Modelling Combined Continuous and Ordinal Outcomes from Developmental Toxicity Studies.

C. Faes [1], H. Geys [1], M. Aerts [1], P. Catalano [2] and G. Molenberghs [1]

[1] Center for Statistics, Limburgs Universitair Centrum, Universitaire Campus, B-3590 Diepenbeek, Belgium
[2] Department of Biostatistics, Harvard School of Public Health and Department of Biostatistical Science, Dana-Farber Cancer Institute, 44 Binney Street, Boston, Massachusetts 02115, U.S.A.

Abstract: 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 quantitative risk assessment. 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 model 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 from a developmental toxicity experiment of diethylene glycol dimetyhl ether in mice.

Keywords: Benchmark Dose, Clustered Data, Plackett-Dale Model, Pseudo-Likelihood, Odds Ratio, Quantitative Risk Assessment.

1 Introduction

Over the last few years society has become increasingly concerned about public health problems. Especially the potential risk of chemical compounds and other environmental agents on fertility, pregnancy, birth defects and developmental abnormalities are of major concern. Regulatory agencies, such as the U.S. Environmental Protection Agency (EPA) and the Food and Drug Administration (FDA) therefore stimulate reproductive and developmental toxicity research. Because of ethical reasons, reliable epidemiological information of adverse effects on

fetal development may often be limited or unavailable. As an alternative, laboratory experiments in small mammalian species can be conducted in advance of human exposure (Williams and Ryan 1996).

Standard experimental protocols for conducting developmental toxicity studies were established by the U.S. FDA. In developmental toxicity studies with a Segment II design, 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. The viable fetuses are measured for birth weight and examined carefully for the presence of malformation.

The analysis of developmental toxicity data raises a number of challenges (Aerts et al. 2002). Indeed, such studies may record both continuous (low birth weight) and ordinal (malformation indicator) outcomes on each embryo. Correlation between these outcomes exists and also correlation between the fetuses within litters is very likely to be present. Since laboratory studies further 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 mixed continuous and discrete outcomes are less familiar. Some attempts have been made towards the joint analysis of a binary and a continuous outcome. A frequent approach is to decompose the joint distribution into a marginal and 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). However, these factorization models may be difficult to apply for quantitative risk assessment (Geys et al. 1999, Regan and Catalano 1999). The model, introduced here, is based on the Plackett-Dale method (Plackett 1965), assuming a Plackett latent variable to model bivariate endpoints in which one component is continuous and the other is discrete. This method can easily be used for estimation of a safe dose of exposure.

In Section 2 we introduce the proposed model, and show how we can deal with the clustering. In Section 3 we apply the model for quantitative risk assessment. Section 4 summarizes some results of an example of developmental toxicity in mice.

2 Model for Data of a Mixed Nature

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 degree 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 degree 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). For simplicity we have omitted the cluster-level index i and the fetus-level index k.

$$F_{W,M}(w,m) = \begin{cases} \frac{1+\left(F_W(w)+F_M(m)\right)\left(\psi-1\right)-S\left(F_W(w),F_M(m),\psi\right)}{2\left(\psi-1\right)} & \text{if } \psi \neq 1, \\ F_W(w)F_M(m) & \text{if } \psi = 1, \end{cases}$$

with

$$S(F_W, F_M, \psi) = \sqrt{\left(1 + (\psi - 1)(F_W + F_M)\right)^2 + 4\psi(1-\psi)F_W F_M}.$$

Based upon this cumulative distribution function $F_{W_{ik},M_{ik}}(w, m)$, we can derive a bivariate Plackett density function $g_{ik}(w, m)$ for mixed continuous-ordinal outcomes.

Dose-response models that incorporate litter- and fetus-specific covariates can be considered for each of the parameters by using appropriate link functions.

Often however, littermates are not independent, but clustered within litters. In the case of clustering, rather than considering the full likelihood contribution for each cluster i, i.e., $f(w_{i1}, \ldots, w_{in_i}, m_{i1}, \ldots, m_{in_i})$, we avoid the computational complexity by replacing the full likelihood by a pseudo-likelihood function that is easier to evaluate. The contribution of the ith litter to the log pseudo-likelihood function is defined as:

$$pl_i = \sum_{k=1}^{n_i} \ln g_{ik}(w,m).$$

With this approach, the correlation 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. Arnold and Strauss (1991) established consistency and asymptotic normality of the pseudo-likelihood

estimators. Another advantage is the close connection of the pseudo-likelihood 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. If the amount of clustering is of interest as well, then above pseudo-likelihood function can be extended to incorporate this as well.

3 Risk Assessment

A primary goal of quantitative risk assessment is to determine a safe level of exposure. Recent techniques for risk assessment in this area 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 the clustering of fetuses within litters, and regulation of exposure is based on the most sensitive outcome. It has been found, however, that a clear pattern of correlation exists between these outcomes (Ryan et al. 1991), so that risk assessment based on a joint model may be more appropriate. The model must both incorporate the correlation between the two outcomes, as well as the correlation due to clustering. For risk assessment purposes, the joint probability that an individual fetus is malformed and/or of low fetal weight must be characterized.

