A Novel Application of a Bivariate Regression Model for Binary and Continuous Outcomes to Studies of Fetal Toxicity

Julie S. Najita, **Yi Li**, and **Paul J. Catalano**†
Department of Biostatistics, Harvard School of Public Health and Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, USA

Summary

Public health concerns over the occurrence of birth defects and developmental abnormalities that may occur as a result of prenatal exposure to drugs, chemicals, and other environmental factors has led to an increasing number of developmental toxicity studies. Because fetal pups are commonly evaluated for multiple outcomes, data analysis frequently involves a joint modeling approach. In this paper, we focus on modelling clustered binary and continuous outcomes in the setting where both outcomes are potentially observable in all offspring but, due to practical limitations, the continuous outcome is only observed in a subset of offspring. The subset is not a simple random sample (SRS) but is selected by the experimenter under a prespecified probability model.

While joint models for binary and continuous outcomes have been developed when both outcomes are available for every fetus, many existing approaches are not directly applicable when the continuous outcome is not observed in a SRS. We adapt a likelihood-based approach for jointly modelling clustered binary and continuous outcomes when the continuous response is missing by design and missingness depends on the binary trait. The approach takes into account the probability that a fetus is selected in the subset. Through the use of a partial likelihood, valid estimates can be obtained by a simple modification to the partial likelihood score. Data involving the herbicide 2,4,5-T are analyzed. Simulation results confirm the approach.

Keywords

Fetal toxicity; Clustering; Dose response modeling; Partial likelihood; Correlated probit model

1. Introduction

Growing concern over the occurrence of birth defects and developmental abnormalities that may occur as a result of prenatal exposure to drugs, chemicals and other environmental factors has led to agencies such as the U.S. Environmental Protection Agency (EPA) and the Food and Drug Administration (FDA) to place emphasis on setting exposure limits in order to protect the public from the adverse effects of these substances. Because data on humans are often unavailable, information from controlled animal experiments accounts for a large proportion of the data on the effects of toxins.

Fetal toxicity studies often involve exposing a pregnant dam to the study agent at the peak of fetal organogenesis. Experiments known as segment II studies, intended to evaluate dose response, follow federal testing guidelines and have been described in detail elsewhere

†Address for correspondence: Julie S. Najita, Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, 44 Binney Street, Boston, MA, 02115, USA jnajita@hsph.harvard.edu.

(Manson, 1994). Briefly, these experiments typically involve 3–4 dose groups including a control group, with each dose group containing 20-30 dams. Just prior to normal delivery, the dams are sacrificed and the uterine contents are examined. Typically, offspring are evaluated on multiple outcomes which may include a mix of discrete and continuous responses. Outcomes include prenatal loss (i.e., resorptions and fetal deaths), and among viable offspring, a number of outcomes including fetal weight, externally visible malformations, and malformations of internal organs and the skeleton. In addition, as it is well known that offspring from the same dam tend to respond more similarly than those from different dams, possibly due to genetic similarity and a shared maternal environment during gestation, clustering within litter is a common characteristic of data from these experiments. The statistical challenges involved have been discussed by others (Ryan, 1992; Aerts et al., 2002).

While much is known about a number of toxins, the biological mechanisms involved are usually not well established. Another area of fetal toxicity studies involves experiments that are intended to evaluate specific mechanisms of the developmental processes that may be interrupted by prenatal exposure to a toxin. Often, these studies require detailed assessments of offspring involving a continuous response, that may reflect minute changes that are not easily determined by gross examination of the fetus. These studies extensively evaluate a small number of animals and are often restricted to two groups, exposed and unexposed. For example, in studies of ocular development, Inagaki & Kotani (2003) reported the rate of microphthalmia and conducted detailed morphometric evaluation of the embryos of rats exposed to soft X-ray irradiation. Examination by stereo-microscope revealed reduced growth of the optic cup in exposed offspring. Others have included detailed evaluation in studies of lung maturation, development of the eye and of the liver (Grasty et al., 2005; Martins et al., 2005; Rogers & Hurley, 1987).

Although these types of studies differ somewhat in their objectives, they have some shared characteristics which includes evaluating multiple outcomes. Some outcomes may be cheap and quick to assess, such as malformation determined by gross examination, while others, such as the size of the ocular cup, are time-consuming and expensive to measure. While there may be sufficient resources to evaluate all offspring in small toxicity studies, in large studies, cost constraints limit the extent to which detailed assessments can be made, so collecting detailed measures on all subjects is unrealistic. To overcome this limitation, some studies evaluate a fixed number of pups per litter (Rogers & Hurley, 1987; Grasty et al., 2005) while others select pups from a fixed number of litters (Kennedy & Elliott, 1986).

If exposure-related changes in the detailed measure are more likely to occur in affected offspring, the experimenter might preferentially select affected offspring in which to observe the detailed measure. While there might be resources to do detailed assessment among a fraction of offspring, a simple random sample (SRS) might not be the best use of limited resources and therefore, selecting a SRS may not be optimal. One approach to selecting a subset is to oversample the affected offspring. Oversampling has arisen before in evaluating eye size and malformation of the eye (Weller et al., 1999).

We consider a case in which a binary trait, such as malformation, is observed in all offspring while a continuous outcome, corresponding to the detailed assessment, is observed in a subset of offspring selected from the main study. The subset is not a SRS in the sense that the probability of selecting an affected offspring is not equal to that of an unaffected offspring. We were motivated by the idea that a probability-based subsample that is not a SRS might be obtained, and sought an approach to analyze these data. Our approach pairs existing methods for mixed effects models, which have been previously applied to analyze fetal toxicity data (Gueorguieva & Agresti, 2001; Dunson, 2000), with the concept of two-phase sampling, which has been used in screening and surveys in psychiatry (Deming, 1977; Shrout & Newman,

1989). In our approach, the screening variable, which is the variable at the first phase of sampling (i.e., malformation), does not necessarily reflect the same phenomenon as the variable at the second phase (the detailed assessment), though the two are expected to be correlated. Our intent, in the application of two-phase sampling, differs somewhat from the screening context in that malformation status itself is of interest because it carries information on exposure. Thus, we were concerned with analyzing the variable at the first phase as well as the variable at the second phase. We refer to the sampling procedure at the second phase as subsampling to reflect this difference in intent. While the application to toxicology is presented as a working example, our approach can be applied more generally to analyze a mix of binary and continuous outcomes when the continuous response is missing by design. We apply our approach to a segment II study involving a number of doses that also allows for estimation of a joint dose response curve.

We begin with a bivariate regression model in Section 2. The model is an extension of the clustered ordinal regression approach of Hedeker & Gibbons (1994) that includes the continuous outcome. To handle subsampling, we then derive a partial likelihood (PL) that is based on the bivariate model, and give an expression for the PL score in Section 3. We show that consistent estimates can be obtained by adjusting the PL score via a simple weighting scheme that accounts for the sampling design. Results from a data analysis involving the herbicide 2,4,5-T are presented in Section 4 and a simulation study confirms the PL approach in Section 5. Some of the technical details are relegated to a separate technical report available from the authors.

2. Bivariate random effect regression model (BRM)

Our approach follows the method of Hedeker & Gibbons (1994) for modelling clustered ordinal outcomes. Central to the model are a normally distributed unobservable latent trait and fixed but unknown threshold values which generate the observed ordinal outcomes. In extending the model of Hedeker & Gibbons (1994) to account for a continuous outcome, the latent trait and continuous outcome are assumed to have a bivariate normal distribution. For the purpose of modelling fetal malformation, attention is restricted to clustered binary outcomes. The bivariate random effect model (BRM) accounts for a binary and a continuous outcome. We assume that mean fetal response depends only on fixed effects so a one-dimensional mean zero random effect for litter is assumed. As the latent trait and the continuous outcome may not be in the same scale, a parameter for each outcome is used to scale the random effect. In addition, consistent with assumptions typical for fetal toxicity studies, no fetus-specific effects are assumed so that only litter-level covariates are considered. Finally, the latent trait and the continuous outcome are assumed to be positively correlated, so that small values of the latent variable, corresponding to malformation, are correlated with lower values of the continuous outcome. We first assume there is a latent variable for malformation, M_{ik}^*, with unknown latent variance σ_1^2, and a continuous outcome, S_{ik}, for fetus k in litter i. The joint model for $(M_{ik}^* \; S_{ik})^T$ is written

$$
\begin{aligned}
M_{ik}^* &= \mathbf{w}_{1i}^T \tilde{\alpha}_1^* + \tilde{\sigma}_{b1}\theta_i + \tilde{\varepsilon}_{1ik} \\
S_{ik} &= \mathbf{w}_{2i}^T \alpha_2 + \sigma_{b2}\theta_i + \varepsilon_{2ik}
\end{aligned}
$$

where the random effect, $\theta_i \overset{iid}{\sim} N(0, 1)$, is independent of the vector of error terms $(\tilde{\varepsilon}_{1ik} \; \varepsilon_{2ik})$ where $(\tilde{\varepsilon}_{1ik} \; \varepsilon_{2ik})^T \sim N(0, \Sigma)$ and

$$\tilde{\sum} = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix}.$$

As M_{ik}^* is unobservable, the unknown latent variance σ_1^2 is not uniquely identifiable. M_{ik}^* can be rescaled by the unknown latent variance producing a second latent variable, Y_{ik}^*, with variance 1. Denoting Y_{ik}^* and S_{ik} for latent malformation and the continuous outcome, respectively, for fetus k in litter i, the joint model for $(Y_{ik}^* S_{ik})^T$, for a single fetus is written

$$\begin{aligned} Y_{ik}^* &= z_{1ik} + \varepsilon_{1ik} \\ S_{ik} &= z_{2ik} + \varepsilon_{2ik} \end{aligned}$$

where $z_{1ik} = \mathbf{w}_{1i}^T \alpha_1^* + \sigma_{b1}\theta_i$, $z_{2ik} = \mathbf{w}_{2i}^T \alpha_2 + \sigma_{b2}\theta_i$, $\alpha_1^* = \tilde{\alpha}_1^*/\sigma_1$. $\sigma_{b1} = \tilde{\sigma}_{b1}/\sigma_1$, and $\varepsilon_{1ik} = \tilde{\varepsilon}_{1ik}/\sigma_1$. $\varepsilon_{ik} = (\varepsilon_{1ik}\ \varepsilon_{2ik})^T$ is a mean zero bivariate normal random variable with covariance matrix

$$\sum = \begin{pmatrix} 1 & \sigma_2\rho \\ \sigma_2\rho & \sigma_2^2 \end{pmatrix}.$$

We assume $\rho > 0$ so that malformations are correlated with lower values of the continuous outcome.

