots, r_2 - 1)$ are cell-specific parameters. Unicity constraints need to be imposed on the row, column, and cell parameters; for instance, $\rho_1 = 0$, $\kappa_1 = 0$, $\tau_{j_1 1} = 0$, $(j_1 = 1, \ldots, r_1 - 1)$, and $\tau_{1 j_2} = 0$, $(j_2 = 1, \ldots, r_2 - 1)$. The higher-order associations usually are assumed to be constant. Parameter estimates are obtained using the maximum likelihood method.

Because this model description yields the BDM for $n = 2$, it follows that the attractiveness and the flexibility of the original two-dimensional version is carried over on its n-dimensional version. But not all properties of the BDM are inherited by the MDM. As mentioned earlier, Palmgren (1989) showed that the estimated marginal and association parameters are orthogonal. This result holds only partially for the MDM; details can be found in the Appendix.

Once the model, the links, and the linear predictors are specified, the model parameters can be estimated by the maximum likelihood method. Using the multivariate Plackett distribution makes it easy to compute both the joint probabilities and their derivatives. A Fisher scoring algorithm is a good choice, as it also provides the asymptotic expected covariance matrix for the model parameters.

When contrasts of log probabilities are used as link functions (e.g., cumulative logit links for the marginal probabilities and log cross-ratios for the associations), the model can be summarized using the terminology of McCullagh and Nelder (1989, pp. 219–221):

$$\boldsymbol{\eta} = \mathbf{C} \ln(\mathbf{L}\mathbf{p}(\mathbf{x})). \quad (4.12)$$

The vector of n-dimensional probabilities $\mathbf{p}(\mathbf{x})$ is expanded to a vector containing all probabilities of dimensions 1 through n by multiplying $\mathbf{p}$ with an appropriate matrix of constants $\mathbf{L}$. Contrasts of the log probabilities, formed by multiplication with the contrast matrix $\mathbf{C}$, are linked to linear predictors. In general, the MDM can include other than log-linear link functions. Although (4.12) provides a simple and elegant description, it provides no shortcut for the computations. In computing the joint probabilities, (4.12) naturally leads to the defining polynomial for the Plackett distribution, which in turn leads to the previously specified equations.

5. MAXIMUM LIKELIHOOD ESTIMATION

For the MDM, a full maximum likelihood estimation program for arbitrary dimension has been written in GAUSS. Despite the fact that the Plackett distribution is known only implicitly, its values can be computed efficiently using numerical algorithms. Further, the derivatives of the Plackett cumulative distribution function can be evaluated in an analytical way, using implicit derivation. Based on these results, the score functions and the expected Fisher information matrix can be used to implement a convenient Fisher scoring algorithm.

We present the basic tools for the computations. We distinguish between the following parts: model description, likelihood function and cell probabilities, and score functions and information matrix.

For convenience, the observations, sharing covariate vector $\mathbf{x}_i$, are combined into an $r_1 \times \cdots \times r_n$ contingency table. The dimension of this table is abbreviated by $\mathbf{r}$. Denote the entries of this table by $y_{i\mathbf{j}}$. Here $\mathbf{j}$ indicates a multi-index: $\mathbf{j} = (j_1, \ldots, j_n)$, $(1 \le j_t \le r_t, t = 1, \ldots, n)$. In vector notation: $\mathbf{1} \le \mathbf{j} \le \mathbf{r}$. A particular table is indicated by $(y_{i\mathbf{j}})_{\mathbf{j}}$.

We assume that the tables are sampled from a multinomial distribution, with cell probabilities $(p_{i\mathbf{j}})_{\mathbf{j}}$, $(i = 1, \ldots, m)$ given by the MDM. These probabilities are derived from the orthant probabilities defined by (4.6). The model is fully

specified by link functions $\eta_{itj} = \eta_{tj}(\mathbf{x}_i)$ given by (4.10) and $\gamma_{t_1t_2}^{j_1j_2}(\mathbf{x}_i)$ given by (4.11), together with the assumption that the higher-order association parameters are constant. If we denote the vector of three- and higher-order associations by ϕ with an appropriate subscript, then we obtain in vector notation $\ln \psi_{\mathbf{h}} = \phi_{\mathbf{h}}$, with $\mathbf{h}$ a vector running through all higher-order associations. The parameters γ and ϕ determine the association structure.

Assume that all parameters form a column vector θ. The log-likelihood takes the form

$$l(\theta) = \sum_{i=1}^{m} \sum_{\mathbf{j}=\mathbf{1}}^{\mathbf{r}} y_{i\mathbf{j}} \ln p_{\mathbf{j}}(\theta, x_i), \tag{5.1}$$

and is fully determined if we indicate in what way the cell probabilities $p_{i\mathbf{j}}(\theta) = p_{\mathbf{j}}(\theta, \mathbf{x}_i)$ arise from the link functions. Let $q_{i\mathbf{j}} = q_{\mathbf{j}}(\mathbf{x}_i)$ denote the n-dimensional cumulative Plackett distribution function F, evaluated in the appropriate links:

$$q_{i\mathbf{j}} = F(\eta_i, \gamma_i, \phi), \tag{5.2}$$

where the arguments are appropriately vectorized forms of the links. Note that $q_{i\mathbf{j}}$ is the orthant probability of $[-\infty, \eta_{i_1j_1}] \times \cdots \times [-\infty, \eta_{i_nj_n}]$. To compute the cell probabilities, write the cutpoints for dimension t as $-\infty = \eta_{it0} < \eta_{it1} < \cdots < \eta_{it,r_t-1} < \eta_{itr_t} = +\infty$. If one or more components j_t of $\mathbf{j}$ equal 0, then the corresponding orthant probability $q_{i\mathbf{j}}$ vanishes. If one or more components of $\mathbf{j}$ equal r_t, then $q_{i\mathbf{j}}$ is an orthant probability of a lower-dimensional marginal distribution.

