1996

Comparison of quantitative risk assessment procedures based on overdispersed data from developmental toxicity studies.

Leonora. Marro
University of Windsor

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

Comparison of Quantitative Risk Assessment Procedures Based on Overdispersed Data From Developmental Toxicity Studies

by

L. Marro

A Thesis
Submitted to the Faculty of Graduate Studies and Research
Through the Department of Mathematics and Statistics
in Partial Fulfillment
of the Requirements for the Degree of
Master of Science
at the University of Windsor

Windsor, Ontario, Canada
1996

Comparison of Quantitative Risk Assessment Procedures Based on Overdispersed Data From Developmental Toxicity Experiments

by

L. Marro

Abstract

Quantitative risk assessment of noncarcinogens has been based on the estimation of the no-observed-adverse-effect-level (NOAEL). The NOAEL procedure has been shown to be unsatisfactory by many scientists. The fitting of dose-response models to teratology data involving littermates in order to estimate a teratogenic risk is becoming more popular as a potential alternative to the traditional approach to risk assessment. In this thesis a comparison of the different methods of risk assessment (NOAEL, ED_α and BMD_α) are conducted through a simulation study and through real life developmental toxicity data. The estimates ED_α and BMD_α are computed with both the Dirichlet-trinomial and beta-binomial variance function, generalized linear model variance function and with the Rao-Scott transformation.

To my father who first interested me in math

ACKNOWLEDGEMENTS

I would like to thank God for His guidance in all of my endeavours.

I would like to express my gratitude to Dr. Fung and Dr. Hlynka, who have encouraged me to pursue my Masters Degree. I would like to especially thank Dr. Fung for giving me the opportunity to work under her capable supervision. Her guidance and support were greatly appreciated. I would also like to thank Dr. D. Krewski from Health Canada for his helpful criticism throughout this project. Finally I would like to thank Dr. S. Paul and Dr. Y. Park for their time and effort as members on my reasearch committee.

To my family for their unfailing love, support, patience and encouragement. Without them I would not be here today. To my mother Rita, and my sisters Anna and Ivana, I would like to say: thank you for everything, and mostly for believing in me.

And to Nathaniel (NCD) who has been through it all with me.

Contents

Chapter 1

INTRODUCTION

Everyday we are exposed to chemical toxicants that may have adverse effects on our health, for example cigarette smoke, household cleaners, contagion in our drinking water, pesticides used to grow our vegetables, preservatives in our foods... We cannot eliminate some of these chemicals completely such as contagions in our drinking water, but how do we determine what is a safe level of these toxicants? Of particular interest is the effects of such compounds on the unborn, since exposure *in utero* has been shown to be a sensitive pathway to serious health problems (U.S. Environmental Protection Agency (EPA),1991; Holmes, 1992). Chemicals that cause adverse reproductive outcomes through prenatal toxicity to the fetus are referred to as developmental toxicants. Exposure to developmental toxicants may cause a wide variety of abnormal developmental outcomes such as miscarriage, fetal death, reduced birth weight, structural malformations, functional deficits, etc.. Since adequate data resources are rarely available from human epidemiological studies, developmental toxicity experiments in laboratory animals play an important role in testing substances with potential danger to developing fetuses. These

experiments are referred to as "teratology" studies and focus on the period during major organogenesis and structural development.

1.1 Quantitative Risk Assessment Procedures for Developmental Toxicity

Quantitative risk assessment for developmental toxicants is still relatively new and many statistical issues remain unsolved. At present, the risk assessment methodology for developmental toxicants has been derived largely from models previously developed for risk assessment of carcinogens. The most comprehensive risk assessment methodology for teratogens to date has been given by the EPA Guidelines for the Health Assessment of Suspect Developmental Toxicants (EPA, 1986). For the statistical determination of teratogenic risk at exposure levels normally encountered in occupational and environmental settings, two types of risk assessment procedures are usually used: 1) the NOAEL-SF approach and 2) mathematical extrapolation methods such as the benchmark dose (BMD). The NOAEL-SF approach assumes the existence of a threshold for teratogenic action, below which no adverse developmental toxicity will occur. Mathematical extrapolation methods do not presuppose any definitive threshold, but provide the dose level that an acceptably low teratogenic risk will occur.

