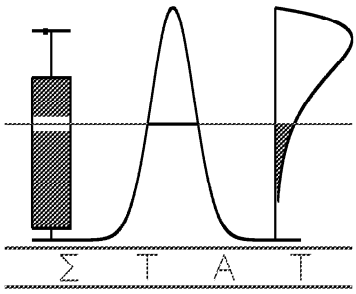
T E C H N I C A L

R E P O R T

0232

Modelling Combined Continuous and Ordinal Outcomes in a Clustered Setting

C. Faes, H. Geys, M. Aerts, G. Molenberghs, and P.J. Catalano



I A P S T A T I S T I C S

N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

Modelling Combined Continuous and Ordinal Outcomes in a Clustered Setting

Christel Faes, Helena Geys, Marc Aerts, Geert Molenberghs
Center for Statistics, Limburgs Universitair Centrum, Diepenbeek, Belgium

and **Paul J. Catalano**
Department of Biostatistical Science, Dana-Farber Cancer Institute and
Department of Biostatistics, Harvard School of Public Health, Boston, MA

Summary

Measurements of both continuous and categorical outcomes appear in many statistical problems. One such example is the study of teratology and developmental toxicity, where both the probability that a live fetus is malformed (ordinal) or of low birth weight (continuous) are important measures in the context of teratogenicity. While multivariate methods of the analysis of continuous outcomes are well understood, methods for jointly continuous and discrete outcomes are less familiar. We propose a likelihood-based method that is an extension of the Plackett-Dale approach. Specification of the full likelihood will be avoided using pseudo-likelihood methodology. The estimation of safe dose levels as part of quantitative risk assessment will be illustrated based on a developmental toxicity experiment of diethylene glycol dimethyl ether in mice.

Keywords: Benchmark Dose; Clustering; Plackett-Dale Model; Pseudo-Likelihood; Quantitative Risk Assessment.

1 Introduction

In this paper, a modeling approach for the analysis of ordinal and continuous outcomes in a clustered data setting has been described. The model is motivated by dose-response modeling of malformation and fetal weight data from developmental toxicology experiments to be used

for quantitative risk assessment. In a typical developmental toxicity study, pregnant animals are exposed during the period of major organogenesis and structural development to a compound of interest. Dose levels for this design typically consist of a control group and three or four exposed groups, each with 20 to 30 pregnant animals. The dams are sacrificed just prior to normal delivery, at which time the uterus is removed and the contents are thoroughly examined for the occurence of defects. Viable fetusus are measured for birth weight and examined carefully for the presence of malformations. Primary goal of these studies is the quantitative risk assessment, i.e., setting a safe dose level of exposure.

The analysis of developmental toxicity data as described above, raises a number of challenges (Aerts et al. 2002). Indeed, these studies often result in multiple outcomes of interest. Developmental toxicity may record both malformation outcomes (ordinal) and birth weight (continuous) on each embryo, as both have been found to be indicative of a toxic effect. In addition, correlation between these outcomes exists (Ryan et al. 1991). Thus, jointly modelling the outcomes might be an appropriate statistical analysis. Further, correlation between the fetuses within litters is very likely to be present because of the genetic similarity and same treatment conditions for offsprings of the same mother. This extra variation must also be taken into account in statistical analysis (Kupper et al. 1986, Chen and Kodell 1989). Since laboratory studies involve considerable amounts of time and money, as well as huge numbers of animals, it is essential that the most appropriate and efficient statistical models are used (Williams and Ryan 1996).

While multivariate methods for the analysis of continuous outcomes are well known (Johnson and Wichern 1992), methods for joint continuous and discrete outcomes are less familiar, especially with clustering. Some attempts have been made towards a joint distribution of binary and continuous responses within a litter. A frequent approach is to apply a conditioning argument that allows the joint distribution to be factorized in a marginal component and a conditional component, where the conditioning can be done on either the binary or continuous response

(Catalano and Ryan 1992, Cox and Wermuth 1992, Cox and Wermuth 1994, Fitzmaurice and Laird 1995, Olkin and Tate 1961). Catalano (1997) extended the idea to model simultaneously ordinal and continuous outcomes. Alternative methods for joint binary and continuous outcomes were proposed by Regan and Catalano (1999a,b) and Geys et al. (2001). The first ones introduce a probit approach, where an underlying continuous variable is assumed for each binary outcome, following a normal distribution. The second approach is based on a Plackett-Dale approach (Plackett 1965), assuming a Plackett latent variable to model bivariate endpoints in which one component is continuous and the other is binary. The bivariate latent variable models are fundamentally different in the way the association between both variables is described. The probit approach uses a correlation coefficient, while the Plackett-Dale approach makes use of an odds ratio. The correlation coefficient of the bivariate normal induces constant local association (Holland and Wang 1987), while the odds ratio is a measure of constant global association (Dale 1986, Lapp, Molenberghs and Lesaffre 1998). However, extensions to joint ordinal and continuous outcomes are in demand.

The Plackett-Dale model has several advantages and additional features compared with the traditionally used factorization models. The main advantage of the Plackett-Dale model is that it lends itself in a natural way to quantitative risk assessment, whereas the factorization models do not provide a clear way to compute a safe dose of exposure. Another advantage of the Plackett distribution is the flexibility with which the marginal densities can be chosen (in this case, a multinomial and a normal distribution are used, both being a common choice for the ordinal and continuous variable, respectively). Furthermore, the odds ratio, being a natural measure of global association (Plackett 1965), is an attractive alternative to the correlation in case of an ordinal variable. Also the generality of the design matrix is an important advantage of this approach, as the assumptions of constant variance and constant association are often not tenable. Indeed, in real data settings, correlation structures are likely to change with exposure

(Kupper, Portier, Hogan and Yamamoto 1986), and ignoring this can lead to bias in the estimates or loss of efficiency (Ryan 2000). The Plackett-Dale model allows to directly model the bivariate intrafetus association, as function of exposure or other covariates of interest, while this is not the case in a factorization model (such as the model of Catalano (1997)). And further, the Plackett-Dale model allows separate dose-response models for malformation and weight outcomes, while taking into account the correlation due to clustering as well as the intrafetus association.