The cell probabilities $p_{i\mathbf{j}}$ can be expressed in terms of $q_{i\mathbf{j}}$: $p_{i\mathbf{j}} = \sum_{\mathbf{h}} (-1)^{s(\mathbf{j},\mathbf{h})} q_{i\mathbf{h}}$. Summation goes over all indices $\mathbf{h}$ satisfying $\mathbf{0} \le \mathbf{j} - \mathbf{h} \le \mathbf{1}$, and the function S is defined by $S(\mathbf{j}, \mathbf{h}) = \sum_{t=1}^{n} j_t - h_t$. The computation of $q_{\mathbf{j}}$ in (5.2) involves the evaluation of the cumulative Plackett distribution. The derivatives are computed by implicit derivation of (4.8).

The derivative of the log-likelihood with respect to a marginal parameter θ can be written as

$$\frac{\partial l}{\partial \theta} = \sum_{i=1}^{m} \sum_{\mathbf{j}=\mathbf{1}}^{\mathbf{r}} y_{i\mathbf{j}} \frac{1}{p_{i\mathbf{j}}} \sum_{t=1}^{n} \sum_{k=1}^{r_t-1} \frac{\partial p_{i\mathbf{j}}}{\partial \eta_{tk}(\mathbf{x}_i)} \frac{\partial \eta_{tk}(\mathbf{x}_i)}{\partial \theta}. \tag{5.3}$$

A few conventions will simplify notation. First, assume that there is only one covariate vector $\mathbf{x}$, thereby dropping the index i. Second, due to model (4.10), a marginal parameter pertains to only one margin, t say. For such a parameter, summation over all $t = 1, \ldots, n$ is replaced by a single t. In principle, we need to distinguish between intercepts α_{tk}, corresponding to only one cutpoint k, and covariate parameters β, common to all cutpoints $k = 1, \ldots, r_t - 1$ of dimension t. But we assume that every marginal parameter pertains to only one cutpoint, k_t say. The correct formula can be obtained by summing over all cutpoints, if needed. In conclusion, t and $k = k_t$ are assumed to be fixed. Finally, note that in most formulas, some indices j_t of $\mathbf{j}$ will play a particular role and should be mentioned explicitly. The remaining indices will be denoted by $\mathbf{j}$. Accordingly, the upper bound is denoted by $\mathbf{r}'$. In subscripts (e.g., $p_{\mathbf{j}}$), only the relevant indices will be mentioned. Applying these conventions to (5.3) yields

$$\frac{\partial l}{\partial \theta} = \frac{\partial \eta_{tk}}{\partial \theta} \sum_{\mathbf{j}'=\mathbf{1}}^{\mathbf{r}'} \left(\frac{y_k}{p_k} - \frac{y_{k+1}}{p_{k+1}} \right) \sum_{\mathbf{h}, h_t = k} (-1)^{s(\mathbf{j}'_k, \mathbf{h})} \frac{\partial q_{\mathbf{h}}}{\partial \eta_{tk}}.$$

For an intercept or covariate parameter in the two-way association model, we deduce

$$\frac{\partial l}{\partial \theta} = \frac{\partial \gamma_{t_1t_2}}{\partial \theta} \sum_{\mathbf{j}} y_{\mathbf{j}} \frac{1}{p_{\mathbf{j}}} \psi_{t_1t_2}^{j_1j_2} \sum_{\mathbf{h}} (-1)^{s(\mathbf{j},\mathbf{h})} \frac{\partial q_{\mathbf{h}}}{\partial \psi_{t_1t_2}}.$$

Note that a similar form obtains for higher-order associations. For a parameter θ in (4.11) pertaining to a row category k, the score equation is

$$\frac{\partial l}{\partial \theta} = \frac{\partial \gamma_{t_1t_2}}{\partial \theta} \sum_{\mathbf{j}'=\mathbf{1}}^{\mathbf{r}'} \left(\frac{y_k}{p_k} - \frac{y_{k+1}}{p_{k+1}} \right) \psi_{t_1t_2}^{kj_2} \sum_{\mathbf{h}, h_{t_1} = k} (-1)^{s(\mathbf{j}'_k, \mathbf{h})} \frac{\partial q_{\mathbf{h}}}{\partial \psi_{t_1t_2}},$$