With respect to noncarcinogens, including those of developmental toxicity, the concept of dose-response modeling is still fairly new and has not been explored as thoroughly. Some work on this include Rodricks et al. (1986), Kimmel (1990), Faustman et al. (1994) and Allen et al. (1994a,b). Models for carcinogenic risk assessment cannot

be directly applied to data from developmental toxicity studies for several reasons as given by Faustman et al. (1989), Kodell et al. (1991) and Krewski and Zhu (1994). Mathematical models for carcinogenic risk assessment exclude the possibility of thresholds below which no carcinogenic risk assessment can occur. Also there is difficulty in such models in handling potential litter effects, which is characteristic of the data from teratological studies. Such litter effects are caused by the tendency for pups in the same litter to respond more similarly than animals from different litters. In addition, since litter size is a random variable, responses might tend to be either increased or decreased depending on the litter size, causing potential litter effect. Intralitter correlation is not encountered in analyzing carcinogenesis bioassay data since litter mates are usually assigned to different treatment groups.

Risk assessment methods for teratogenic hazards include several special features not generally considered in conventional risk assessments. Simple measures of the level and duration of maternal exposure to a teratogenic substance do not necessarily provide reliable predictor of the magnitude of the teratogenic response at target sites within the developing fetuses, because the dose-response relationship is closely related to the stage of development. Thus a single brief exposure may be sufficient to produce adverse developmental effects during a critical period in the stage of development, while doses that produce chronic toxicity in the mother may not produce any adverse teratogenic outcome (WHO, 1984).

Risk assessment for reproductive and developmental effects has traditionally followed the approach of dividing a no-observed-adverse-effect-level (NOAEL) by a

safety factor (SF) in order to determine acceptable low risk dose levels. The NOAEL is the experimental dose level immediately below the lowest dose that produces a statistically significant increase in the rate of adverse effects over the control group. The experimental dose that is chosen as the NOAEL is then divided by a safety factor usually being a factor of 10 (Barnes and Dourson, 1988), to extrapolate the safe dose level from animals to humans. The NOAEL-SF procedure assumes the existence of a threshold for noncarcinogenic toxic effects, below which no adverse developmental toxicity will occur. There are many drawbacks to this method, as will be demonstrated in the section 2.1.

A second approach involves fitting a dose-response model and then applying mathematical extrapolation methods to estimate a virtually safe dose. These methods do not presuppose any definite threshold, but provide the dose level where an acceptably low risk will occur. The most favourable approach to date is the estimation of the benchmark dose (BMD), first proposed by Crump (1984). The BMD approach is commendatory over the conventional NOAEL because it provides an indication of the potential risk associated with exposures near the NOAEL, taking into account both the experimental error and the shape of the dose response curve (Krewski and Zhu, 1994). The U.S. EPA (1991) has recently recommended the BMD method as an alternative or adjunct to the traditional NOAEL for arriving at a safe dose.

Other approaches utilizing dose-response modeling have also been proposed in the literature (e.g., Gaylor, 1983, 1988; Dourson, 1986). Evidence of a dose-response relationship is important and necessary in the assessment of risk. Unfortunately, animal teratogenicity data do not always follow a classical dose-response relationship for

developmental endpoints due to multiplicity of potential endpoints and the effects of competing risks (malformation versus fetal death). In addition, because the endpoints involve continuous measurements, counts and binary data, no single statistical procedure can be recommended to be used routinely for all types of models (IRLG, 1986).

1.2 Developmental Toxicity Experiment

The objective of developmental toxicity studies is to investigate developmental and structural abnormalities caused by potential test agents, with levels of exposure chosen with respect to the intended route(s) to humans (EPA, 1991). Test animals are selected based on consideration of species, strains, age, weight and health status. Two species of rodents (mice and rats) or rabbits are usually chosen as the experimental units because they breed easily and have a short gestation. A typical teratology study is designed with groups of 20 to 30 mated females (dams) exposed to one of 4-5 doses of the toxicant, including an unexposed control, during the critical period of major organogenesis (days 6-16 in rats, 6-15 in mice and 6-19 in rabbits), with the day of insemination considered as day zero of pregnancy (cf. International Life Sciences Institute, 1989). The highest dose in the experiment is chosen to produce no more than 10% mortality (EPA, 1991). The dams are then sacrificed just prior to normal delivery at which time the uterus is removed and the contents are thoroughly examined for multiple endpoints of developmental abnormalities. The usual endpoints of interest include the number of resorptions; the number of late fetal deaths; the number of live offsprings, the number of malformed live births and fetal weight. For live births, specific type(s) of malformations are also

recorded. Figure 1 shows a schematic representation of the observable experimental outcomes having a natural heirarchy (Krewski et al., 1994).