A malformation occurs when the latent variable falls below the fixed threshold, γ_1. The observable pair of variables is (O_{ik}, S_{ik}) where $O_{ik} = 1$, a fetal malformation, if $\gamma_0 < Y_{ik}^* \le \gamma_1$ and $O_{ik} = 2$ (no malformation) if $\gamma_1 < Y_{ik}^* \le \gamma_2$. We take $\gamma_0 = -\infty$ and $\gamma_2 = \infty$. When a binary outcome is generated from the unobservable latent variable, it is well known that the unknown threshold is not individually estimable. Therefore, while α_1^* is not estimable, a parameter, α_1, that is a transformation of α_1^* that depends on γ_1 can be estimated. For example, in the simplified case where there is no clustering, if the latent mean is $\alpha_{11}^* + \alpha_{12}^* d$, only $\alpha_{11} = \gamma_1 - \alpha_{11}^*$ and $\alpha_{12} = -\alpha_{12}^*$ are estimable since only the probability of malformation can be estimated from the data. For the data analysis, we were primarily interested in modeling the dose effect of a toxin. For convenience, we chose to set $\gamma_1 = 0$, although other values could be chosen, which would result in shifting the intercept term for malformation.

Supressing the subscript for litter, i, temporarily, the marginal between-fetus and within-fetus correlations are found to be

$$\begin{aligned} \rho_Y &= \mathrm{cor}(Y_k^*, Y_{k'}^*) = \frac{\sigma_{b1}^2}{\sigma_{b1}^2 + 1} \\ \rho_S &= \mathrm{cor}(S_k, S_{k'}) = \frac{\sigma_{b2}^2}{\sigma_{b2}^2 + \sigma_2^2} \\ \rho_{YS} &= \mathrm{cor}(Y_k^*, S_k) = \frac{\sigma_2\rho + \sigma_{b1}\tilde{\sigma}_{b2}}{\sqrt{(\sigma_{b1}^2 + 1)(\sigma_{b2}^2 + \sigma_2^2)}} \\ \rho_{YS'} &= \mathrm{cor}(Y_k^*, S_{k'}) = \frac{\sigma_{b1}\sigma_{b2}}{\sqrt{(\sigma_{b1}^2 + 1)(\sigma_{b2}^2 + \sigma_2^2)}}. \end{aligned}$$

This marginal correlation structure is represented in Figure 1. The marginal joint distribution is

$$\begin{pmatrix} \mathbf{Y}^* \\ \mathbf{S} \end{pmatrix} \sim N_{2n} \left(\begin{pmatrix} \mathbf{w}_1^T \alpha_1 \mathbf{1}_n \\ \mathbf{w}_2^T \alpha_2 \mathbf{1}_n \end{pmatrix}, V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix} \right)$$

where $V_{11} = I_n + \sigma_{b1}^2 J_n$, $V_{21} = V_{12} = \sigma_2 \rho I_n + \sigma_{b2}^2 J_n$, and $V_{22} = \sigma_2^2 I_n + \sigma_{b2}^2 J_n$. V can be rewritten

$$V = \begin{pmatrix} \sigma_Y^2[(1-\rho_Y)I_n + \rho_Y J_n] & \sigma_Y \sigma_S[(\rho_{YS} - \rho_{YS'})I_n + \rho_{YS'} J_n] \\ \sigma_Y \sigma_S[(\rho_{YS} - \rho_{YS'})I_n + \rho_{YS'} J_n] & \sigma_S^2[(1-\rho_S)I_n + \rho_S J_n] \end{pmatrix}.$$

where $\sigma_Y^2 = \sigma_{b1}^2 + 1$ and $\sigma_S^2 = \sigma_{b2}^2 + \sigma_2^2$. Of interest is to estimate $\eta = (\alpha_1 \ \alpha_2 \ \sigma_{b1} \ \sigma_{b2} \ \sigma_2 \ \rho)^T$. The mean parameters, α_1 and α_2, and the within-litter variance for fetal weight, σ_2, are uniquely identifiable as they can be estimated by analyzing the outcomes separately. As the marginal correlations, ρ_Y, ρ_S, and ρ_{YS} can be individually estimated, σ_{b1}, σ_{b2}, and ρ are also uniquely identifiable which is seen by rewriting the expressions relating model parameters and the marginal correlations (see Appendix A).

2.1. Estimation for complete data pairs

If (o_{ik}, s_{ik}) are observed for every fetus $k = 1, \cdots, n_i$ and litter $i = 1, \cdots, N$, the marginal joint likelihood is given by

$$L(\eta) = \prod_{i=1}^{N} h(\mathbf{o}_i, \mathbf{s}_i; \eta) = \prod_{i=1}^{N} \int P(\mathbf{o}_i, \mathbf{s}_i|\theta_i) g(\theta_i) d\theta_i \tag{1}$$

where $P(\mathbf{o}_i, \mathbf{s}_i|\theta_i) = \prod_{k=1}^{ni} P(o_{ik}, s_{ik}|\theta_i)$. Estimation proceeds via Fisher scoring in a fashion similar to the univariate ordinal model. Estimates are obtained by solving the score equation

$$U(\eta) = \sum_{i=1}^{N} h(\mathbf{o}_i, \mathbf{s}_i; \eta)^{-1} \frac{\partial}{\partial \eta} h(\mathbf{o}_i, \mathbf{s}_i; \eta)$$

where derivatives of the log likelihood are obtained by directly differentiating the log of (1)

$$\frac{\partial h(\mathbf{o}_i, \mathbf{s}_i; \eta)}{\partial \eta} = \int \sum_{k=1}^{n_i} \left[\frac{\frac{\partial}{\partial \eta} P(s_{ik}|\theta_i)}{P(s_{ik}|\theta_i)} + \sum_{j=1}^{2} 1(o_{ik} = j) \frac{\frac{\partial}{\partial \eta} P(O_{ik} = j|s_{ik}, \theta_i)}{P(O_{ik} = j|s_{ik}, \theta_i)} \right] P(\mathbf{o}_i, \mathbf{s}_i|\theta_i) g(\theta_i) d\theta_i \tag{2}$$

where $P(s_{ik}|\theta_i) = \phi(s'_{ik})/\sigma_{2i}$, $P(O_{ik} = 1|s_{ik}, \theta_i) = \Phi(v_{ik})$, $P(O_{ik} = 2|s_{ik}, \theta_i) = 1 - \Phi(v_{ik})$,

$v_{ik} = -m_{ik}/\sqrt{1-\rho^2}$, $m_{ik} = z_{1ik} + \rho(s_{ik} - z_{2ik})/\sigma_2$, and $s'_{ik} = (s_{ik} - z_{2ik})/\sigma_2$. $\varphi()$ and $\Phi()$ denote, respectively, the probability density function and cumulative distribution function of the normal distribution. Substituting terms, the right hand side of (2) is written

$$\int \sum_{k=1}^{n_i} \left[\frac{\partial}{\partial \eta} \log P(s_{ik}|\theta_i) + \left(\frac{1(o_{ik}=1)}{\Phi(v_{ik})} - \frac{1(o_{ik}=2)}{1-\Phi(v_{ik})} \right) \frac{\partial \Phi(v_{ik})}{\partial \eta} \right] P(\mathbf{o}_i, \mathbf{s}_i|\theta_i) g(\theta_i) d\theta_i. \tag{3}$$

Standard error estimates are based on the inverse observed information. Marginal quantities are approximated by numerical quadrature (Najita, 2006).

3. Partial likelihood approach

To account for subsampling, the mechanism generating the observed data is viewed in terms of two steps: a sampling step and an additional selection step. In the sampling step, a fetus potentially observable for the data pair, (*o*, *s*), is sampled at random from the population, depicted on the left in Figure 2. This step corresponds to viable offspring observed when the uterine contents are examined. The sampled data pair falls in one of two strata, based on malformation status. At the selection step, a subsample is obtained where offspring are selected from stratum 1 (abnormality or affected) with probability p_1 and from stratum 2 (healthy or unaffected) with probability p_2. The continuous outcome, *S*, is observed only for pups selected at the second step. Thus, in affected pups, a data pair is observed with probability p_1 while in unaffected pups, the data pair is observed with probability p_2. For pups not selected at the second step, only malformation status is observed.