whereas for a cell-specific parameter we find

$$\frac{\partial l}{\partial \theta} = \frac{\partial \gamma_{t_1t_2}}{\partial \theta} \sum_{\mathbf{j}'=\mathbf{1}}^{\mathbf{r}'} \left(\frac{y_{k_1k_2}}{p_{k_1k_2}} - \frac{y_{k_1+1,k_2}}{p_{k_1+1,k_2}} - \frac{y_{k_1,k_2+1}}{p_{k_1,k_2+1}} + \frac{y_{k_1+1,k_2+1}}{p_{k_1+1,k_2+1}} \right)$$
$$\times \psi_{t_1t_2}^{k_1k_2} \sum_{\mathbf{h}, h_{t_1}=k_1, h_{t_2}=k_2} (-1)^{s(\mathbf{j}_{k_1k_2}, \mathbf{h})} \frac{\partial q_{\mathbf{h}}}{\partial \psi_{t_1t_2}}.$$

Straightforward but lengthy computations lead to expressions for the elements of the expected information matrix. We do not present them here. They are used to implement a Fisher scoring algorithm, to maximize (5.1). Full details are described in a technical report that can be obtained from the authors.

6. ANALYSIS OF THE EXAMPLES

6.1 The Primary Dysmenorrhea Data

6.1.1 Modeling Crossover Data. Consider a crossover trial where each patient subsequently receives each of three treatments (A, B, C) in a random order. There are six treatment sequences: ABC, ACB, BAC, BCA, CAB, and CBA. Suppose that the outcome at time i (corresponding to treatment j) is an ordered categorical variable Y_{ij} with t levels. Then a $t \times t \times t$ table is assigned to each sequence, containing the joint outcomes for the patients allocated to that particular sequence. The MDM can be used to fit such data. The marginal parameters are used to describe the overall treatment effects, the period and the carry-over effects. The cross-ratios play a role, similar to the subject specific parameters in the paper of Kenward and Jones (1991).

Given a particular sequence s, let $L_{ijk}^{s} = \text{logit}(P(Y_{ij} \le k))$ be the cumulative logit for cutpoint k ($k = 1, \ldots, t-1$), and time i which, for sequence s, corresponds to treatment j. In full detail, we have L_{11k}^{ABC}, L_{22k}^{ABC}, L_{33k}^{ABC}; L_{11k}^{ACB}, L_{23k}^{ACB}, L_{32k}^{ACB}; L_{12k}^{BAC}, L_{21k}^{BAC}, L_{33k}^{BAC}; L_{12k}^{BCA}, L_{23k}^{BCA}, L_{31k}^{BCA}; L_{13k}^{CAB}, L_{21k}^{CAB}, L_{32k}^{CAB}; L_{13k}^{CBA}, L_{22k}^{CBA}, L_{31k}^{CBA}.

The following model for the logits is adopted: $L_{ijk}^{s} = \mu_k + \tau_j + \rho_i + \lambda_{s(i-1)}$, where μ_k are intercept parameters, τ_j are treatment effects, ρ_i are period effects, and $\lambda_{s(i-1)}$ stands for the carry-over effect, corresponding to the treatment at time $i-1$ in sequence s. Given for instance sequence CAB, we get $L_{13k} = \mu_k + \tau_3 + \rho_1$, $L_{21k} = \mu_k + \tau_1 + \rho_2 + \lambda_3$, and L_{32k}

Table 1. The Primary Dysmenorrhea Data: Selection of Effects

	Effects	Log-likelihood	vs.	G^2	df	P value
Marginal effects						
1	μ_k	−279.74				
2	μ_k, τ_i	−245.53	1	68.42	2	<.0001
3	μ_k, τ_i, ρ_j	−243.78	2	3.50	2	.1740
4	$\mu_k, \tau_i, \lambda_{s(i-1)}$	−245.40	2	.26	2	.8790
Model 2 + association effects						
5	$\mu_k, \tau_i; \mu, \psi_{123}$	−244.40				
6	$\mu_k, \tau_i; \mu, \tau_{ii'}, \psi_{123}$	−239.54	5	9.66	2	.0080
7	$\mu_k, \tau_i; \mu, \rho_{jj'}, \psi_{123}$	−239.50	5	9.73	2	.0077
8	$\mu_k, \tau_i; \mu, \tau_{ii'}, \rho_{jj'}, \psi_{123}$	−236.44	5	15.87	4	.0032
			6	6.21	2	.0448
			7	6.14	2	.0465

NOTE: The columns describe the model number, the effects included, the log-likelihood of the model, the number of the model to which this model is compared, the G^2 statistics with the number of degrees of freedom, and the corresponding P value.

$= \mu_k + \tau_2 + \rho_3 + \lambda_1$. To avoid overparameterization, the following unicity constraints are set: $\tau_1 = \rho_1 = \lambda_1 = 0$.