The following notation will be used throughout the thesis. Let n_i be the number of pregnant animals exposed to dose level d_i , (i=1,...,t; d_1=0) in the experiment. Then for the j^{th} litter in the i^{th} dose group we have y_{ij} malformed fetuses (with specific or any type of malformation, depending on the protocol of the study), r_{ij} prenatal deaths (including both resorbed implants and dead pups), s_{ij} live births and m_{ij} implants in each litter. The data can be summarized in the form of multinomial counts (y_{ij}, r_{ij}, s_{ij}, m_{ij}). Without loss of generality we drop the subscripts (ij) for convenience of notation. Note that the number of implants m is the sum of the number of live pups and the number of dead pups in each litter i.e., m=s+r.

Since the dams are exposed to the test agent during day 6 in gestation, the number of implants m is not affected by and does not contain any information about the dose-response relationship (Krewski and Zhu, 1994). Therefore statistical analyses of teratological studies is usually carried out conditional on the observed implantation number m. The response (y,r,s-y), given the number of implants m, is a trinomial count; where each implant in the litter is classified into one of the three mutually exclusive categories, malformed, dead, or normal (Krewski et al., 1994).

1.3 Scope and Organization

In this thesis our main objective is to compare the different risk assessment methods using the traditional NOAEL, separate modeling of the malformation rate y/s, prenatal

death rate r/m and overall toxicity rate $(y+r)/m$, and joint analyses of the endpoints prenatal death and malformation. Estimates of risk calculated from the Rao-Scott transformed data are also compared with those obtained from the original data. In chapter 2 we present these different methodologies that will be used for quantitative risk assessment in this thesis. In chapter 3 statistical models for teratogenic endpoints will be considered. Methods of model fitting for both separate and joint analysis for the toxic endpoints will also be explained, with subsequent estimation of the benchmark dose. The Rao-Scott transformation is discussed for both binary and trinomial outcomes in chapter 4. A large simulation study is conducted to compare the different methods of quantitative risk assessment and its results will be examined in chapter 5. In chapter 6 we apply these same methods for quantitative risk assessment to 28 sets of real life developmental toxicity data. A summary of results and discussion for further research are presented in chapter 7.

Chapter 2

TOXICOLOGICAL RISK ASSESSMENT PROCEDURES FOR DEVELOPMENTAL TOXICITY

2.1 NOAEL-SF Procedure

The present guidelines for arriving at a "safe" dose for humans from developmental toxicity studies is based on a no-observed-adverse-effect-level (NOAEL) divided by a safety factor (SF). NOAEL is the highest exposure level in an experiment producing no statistically significant increase in adverse effects between the dose groups and the control. The usual method of determining the NOAEL for dichotomous data is to perform sequential pairwise comparisons between the proportions of a particular response (e.g. malformation, prenatal death or overall toxicity) in the control group with those of the other dose levels.

Let x_{ij} be the specific toxic endpoint that is observed in an animal that is chosen at random in the j^{th} litter and exposed to dose d_i, (i=1,...,t, d_1=0; j=1,...,n_i). The proportion of pups in the i^{th} dose exhibiting the specified adverse effect is given by

$\hat{p}_i = x_{i.}/m_{i.} = \sum_{j=1}^{n_i} x_{ij} \Big/ \sum_{j=1}^{n_i} m_{ij}$. In order to calculate the NOAEL, hypotheses involving pairwise differences between p_1 and p_i (i=2,...,t) are tested using Fisher's exact test sequentially upward starting from the lowest dose until the proportion of outcomes is significantly different from the control group. The hypotheses to be tested at the i^{th} (i=2,...,t) stage is,

$$H_0^i: p_1 = p_i \text{ versus the alternative } H_a^i: p_1 < p_i, \quad (i=2,...,t).$$

If H_0^i is rejected, then the NOAEL is set to dose level d_{i-1}, otherwise we test H_0^{i+1}. The process continues until the NOAEL is found. If the lowest experimental dose produces a significant teratogenic effect, the NOAEL is not assigned to dose level d_1=0. This lowest dose is defined as the lowest-observable-adverse-effect-level (LOAEL). An additional safety factor of 10 is usually used to calculate the upper exposure limit, if LOAEL is used in place of NOAEL (Kimmel, 1990). Following Leisenring and Ryan (1992), we also let NOAEL equal to $d_2/10$ in this situation.