Section 2 introduces a motivating data set. In Section 3 a joint continuous-ordinal model based on a Plackett distribution is proposed. In a first step, the model is described under an independence assumption, and is then extended to deal with the clustering of outcomes within litters. The ultimate goal of developmental toxicity studies is to perform risk assessment, i.e., to set safe limits of human exposure, based on the fitted model (Crump 1984). This will be discussed in Section 4. In Section 5, our method is applied to the developmental toxicity data introduced in Section 2.

2 Developmental Toxicity of DYME in Mice

The developmental toxicity of diethylene glycol dimethyl ether (DYME) in mice has been described by Price et al. (1987). DYME is a component of widely used industrial solvents, used in the manufacture of protective coatings such as lacquers, metal coatings, baking enamels, etc. Although to date, several attempts have proven inadequate to evaluate the potential of glycol ethers to produce human reproductive toxicity, structurally related compounds have been identified as reproductive toxicants in several mammalian species, producing testicular toxicity and embryotoxicity. In this study, DYME was administered by gavage in distilled water to timed-pregnant mice during major organogenesis (gestational days 8 through 15). The doses selected for the study were 0, 62.5, 125, 250, and 500 mg/kg/day with 21, 20, 24, 23 and 23 pregnant dams assigned to each of these dose groups, respectively. For each live fetus, the fetal weight (continuous)

Table 1: Summary Data from a DYME Experiment in Mice

Dose	Dams	Live	Litter Size	Weight	
(mg/kg/day)			Mean	Mean	sd
0.0	21	282	13.4	1.000	0.109
62.5	20	225	11.3	0.967	0.116
125	24	290	12.1	0.910	0.113
250	23	261	11.3	0.793	0.097
500	23	141	6.1	0.562	0.083

Dose	Minor Malformation		Full Malformation		All Malformations	
(mg/kg/day)	Number	%	Number	%	Number	%
0.0	7	2.5	1	0.4	8	2.9
62.5	9	4.0	0	0.0	9	4.0
125	31	10.7	7	2.4	38	13.1
250	27	10.3	59	22.6	86	32.9
500	0	0.0	132	93.6	132	93.6

and malformation status (normal, minor malformation, full malformation) is recorded. Table 1 summarizes the outcomes for this experiment.

The data show clear dose-related trends for both outcomes. There is a pronounced dose-related reduction in fetal weight, with the highest administration of DYME resulting in roughly half of the mean weight in control animals. The malformation data also exhibit trends with dose. The rate of a minor malformation increases at the lower doses, while the rate of full malformations shows strong increase at the highest doses. The rate of all malformations (minor and full) increases monotonically with dose, ranging from 2.9% in the control group to 93.6% in the highest dose group.

3 A Model for Data of a Mixed Nature

In this section we describe a model for simultaneously observed continuous and ordinal endpoints, based on a Plackett-Dale approach. The Plackett-Dale idea has been used by Molenberghs, Geys and Buyse (2001) to assess the validation of surrogate endpoints in randomized experiments with a binary surrogate and a continuous true endpoint or vice versa. Geys et al. (2001) extended this idea to the context of teratology experiments where rodents are clustered within litters.

Consider an experiment involving N clusters, the ith of which contains n_i individual fetuses. Each of the individuals are examined for two outcomes, the degree of malformation (e.g. none, minor, severe) and the fetal weight. Let M_{ik} be the random variable representing the status of malformation ($m = 1, 2, \ldots, c$) of the kth individual in litter i, and W_{ik} the continuous weight outcome. Together with this vector of two responses $Z_{ik} = (W_{ik}, M_{ik})^T$, a vector of covariates X_{ik} is observed.

First, suppose that all littermates are independent. Let us denote the continuous cumulative distribution of the weight outcome as $F_{W_{ik}}$, and the discrete cumulative distribution of the malformation outcome as $F_{M_{ik}}$. We assume a normal distribution for the continuous outcome W_{ik} with mean μ_{ik} and variance σ^2_{ik}, and a multinomial distribution for the ordinal outcome M_{ik} with $\pi_{l,ik}$ the cumulative probability $P(M_{ik} \leq l)$ of observing a malformation of degree smaller or equal to l. The dependence between malformation status and fetal weight can be defined using a global cross-ratio at cutpoint (w, m):

$$\psi_{ik}(w,m) = \frac{F_{W_{ik},M_{ik}}(w,m)\{1 - F_{W_{ik}}(w) - F_{M_{ik}}(m) + F_{W_{ik},M_{ik}}(w,m)\}}{\{F_{W_{ik}}(w) - F_{W_{ik},M_{ik}}(w,m)\}\{F_{M_{ik}}(m) - F_{W_{ik},M_{ik}}(w,m)\}}.$$

Using this relationship, the joint cumulative distribution $F_{W_{ik},M_{ik}}$ can be written as function of the marginal distributions and the global cross-ratio (Plackett 1965):

$$F_{W_{ik},M_{ik}}(w,m) = \begin{cases} \frac{1+\left(F_{W_{ik}}(w)+F_{M_{ik}}(m)\right)\left(\psi_{ik}-1\right)-S\left(F_{W_{ik}}(w),F_{M_{ik}}(m),\psi_{ik}\right)}{2\left(\psi_{ik}-1\right)} & \psi_{ik} \neq 1, \\ F_{W_{ik}}(w)F_{M_{ik}}(m) & \psi_{ik} = 1, \end{cases}$$

with

$$S(F_{W_{ik}}, F_{M_{ik}}, \psi_{ik}) = \sqrt{\big(1 + (\psi_{ik} - 1)(F_{W_{ik}} + F_{M_{ik}})\big)^2 + 4\psi_{ik}(1 - \psi_{ik})F_{W_{ik}}F_{M_{ik}}}.$$

Note that, for every cutpoint (w, m), a global cross-ratio ψ_{ik} is obtained (m=1, ... ,c-1). And thus, for every cutpoint (w, m) a different underlying Plackett distribution is assumed. Assuming a constant odds ratio $\psi_{ik}(w, m) \equiv \psi_{ik}$, there is a single underlying Plackett distribution.