Let $R^s_{ij,i'j'} = \ln \psi^s_{ij,i'j'}$ be the log cross-ratio for the marginal $t \times t$ table, formed by the responses at times i and i for sequence s (corresponding to treatments j and j). The simplest model for the cross-ratios is given by $R^s_{ij,i'j'} = \mu$. The most complex model assumes all 18 cross-ratios to be different, which was done by Jones and Kenward (1989) and by Becker and Balagtas (1993). In between those two models there is room for modeling. One can think of the following linear models in the log cross-ratios:

$$R^s_{ij,i'j'} = \mu + \tau_{jj'}, \tag{6.1}$$

$$R^s_{ij,i'j'} = \mu + \rho_{ii'}, \tag{6.2}$$

and

$$R^s_{ij,i'j'} = \mu + \tau_{jj'} + \rho_{ii'}, \tag{6.3}$$

where μ is an intercept parameter, $\tau_{jj'}$ are parameters for the joint (j, j')th treatments effects, and $\rho_{ii'}$ describe effects for periods i and i'. In model (6.1) the log cross-ratio depends only on the treatments, irrespective of their order and the periods in which they were administered. In model (6.2) only the periods are of importance. In model (6.3) the two effects are combined linearly. For instance, for sequence CAB we get $R_{13,21} = \mu + \tau_{13} + \rho_{12}$, $R_{13,32} = \mu + \tau_{23} + \rho_{13}$, and $R_{21,32} = \mu + \tau_{12} + \rho_{23}$. Possible unicity constraints are $\tau_{12} = \rho_{12} = 0$. Model (6.2) corresponds to the model introduced in Section 4.3. In models (6.1) and (6.3) the two-way cross-ratios change with the treatment combination, which is a time-dependent covariate. Finally, in all six cases the three-way association depends on the same periods and treatments, the only difference being the order in which the treatments occur. So the most natural choice is $R^s_{123} = \mu + \mu^s$, $(\mu^{ABC} = 0)$; however, in most cases it is reasonable to assume that $R^s_{123} = R_{123}$ constant over sequences.

No carry-over effects are incorporated in the cross-ratios, as the marginal carry-over parameters have no straightforward generalization. As usual, the different nested models can be tested using the likelihood ratio G^2 statistic.

6.1.2 Analysis of the Primary Dysmenorrhea Data

Table 1 gives the details concerning the selection of effects for the primary dysmenorrhea data. As can be seen from this table, the marginal logit modeling yields a highly significant treatment effect. The period and carry-over effects are not significant. The model retained (model I in Table 2) consists of two cutpoints μ_k and two treatment parameters τ_i; the estimates are shown in Table 2. Up to now, no two-way or three-way association is assumed.

The next step is to model the association structure; the three-way association is assumed constant in all cases. First the minimal model is fitted. This model will serve as the basic model against which the other models will be compared. Models (6.1), (6.2), and (6.3) were fitted to the data. There seems to be evidence that both the treatment terms as well as the period terms are necessary. The maximal model (i.e., with 18 cross-ratios) has a G^2 statistic of 16.27 (df = 13, P = .2349) compared with model III. Model II in Table 2 shows the parameter estimates when treatment parameters are included in the two-way cross-ratios. Model III contains as association parameters the intercept μ, treatment effects $\tau_{ii'}$, period parameters $\rho_{jj'}$ and the three-way interaction $\ln \psi_{123}$. This model will be chosen.

Parameter interpretation is as follows. The odds of observing $Y_{ij} \leq k$ $(k = 1, 2)$ decreases with factor $\exp(-1.98)$ when the patient is treated with the analgesic at low dose rather than with placebo. A further decrease with factor $\exp(-2.37 + 1.98)$ is observed if the patient is treated with

Table 2. Models Fitted to the Primary Dysmenorrhea Data

	Model I	Model II	Model III
Marginal effects			
μ_1	1.07 (.25)	1.07 (.24)	1.08 (.24)
μ_2	2.71 (.29)	2.70 (.29)	2.72 (.29)
τ_2	−2.03 (.33)	−2.02 (.35)	−1.98 (.34)
τ_3	−2.41 (.33)	−2.37 (.36)	−2.37 (.35)
Two-way association effects			
μ	0 (—)	−.62 (.47)	−.46 (.56)
τ_{13}	0 (—)	−.16 (.65)	−.10 (.58)
τ_{23}	0 (—)	1.51 (.64)	1.32 (.61)
ρ_{13}	0 (—)	0 (—)	−1.12 (.55)
ρ_{23}	0 (—)	0 (—)	.51 (.66)
Three-way association	1 (—)	1.59 (.75)	.63 (.88)
Log-likelihood	−245.53	−239.54	−236.43

NOTE: Each entry represents the parameter estimates (standard error). The absence of a standard errors corresponds to a preset value.

the analgesic at high dose. Further, the association between responses is higher if they are close to each other in time ($\hat{\rho}_{13} = -1.12$). Also, responses from the two analgesic treatments are more associated than responses from one analgesic treatment and placebo ($\hat{\tau}_{23} = 1.32$).

Thus our analysis confirms the results found by Kenward and Jones (1991). But the marginal approach here allows the estimation of treatment effects which now are easily interpretable, in contrast with Kenward and Jones (1991) and with the conditional approach in Jones and Kenward (1989). Confidence intervals for the effects can be found from the estimated standard errors, shown in Table 2 for model III. Finally, the method allows flexible modeling of the association.