At present, no quantitative criteria are specified by the EPA for calculation of safety factors for teratogens. The EPA stated that the safety factors themselves are inconsistent, and should be determined on a case-by-case basis, according to scientific judgement. The FDA (Frankos, 1985) addressed this problem for determining safety factors by classifying substances according to the developmental specificity of the teratogenic endpoints and the embryotoxic selectivity of the teratogenic dose-response relationship.

Many scientists have expressed concern that the use of the NOAEL approach has a number of limitations (Gaylor, 1983, 1989; Kimmel and Gaylor, 1988). NOAELs are dependent on: 1) the background incidence of the effect in nonexposed animals, and 2) the spacing of the doses in the experiment (Barnes et al., 1995). The NOAEL is restricted to one of the experimental dose levels. It does not take sample size into consideration, for instance, smaller and less sensitive experiments lead to higher NOAELs than larger studies. The methodology ignores the shape of the dose-response curve, and in general does not make complete use of the available data. For instance the approach does not take into consideration the intra-litter correlation between the pups in a litter that is characteristic of developmental toxicity data. By studying the statistical properties of the NOAEL, Leisenring and Ryan (1992) found that the NOAEL may identify a dose level associated with an unacceptably high risk with a reasonably high probability. Given these limitations associated with the NOAEL, the EPA (1991) is considering adopting an alternative procedure called the benchmark dose (BMD) for deriving an acceptable dose level in developmental toxicity studies.

2.2 Mathematical Modeling Approaches

A quantitative approach to estimating a low dose risk involves fitting a suitable dose-response model to data from a developmental toxicity experiment and providing the dose level where an acceptably low teratogenic risk will occur. This method makes use of all the available data and takes into account the intra-litter correlation present in the data.

A number of approaches utilizing dose-response modeling have been proposed in the literature (e.g., Gaylor, 1983, 1988; Crump, 1984; Dourson, 1986). The most favourable approach to date is the estimation of the benchmark dose (BMD) proposed by Crump (1984). The BMD is defined as the lower 95% confidence limit of the effective dose d_α, that induces α-percent increase in risk compared to background (Dourson, 1986). This effective dose, d_α or ED_α, is defined as the solution to

$$\frac{\pi(d_\alpha) - \pi(0)}{1 - \pi(0)} = \alpha ,$$

where $\pi(d)$ represents the probability of a response at dose d based on an appropriate dose-response model for a particular endpoint. If we solve for d_α in $\pi(d_\alpha) - \pi(0) = \alpha$, then the α–percent refers to excessive risk. In the definition given above for the solution of d_α, the α-percent refers to relative risk, and takes into account the background risk in the absence of exposure i.e., an anomaly occurring spontaneously (Zhu and Fung, 1996). When the background risk is $\pi(0)=0$ then the two measures of risk are equivalent. The benchmark dose is then computed as the lower 95% confidence limit of $\hat{ED}_\alpha$,

$$\hat{BMD}_\alpha = \hat{ED}_\alpha - 1.645\sqrt{\hat{\text{var}}(\hat{ED}_\alpha)} ,$$

where 1.645 corresponds to the upper 95 percentile of the standard normal distribution.

Different mathematical models have been proposed for use in the benchmark dose approach. Theoretically, the choice of the model should not be critical, as long as it fits the data well, since estimation is within the observed dose range for most quantal endpoints. However, factor of covariates relating specifically to developmental toxicity experiments or outcomes (e.g., litter size, litter effect) should be incorporated into the

model as much as possible, to account for variability in the data. Model fitting will be discussed in detail in chapter 3.