The case $p_1 = p_2 = 1$ corresponds to the complete data setting where the selection step is absent. $p_1 = p_2 = p < 1$ corresponds to the SRS case where the population-based relative frequencies are retained in the subsample. $p_1 > p_2$ corresponds to oversampling malformations while $p_1 < p_2$ corresponds to undersampling. When there is subsampling, the likelihood involves a normalizing constant that depends on the unknown probability of malformation, so the full likelihood is complex. We base inference on a PL in which the selection process is not formally modeled.

3.1. Partial likelihood under subsampling

We now derive expressions for the PL and the PL score. Let, Δ_{ik} be the variable indicating whether the continuous outcome is observed for the kth fetus where $P(\Delta_{ik}=1|O_{ik}=o_{ik})=p_1^{1(o_{ik}=1)} p_2^{1(o_{ik}=2)}$. Write the data available for the kth fetus as $\mathbf{d}'_{ik}=(o_{ik}, \delta_{ik})$ if $\delta_{ik}=0$ and $\mathbf{d}'_{ik}=(o_{ik}, s_{ik}, \delta_{ik})$ if $\delta_{ik}=1$. Conditional on the random effect for litter, the distribution of the observed data for a single fetus is $P(\mathbf{d}'_{ik}|\theta_i)=P(o_{ik}, \delta_{ik}|\theta_i)^{1-\delta_{ik}} P(o_{ik}, s_{ik}, \delta_{ik}|\theta_i)^{\delta_{ik}}$ which we write as

$$P(\mathbf{d}'_{ik}|\theta_i)=\prod_{j=1}^{2} [(1-p_j) P(O_{ik}=j|\theta_i)]^{d_{0jik}} [p_j P(O_{ik}=j, s_{ik}|\theta_i)]^{d_{1jik}} \tag{4}$$

where $d_{\ell jik} = 1(\delta_{ik} = \ell, o_{ik} = j)$, $\ell = 0, 1$ and $j = 1, 2$. Setting $P(\mathbf{d}'_i|\theta_i)=\prod_{k=1}^{n_i} P(\mathbf{d}'_{ik}|\theta_i)$ and integrating over the random effect gives the PL

$$\mathrm{PL}(\eta)=\prod_{i=1}^{N} \int P(\mathbf{d}'_i|\theta_i) g(\theta_i) d\theta_i. \tag{5}$$

Setting $\mathrm{PL}_i = \int P(\mathbf{d}'_i|\theta_i)g(\theta_i)d\theta_i$, the PL score is obtained by substituting (4) in (5) and directly differentiating the log PL:

$$U(\eta) = \sum_{i=1}^{N} \frac{1}{\mathrm{PL}_i} \int D_1(\mathbf{d}'_i; \theta_i, \eta) P(\mathbf{d}'_i|\theta_i) g(\theta_i) d\theta_i,$$

(6)

where

$$D_1(\mathbf{d}'_i; \theta_i, \eta) = \sum_{k=1}^{n_i} \sum_{j=1}^{2} \left[d_{0jik} \frac{\frac{\partial}{\partial \eta} P(O_{ik}=j|\theta_i)}{P(O_{ik}=j|\theta_i)} + d_{1jik} \frac{\frac{\partial}{\partial \eta} P(O_{ik}=j, s_{ik}|\theta_i)}{P(O_{ik}=j, s_{ik}|\theta_i)} \right].$$

(7)

Writing $P(O_{ik}=j|\theta_i) = \Phi(u_{ik})^{1(j=1)}(1-\Phi(u_{ik}))^{1(j=2)}$ where $u_{ik} = -z_{1ik}$, the derivatives in (7) can be expressed as

$$\frac{\frac{\partial}{\partial \eta} P(O_{ik}=j|\theta_i)}{P(O_{ik}=j|\theta_i)} = \frac{(-1)^{1(j=2)} \frac{\partial}{\partial \eta} \Phi(u_{ik})}{\Phi(u_{ik})^{1(j=1)}[1-\Phi(u_{ik})]^{1(j=2)}}$$

$$\frac{\frac{\partial}{\partial \eta} P(O_{ik}=j, s_{ik}|\theta_i)}{P(O_{ik}=j, s_{ik}|\theta_i)} = \frac{(-1)^{1(j=2)} \frac{\partial}{\partial \eta} \Phi(v_{ik})}{\Phi(v_{ik})^{1(j=1)}[1-\Phi(v_{ik})]^{1(j=2)}} + \frac{\partial}{\partial \eta} \log P(s_{ik}|\theta_i)$$

with $P(s_{ik}|\theta_i)$ and v_{ik} as defined in Section 2.1. Substituting terms, (7) can be written as

$$D_1(\mathbf{d}'_i; \theta_i, \eta) = \sum_{k=1}^{n_i} \sum_{j=1}^{2} \left[\frac{d_{0jik}(-1)^{1(j=2)} \frac{\partial}{\partial \eta} \Phi(u_{ik})}{\Phi(u_{ik})^{1(j=1)}[1-\Phi(u_{ik})]^{1(j=2)}} \right.$$
$$\left. + d_{1jik} \left(\frac{(-1)^{1(j=2)} \frac{\partial}{\partial \eta} \Phi(v_{ik})}{\Phi(v_{ik})^{1(j=1)}[1-\Phi(v_{ik})]^{1(j=2)}} + \frac{\partial}{\partial \eta} \log P(s_{ik}|\theta_i) \right) \right].$$

(8)

It can be shown that when selection occurs at rates that differ by malformation status, the expected value of (6) is non-zero for α_2, the component corresponding to the mean continuous outcome. To correct for bias, the PL score is modified by sampling weights. The weighted PL score, $\tilde{U}(\eta)$, can be written $\tilde{U}(\eta) = \sum_{i=1}^{N} \tilde{U}_i(\eta)$, where

$$\tilde{U}_i(\eta) = \frac{1}{\mathrm{PL}_i} \int D_w(\mathbf{d}'_i; \theta_i, \eta) P(\mathbf{d}'_i|\theta_i) g(\theta_i) d\theta_i.$$

(9)

D_w has the form of D_1 in (7) with $w_{\ell j} d_{\ell jik}$ substituted for $d_{\ell jik}$ where $w_{1j} = 1/p_j$ and $w_{0j} = 1/(1-p_j), j = 1, 2$. It is straightforward to show that the weighted PL score equation is an unbiased estimating equation by conditioning on the random effect and taking expectations (Najita, 2006). PL estimates (PLE's) are obtained by solving the weighted score equation $\tilde{U}(\eta) = 0$.

Standard error estimates are based on the robust sandwich estimator. To obtain an expression for the robust SE, we write the estimating function $G_N(\eta) = \frac{1}{N}\sum_{i=1}^{N}\Psi(\mathbf{d}_i', \eta)$ where $\Psi(\mathbf{d}_i', \eta) = \tilde{U}_i(\eta)$. Expansion of G_N around η_0 gives, for large N,

$$\sqrt{N}(\widehat{\eta} - \eta_0) \approx -\dot{G}_N(\eta_0)\sqrt{N}G_N(\eta_0)$$

where $\dot{G}_N(\eta_0)$ denotes $\frac{\partial G_N}{\partial \eta^T} = \frac{\partial}{\partial \eta^T}\frac{1}{N}\sum_{i=1}^{N}\tilde{U}_i(\eta)$ evaluated at $\eta = \eta_0$. Following general results for M-estimators (Huber, 1967), PL estimates are asymptotically normal, so, for large N, $\widehat{\eta} \approx N(\eta_0, 1/N\,A(\eta_0)^{-1}B(\eta_0)\,[A(\eta_0)^{-1}]^T)$ where $A(\eta_0) = \mathrm{E}\left[-\frac{\partial}{\partial \eta^T}\tilde{U}_i(\eta_0)\right]$ and $B(\eta_0) = \mathrm{E}\,[\tilde{U}_i(\eta_0)\tilde{U}_i(\eta_0)^T]$. Therefore, we estimate the robust sandwich variance by

$$\left(\frac{1}{N}\sum_i\frac{\partial\tilde{U}_i}{\partial\eta}\right)^{-1}\frac{1}{N}\sum_i\tilde{U}_i(\eta)\tilde{U}_i(\eta)^T\left[\left(\frac{1}{N}\sum_i\frac{\partial\tilde{U}_i}{\partial\eta}\right)^{-1}\right]^T\Bigg|_{\eta=\widehat{\eta}}.$$

(10)

Expressions for the derivatives are obtained by direct differentiation (see Appendix B).

4. Application to 2,4,5-T data

We have confirmed the PL method via extensive simulations (Najita (2006) and summarized in Section 5 below) and here we show results from an analysis of a large toxicity study. For the data analysis, we wanted to evaluate the PL approach by comparing PLE's to MLE's as we wanted to directly compare differences due only to subsampling. Our motivation in this approach was to compare MLE's obtained from an "expensive" study, in which all offspring are evaluated, with PLE's that could be obtained had the study been implemented by oversampling affected offspring, a potentially less expensive study.

In order to evaluate the PLE's directly against MLE's, we focused on live outcomes. We believe that limiting the analysis to live outcomes was the simplest approach to make the comparison as the subsampling is intended for live outcomes only. Thus, our analysis excludes non-live outcomes and while experimenters who are interested in risk assessment (evaluating safe doses) may eventually wish to include non-live outcomes, existing approaches to augment the likelihood can be easily applied to the PL as well (Catalano & Ryan, 1992; Regan & Catalano, 1999).