6.2 The Psychiatric Study

The relationship between the severity of the side effects at the three visits and some baseline characteristics of the patients was established. The response is a trivariate ordered categorical vector with four classes, measured at three visits. For the selection of significant predictors of the response, age and sex were fixed into the model. The other baseline characteristics were then considered for selection. Only the duration (months) of the disease and the initial severity (measured on a seven-point scale) turned out to significantly influence the severity of side effects.

At the second and third visit, a nonnegligible portion of the patients (20%) dropped out from the study. An ordinary contingency table analysis, as well as a logistic regression of the variable dropout on potential covariates, showed that the dropout mechanism is heavily dependent on the severity of the side effect reported at the preceding visit. We cannot claim that the missing data are missing completely at random. Thus the assumption that the data are missing at random is plausible (see Little and Rubin 1987, chap. 5). These authors showed that for this pattern of missing data, valid inferences can still be drawn if the analysis is based on likelihood methods. If we want to take the possibility into account that the missing data mechanism cannot be ignored, then this mechanism must be modeled explicitly. This extension will be the goal of future research. If we assume missing at random, then our analysis is valid despite the dropouts, a property the GEE (Liang and Zeger 1986) does not possess. We argue that this is an important advantage of our method over the now-popular GEE method.

From the parameter estimates shown in Table 3 (model I), it is seen that the effect of some covariates is almost constant over time. The G^2 test statistic for the hypothesis that both the intercepts and parameters for age and sex are time-invariant is 5.37 (df = 10, $P = .8654$). But duration and initial severity depend on time ($G^2 = 37.58$, df = 4, $P < .0001$). This leads to a more parsimonious model II. The odds of observing high side effects increases with age and duration and decreases with initial severity. The influence of initial severity increases over time. There is a strong association between side effects measured at successive visits. Although significant, the association is less strong between the first and third visit.

Table 3. Analysis of the Psychiatric Study Data

Model I			
Marginal Parameters			
Parameter	*Side 1*	*Side 2*	*Side 3*
μ_1	−.41 (.90)	−.45 (.95)	−.79 (1.06)
μ_2	1.78 (.90)	1.64 (.96)	1.64 (1.07)
μ_3	2.94 (.92)	2.97 (.99)	2.85 (1.13)
Age	−.19 (.09)	−.22 (.09)	−.25 (.10)
Duration	−.14 (.05)	−.20 (.05)	−.24 (.06)
Initial severity	.29 (.14)	.28 (.15)	.42 (.17)
Sex	−.23 (.24)	.09 (.24)	.16 (.27)
Association Parameters			
12	*13*	*23*	*123*
3.20 (.27)	2.49 (.28)	3.71 (.33)	−.38 (.76)
Model II			
Marginal Parameters			
Parameter	*Side 1*	*Side 2*	*Side 3*
μ_1		−.52 (.82)	
μ_2		1.67 (.82)	
μ_3		2.89 (.84)	
Age		−.21 (.07)	
Duration	−.14 (.05)	−.21 (.05)	−.24 (.06)
Initial severity	.27 (.13)	.33 (.13)	.42 (.13)
Sex		−.06 (.22)	
Association Parameters			
12	*13*	*23*	*123*
3.13 (.26)	2.43 (.27)	3.74 (.33)	−.29 (.74)

NOTE: The side effects at three successive times are regressed over age, duration, initial severity, and sex. In model I the parameters are assumed to be different over time. In model II only duration and initial severity have a time-dependent effect. The entries represent the parameter estimates (standard errors).

6.3 The POPS Example

From the eight candidate predictor variables, neonatal seizures (NSZ), congenital malformation (CGM), and highest bilirubin value since birth (BIL) were retained for analysis. They were selected using a stepwise logistic analysis for each response separately, at significance level .05. The first two regressors are dichotomous; the third is continuous.

We fitted the trivariate Dale model (TDM) with both normal (N) and logistic (L) margins. We also fitted the trivariate probit model (TPM). (See, for example, Lesaffre and Molenberghs 1991 for an extensive description of this approach.) Table 4 contains the estimated parameters under the TPM, TDM-N, and TDM-L. It is seen that the presence of neonatal seizures and/or congenital malformation significantly decreases the probability of successfully performing any of the three ability tests. A similar effect of BIL on $ABIL_1$ and $ABIL_2$ is observed.

Based on the log-likelihood, the TDM is slightly preferable. The association is given by means of correlations for the TPM and cross-ratios for the Dale models. There is a strong association between each pair of dichotomous responses, but no significant three-way association.

Table 4. The POPS Study: Parameter Estimates (Standard Errors) for the Trivariate Models