Based upon the cumulative distribution function $F_{W_{ik},M_{ik}}(w, m)$, a bivariate Plackett density function $g_{ik}(w, m)$ for joint continuous-ordinal outcomes is derived. Put $f_m(w) = \partial P(W_{ik} \leq w | M_{ik} = m)/\partial w$ for every $m = 1, \ldots, c$. If we define $g_{ik}(w, m) = f_m(w)P(M_{ik} = m)$, or

$$g_{ik}(w, m) = \begin{cases} \frac{\partial}{\partial w}\big(F_{W_{ik},M_{ik}}(w, 1)\big) & m = 1, \\ \frac{\partial}{\partial w}\big(F_{W_{ik},M_{ik}}(w, m)\big) - \frac{\partial}{\partial w}\big(F_{W_{ik},M_{ik}}(w, m-1)\big) & m = 2, \ldots, c-1, \\ f_{W_{ik}}(w) - \frac{\partial}{\partial w}\big(F_{W_{ik},M_{ik}}(w, c-1)\big) & m = c, \end{cases}$$

then this leads to specifying the density function $g_{ik}(w, m)$ by:

$$g_{ik}(w, m) = \begin{cases} \frac{f_{W_{ik}}(w)}{2}\Big[1 - d(w, m)\Big] & m = 1, \\ \frac{f_{W_{ik}}(w)}{2}\Big[d(w, m-1) - d(w, m)\Big] & m = 2, \ldots, c-1, \\ \frac{f_{W_{ik}}(w)}{2}\Big[1 + d(w, m-1)\Big] & m = c. \end{cases}$$

with

$$d(w, m) = \frac{1 + F_{W_{ik}}(w)(\psi_{ik} - 1) - F_{M_{ik}}(m)(\psi_{ik} + 1)}{S(F_{W_{ik}}(w), F_{M_{ik}}(m), \psi_{ik})}.$$

One can show that the function $g_{ik}(w, m)$ satisfies the classical density properties:

(i) $g_{ik}(w, m) \geq 0$ for all possible values of w and m,

(ii) $\int \sum_{m=1}^{c} g_{ik}(w, m)dw = \int f_{W_{ik}}(w)dw = 1.$

Further, note that $g_{ik}(w,m)$ factorizes as a product of the marginal density $f_{W_{ik}}(w)$ and the conditional density $f_{M_{ik}|W_{ik}}(m|w)$. Some interesting special cases are obtained when the two outcomes are independent ($\psi_{ik}=1$), perfectly negatively associated ($\psi_{ik}=0$) and perfectly positively associated ($\psi_{ik}=\infty$). In case weight and malformation are independent, the function $g_{ik}(w,m)$ reduces to $f_{W_{ik}}(w)f_{M_{ik}}(m)$.

Dose-response models that incorporate litter- and fetus-specific covariates can be considered for each of the parameters by using appropriate link functions. The parameters $\mu_{ik}, \sigma^2_{ik}, \pi_{l,ik}$ and ψ_{ik} for individual k in cluster i can be modelled by:

$$\eta_{ik} \equiv \begin{pmatrix} \mu_{ik} \\ \ln(\sigma^2_{ik}) \\ \text{logit}(\pi_{1,ik}) \\ \vdots \\ \text{logit}(\pi_{c-1,ik}) \\ \ln(\psi_{ik}) \end{pmatrix} = \mathbf{X}_{ik}\beta, \tag{1}$$

where $\mathbf{X}_{ik}$ is a design matrix for the kth fetus in the ith cluster and β is a vector of unknown regression parameters. The generality of the design matrix is an important advantage of this approach. Estimates are obtained by solving the estimating equations $U(\beta)=0$. Grouping all parameters $\mu_{ik}, \sigma^2_{ik}, \pi_{l,ik}$ and ψ_{ik} for individual k in cluster i in a vector θ_{ik}, and grouping all vectors θ_{ik} and η_{ik} for the ith cluster in θ_i and η_i, respectively, the estimating equations can than be written as:

$$\mathbf{U}(\beta) = \sum_{i=1}^{N} U_i(\beta) = \sum_{i=1}^{N}\sum_{k=1}^{n_i} \Big(\frac{\partial \eta_i}{\partial \beta}\Big)^T \Big(\frac{\partial \eta_i}{\partial \theta_i}\Big)^{-T} \Big(\frac{\partial \ln g_{ik}(w,m)}{\partial \theta_i}\Big) = 0.$$

Expressions for the derivatives are deferred to the Appendix.

Often, littermates are not independent, but clustered within litters. In the case of clustering, we can use a pseudo-likelihood function, rather than considering the full likelihood. The pseudo-likelihood approach was proposed by Arnold and Strauss (1991), and also found in Connolly and Liang (1988), Liang and Zeger (1986) and Le Cessie and Van Houwelingen (1994).

The principal idea is to replace a numerically challenging joint density by a simpler function that is a suitable product of ratios of likelihoods of subsets of the variables. As such, we avoid the computational complexity of the full likelihood distribution of each cluster i, i.e., $f(w_{i1}, \ldots, w_{in_i}, m_{i1}, \ldots, m_{in_i})$. In addition, the pseudo-likelihood method provides a way to deal with nuisance parameters (Liang and Zeger 1989, Arnold and Strauss 1991). Arnold and Strauss (1991) established consistency and asymptotic normality of the pseudo-likelihood estimator. Thus, valid inference can be obtained from such models.

In a first step, the association between weight and malformation outcomes for an individual fetus is modeled explicitely, but for outcomes from different littermates independence is taken as a working assumption.

$$pl = \sum_{i=1}^{N} \sum_{k=1}^{n_i} \ln g(w_{ik}, m_{ik}). \tag{2}$$

This approach acknowledges the fact that, while the association between different outcomes on the same littermate is often of scientific interest, the association due to clustering within litters is usually considered a nuisance. Indeed, in quantitative risk assessment primary interest lies in the probability that a fetus is affected, either by malformation or by low birth weight. This probability is a function only of the mean parameters and the bivariate intrafetus association. A sandwich variance estimator is then used to adjust for potential bias in the variance estimator. Arnold and Strauss (1991) showed that under regularity conditions, the pseudo-likelihood estimator $\hat{\beta}$, obtained by maximizing the log-pseudo likelihood function (1) is consistent and asymptotically normal with estimated covariance matrix:

$$\widehat{\mathrm{Cov}}(\hat{\beta}) = \Big(\sum_{i=1}^{N} \frac{\partial U_i}{\partial \beta}\Big)^{-1} \Big(\sum_{i=1}^{N} U_i(\beta) U_i(\beta)^T\Big) \Big(\sum_{i=1}^{N} \frac{\partial U_i}{\partial \beta}\Big)^{-1}\Big|_{\beta=\hat{\beta}}.$$