For live outcomes, we focused on malformation and fetal weight which are typical primary endpoints for live offspring. Given a subsampling of fetal weight, we wanted to show that valid estimates could still be obtained when the subsample is not a SRS, and because fewer pups would be evaluated, there is potential for cost-saving. For continuous outcomes that are more expensive to measure than fetal weight, such as the size of the ocular cup or eye size, the potential for cost-savings would be greater.

In applying the PL approach, we sought a single data set in which we could compare MLEs with PLE's. To make the comparison, we began with data containing information on all offspring (complete pairs). A second data set was derived by oversampling malformations from the complete pairs data set (subsampling data set). Because we were interested in making a

direct comparison between the MLE's and PLE's, a subsampling data set alone was not sufficient for our purposes. We chose to analyze data from a large developmental toxicity study of the herbicide 2,4,5-tricholorophenoxyacetic acid (2,4,5-T) in C57BL/6 mice (Holson et al., 1992).

Between gestational days 4–16, dams were exposed to 2,4,5-T at one of seven dose levels (0, 15, 30, 45, 60, 75, 90 mg/kg/day). The number of pregnant dams at each dose level varied by dose group by design (13-66 dams per dose). The main study was comprised of 367 litters and we focus on the 337 litters containing at least one viable pup. While the most pronounced effects were reduction in fetal weight and the incidence of cleft palate, our analysis involved malformation of any type. Fetal weight and malformation status were available for 2,295 fetuses, which comprise the complete pairs data set of 337 litters. Dose effects of 2,4,5-T were evident in fetal weight and malformation. The proportion of fetal malformations increased steadily with dose ranging from a background rate of approximately 1% to 37% at the highest dose (see left panel, Figure 3). A plot of malformation rate transformed on a normal quantile scale suggested a linear trend in dose.

Fetal weight declined with dose, from 815 mg at the control dose to 548 mg at the highest dose. As there appeared to be some deviation from this trend at 15 mg/kg (see right panel, Figure 3), attempts were made to fit a quadratic term for dose in a univariate model for fetal weight in order to determine whether non-linear terms were appropriate before fitting the bivariate model. While the linear term for dose was significant, the quadratic term was not significant (possibly due to a smaller number of litters at the 15 mg/kg dose level). We fit the regression model

$$Y_{ik}^* = \alpha_{11} + \alpha_{12} d_i + \sigma_{b1} \theta_i + \varepsilon_{1ik} \tag{11}$$

$$S_{ik} = \alpha_{21} + \alpha_{22} d_i + \sigma_{b2} \theta_i + \varepsilon_{2ik} \tag{12}$$

$$\mathrm{var}(\varepsilon_{ik}) = \begin{pmatrix} 1 & \sigma_2 \rho \\ \sigma_2 \rho & \sigma_2^2 \end{pmatrix} \tag{13}$$

where σ_{b1}, σ_{b2}, σ_2, and ρ were included as fixed constants. A summary of dose levels, malformation rates and fetal weight are presented in Table 1.

To apply the PL approach, the subsampling data set was created from the complete data set by oversampling malformations ($p_1 = 0.9$, $p_2 = 0.6$). The subsampling data set consisted of data on fetal malformation (n=2295) and fetal weight (n=1469) for pups from the 337 litters. Seven litters contained no fetal weight observations while fetal weight was observed for a single pup in 21 litters. Among the remaining 309 litters, there was information on both outcomes for at least two pups. As a result of oversampling malformations, mean fetal weight in the subsampling data set was generally lower than that in the complete data set. MLE's were obtained from the complete data set and PLE's were obtained from the subsampling data set. The regression estimates were used to calculate marginal fetus-level correlations as described in Section 2.

A comparison of parameter estimates is presented in Table 2. Overall, regression estimates were similar (see Figure 3). MLE's of the within litter variance of fetal weight and the PLE's

were similar (MLE: $\sigma_2 = 0.073$ versus PLE: $\sigma_2 = 0.072$). Estimates of the random effect SD's were also similar (MLE: $\sigma_{b1}=0.48$, $\sigma_{b2}=0.09$ versus PLE: $\sigma_{b1}=0.54$, $\sigma_{b2}=0.10$). As a result, there was very good agreement in estimates of marginal correlations within litter (MLE: $\rho_Y=0.19$, $\rho_S=0.62$ versus PLE: $\rho_Y=0.23$, $\rho_S=0.67$). In addition, the within-fetus correlations between fetal weight and latent malformation were also similar (MLE: $\rho=0.40$ versus PLE: $\rho=0.39$) and this led to similar estimates of the marginal within-fetus correlations between outcomes (MLE: $\rho_{YS}=0.56$ versus PLE: $\rho_{YS}=0.59$). Between-fetus correlations between outcomes were comparable (MLE: $\rho_{YS'}=0.34$ versus PLE: $\rho_{YS'}=0.39$). As expected, robust SE's tended to be larger than ML SE's. Although robust SE's for latent malformation were twice that of the ML SE's and the robust SE's for fetal weight were 3–4 times the ML SE's, all model parameters remained highly significant and qualitatively, the overall conclusions were unaffected by the larger robust SE's. The smaller sample size of fetal weight may be one explanation for the larger robust SE's.

We chose to model the variance and correlation terms, σ_{b1}, σ_{b2}, σ_2, and ρ, as constants because this is frequently assumed in developmental toxicity studies and because it is consistent with a prior analysis of the data (Holson et al., 1992). In other toxicity studies, the experimenter's prior knowledge of the test agent's effects on the animal species selected for testing is expected to play a large role in specifying a model for these second order terms. The model is generalizable in the sense that analysts who may not want to assume homogeneity in the variances and correlations can specify dose-dependence. This is possible when studies involve a small number of doses and many replicates at each dose level, however, the precision of estimates would be limited by the sample size at each dose level.

In developmental toxicity studies, litter size often carries information about the live outcomes. For example, for fetal weight, it is well known that weight can be inversely related to litter size as there are fewer pups competing for a fixed amount of nutritional resources when the litter size is smaller. The dose effect can be underestimated if the test agent induces embryolethality and therefore smaller litters. This bias has been described previously by Romero et al. (1992). Approaches have proposed adjusting for litter size by including it as a covariate (Catalano & Ryan, 1992; Regan & Catalano, 2000; Gueorguieva & Agresti, 2001; Chen, 1993) while others have proposed jointly modeling litter size and the live outcomes (Dunson et al., 2003; Catalano et al., 1993). In the data analysis, we focused on modelling live outcomes, and therefore condition on the litter size. Thus, there is the potential for biased inference of the dose effect. However, here we found that estimates of the dose effect did not change when litter size was included in the model, hence we chose to fit the simpler model including only dose effects in our bivariate model. The model we fit is also consistent with a prior analysis of the data.

5. Simulation Study

In applying the PL approach to the 2,4,5-T data, application was limited to the oversampling case ($p_1 > p_2$). In this section, we give results from a simulation study where we compared MLE's and PLE's under a range of subsampling scenarios. For the purpose of the study, simulations involved binary and continuous outcomes with characteristics similar to the 2,4,5-T data and while there are differences in some of the details, the simulated data are consistent with data commonly seen in practice. Specifically, simulated experiments included four doses and a control, assuming the rate of malformation, transformed by the inverse probit function, increased linearly with dose while mean fetal weight decreased linearly with dose. The malformation rate ranged from 7% (background) to 69% at the highest dose. Comparing controls and those exposed to the highest dose, mean fetal weight was reduced by 300 mg. For each outcome, within-litter correlations were modest ($\rho_Y=0.10$, $\rho_S=0.10$) while a moderately high within-fetus correlation was assumed ($\rho_{YS}=0.60$). Correlations were held constant over

dose. Therefore, we fit the models in Equations (11)–(13) with $\alpha_1 \mathbf{w}_{1i} = 1.5 - 2d_i$, $\alpha_2 \mathbf{w}_{2i} = 5 - 3di$, $\sigma_{b1} = 0.33$, $\sigma_{b2} = 0.33$, $\sigma_2 = 1.0$, and $\rho = 0.56$. Dose levels were spaced equally between the control dose and the maximum (i.e., set to 0, 0.25, 0.50, 0.75, and 1.0). The number of live offspring in each litter was fixed at 10 and the number of dams assigned to each dose level was fixed at 30, corresponding to study sizes of 1,500 pups.

Subsampling scenarios were achieved by varying the sampling rates in malformation strata (p_1 and p_2). Specifically, for the SRS case, we sampled fetal weight from 90% of fetuses ($p1 = 0.9, p_2 = 0.9$). For the oversampling case, we sampled fetal weight from 90% affected offspring while only 60% of healthy fetuses were sampled ($p_1 = 0.9, p_2 = 0.6$). These sampling rates were exchanged ($p_1 = 0.6$, $p_2 = 0.9$) for the undersampling case. To demonstrate the robustness of PLE's over a range of subsampling rates, we also sampled fetal weight at rates with greater discrepancy as we expected bias to occur only when subsampling rates were different, and that the size of the bias would increase as the difference in sampling rates increased. We fixed the sampling rate at 0.90 for one stratum (e.g., affected offspring) and sampled at lower rates in the other ($p = 0.30, 0.40, 0.50$). While our intent was to confirm the PL method for bias correction, we also calculated robust SE's for comparison with simulation-based estimates.

To evaluate the PL approach, we compared the PLE's to the true parameter values under oversampling and undersampling. To understand the potential bias in the over- and undersampling cases, in addition to PLE's, we also calculated uncorrected estimates for the oversampling and undersampling cases, which were obtained by maximizing the PL without adjustment by sampling weights. Finally, for comparison purposes, we calculated MLE's when there were complete data or the subset was a SRS. For each case, 100 experiments were simulated.