	TPM	TDM-N	TDM-L
Association 12	.73	17.37	17.35
	1.85 (.23)	2.85 (.30)	2.85 (.30)
Association 13	.81	30.64	30.61
	2.27 (.25)	3.42 (.32)	3.42 (.32)
Association 23	.72	17.70	17.65
	1.83 (.23)	2.87 (.31)	2.87 (.31)
Association 123	—	.91	.92
	—	−.09 (.76)	−.09 (.76)
Log-likelihood	−570.69	−567.11	−567.09
	Parameters for ABIL1		
CONST	2.01 (.26)	2.03 (.27)	3.68 (.52)
NSZ	−1.12 (.26)	−1.16 (.26)	−2.06 (.44)
CGM	−.61 (.18)	−.62 (.18)	−1.17 (.33)
BIL (×100)	−.32 (.14)	−.32 (.14)	−.64 (.27)
	Parameters for ABIL2		
CONST	2.19 (.27)	2.21 (.27)	4.01 (.54)
NSZ	−1.27 (.26)	−1.29 (.26)	−2.28 (.44)
CGM	−.56 (.19)	−.59 (.19)	−1.11 (.34)
BIL (×100)	−.42 (.14)	−.41 (.14)	−.80 (.27)
	Parameters for ABIL3		
CONST	1.84 (.27)	1.91 (.27)	3.49 (.54)
NSZ	−.88 (.27)	−.93 (.27)	−1.70 (.46)
CGM	−.47 (.19)	−.49 (.19)	−.96 (.35)
BIL (×100)	−.21 (.14)	−.24 (.14)	−.49 (.28)

NOTE: For the associations, two entries are given: correlations and transformed correlations ($\varphi = \ln((1+\rho)/(1-\rho))$) for the *TPM* and cross-ratios and log cross-ratios for the *TDM*.

Note that the coefficients associated with the marginal risk probabilities are close to each other for all three models if one multiplies the coefficients of the BDM-L with the well-known factor $\sqrt{\pi}/3$.

Another feature of the likelihood method is that calculation of individual probabilities can be performed. For example, the method allows one to calculate the joint probability of failing at the three tests. This can be quite different from the joint probability obtained by assuming independent responses, as is shown in Figure 1, where the probability that the child will fail on all three ability scores is calculated for different bilirubin values, given that both CGM and NSZ are 1.

7. DISCUSSION

A model has been proposed for the analysis of dependent ordinal categorical data using an underlying Plackett distribution. As for the bivariate global cross-ratio model, this assumption is not essential. In our case it was the vehicle to generalize the BDM to any dimensions. For a model with association depending on cutpoints $\mathbf{j}$ and covariates $\mathbf{x}$, the Plackett distribution changes accordingly with $\mathbf{j}$ and $\mathbf{x}$. Thus the model can be used when the assumption of a single underlying distribution does not hold. There is no claim that using our model is physically or logically more justified than using any other multivariate model. But as it is usually the case, our model is a good candidate for modeling categorical dependent variables because of its elegant statistical properties. For instance, the flexibility with which the marginal and association structures can be modeled is a great advantage. Furthermore, because some investigators may find the odds ratio easier to interpret, this model has also some interpretative advantages over, say, the multivariate probit model. (See also Lipsitz et al. 1991 for a similar argument in the related GEE methodology.) Finally, the philosophy underlying the model was recently and independently touched on, although from a different angle, by Liang et al. (1992).

The multivariate Dale approach does not support the analysis of nominal categorical data. Although this is a limitation, we claim that the model covers the most interesting applications, certainly in the area of clinical trials. Furthermore, the model can be easily adapted to cover mixtures of continuous and ordinal responses. This will be investigated in the future.

At several occasions in the analysis of the examples, our approach is compared with other existing approaches. Thereby we showed the advantages of our likelihood method over the GEE approach. These advantages are shared by other likelihood-based methods. For binary responses the model specification is "close" to the GEE approach. In GEE1 only the margins are modeled, whereas in GEE2 the second-order cross-ratios are modeled as well. In both cases the specifications are the same as in our model. But one can see that even if the higher-order cross-ratios are all set equal to 1, they cannot be left out from the model if a full likelihood method is envisaged. So the difference between our method and the GEE approach is that in the former model all higher-order cross-ratios are kept in the model even if they are not important. Despite the stated advantages, however, we recognize that the GEE approach has appealing properties. The most important one is possibly the consistency in GEE1 of the marginal parameter estimates, even under misspecification of the association structure. Theoretical consistency

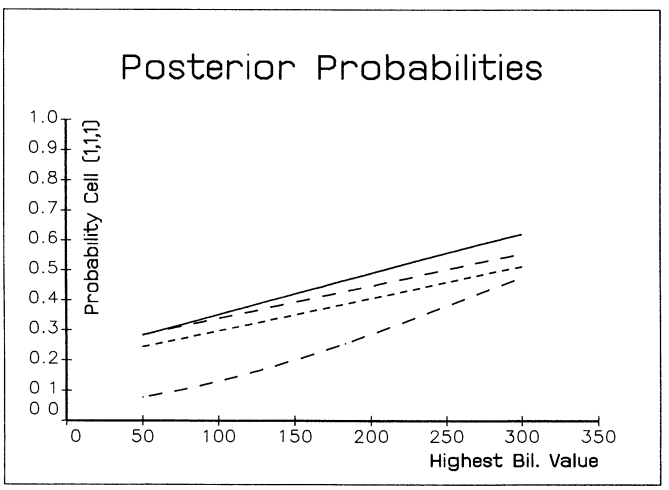

Figure 1. POPS Study: probability that a child fails on all three ability scores for a range of bilirubin values, evaluated under four fitted models: the trivariate Dale model (TDM), with logistic (L; solid line) and Normal (N; large-dashed line) margins; the trivariate probit model (TPM; small-dashed line), and the model assuming independent responses (three logistic regressions; dashed and dotted line).

and robustness properties still have to be investigated for the n-way Dale model, but some empirical investigations are very promising.