However, if one is interested in the amount of clustering as well, this pseudo-likelihood function can be extended by including the products of the bivariate probabilities of (i) two weight outcomes for two different individuals in the same cluster, (ii) two malformation outcomes for

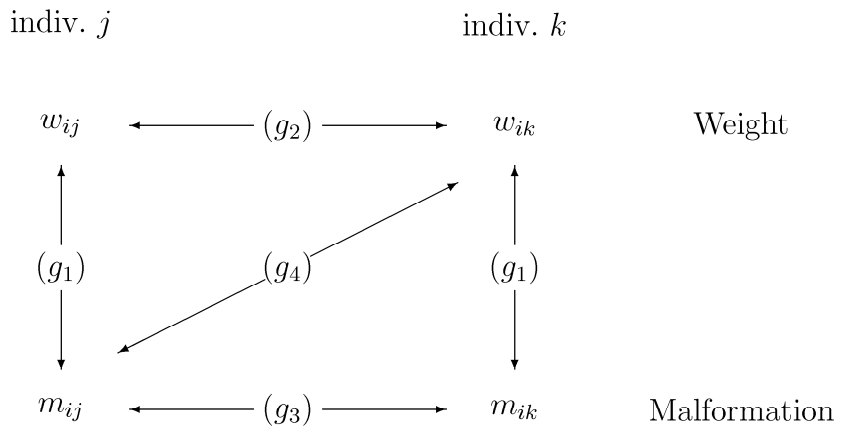


Figure 1: *Four different types of contributions for the pseudo-likelihood.*

two different individuals in the same cluster and (iii) a weight and malformation outcome for two different individuals in the same cluster. This leads to the following log-pseudo likelihood function:

$$\begin{aligned} pl &= \sum_{i=1}^{N}\sum_{k=1}^{n_i} \ln g_1(w_{ik}, m_{ik}) + \sum_{i=1}^{N}\sum_{k<j} \ln g_2(w_{ik}, w_{ij}) \\ &\quad + \sum_{i=1}^{N}\sum_{k<j} \ln g_3(m_{ik}, m_{ij}) + \sum_{i=1}^{N}\sum_{j\neq k}^{n_i} \ln g_4(w_{ik}, m_{ij}), \end{aligned}$$

with g_1, g_2, g_3, g_4 for example bivariate Plackett distributions, characterized by potentially different odds ratios. The four different types of contributions captured in the model are depicted in Figure 1.

The pseudo-likelihood methodology is very general and flexible. It can be found in many applications and fields of interest. It has been most advantageously used in the spatial data context, where the full likelihood distribution is typically cumbersome (Hjort, 1993, Guyon, 1995). But also in this context, maximum likelihood methods are not feasible, due to excessive computational requirements. As a consequence, alternative methods have been in demand and one has

to rely on non-likelihood methods such as generalized estimating equations or pseudo-likelihood methods. An important advantage of the pseudo-likelihood approach is the close connection with likelihood, which enabled Geys, Molenberghs and Ryan (1999) to construct pseudo-likelihood ratio test statistics that have easy-to-compute expressions and intuitively appealing limiting distributions. Geys, Molenberghs and Lipsitz (1998) compared pairwise likelihood with other estimating equations approaches (GEE1 and GEE2) in marginally specified odds ratio models with exchangeable association structure. The efficiency of the pseudo-likelihood estimators for the main effects is comparable to the efficiency of GEE estimators. Yet, pseudo-likelihood allows the estimation of both main effect parameters and association parameters, whereas GEE1 is restricted to main effect parameters. While GEE2 includes second order association parameters as well and is slightly more efficient than both GEE1 and PL, it is computationally much more complex and becomes cumbersome for large cluster sizes. In contrast, pseudo-likelihood can be used with very large clusters.

4 Risk Assessment

In the area of developmental toxicity, an important goal is the quantitative risk assessment, i.e., deriving a safe dose of exposure. Recent techniques for risk assessment are based on fitting dose-response models and estimating the dose corresponding to a certain increase in risk of an adverse effect over background, i.e., benchmark dose. In case of multiple outcomes, the outcomes are often examined individually, using appropriate methods to account for correlation, and regulation of exposure is then based on the most sensitive outcome. This approach assumes that protecting against the most sensitive outcomes protects against all adverse outcomes. It has been found, however, that a clear pattern of correlation exists between weight and malformation outcomes (Ryan et al. 1991), so that risk assessment based on a joint model may be more appropriate. For risk assessment purposes, the joint probability that an individual fetus is malformed and/or

of low fetal weight must be characterized.

The standard approach to quantitative risk assessment based on dose-response modelling requires the specification of an adverse event, along with its risk expressed as a function of dose. The risk function $r(d)$ can be defined as the probability that a fetus has a high malformation level or a low birth weight at dose level d. In other words, for the kth fetus in the ith cluster:

$$r(d) = P(W_{ik} \leq W_c \text{ or } M_{ik} \geq M_c | d),$$

where W_c and M_c, respectively, denote some cutoff values that determines fetal weight low enough and malformation severe enough to be considered adverse. Based on this probability, a common measure for the excess risk over background is given by

$$r^*(d) = \frac{r(d) - r(0)}{1 - r(0)},$$

where greater weights are given to outcomes with larger background risk. The benchmark dose is then defined as the dose corresponding to a small increase in risk over background. More formally, the benchmark dose (BMD_q) is defined as the dose satisfying $r^*(d) = q$, where q corresponds to a pre-specified level of increased response and is typically specified as $0.01, 1, 5$ or 10% (Crump 1984).