Parameter estimates are summarized in Tables 3, 5 and 6. Simulations confirmed MLE's when fetal malformation and weight were available for all fetuses. When fetal weight was available in an SRS, evidence of bias was not observed. However, as expected for the unweighted PL, dose response was overestimated when malformations were oversampled (see uncorrected estimates, Table 3). Similarly, with the unweighted estimates, the dose effect was underestimated when malformations were undersampled, demonstrating the importance of including sampling weights. The magnitude of bias in the unweighted estimates increased as the discrepancy between the sampling rates increased and was greatest when the sampling rate was 0.30 (Tables 5 and 6). Agreement between PLE's and the true parameter values was very good. Model-based SE's agreed well with empirical estimates when there were complete data or the subsample was a SRS (Table 4). Robust SE's were similar to empirical estimates and robust SE's increased as the discrepancy in sampling rates increased (see Tables 7 and 8).

We elected not to sample at rates lower than 0.30 for practical reasons. First, it was thought to be unlikely that experimenters would choose rates lower than 0.30 as there might be interest in evaluating the continuous outcome alone and a small subsample might be viewed as having limited utility on its own. Second, while the simulations were designed to estimate the dose effect assuming all data pairs were available, subsampling at lower rates was thought to impact sample size sufficiently to reduce power. Hence subsampling at rates lower than 0.30 was not thought to be of practical interest. Third, simulation runs became impractical at rates lower than 0.30 because convergence was substantially slower so fitting the data required many more iterations than when $p_2 = 0.50$ or 0.60. Although infrequent, when fitting the simulated data, we observed sensitivity to the starting value when the sampling rate was 0.30. While we hypothesize this sensitivity to become more frequent with subsampling rates lower than 0.30, futher study would be needed to determine whether our results are unique to the experiment we have simulated.

6. Discussion and Future Work

Accounting for multiple endpoints may be appropriate for characterizing overall toxicity compared with approaches that focus on the most sensitive outcome thus motivating a joint modelling approach. In this paper, we have presented a joint model for binary and continuous outcomes that are clustered within litter and, using a PL, an approach when the continuous outcome is missing by design. To account for the non-representative subsample, the approach weights fetus-level terms by the inverse of subsampling probabilities. As methods that specify the joint distribution directly are limited by the lack of a natural joint distribution for a mix of discrete and continuous outcomes, a latent variable framework provides a convenient means with which to model this mix of outcomes, as others have done (Catalano & Ryan, 1992; Regan & Catalano, 1999; Gueorguieva & Agresti, 2001; Faes et al., 2004; Dunson, 2000). In contrast to approaches in which estimates depend on a litter-level statistic (Ochi & Prentice, 1984; Regan & Catalano, 1999), the mixed model formulation naturally lends itself to incorporating fetus-specific sampling weights because estimators can be expressed in terms of fetus-level responses. For segment II studies, prior approaches have incorporated non-live outcomes such as prenatal loss in evaluating dose response (Catalano & Ryan, 1992; Ryan, 1992; Dunson, 2000; Gueorguieva, 2005; Faes et al., 2006). While others have incorporated non-live outcomes, we were interested in modeling the live outcomes because we were interested in evaluating the PL approach, an approach intended for live outcomes. Although our method does not handle non-live outcomes, adjustments described by Regan & Catalano (2000) are possible. The PL approach is easily paired with a model for prenatal loss because our approach conditions on the number of live offspring. This can be accomplished by fitting a separate model for non-live outcomes followed by the PL approach for live outcomes.

In our model, outcomes within offspring are linked through a one-dimensional random effect corresponding to a random intercept model. We chose to analyze the 2,4,5-T data assuming a random intercepts model as a prior analysis found this assumption to be appropriate because with C57BL/6 mice, doses in this range had only a small negative effect on net maternal weight gain that was not statistically significant and decrement in maternal weight gain had a relatively small effect on fetal weight (Holson et al., 1992). In other experiments, a toxin could adversely impact dams and therefore, the maternal environment. In this case, it would be reasonable to include a multivariate random effect in order to account for a random intercept and random slope. One way to link outcomes with a multivariate random effect is to include a linear combination of random effects for each outcome, where the random effect components are related through a joint normal distribution, as suggested by Gueorguieva & Agresti (2001). In our approach, because the estimation requires integrating over the random effect, there would be practical limits as approximating the integral by numerical quadrature can become computationally challenging as the dimension of the random vector increases.

While we apply the PL approach to data from a segment II study, the method can be used in other settings as well. The PL approach might aid researchers who wish to conduct experiments in order to understand the complex processes involved in fetal development and the role that toxins play in disrupting normal development during organogenesis. If the continuous outcome is important for understanding the processes that lead to affected offspring, such as malformations, and affected offspring are relatively less common, then oversampling the affected pups may confer cost savings compared with approaches that assess all pups or a fixed proportion of pups. For example, if 10% of pups are affected, sampling 90% of affected pups and 30% of unaffected pups would lead to evaluating slightly more than one-third of all offspring, compared with approaches that sample one-half of all fetuses, for example. Greater savings are achievable by reducing the fraction of unaffected pups that are sampled. Our approach might also be of interest in the context of epidemiologic investigations. For example, in studies of respiratory health in children, the impact of indoor environmental exposures on

childhood asthma may be of interest. Measures of respiratory health in children who share the same household environment may be correlated so that accounting for clustering within household when exposure occurs at the household level may be important. While self-reported information on physician-diagnosed asthma may be easily obtained for all subjects, it may not be feasible to do detailed assessment for every participant if the assessments involve the use of tissue or serum.

In the PL approach, essential parameters are the subsampling probabilities which are specified at the time of study design. Future work with this approach would include a method for choosing the probabilities in order to achieve a desired precision for a parameter of interest or an overall cost requirement. While we have presented simulation results for one experiment with several choices of p_1 and p_2, further study is needed to understand more generally how to choose sampling rates during the planning phase of a study.

Appendix A Estimability of parameters in the joint model

As the mean parameters and the variance for fetal weight are estimable, it remains to show that the remaining parameters, σ_{b1}, σ_{b2}, and ρ are uniquely identifiable. Estimability of these parameters can be seen by rewriting the expressions relating the marginal correlations to the model parameters as follows:

$$\sigma_{b1} = \sqrt{\rho_Y/(1-\rho_Y)} \tag{14}$$

$$\sigma_{b2} = \sqrt{\rho_S \sigma_2^2/(1-\rho_S)} \tag{15}$$

$$\rho = \left(\rho_{YS} \sqrt{(\sigma_{b1}^2+1)(\sigma_{b2}^2+\sigma_2^2)} - \sigma_{b1}\sigma_{b2} \right)/\sigma_2 \tag{16}$$

The marginal correlation ρ_Y can be estimated from the data by fitting the binary outcome alone. The estimate for σ_{b1} is obtained by substituting the estimate for ρ_Y in (14). Similarly, the estimate for σ_{b2} is obtained by substituting estimates for ρ_S and σ_2 in the expression for σ_{b2} in (15). Finally, as the marginal correlation σ_{YS} is also estimable from the data, estimates of parameters on the right-hand side of (16) can be substituted to estimate ρ.

Appendix B BRM partial likelihood robust standard error estimates

Let $G_N(\eta) = \frac{1}{N} \sum_{i=1}^{N} \tilde{U}_i(\eta)$ be the scaled modified PL score function. The derivative of the estimating function is given by

$$\dot{G}_N(\eta) = \frac{\partial}{\partial \eta} \frac{1}{N} \sum_{i=1}^{N} \tilde{U}_i(\eta^T) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{L_i} \frac{\partial^2 \tilde{L}_i}{\partial \eta \partial \eta^T} - \frac{1}{L_i^2} \frac{\partial L_i}{\partial \eta} \frac{\partial \tilde{L}_i}{\partial \eta^T}.$$

Writing the derivative $\frac{\partial \tilde{L}_i}{\partial \eta} = \int D_w(\mathbf{d}'_i; \theta_i, \eta) P(\mathbf{d}'_i | \theta_i) g(\theta_i) d\theta_i$ and recalling