For categorical responses the global cross-ratio was chosen as association parameter. Other choices are possible, such as the local cross-ratio. There is some debate about the choice between a local or a global cross-ratio model. We do not want to add anything to this dispute. But it should be noted that similar statistical arguments can be given to defend both models, which is far from saying that both approaches are suitable in all circumstances.

Possibly a disadvantage of our model is that it is only implicitly defined. Only parts of the model are "visible"—the marginal probabilities and the association structure (again a similarity with the GEE approach). The consequence is that some statistical properties of the model are "hidden." But this was no obstacle to writing a full maximum likelihood program based on the Fisher scoring algorithm. The program, written in GAUSS, is available from the authors on request. The derivation of the score equations and the expected Fisher information matrix, which forms the basis for the Newton–Raphson fitting algorithm, is summarized in a technical report that also can be obtained from the authors.

An important advantage of the MDM is that classical goodness-of-fit tests, available for likelihood methods, can be performed. Pearson's X^2 statistic can be used. Furthermore, hypotheses formulated in terms of parameters and/or multivariate cell probabilities can be tested using the likelihood ratio G^2 statistic.

A formal comparison between the Dale model and a version of GEE for the analysis of multivariate ordinal data will be the subject of a separate publication.

APPENDIX: ORTHOGONALITY

Palmgren (1989) showed that for the BDM, the marginal parameters and the global cross-ratio are orthogonal, in the sense that the corresponding expected correlations vanish. This section presents an analogous result for the MDM.

Suppose that there are m groups (covariate combinations), with an $r_1 \times r_2 \times \cdots \times r_n$ contingency table corresponding to each group. We denote the jth cell of the ith table ($i = 1, \ldots, m$ and $j = 1, \ldots, R$ with $R = \prod_{s=1}^{n} r_s$) by y_{ij} and denote the total number of observations in the ith table by y_i. Let γ be the part of the parameter vector corresponding to the n-way association $\psi_{12\ldots n} = \psi$, and let α be the parameters corresponding to the cumulative logits and the lower-order associations. In that case, the log-likelihood for the MDM is $l = \sum_{i=1}^{m} \sum_{j=1}^{R} y_{ij} \ln p_{ij}(\alpha, \gamma) = \sum_{i=1}^{m} l_i(\alpha, \gamma)$. We prove the following result.

Theorem 1. Assume γ and α are disjoint. Choose parameters $\gamma \in \gamma$ and $\alpha \in \alpha$, with α corresponding to A, where A is either a logit or a lower order association. Then,

$$E\left(\frac{\partial^2 l}{\partial\alpha\partial\gamma}\right) = E\left(-\sum_{i=1}^{m} \frac{\partial l_i}{\partial\alpha}\frac{\partial l_i}{\partial\gamma}\right) = 0. \tag{A.1}$$

Proof. The right side of (A.1) can be rewritten as

$$E\left(-\sum_{i=1}^{m} \frac{\partial l_i}{\partial\alpha}\frac{\partial l_i}{\partial\gamma}\right) = -\sum_{i=1}^{m} \frac{\partial\psi}{\partial\gamma}\frac{\partial A}{\partial\alpha} y_i \sum_{j=1}^{R} \frac{1}{p_{ij}}\frac{\partial p_{ij}}{\partial\psi}\frac{\partial p_{ij}}{\partial A}. \tag{A.2}$$

In the expression for the cell probabilities p_{ij}, $F_{12\ldots n} = F$ depends on ψ, but the lower-order marginal distributions are independent of ψ. Thus

$$\frac{\partial p_{ij}}{\partial\psi} = \operatorname{sgn}(j)\frac{\partial F}{\partial\psi}, \tag{A.3}$$

where $\operatorname{sgn}(j)$ denotes the sign of F in the expression for p_{ij}. Setting $V_i = (\partial\psi/\partial\gamma)(\partial A/\partial\alpha)(\partial F/\partial\psi) y_i$ and combining (A.2) and (A.3) yields

$$\begin{aligned} E\left(-\sum_{i=1}^{m} \frac{\partial l_i}{\partial\alpha}\frac{\partial l_i}{\partial\gamma}\right) &= \sum_{i=1}^{m} V_i \sum_{j=1}^{R} \operatorname{sgn}(j)\frac{1}{p_{ij}}\frac{\partial p_{ij}}{\partial A} \\ &= \sum_{i=1}^{m} V_i \sum_{j=1}^{R} \frac{\partial}{\partial A}(\operatorname{sgn}(j)(\ln p_{ij})) \\ &= \sum_{i=1}^{m} V_i \frac{\partial}{\partial A}(\ln\psi) \equiv 0, \end{aligned}$$

completing the proof.