Because the dose-response curve is estimated from the data and has inherent variability, the benchmark dose itself is an estimate of the true dose that would result in the corresponding level of excess risk. This sampling uncertainty for the model on which the benchmark dose is based can be acknowledged, by replacing the benchmark dose by a lower confidence limit. Several approaches exist (Williams and Ryan 1996, Kimmel and Gaylor 1988, Crump and Howe 1983). A well known approach is the use of the lower effective dose, where an upper limit for the risk function is used to determine a safe dose level. The lower effective dose (LED_q) is defined as the solution of

$$\hat{r}^*(d) + 1.645\sqrt{\widehat{\text{Var}}\big(\hat{r}^*(d)\big)} = q,$$

where q corresponds to the pre-specified level of increased response, and the variance of the estimated increased risk function $\hat{r}^*(d)$ is estimated as

$$\widehat{\mathsf{Var}}\big(\hat{r}^*(d)\big) = \Big(\frac{\partial r^*(d)}{\partial \boldsymbol{\beta}}\Big)^T \widehat{\mathsf{Cov}}(\hat{\boldsymbol{\beta}}) \Big(\frac{\partial r^*(d)}{\partial \boldsymbol{\beta}}\Big)\Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}},$$

with $\widehat{\mathsf{Cov}}(\hat{\boldsymbol{\beta}})$ the estimated covariance matrix of $\hat{\boldsymbol{\beta}}$.

5 Analysis of DYME Data

In this section we illustrate the methods described through the analysis of developmental toxicity data from a study of diethylene glycol dimethyl ether (DYME) in mice. Scientific interest lies in the effects of dose on the overall risk due to malformation and low birth weight, i.e. the probability that an individual fetus is malformed or of low birth weight.

For risk assessment to be reliable, models should fit the data well. In order to select the most parsimonious model for the DYME data, we rely on the adjusted pseudo-likelihood ratio test statistic, defined by Geys, Molenberghs and Ryan (1999). We start from a complex model with a linear dose trend on the mean weight outcome (μ) and on the logit of the cumulative malformation probabilities (logit(π_1),logit(π_2)). The observed table (Table 1) suggests a decreasing litter size with increasing dose level. In addition, fetal weight and malformation are often affected by litter size. Therefore, to adjust for the litter size, a covariate for the deviation of the overall average litter size $(n_i - n)$ is incorporated into the model:

$$\begin{aligned}
\mu_i &= \alpha_0 + \alpha_1 d_i + \alpha_2 (n_i - \bar{n}), \\
\mathsf{logit}(\pi_1) &= \beta_0 + \beta_1 d_i + \beta_2 (n_i - \bar{n}), \\
\mathsf{logit}(\pi_2) &= \gamma_0 + \gamma_1 d_i + \gamma_2 (n_i - \bar{n}).
\end{aligned}$$

Note that this model is more flexible than the standard proportional odds model. However, a careful modelling approach has to be considered with constraints on the parameters in order to

Table 2: Model Selection for DYME Study in Mice.

Model	Description	# pars.	pl
1.	Different weight variances across doses;	17	-553.48
	Common d trend on ψ; Different ψ depending on cutpoint m		
2.	Linear d trend on σ^2;	14	-551.09
	Common d trend on ψ; Different ψ depending on cutpoint m		
3.	Constant σ^2;	13	-540.73
	Common d trend on ψ; Different ψ depending on cutpoint m		
4.	Linear d trend on σ^2; Different ψ depending on cutpoint m	13	-549.99
5.	Linear d trend on σ^2; Constant ψ	12	-546.53

Comparison	df	$\bar{\lambda}$	G_a^{2*}	(p-value)
1-2	4	6.46	0.74	(0.946)
2-3	1	6.64	3.12	(0.077)
2-4	1	0.74	0.73	(0.393)
4-5	1	2.97	9.16	(0.002)

ensure that $\pi_1 \leq \pi_2$. Different models on the fetal weight variance ($\ln(\sigma^2)$) and on the log odds ratio ($\ln(\psi)$) between weight and malformation are considered. Table 2 shows the different fitted models, and a summary of model selection.

The weight variances were initially fit separately, but a more parsimonious model with a linear dose trend on the variances (with ln link function) is accepted (Model 2). The dose effect parameter on the variances is only borderline significant, but will be kept in the model. The linear dose trend on the log odds ratio can be deleted without substantial decrease in fit. However, use of different log odds ratios, depending on the malformation cutpoint are significant ($\ln(\psi) = \zeta_0 + \zeta_1 I(m = 2)$). Hence, the model can be reduced to Model 4. Other parameters

cannot be removed without a substantial decrease in fit. Therefore, we accept Model 4 as final model on fetal weight variance and odds ratio.

The results of fitting the clustered bivariate model to the DYME data, using the final Model 4, are displayed in Table 3, in the column labeled $Plackett\text{-}Dale$. The table tabulates the parameter estimates, standard errors and Z scores for the average weight (μ) and cumulative malformation probabilities (π) and the weight variance (σ^2) and odds ratio (ψ). For fetal weight, the dose coefficient is significantly negative, but there appears to be little effect of litter size on weight. The fetal weight variance decreases with dose. Because of the working independence assumption, there is no estimated intralitter correlation for fetal weight. The dose coefficient for the cumulative malformation probabilities is significantly negative, and the significantly positive coefficient of litter size suggests that larger litters have a smaller malformation risk. Again, because of the working independence assumption, there is no estimated intralitter correlation for malformation. The estimated odds ratios are less than 1 because of the negative association between weight and malformation; the small value indicates the strength of the association.

Observed and predicted values in Figure 2 (a)-(b) show how the model fits the dose-specific averages for both univariate outcomes. The predicted weight outcome fits the data quite well. The lower points and curve in the malformation portion of the figure denote, respectively, the observed and predicted probability of malformation and the upper points and curve correspond to the probability of minor or full malformation. Notice that in the malformation status plot the highest dose has only one observed value since all events were classified as full malformations. Estimates for both malformation outcomes are in strong agreement with the observed probabilities.

Results can be compared with these of Catalano (1997), which are displayed in Table 3, in the column labeled $Cond\text{-}C$. Catalano (1997) used a factorization model that conditions on the continuous outcome. Three main differences are noted. First, in this model, different

Figure 2: *DYME Study: Estimated and Observed Outcomes. From Top to Bottom, (a) Fetal Weight Model, (b) Malformation Status Model, (c) Joint Weight-Malformation Risk Model*

intercepts and common slope parameters for the malformation model are assumed. Different slopes on the probit scale would complicate the model. Secondly, the regression parameters of the malformation model do not have a direct marginal interpretation because of the non-linearity of the link function relating the conditional mean of the ordinal response to the covariates. The Plackett-Dale model allows separate dose-response models for malformation and weight outcomes, while taking into account the correlation due to clustering as well as the intrafetus association. Finally, the model induces association by adding the residuals from the marginal model as covariates in the conditional model, however these association parameters are not directly interpretable. Thus, while taking into account the dependence between weight and malformation, the conditional models do not directly specify a measure of association. To overcome this problem, one needs

a joint model that incorporates the correlation between outcomes directly, such as the proposed Plackett-Dale model. The Plackett-Dale model uses the odds ratio as intrafetus measure of association, which is readily interpretable.