$\frac{\partial}{\partial \eta} P(\mathbf{d}'_i | \theta_i) = D_1(\mathbf{d}'_i; \theta_i, \eta) P(\mathbf{d}'_i | \theta_i)$, the second derivative is obtained by direct differentiation

$$\frac{\partial^2 \tilde{L}_i}{\partial \eta \partial \eta^T} = \frac{\partial}{\partial \eta} \int D_w(\mathbf{d}'_i; \theta_i, \eta^T) P(\mathbf{d}'_i | \theta_i) g(\theta_i) d\theta_i$$

$$= \int \left[\frac{\partial}{\partial \eta} D_w(\mathbf{d}'_i; \theta_i, \eta^T) + D_1(\mathbf{d}'_i; \eta) D_w(\mathbf{d}'_i; \theta_i, \eta^T) \right] P(d'_i | \theta_i) g(\theta_i) d\theta_i$$

where

$$\frac{\partial}{\partial \eta} D_w(\mathbf{d}'_i; \theta_i, \eta^T) = \sum_{k=1}^{n_i} \left[-\left(\frac{w_{01} d_{11ik}}{\Phi(u_{ik})^2} + \frac{w_{02} d_{12ik}}{[1-\Phi(u_{ik})]^2} \right) \frac{\partial \Phi(u_{ik})}{\partial \eta} \frac{\partial \Phi(u_{ik})}{\partial \eta^T} \right.$$
$$+ \left(\frac{w_{01} d_{01ik}}{\Phi(u_{ik})} - \frac{w_{02} d_{02ik}}{1-\Phi(u_{ik})} \right) \frac{\partial^2 \Phi(u_{ik})}{\partial \eta \partial \eta^T} - \left(\frac{w_{11} d_{11ik}}{\Phi(v_{ik})^2} + \frac{w_{12} d_{12ik}}{[1-\Phi(v_{ik})]^2} \right) \frac{\partial \Phi(v_{ik})}{\partial \eta} \frac{\partial \Phi(v_{ik})}{\partial \eta^T}$$
$$+ \left. \left(\frac{w_{11} d_{11ik}}{\Phi(v_{ik})} - \frac{w_{12} d_{12ik}}{1-\Phi(v_{ik})} \right) \frac{\partial^2 \Phi(v_{ik})}{\partial \eta \partial \eta^T} + (w_{11} d_{11ik} + w_{12} d_{12ik}) \frac{\partial^2 \log P(s_{ik}|\theta_i)}{\partial \eta \partial \eta^T} \right] \tag{17}$$

Expressions for derivatives of $\Phi(u_{ik})$, $\Phi(v_{ik})$ and $\log P(s_{ik}| \theta_i)$ are provided in Appendix C.

Appendix C Derivatives useful for robust standard error estimates

The derivatives in (17) take the form

$$\frac{\partial^2 \Phi(v_{ik})}{\partial \eta \partial \eta^T} = -v_{ik} \phi(v_{ik}) \frac{\partial v_{ik}}{\partial \eta} \frac{\partial v_{ik}}{\partial \eta^T} + \phi(v_{ik}) \frac{\partial^2 v_{ik}}{\partial \eta \partial \eta^T},$$
$$\frac{\partial^2 \Phi(u_{ik})}{\partial \eta \partial \eta^T} = -u_{ik} \phi(u_{ik}) \frac{\partial u_{ik}}{\partial \eta} \frac{\partial u_{ik}}{\partial \eta^T}$$

because the second derivatives of u_{ik} are zero. In addition, with

$$\frac{\partial}{\partial \eta_r} \log P(s_{ik}|\theta_i) = -\frac{1}{\sigma_{2i}} \frac{\partial \sigma_{2i}}{\partial \eta_r} - s'_{ik} \frac{\partial s'_{ik}}{\partial \eta_r},$$

we have

$$\frac{\partial^2 \log P(s_{ik}|\theta_i)}{\partial \eta \partial \eta^T} = \frac{1}{\sigma^2_{2i}} \frac{\partial \sigma_{2i}}{\partial \eta} \frac{\partial \sigma_{2i}}{\partial \eta^T} - \frac{1}{\sigma_{2i}} \frac{\partial^2 \sigma_{2i}}{\partial \eta \partial \eta^T} - \frac{\partial s'_{ik}}{\partial \eta} \frac{\partial s'_{ik}}{\partial \eta^T} - s'_{ik} \frac{\partial^2 s'_{ik}}{\partial \eta \partial \eta^T}.$$

The non-zero first derivatives are

$$\frac{\partial u_{ik}}{\partial \alpha_1} = -\mathbf{w}_{1i}, \quad \frac{\partial u_{ik}}{\partial \sigma_{b1}} = -\theta_i, \quad \frac{\partial v_{ik}}{\partial \alpha_1} = -(1-\rho^2)^{-1/2} \mathbf{w}_{1i}, \quad \frac{\partial v_{ik}}{\partial \alpha_2} = \rho \sigma_2^{-1} (1-\rho^2)^{-1/2} \mathbf{w}_{2i},$$

$$\frac{\partial v_{ik}}{\partial \sigma_{b1}} = -(1-\rho^2)^{-1/2}\theta_i, \qquad \frac{\partial v_{ik}}{\partial \sigma_{b2}} = \rho\sigma_2^{-1}(1-\rho^2)^{-1/2}\theta_i,$$

$$\frac{\partial v_{ik}}{\partial \rho} = -s'_{ik}(1-\rho^2)^{-1/2} - \rho(1-\rho^2)^{-3/2}m_{ik}. \qquad \frac{\partial v_{ik}}{\partial \sigma_2} = \rho(1-\rho^2)^{-1/2}s'_{ik}\sigma_2^{-1}$$

$$\frac{\partial s'_{ik}}{\partial \alpha_2} = -\sigma_2^{-1}\mathbf{w}_{2i}, \qquad \frac{\partial s'_{ik}}{\partial \sigma_{b2}} = -\sigma_2^{-1}\theta_i,$$

$$\frac{\partial s'_{ik}}{\partial \sigma_2} = -\sigma_2^{-1}s'_{ik}.$$

In addition, non-zero second derivatives of v_{ik} are

$$\frac{\partial^2 v_{ik}}{\partial \alpha_1 \partial \rho^T} = -\rho(1-\rho^2)^{-3/2}\mathbf{w}_{1i}, \qquad \frac{\partial^2 v_{ik}}{\partial \alpha_2 \partial \sigma_2^T} = -\rho\sigma_2^{-2}(1-\rho^2)^{-1/2}\mathbf{w}_{2i},$$

$$\frac{\partial^{-2} v_{ik}}{\partial \alpha_2 \partial \rho^T} = \left(\rho^2(1-\rho^2)^{-3/2} + (1-\rho^2)^{-1/2}\right)\sigma_2^{-1}\mathbf{w}_{2i}, \qquad \frac{\partial^2 v_{ik}}{\partial \sigma_{b1} \partial \rho^T} = -\rho(1-\rho^2)^{-3/2}\theta_i,$$

$$\begin{aligned}
\frac{\partial^2 v_{ik}}{\partial \sigma_{b2} \partial \sigma_2^T} &= -\rho(1-\rho^2)^{-1/2}\sigma_2^{-2}\theta_i, \\
\frac{\partial^2 v_{ik}}{\partial \sigma_{b2} \partial \rho^T} &= \left((1-\rho^2)^{-1/2} + \rho^2(1-\rho^2)^{-3/2}\right)\sigma_2^{-1}\theta_i \\
\frac{\partial^2 v_{ik}}{\partial \sigma_2 \partial \rho^T} &= \left(\rho^2(1-\rho^2)^{-3/2} + (1-\rho^2)^{-1/2}\right)\sigma_2^{-1}s'_{ik} \\
\frac{\partial^2 v_{ik}}{\partial \sigma_2^2} &= -2\rho(1-\rho^2)^{-1/2}\sigma_2^{-2}s'_{ik} \\
\frac{\partial^2 v_{ik}}{\partial \rho^2} &= -2s'_{ik}\rho(1-\rho^2)^{-3/2} - m_{ik}(1-\rho^2)^{-5/2} - 3m_{ik}\rho^2(1-\rho^2)^{-5/2}
\end{aligned}$$

The non-zero second derivatives of s'_{ik} are

$$\frac{\partial^2 s'_{ik}}{\partial \sigma_{b2} \partial \sigma_2} = \sigma_{2i}^{-2}\theta_i, \qquad \frac{\partial^2 s'_{ik}}{\partial \alpha_2 \partial \sigma_2} = \sigma_{2i}^{-2}\mathbf{w}_{2i}, \qquad \frac{\partial^2 s'_{ik}}{\partial \sigma_2^2} = 2\sigma_{2i}^{-2}s'_{ik}.$$

References

Aerts, M.; Geys, H.; Molenberghs, G.; Ryan, L. Topics in Modelling of Clustered Data. Chapman & Hall; 2002.

Catalano P, Ryan L. Bivariate latent variable models for clustered discrete and continuous outcomes. JASA 1992;87:651–658.

Catalano P, Scharfstein D, Ryan L, Kimmel C, Kimmel G. Statistical model for fetal death, fetal weight, and malformation in developmental toxicity studies. Teratology 1993;47:281–290. [PubMed: 8322222]

Chen J. A malformation incidence dose-response model incorporating fetal weight and/or litter size as covariates. Risk Analysis 1993;13:559–564. [PubMed: 8259446]

Deming W. An essay on screening, or on two-phase sampling, applied to surveys of a community. International Statistical Review 1977;45:29–37.

Dunson D. Bayesian latent variable models for clustered mixed outcomes. J R Statist Soc B 2000;62:355–366.

Dunson D, Chen Z, Harry J. A bayesian approach for joint modeling of cluster size and subunit-specific outcomes. Biometrics 2003;59:521–530. [PubMed: 14601753]

Faes C, Geys H, Aerts M, Molenberghs G. A hierarchical modeling approach for risk assessment in developmental toxicity studies. Computational Statistics and Data Analysis 2006;51:1848–1861.

Faes C, Geys H, Aerts M, Molenberghs G, Catalano P. Modeling combined continuous and ordinal outcomes in a clustered setting. Journal of Agricultural, Biological, and Environmental Statistics 2004;9:515–530.

Grasty R, Bjork J, Wallace K, Lau C, Rogers J. Effects of prenatal perfluorooctane sulfonate (PFOS) exposure on lung maturation in the perinatal rat. Birth Defects Research (Part B) 2005;74:405–416.

Gueorguieva R. Comments about joint modeling of cluster size and binary and continuous subunit-specific outcomes. Biometrics 2005;61:862–867. [PubMed: 16135040]

Gueorguieva R, Agresti A. A correlated probit model for joint modeling of clustered binary and continuous responses. Journal of the American Statistical Association 2001;96:1102–1112.