Theorem 1 states that the expected asymptotic covariance matrix has a block structure. The block corresponding to covariances between the n-way association parameters and the other parameters is 0. Note that this is the only block with this property. It does not hold for the blocks pertaining to the marginal parameters and the lower order association parameters. A counterexample is provided by the TPM. In this case, there are three marginal parameter vectors (α_1, α_2, and α_3), three bivariate cross-ratio parameter vectors (γ_{12}, γ_{13}, and γ_{23}), and a trivariate association parameter vector γ_{123}. Choose $\gamma \in \gamma_{12}$ and $\alpha \in \alpha_t$ for some $t = 1, 2, 3$. The left side of (A.1) becomes

$$E\left(-\sum_{i=1}^{m} \frac{\partial l_i}{\partial\alpha}\frac{\partial l_i}{\partial\gamma}\right) = \sum_{i=1}^{m} V_i\left[G_1\frac{\partial}{\partial A}\ln\psi_{12,(j=1)} - G_2\frac{\partial}{\partial A}\ln\psi_{12,(j=2)}\right], \tag{A4}$$

where

$$G_j = \frac{\partial p_{11j}}{\partial\psi_{12}} = \frac{\partial p_{22j}}{\partial\psi_{12}} = -\frac{\partial p_{12j}}{\partial\psi_{12}} = -\frac{\partial p_{21j}}{\partial\psi_{12}}, \qquad (j = 1, 2).$$

(A.4) vanishes if and only if either $G_1 = G_2$ or $\psi_{12,(j)}$ is modeled separately, contradicting the nature of the Plackett distribution. But, empirical calculations showed that the correlations between estimated marginal and association parameters are usually small, when compared to the other correlations.

[Received September 1992. Revised March 1993.]

REFERENCES

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

Ashby, M., Neuhaus, J. M., Hauck, W. W., Bacchetti, P., Heilbron, D. C., Jewell, N. P., Segal, M. R., and Fusaro, R. E. (1992), "An Annotated Bibliography of Methods for Analyzing Correlated Categorical Data," *Statistics in Medicine,* 11, 67–99.

Ashford, J. R., and Sowden, R. R. (1970), "Multivariate Probit Analysis," *Biometrics,* 26, 535–546.

Becker, M. P., and Balagtas, C. C. (1993), "Marginal Modeling for Binary Cross-Over Data," *Biometrics,* 49, 997–1009.

Chemical Rubber Co. (1972), Standard Mathematical Tables, Cleveland: Author.

Dale, J. R. (1986), "Global Cross-Ratio Models for Bivariate, Discrete, Ordered Responses," *Biometrics,* 42, 909–917.

Edlefsen, L. E., and Jones, S. D. (1992), GAUSS. Kent, WA: Aptech Systems.

Fréchet, M. (1951), "Sur les Tableaux de Corrélation Dont les Marges Sont Données," *Annals Université Lyon,* Sec. A, Ser. 3, 14, 53–77.

Goodman, L. A. (1981), "Association Models and Canonical Correlation in the Analysis of Cross-Classifications Having Ordered Categories," *Journal of the American Statistical Association,* 76, 320–334.

Jones, B., and Kenward, M. G. (1989), *The Analysis of Cross-Over Studies,* London: Chapman and Hall.

Jennrich, R. I., and Schluchter, M. D. (1986), "Unbalanced Repeated-Measures Model With Structured Covariance Matrix," *Biometrics,* 42, 805–820.

Kenward, M. G., and Jones, B. (1991), "The Analysis of Categorical Data From Cross-Over Trials Using a Latent Variable Model," *Statistics in Medicine,* 10, 1607–1619.

Koch, G. G., Landis, J. R., Freeman, J. L., Freeman, D. H., and Lehnen, R. G. (1977), "A General Methodology for the Analysis of Experiments With Repeated Measurements of Categorical Data," *Biometrics,* 33, 133–158.

Lesaffre, E., and Molenberghs, G. (1991), "Multivariate Probit Analysis: A Neglected Procedure in Medical Statistics," *Statistics in Medicine,* 10, 1391–1403.

Lipsitz, S. R., Laird, N. M., and Harrington, D. P. (1991), "Generalized Estimating Equations for Correlated Binary Data: Using the Odds Ratio as a Measure of Association," *Biometrika,* 78, 153–160.

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

Liang, K.-Y., and Zeger, S. L. (1986), "Longitudinal Data Analysis Using Generalized Linear Models," *Biometrika,* 73, 13–22.

Liang, K.-Y., Zeger, S. L., and Qaqish, B. (1992), "Multivariate Regression Analyses for Categorical Data," *Journal of the Royal Statistical Society,* Ser. B, 54, 3–40.

Mardia, K. V. (1970), *Families of Bivariate Distributions,* London: Griffin.

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

Palmgren, J. (1989), "Regression Models for Bivariate Responses," Technical Report 101, School of Public Health and Community Medicine, Dept. of Biostatistics.

Plackett, R. L. (1965), "A Class of Bivariate Distributions," *Journal of the American Statistical Association,* 60, 516–522.

Verloove-Vanhorick, S. P., Verwey, R. A., Brand, R., Bennebroek Gravenhorst, J., Keirse, M. J. N. C., and Ruys, J. H. (1986), "Neonatal Mortality Risk in Relation to Gestational Age and Birthweight," *Lancet,* 1, 55–57.

Ware, J. H. (1985), "Linear Models for the Analysis of Longitudinal Studies," *The American Statistician,* 39, 95–101.