In order to calculate a benchmark dose based on the joint model, we first need to specify the risk of an adverse effect, i.e. the probability that an individual fetus is malformed or of low birth weight. Therefore, we need to define a weight below which a fetus can be considered as being of "low fetal weight". However, there is no standard definition. For simplicity, we specify the cutoff point W_c as two standard deviations below the control average fetal weight. As such, the cutoff level for determining low fetal weight is equal to $W_c = 0.7816$, corresponding to a $1,77\%$ low birth weight rate in control animals. Further, we consider two definitions of risk, depending on the cutpoint M_c for what is considered as a "malformed" fetus. Either we define it as the probability that a fetus has a minor or full malformation ($M_c = 2$), or a low fetal weight. Alternatively, we define it as the probability that a fetus has a full malformation ($M_c = 3$), or has a low fetal weight. These two risk functions are displayed in Figure 2 (c) together with the data. The risks are evaluated at the average litter size ($n_i = \bar{n}$). The lower points and curve in this graph correspond to the risk of a fetus with low fetal weight and/or minor or full malformation. The higher points and curve correspond to the second definition of a risk. In case of the first definition, the estimated risk function is higher and steeper compared with risk when defined under the second definition. But, for both definitions of "risk", the risk functions seem to fit the data very well.

Table 4 shows the benchmark doses corresponding to the 1% and 10% excess risk for Model 4, as well as the 1% and 10% lower limit LED. We also added the corresponding quantities, calculated from univariate versions of the model. We can compare the joint modelling approach with the traditional approach for multiple outcomes in which the lower of the individual malformation and fetal weight LEDs is used as an overall LED. The miminum of the two LEDs

is more than 20% higher than those obtained using the bivariate methods that incorporate the relationship between the two outcomes. Since both univariate outcomes suffer from a substantial risk, focusing attention to a single response or a collapsed outcome would overestimate the safe dose. The joint model yields higher risks, since it accounts for the correlation between both outcomes. Thus, ignoring the correlation between the two outcomes leads to too high and hence inappropriate safe doses. Note that joint outcome models based on factorization, such as the model of Catalano (1997), do not provide a clear way to compute joint benchmark doses (Geys et al. 2001, Regan and Catalano 1999). This is the major drawback of the conditional models, since the joint benchmark doses are of primary interest.

6 Discussion

In this paper, a modelling approach for the analysis of clustered data with both continuous and ordinal outcomes has been considered. The model was applied to a developmental toxicity study (DYME in mice), and used for quantitative risk assessment. The Plackett-Dale method uses a global odds ratio as intrafetus measure of association. The association between fetal weight and malformation are directly modelled, and can be modelled in a general way, including covariate information.

A problem in the quantitative risk assessment, that has received only minor attention, is the choice of suitable cutpoints. For binary outcomes, the definition of an adverse health effect is intuitively clear. This is less the case for continuous and/or ordinal responses. We considered a dichotomized version of the continuous outcomes to determine a benchmark dose, after fitting the dose-response model based on the continuous outcome. The risk of low fetal weight was based on a cutoff level for determining a low weight extreme enough to be considered an adverse event. Because of arbitrariness of the cutpoint, estimating a BMD from a continuous response has led to much discussion (Bosch et al. 1996, Crump 1984, 1995, Gaylor and Slikker 1990,

Kavlock et al. 1995, Kodell and West 1993). Several efforts have been made to develop risk assessment for continuous outcomes (West and Kodell 1993). However, the used definition of risk, defined in terms of the tail of the background (control) distribution, seems quite plausible.

In this paper, the response rates are allowed to depend on litter size and a safe dose is then calculated at an "average" litter size. As observed with the DYME data, the number of viable fetuses in a dam, i.e., the litter size, decreases with increasing dose levels. Thus, a method that in addition acknowledges the stochastic nature of the litter size would be appropriate. This is the subject of further research.

Although the method is presented in the specialized field of developmental toxicity, the methodology is applicable in a general clustered or even correlated data setting with a continuous and ordinal outcome. Thus, use of the proposed modelling approach is far beyond the developmental toxicity context.

Acknowledgment

The first two authors gratefully acknowledge support from the Institute for the Promotion of Innovation by Science and Technology (IWT) in Flanders, Belgium. The second author also acknowledges support from "Fonds voor Wetenschappelijk Onderzoek" (FWO), Flanders, Belgium. In addition, the authors acknowledge the Interuniversity Poles of Attraction (IUAP).

References

Aerts, M., Geys, H., Molenberghs, G., and Ryan, L. (2002). *Topics in Modelling of Clustered Data.*, London: CRC Press.

Arnold, B.C. and Strauss, D. (1991). Pseudolikelihood estimation : Some examples. *Sankhya B* **53**, 233–243.

Bosch, R.J., Wypij, D., and Ryan, L.M. (1996). A semiparametric approach to risk assessment

for quantitative outcomes. *Risk Analysis* **16**, 657–665.

Catalano, P.J. (1997). Bivariate modelling of clustered continuous and ordered categorical outcomes. *Statistics in Medicine* **16**, 883–900.

Catalano, P.J. and Ryan, L.M. (1992). Bivariate latent variable models for clustered discrete and continuous outcomes. *Journal of the American Statistical Association* **87**, 651–658.

Chen, J.J and Kodell, R.L. (1989). Quantitative risk assessment for teratologic effects. *Journal of the American Statistical Association* **84**, 966–971.

Connolly, M.A. and Liang, K.Y. (1988). Conditional logistic regression models for correlated binary data. *Biometrika* **75**, 501–506.

Cox, D.R. and Wermuth, N. (1992). Response models for mixed binary and quantitative variables. *Biometrika* **79**, 441–461.

Cox, D.R. and Wermuth, N. (1994). A note on the quadratic exponential binary distribution. *Biometrika* **81**, 403–408.