Hedeker D, Gibbons R. A random-effects ordinal regression model for multilevel analysis. Biometrics 1994;50:933–944. [PubMed: 7787006]

Holson J, Gaines T, Nelson C, LaBorde J, Gaylor D, Sheehan D, Young J. Developmental toxicity of 2,4,5-trichlorophenoxyacetic acid (2,4,5-T). I. Multireplicated dose-response studies in four inbred strains and one outbred stock of mice. Toxicological Sciences 1992;19:286–297.

Huber P. The behavior of maximum likelihood estimates under non-standard conditions. Procedings of the Fifth Berkeley Symposium on Mathematics, Statistics, and Probability 1967;I:221–233.

Inagaki S, Kotani T. Lens formation in the absence of optic cup in rat embryos irradiated with soft x-ray. Veterinary Ophthalmology 2003;6:61–66. [PubMed: 12641845]

Kennedy L, Elliott M. Ocular changes in the mouse embryo following acute maternal ethanol intoxication. International Journal of Developmental Neuroscience 1986;4:311–317. [PubMed: 3455592]

Manson, J. Developmental Toxicology. In: Kimmel, C.; Buelke-Sam, J., editors. Testing of Pharmaceutical Agents for Reproductive Toxicity. 2nd. Vol. chapter 15. Raven Press; 1994. p. 379-402.

Martins A, Azoubel R, Lopes R, di Matteo MS, de Arruda JF. Effect of sodium cyclamate on the rat fetal liver: a karyometric and stereological study. International Journal of Morphology 2005;23:221–226.

Najita, J. Technical Report 1155Z. Department of Biostatistics and Computational Biology, Dana Farber Cancer Institute; 2006. Technical report for "A new class of bivariate regression models for mixed binary and continuous outcomes: an application in developmental toxicology".

Ochi Y, Prentice R. Likelihood inference in a correlated probit regression model. Biometrika 1984;71:531–543.

Regan M, Catalano P. Likelihood models for clustered binary and continuous outcomes: application to developmental toxicology. Biometrics 1999;55:760–768. [PubMed: 11315004]

Regan M, Catalano P. Regression models and risk estimation for mixed discrete and continuous outcomes in developmental toxicology. Risk Analysis 2000;20:363–376. [PubMed: 10949415]

Rogers J, Hurley L. Effects of zinc deficiency on morphogenesis of the fetal rat eye. Development 1987;99:231–238. [PubMed: 3652998]

Romero A, Villamayor F, Grau M, Sacristan A, Ortiz J. Relationship between fetal weight and litter size in rats: application to reproductive toxicology studies. Reproductive Toxicology 1992;6:453–456. [PubMed: 1463926]

Ryan L. Quantitative risk assessment for developmental toxicity. Biometrics 1992;48:163–174. [PubMed: 1581483]

Shrout P, Newman S. Design of two-phase prevalence surveys of rare disorders. Biometrics 1989;45:549–555. [PubMed: 2765638]

Weller E, Long N, Smith A, Williams P, Ravi S, Gill J, Henessey R, Skornik W, Brain J, Kimmel C, Kimmel G, Holmes L, Ryan L. Dose-rate effects of ethylene oxide exposure on developmental toxicity. Toxicological Sciences 1999;50:259–270. [PubMed: 10478863]

$$
\begin{array}{ccccc}
& \text{fetus } k & & & \text{fetus } k' \\
\text{malformation} & Y_k & \longleftarrow \rho_Y \longrightarrow & & Y_{k'} \\
& \uparrow \quad \searrow & & \nearrow \quad \uparrow \\
& \rho_{YS} & \rho_{YS'} & & \rho_{YS} \\
& \downarrow \quad \swarrow & & \searrow \quad \downarrow \\
\text{weight} & S_k & \longleftarrow \rho_S \longrightarrow & & S_{k'}
\end{array}
$$
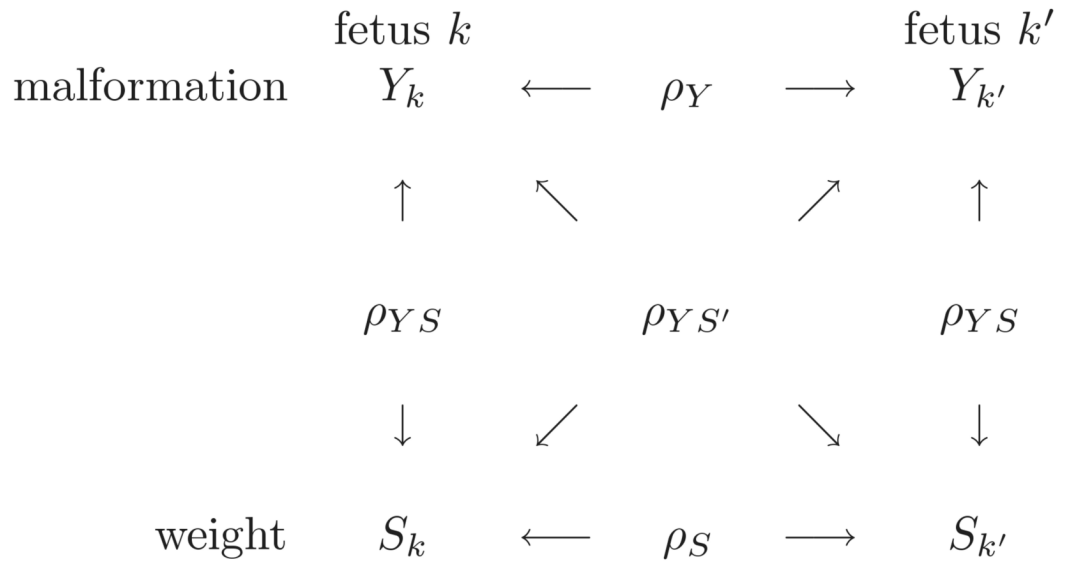
Fig. 1.
Marginal correlation structure of the bivariate model.
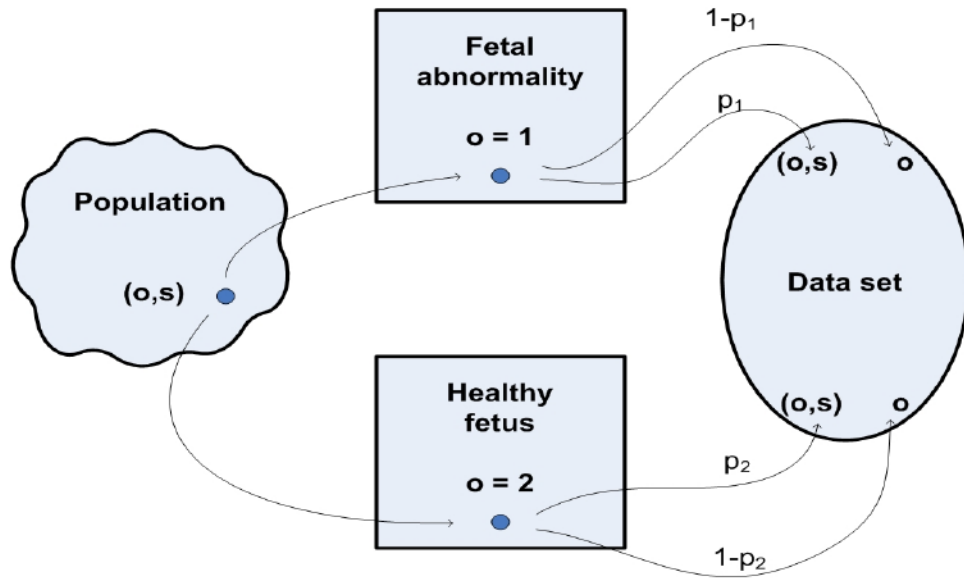
Fig. 2.
Selection model in the subsampling setting.
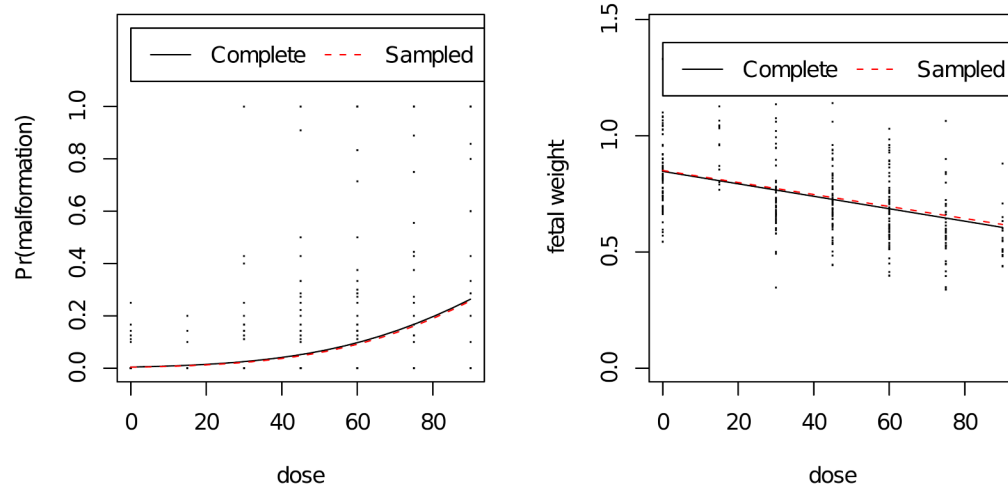
Fig. 3.
Estimated dose response for fetal malformation and mean fetal weight based on fitting the joint model to the complete data set and the sampled data set. Left panel: litter-specific proportion of fetal malformation plotted by dose. Right panel: litter-specific mean fetal weight in grams plotted by dose. Estimate based on the complete data set (MLE) is indicated by the solid line. Estimate based on the sampled data set (PLE) is indicated by the dashed line.