The risk function $r(d)$ representing the probability that the kth fetus in the ith cluster has a high malformation level of a low birth weight at dose level d can be written as:

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

where W_c and M_c respectively denote some cutoff value that determines fetal weight low enough and malformation severe enough to be considered adverse. This expression can be rewritten using the univariate discrete distribution function $F_{W_{ik}}$ and the joint continuous-discrete distribution function $F_{W_{ik},M_{ik}}$:

$$r(d) = 1 - F_{M_{ik}}(M_c - 1) + F_{W_{ik},M_{ik}}(W_c, M_c - 1).$$

Based on this probability, a measure for the excess risk $r^*(d)$ over background can be specified, and a safe dose of exposure can be defined. The benchmark dose (BMD) is the dose d satisfying $r^*(d) = q$, where q is a pre-specified level of increased response (typically specified as 1, 5 or 10%). To allow for estimation variability, a 95% lower confidence limit of the BMD such as the lower effective dose (LED) can be used for determining an acceptable low-risk exposure level (Crump 1984, Kimmel and Gaylor 1988).

4 Developmental Toxicity of DYME in Mice

Price et al. (1987) describe a study in which diethylene glycol dimethyl ether (DYME) was administered to timed-pregnant mice during major organogenesis. DYME is a component of widely used industrial solvents, used in the manufacture of protective coatings such as lacquers, metal coatings, baking enamels, etc. 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. Table 1 summarizes the data. Scientific interest lies in 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.

TABLE 1. Summary Data from a DYME Experiment in Mice

Dose	Dams	Live	Litter Size (mean)	Weight (mean)	Malformations %		
					none	minor	severe
0.0	21	282	13.4	0.9998	97.2	2.5	0.4
62.5	20	225	11.3	0.9673	96.0	4.0	0.0
125	24	290	12.1	0.9102	86.9	10.7	2.4
250	23	261	11.3	0.7928	67.0	10.3	22.6
500	23	141	6.1	0.5617	6.4	0.0	93.6

Risk assessment is based on dose-response modelling. Selection of a parsimonious model relied on the pseudo-likelihood ratio test statistic. The final dose-response model that we considered here was characterized by

$$\begin{aligned} \mu_{ik} &= \alpha_0 + \alpha_1 d_i \\ \text{logit}(\pi_{1,ik}) &= \beta_0 + \beta_1 d_i \\ \text{logit}(\pi_{2,ik}) &= \gamma_0 + \gamma_1 d_i \\ \ln(\sigma^2_{ik}) &= \varsigma_0 \\ \ln(\psi_{ik}) &= \eta_0 + \eta_1 d_i \end{aligned}$$

with d_i the dose. For this model, the fitted values were close to the observed ones. There is a significantly negative dose-effect for fetal weight. The dose coefficients for the malformation probability are also significant. The estimated odds ratio is less than 1, indicating the negative association between weight and malformation.

In order to calculate a benchmark dose based on this 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". Because of the arbitrariness of the cutpoint, estimating a benchmark dose from a continuous response has led to much discussion (Bosch et al. 1996, Crump 1984). We specify the cutoff point W_c as two standard errors below the control average fetal

weight. By means of this definition, fetuses that weighted less than 0.7816g are considered to be of low fetal weight, which corresponds to a 1,8% rate in the 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 severe malformation ($M_c = 2$), or a low fetal weight. Alternatively, we define it as the probability that a fetus has a severe malformation ($M_c = 3$), or has a low fetal weight.

Table 2 shows for both types of definitions the estimated benchmark doses (in mg/kg/day) corresponding to the 1% and 10% excess risk, as well as the corresponding 95% lower limit LED (lower effective dose). We also added the corresponding quantities, calculated from univariate versions of the model.

TABLE 2. Risk Assessment for DYME Study in Mice.

		$M_c = 2$		$M_c = 3$	
q	Model	$\widehat{BMD}_q$	$\widehat{LED}_q$	$\widehat{BMD}_q$	$\widehat{LED}_q$
1%	Joint	14.31	12.71	20.79	17.75
	Continuous	25.09	24.47	25.09	24.47
	Ordinal	26.54	25.67	67.13	60.17
10%	Joint	86.56	80.81	104.07	96.22
	Continuous	115.65	114.73	115.65	114.73
	Ordinal	135.47	132.82	198.16	191.60

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 which incorporate the relationship between the two outcomes. Thus, ignoring the correlation between the two outcomes leads to too high and hence unscientific safe doses.

Acknowledgments: We gratefully acknowledge support from the Institute for the Promotion of Innovation by Science and Technology (IWT) in Flanders, Belgium.

References

Aerts, M., Geys, H., Molenberghs, G. and Ryan, L.M. (2002). *Topics in Modelling of Clustered Data*, Chapmann and Hall.

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. and Ryan, L.M. (1992). Bivariate Latent Variable Models for Clustered Discrete and Continuous Outcomes. *Journal of the American Statistical Association*, **87**, 651–971.

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.

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.

Geys, H., Molenberghs, G. and Ryan, L. (1999). Pseudo-likelihood modelling of multivariate outcomes in developmental toxicity. *Journal of American Statistical 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.

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

Kimmel, C.A. and Gaylor, D.W. (1988). issues in Qualitative and Quantitative Risk Analysis for Developmental Toxicology, *Risk Analysis*, **8**, 15–20.

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. (1999). Likelihood Models for Clustered Binary and Continuous Outcomes: Application to Developmental Toxicology. *Biometrics*, **55**, 760–768.

Ryan, L.M., Catalano, P.J., Kimmel, C.A., Kimmel, G.L. (1991). Relationship between Fetal Weight and Malformation in Developmental Toxicity Studies, *Teratology*, **44**, 215–223.

Williams, P.L., Ryan, L.M. (1996). Dose-Response Models for Developmental Toxicology, in R.D. Hood (ed.), *Handbook of Developmental Toxicology*, New York: CRC Press, pp. 635–666.