Crump, K. (1984). A new method for determining allowable daily intakes. *Fundamental and Applied Toxicology* **4**, 854–871.

Crump, K. (1985). Calculation of benchmark dosis from continuous data. *Risk Analysis* **15**, 79–89.

Crump, K.S. and Howe, R.B. (1983). A review of methods for calculating statistical confidence limits in low dose extrapolation, in Clayson, D.B., Krewski, D. and Munro, I. (eds), *Toxicological Risk Assessment. Volume I: Biological and Statistical Criteria*, Boca Raton: CRC Press, pp. 187–203.

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

Fitzmaurice, G.M. and Laird, N.M. (1995). Regression models for a bivariate discrete and continuous outcome with clustering. *Journal of the American Statistical Association* **90**, 845–852.

Food and Drug Administration (1966). Guidelines for developmental toxicity and risk assessment. *Fed. Regist.* **56**, 63798.

Gaylor, D.W. and Slikker, W. (1990). Risk assessment for neurotoxic effects. *NeuroToxicology* **11**, 211–218.

Geys, H., Molenberghs, G., Ryan, L. (1999). Pseudo-likelihood modelling of multivariate outcomes in developmental toxicology. *Journal of the American Statistcal Association* **94**, 734–745.

Geys, H., Regan, M., Catalano, P., and Molenberghs, G. (2001). Two latent variable risk assessment approaches for mixed continuous and discrete outcomes from developmental toxicity data. *Journal of Agricultural Biological and Environment Statistics* **6**, 340-355.

Holland, P.W. and Wang, Y.J. (1987). Dependence function for continuous bivariate densities. *Communications in Statistics - Theory and Methods* **2**, 863-876.

Johnson, R.A. and Wichern, D.W. (1992). *Applied Multivariate Statistical Analysis*, 3rd Edn. New Jersey: Prentice Hall.

Kavlock, R.J., Allen, B.C., Faustman, E.M. and Kimmel, C.A. (1995). Dose-response assessment for developmental toxicity. IV. Benchmark dose for fetal weight changes. *Fundamental and Applied Toxicology* **26**, 211–222.

Kimmel, C. (1990). Quantitative approaches to human risk assessment for noncancer health effects. *Neurotoxicology* **11**, 201.

Kimmel, C.A. and Gaylor, D.W. (1988). Issues in qualitative and quantitative risk analysis for developmental toxicology. *Risk Analysis* **8**, 15–20.

Kodell, R.L. and West, R.W. (1993). Upper confidence limits on excess risk for quantitative responses. *Risk Analysis* **13**, 177–182.

Kupper, L.L., Portier, C., Hogan, M.D. and Yamamoto, E. (1986). The impact of litter effects on dose-response modeling in teratology. *Biometrics* **42**, 84–98.

Lapp, K., Molenberghs, G. and Lesaffre, E. (1998). Local and global cross ratios to model the association between ordinal variables. *Computational Statistics and Data Analysis* **28**, 387–412.

Le Cessie, S. and Van Houwelingen, J.C. (1994). Logistic regression for correlated binary data. *Applied Statistics* **43**, 95–108.

Liang, K.Y. and Zeger, S. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.

Molenberghs, G., Geys, H. and Buyse, M. (2001). Evaluations of surrogate enpoints in randomized experiments with mixed discrete and continuous outcomes. *Statistics in Medicine*, **20**, 3023–3038.

Olkin, I. and Tate, R.F. (1961). Multivariate correlation models with mixed discrete and continuous variables. *Annals of Mathematical Statistics* **32**, 448–465 (with correction in **36**, 343–344).

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

Price, C.J., Kimmel, C.A., George, J.D. and Marr, M.C. (1987). The developmental toxicity of diethylene glycol dimethyl ether in mice. *Fundamental and Applied Toxicology* **8**, 1033–1048.

Regan, M.M. and Catalano, P.J. (1999a). Likelihood models for clustered binary and continuous outcomes: Application to developmental toxicology. *Biometrics* **55**, 760–768.

Regan, M.M. and Catalano, P.J. (1999b). Bivariate dose-response modeling and risk estimation in developmental toxicology. *Journal of Agricultural, Biological and Environmental Statistics* **4**, 217–237.

Rodricks, J., Tardiff, R., Brett, S., Putzrath, R. and St. Hilaire, C. (1986). Elements of Toxicology and Chemical Risk Assessment. *Environ Corporation*, Washington, D.C.

Ryan, L.M., Catalano, P.J., Kimmel, C.A., and Kimmel, G.L. (1991). Relationship between fetal weight and malformation in developmental toxicity studies, *Teratology* **44**, 215–223.

Schwetz, B. and Harris, M. (1993). Developmental toxicology: Status of the field and contribution of the National Toxicology Program. *Environmental Health Perspective* **100**, 269.

West, R.W. and Kodell, R.L. (1993). Statistical methods of risk assessment for continuous variables. *Communications in Statistics - Theory and Methods* **22**, 3363–3376.

Williams, P.L., Molenberghs, G. and Lipsitz, S.R. (1996). Analysis of multiple ordinal outcomes in developmental toxicity studies. *Journal of Agricultural, Biological, and Developmental Toxicity Studies* **1**, 250–274.

Williams, P.L. and Ryan, L.M. (1996). Dose-response models for developmental toxicology, in R.D. Hood (ed.). *Handbook of Developmental Toxicology*, New York: CRC Press, 635–666.