Table 1

Summary of malformation rate and fetal weight observed in C57BL/6 mice in the 2,4,5-T study. All viable litters (litters with at least one live pup) were included. Number of viable pups per litter refers to the average number over all litters at the specified dose. Malformation rate refers to the proportion of malformation over all pups at the given dose. Mean fetal weight per litter (in grams) refers to the mean fetal weight in the litter averaged over all litters at the given dose.

Response	0	15	30	45	60	75	90	Total
				2,4,5-T dose level (mg/kg/day)				
Total litters	71	13	56	75	73	47	32	367
Viable litters	71	13	54	74	69	37	19	337
Viable pups	494	95	364	526	463	247	106	2295
# Viable pups per litter	7.13	7.31	7.2	7.28	6.97	6.83	6.8	
# Malformations	7	3	24	50	69	64	38	
% Malformation	0.014	0.032	0.067	0.095	0.150	0.260	0.373	
Mean fetal weight per litter	0.81	0.94	0.73	0.71	0.67	0.61	0.55	
SD	0.07	0.12	0.06	0.06	0.06	0.06	0.05	

Table 2

Comparison of MLE's and PLE's for the complete data set and the subsampling data set in the 2,4,5-T study. RE SD refers to the SD of the random effect for litter. Fetal weight SD refers to the SD in fetal weight within a litter. The within-fetus correlation refers to the correlation between latent malformation and fetal weight conditional on the litter.

Term	Parameter	complete pairs n= 2295		sampled fetal weight n= 1469	
		Est	SE×10²	Est	Robust SE ×10²
Malformation					
Int	α_{11}	2.626	8.63	2.697	16.31
Dose	α_{12}	-0.0222	0.13	-0.0227	0.25
Fetal weight					
Int	α_{21}	0.8465	0.24	0.850	0.73
Dose	α_{22}	-0.0027	0.005	-0.0026	0.02
SD	σ_2	0.0730	0.06	0.0722	0.24
RE SD_{Y^*}	σ_{b1}	0.4848	2.41	0.5390	8.02
Within-fetus corr	ρ	0.3958	2.65	0.3922	6.52
RE SD_S	σ_{b2}	0.0930	0.10	0.1020	0.29

Table 3

Simulation results: comparison of MLE's and PLE's. Uncorrected estimates refer to parameter values that maximize the PL without accounting for sampling weights. Estimates were averaged over 100 replications.

parameter	true	Complete $(p_1=p_2=1)$ MLE	SRS $(p_1=p_2=0.90)$ MLE	Oversampling $(p_1=0.90, p_2=0.60)$ uncorrected	PLE	Undersampling $(p_1=0.60, p_2=0.90)$ uncorrected	PLE
α_{11}	1.50	1.5005	1.5098	1.1801	1.4933	1.7508	1.4882
α_{12}	-2.00	-2.0077	-2.0167	-2.0980	-1.9886	-1.8391	-1.9892
α_{21}	5.00	4.9934	4.9982	4.8472	4.9875	5.0830	4.9896
α_{22}	-3.00	-2.9950	-2.9969	-3.1008	-2.9785	-2.8262	-2.9861
σ_{b1}	0.33	0.3221	0.3296	0.3320	0.3210	0.3052	0.3190
σ_2	1.00	0.9994	0.9992	1.0039	0.9991	0.9926	0.9987
ρ	0.56	0.5597	0.5615	0.5659	0.5619	0.5543	0.5600
σ_{b2}	0.33	0.3196	0.3224	0.3409	0.3318	0.3131	0.3276

Table 4

Simulation results: comparison of model-based and robust standard error estimates with empirical estimates. Robust standard error estimates are based on bias-corrected PL approach.

parameter	Complete $(p_1=p_2=1)$		SRS $(p_1=p_2=0.90)$		Oversampling $(p_1=0.90, p_2=0.60)$		Undersampling $(p_1=0.60, p_2=0.90)$	
	model SE	emp SE	model SE	emp SE	robust SE	emp SE	robust SE	emp SE
α_{11}	0.092	0.090	0.122	0.143	0.103	0.107	0.117	0.110
α_{12}	0.142	0.140	0.186	0.251	0.167	0.172	0.175	0.171
α_{21}	0.066	0.064	0.080	0.098	0.076	0.082	0.074	0.076
α_{22}	0.107	0.112	0.128	0.181	0.126	0.143	0.123	0.134
σ_b	0.058	0.054	0.089	0.081	0.063	0.063	0.064	0.062
σ_2	0.020	0.020	0.029	0.028	0.023	0.023	0.020	0.023
ρ	0.029	0.027	0.042	0.043	0.032	0.033	0.033	0.032
σ_{b2}	0.040	0.041	0.059	0.057	0.044	0.043	0.042	0.040

Table 5

Simulations results: comparison of uncorrected estimates and PLE's when oversampling over a range (p_1=0.90 and p_2=0.30, 0.40, 0.50).

		Oversampling					
		(p_1=0.9, p_2=0.50)		(p_1=0.90, p_2=0.40)		(p_1=0.90, p_2=0.30)	
parameter	true	uncorr	PLE	uncorr	PLE	uncorr	PLE
α_{11}	1.50	1.0454	1.5093	0.8317	1.5013	0.5771	1.4918
α_{12}	-2.00	-2.1495	-2.0175	-2.1465	-2.0205	-2.1227	-2.0168
α_{21}	5.00	4.7817	5.0024	4.6743	4.9957	4.5195	4.9947
α_{22}	-3.00	-3.1337	-3.0056	-3.1425	-3.0074	-3.0921	-3.0019
σ_{b1}	0.33	0.3281	0.3193	0.3289	0.3170	0.3153	0.3133
σ_2	1.00	1.0013	0.9989	0.9995	0.9965	1.0006	0.9946
ρ	0.56	0.5662	0.5635	0.5689	0.5618	0.5761	0.5649
σ_{b2}	0.33	0.3275	0.3222	0.3217	0.3182	0.3155	0.3213

Table 6

Simulations results: comparison of uncorrected estimates and PLE's when undersampling over a range (p_1=0.30, 0.40, 0.50 and p_2=0.90).

		Undersampling					
		(p_1=0.50, p_2=0.90)		(p_1=0.40, p_2=0.90)		(p_1=0.30, p_2=0.90)	
parameter	true	uncorr	PLE	uncorr	PLE	uncorr	PLE
α_{11}	1.50	1.8865	1.5150	2.0014	1.5260	2.1615	1.5583
α_{12}	-2.00	-1.7850	-2.0008	-1.7083	-2.0261	-1.5962	-2.0676
α_{21}	5.00	5.1232	4.9976	5.1433	5.0008	5.1575	5.0127
α_{22}	-3.00	-2.7662	-2.9895	-2.6768	-3.0000	-2.5467	-3.0159
σ_{b1}	0.33	0.3001	0.3225	0.2896	0.3157	0.2827	0.3165
σ_2	1.00	0.9863	0.9980	0.9820	0.9995	0.9742	0.9976
ρ	0.56	0.5518	0.5643	0.5492	0.5660	0.5411	0.5677
σ_{b2}	0.33	0.3003	0.3231	0.2922	0.3231	0.2787	0.3231

Table 7

Oversampling simulations. Robust SE estimates.

	Oversampling					
	$(p_1=0.9, p_2=0.50)$		$(p_1=0.90, p_2=0.40)$		$(p_1=0.90, p_2=0.30)$	
parameter	robust SE	emp SE	robust SE	emp SE	robust SE	emp SE
α_{11}	0.109	0.106	0.114	0.112	0.126	0.123
α_{12}	0.181	0.196	0.194	0.190	0.223	0.243
α_{21}	0.081	0.077	0.086	0.086	0.098	0.101
α_{22}	0.133	0.152	0.143	0.145	0.163	0.184
σ_{b1}	0.071	0.071	0.073	0.073	0.082	0.077
σ_2	0.025	0.025	0.027	0.028	0.030	0.031
ρ	0.034	0.036	0.036	0.036	0.038	0.039
σ_{b2}	0.048	0.046	0.052	0.052	0.058	0.058

NIH-PA Author Manuscript   NIH-PA Author Manuscript   NIH-PA Author Manuscript

Table 8

Undersampling simulations. Robust SE estimates.

	Undersampling					
	(p_1=0.50, p_2=0.90)		(p_1=0.40, p_2=0.90)		(p_1=0.30, p_2=0.90)	
parameter	robust SE	emp SE	robust SE	emp SE	robust SE	emp SE
α_{11}	0.133	0.131	0.146	0.155	0.170	0.190
α_{12}	0.194	0.205	0.212	0.196	0.244	0.236
α_{21}	0.078	0.079	0.082	0.085	0.090	0.099
α_{22}	0.130	0.152	0.139	0.138	0.155	0.160
σ_{b1}	0.073	0.076	0.079	0.076	0.087	0.080
σ_2	0.023	0.023	0.025	0.025	0.027	0.027
ρ	0.035	0.039	0.038	0.037	0.042	0.040
σ_{b2}	0.045	0.045	0.047	0.050	0.052	0.057