Appendix

For simplicity, we have omitted the cluster-level index i and the fetus-level index k:

Some first derivatives

$$
\begin{aligned}
\frac{\partial f_W(w)}{\partial \mu} &= f_W(w)\frac{(w-\mu)}{\sigma^2} \\
\frac{\partial f_W(w)}{\partial \sigma^2} &= \frac{f_W(w)}{2}\left(\frac{(w-\mu)^2-\sigma^2}{\sigma^4}\right) \\
\frac{\partial F_W(w)}{\partial \mu} &= \int_{-\inf}^{x} f_W(w)\frac{(w-\mu)}{\sigma^2}dw \\
&= -f_W(w) \\
\frac{\partial F_W(w)}{\partial \sigma^2} &= \int_{-\inf}^{x} \frac{f_W(w)}{2}\left(\frac{(w-\mu)^2-\sigma^2}{\sigma^4}\right)dw \\
&= -\left(\frac{(w-\mu)}{\sigma^2}\right)\frac{f_W(w)}{2} \\
\frac{\partial F_M(m)}{\partial p_k} &= \begin{cases} 1 & \text{if } k \le m \\ 0 & \text{if } k > m \end{cases}
\end{aligned}
$$

Derivatives of $S(F_W, F_M, \psi)$

$$
\begin{aligned}
S &= \left[(1+(\psi-1)(F_W+F_M))^2 + 4\psi(1-\psi)F_W F_M\right]^{1/2} \\
\frac{\partial S}{\partial \mu} &= \frac{1}{2S}\Big[2[1+(\psi-1)(F_W+F_M)](\psi-1)\frac{\partial F_W}{\partial \mu} + 4\psi(1-\psi)\frac{\partial F_W}{\partial \mu}F_M\Big] \\
\frac{\partial S}{\partial \sigma^2} &= \frac{1}{2S}\Big[2[1+(\psi-1)(F_W+F_M)](\psi-1)\frac{\partial F_W}{\partial \sigma^2} + 4\psi(1-\psi)\frac{\partial F_W}{\partial \sigma^2}F_M\Big] \\
\frac{\partial S}{\partial p_k} &= \frac{1}{2S}\Big[2[1+(\psi-1)(F_W+F_M)](\psi-1)\frac{\partial F_M}{\partial p_k} + 4\psi(1-\psi)F_W\frac{\partial F_M}{\partial p_k}\Big] \\
\frac{\partial S}{\partial \psi} &= \frac{1}{2S}\Big[2[1+(\psi-1)(F_W+F_M)](F_W+F_M) + 4(1-\psi)F_W F_M - 4\psi F_W F_M\Big]
\end{aligned}
$$

Derivatives of $d(w, m)$

$$
\begin{aligned}
d(w,m) &= \frac{f_W(w)}{2}\left\{1 - \frac{1+F_W(w)(\psi-1)-F_M(m)(\psi+1)}{S(F_W(w),F_M(m),\psi)}\right\} \\
\frac{\partial d(w,m)}{\partial \mu} &= \frac{1}{2}\frac{\partial f_W(w)}{\partial \mu}\left\{1 - \frac{1+F_W(w)(\psi-1)-F_M(m)(\psi+1)}{S(F_W(w),F_M(m),\psi)}\right\} - \frac{f_W(w)}{2S}\left\{\frac{\partial F_W(w)}{\partial \mu}(\psi-1)\right\} \\
&\quad + \left\{1 + F_W(w)(\psi-1) - F_M(m)(\psi+1)\right\}\frac{f_W(w)}{2S^2}\frac{\partial S}{\partial \mu} \\
\frac{\partial d(w,m)}{\partial \sigma^2} &= \frac{1}{2}\frac{\partial f_W(w)}{\partial \sigma^2}\left\{1 - \frac{1+F_W(w)(\psi-1)-F_M(m)(\psi+1)}{S(F_W(w),F_M(m),\psi)}\right\} - \frac{f_W(w)}{2S}\left\{\frac{\partial F_W(w)}{\partial \sigma^2}(\psi-1)\right\} \\
&\quad + \left\{1 + F_W(w)(\psi-1) - F_M(m)(\psi+1)\right\}\frac{f_W(w)}{2S^2}\frac{\partial S}{\partial \sigma^2} \\
\frac{\partial d(w,m)}{\partial p_k} &= \frac{f_W(w)}{2S}\frac{\partial f_M(m)}{\partial p_k}(\psi+1) + \frac{f_W(w)}{2S^2}\left\{1 + F_W(w)(\psi-1) - F_M(m)(\psi+1)\right\}\frac{\partial S}{\partial p_k} \\
\frac{\partial d(w,m)}{\partial \psi} &= -\frac{f_W(w)}{2S}\left\{F_W(w) - F_M(m)\right\} + \frac{f_W(w)}{2S^2}\left\{1 + F_W(w)(\psi-1) - F_M(m)(\psi+1)\right\}\frac{\partial S}{\partial \psi}
\end{aligned}
$$

Table 3: Model Fits for DYME Study in Mice.

		Plackett-Dale		Cond-C	
	Coefficient	Estimate	(s.e.; Z-score)	Estimate	(s.e.; Z-score)
Weight Model:					
	Intercept	1.014	(0.014; 72.4)	1.024	(0.019;55.3)
	Dose	-0.444	(0.028;-15.9)	-0.472	(0.035;-13.6)
	Litter Size	0.0002	(0.003; 0.1)	-0.002	(0.004;-0.6)
Malformation Status Model:					
π_1:	Intercept	3.462	(0.357; 9.7)	2.30	(0.207;11.12)
	Dose	-5.677	(0.987;-5.8)	-4.05	(0.452;-4.47)
	Litter Size	0.106	(0.060; 1.8)	0.076	(0.039;1.91)
π_2:	Intercept	5.021	(0.323; 15.5)	2.78	(0.215;12.90)
	Dose	-7.236	(0.623;-11.6)	-4.05	(0.452;-4.47)
	Litter Size	0.106	(0.100; 1.1)	0.076	(0.039;1.91)
Average weight residual		-	-	3.93	(1.200;3.27)
Weight residual		-	-	2.12	(1.091;1.94)
Average weight residual $\times$ litter size		-	-	0.068	(0.390;0.17)
Fetal Weight Variance:					
Dose	0.000	0.014	(0.002; 6.1)	-	-
	0.125	0.013	(0.002; 7.6)	-	-
	0.250	0.012	(0.001; 9.4)	-	-
	0.500	0.010	(0.001; 9.4)	-	-
	1.000	0.007	(0.002; 4.1)	-	-
Fetal Weight / Malformation Association:					

Table 4: Risk Assessment for DYME Study in Mice.

		$M_C = 2$		$M_C = 3$	
q	Model	BMD_q	LED_q	BMD_q	LED_q
1%	Joint	13.91	12.66	20.05	17.72
	Continu	23.70	23.37	23.70	23.37
	Ordinal	26.26	25.52	66.90	60.26
10%	Joint	88.64	83.88	108.13	101.85
	Continu	118.99	118.26	118.99	118.26
	Ordinal	136.00	133.71	199.59	193.29