2.3 Rao-Scott Transformation

Krewski and Zhu (1995) proposed a simple data transformation for estimating benchmark doses in developmental toxicity experiments. The transformation is based on the concept of generalized design effects due to Rao and Scott (1992) and will be discussed in detail in chapter 4. After the data transformation, standard statistical methods for analysis of uncorrelated multinomial data can be applied. Not only is this method simple to use, but is also robust against the misspecification of the covariance structure for the underlying distribution of the data. In this thesis, we demonstrate that the Rao-Scott transformation gives results that are closely comparable to those obtained from direct analysis of untransformed data both in terms of model fitting and benchmark dose calculation. This agreement indicates that the Rao-Scott transformation gives satisfactory results in developmental toxicity risk assessment in practice.

Chapter 3

STATISTICAL MODELS FOR TERATOGENIC ENDPOINTS

Toxicologists generally agree that the incidence of prenatal death (r) and fetal malformation (y) are both very important endpoints for assessing developmental toxicity (EPA, 1986, 1989; Ryan, 1992) because they measure embryolethality and teratogenicity, respectively. Other variables such as fetal weight, or specific types of malformation are also of interest, but for purposes of this thesis we concentrate on modeling the dose-response relationship and subsequent estimation of BMDs for fetal malformation and prenatal death only.

3.1 Generalized Binomial and Multinomial Models

Traditionally, statistical analysis of developmental toxicity data has been carried out on separate examination of teratogenic endpoints. A number of generalized binomial models have been developed to incorporate the extra-binomial variation due to clustering that is present in the data. A good review of the analysis of proportions with extra-binomial variation can be found in Haseman and Kupper (1979). The beta-binomial model has

been widely used for fitting dose-response relationships to malformation rates (Klienman, 1973; Williams, 1975; Crowder, 1978; Kupper et al., 1986). The superiority of the beta-binomial model for the analysis of proportions has been shown by Paul (1982) and Pack (1986).

The method of generalized estimating equations (GEE) (cf. Liang and Zeger, 1986; Zeger and Liang, 1986) may be used in order to avoid specification of a full likelihood function. Williams (1982) proposed a quasi-likelihood method for extra-binomial variation to fit fetal malformation rate y/s under a linear logistic model. This is a semi-parametric approach based on the first two moments of the beta-binomial distribution. Others have modeled overdispersion by introducing a linear predictor of the dispersion via likelihoods and extended quasi-likelihood models (Efron, 1986; Aitkin, 1987; Nelder and Pregibon, 1987; Smyth, 1989; McCullaugh and Nelder, 1989). With the extended quasi-likelihood approach, the dispersion parameters can be modeled with respect to covariates in a manner similar to that of the mean parameters without the specification of a likelihood function. This will facilitate the simultaneous estimation of the mean and dispersion parameters.

Several authors have discussed the desirability of a joint analysis of the correlated endpoints malformation and prenatal death. Generalized multinomial models provide a method for estimating virtually safe doses with respect to the combined risk of fetal death and malformation. Chen et al. (1991) used a Dirichlet-trinomial distribution to characterize the conditional distribution of the number of prenatal deaths and fetal malformations given the total number of implants. Since the test agent is normally

administered to the dam after mating, the number of implants is not affected. It is treated as an ancillary statistic containing no information about the dose response relationship for developmental toxicity. The Dirichlet-multinomial is a mixture of a multinomial and a Dirichlet distribution. The advantage of using the Dirichlet-multinomial over the multinomial distribution alone is its ability in handling extra-multinomial variation induced by the positive intralitter correlation. The Dirichlet-trinomial distribution can be factored as the product of two beta-binomial distributions, one characterizing fetal malformation conditional on the number of live pups ($y|s$), and the other characterizing prenatal death conditional on the total litter size ($r|m$) (Krewski and Zhu, 1994).

Ryan (1992) used quasi-likelihood to jointly model fetal malformation and prenatal death using extreme-value distribution using a generalized linear model variance-covariance function. The method of moments based on Pearson's χ^2 statistic was used by Williams (1982), Ryan (1992) and Krewski and Zhu (1994) to estimate the overdispersion parameters involved in the variance functions. Quadratic GEEs in conjunction with an extended Dirichlet-trinomial covariance function for joint dose response modeling of correlated trinomial data were used by Zhu et al. (1994) and Krewski and Zhu (1995).

Zhu and Fung (1996) comment on several advantages of GEEs. First, since only the first two moments are required, the method of GEEs is robust against misspecification of the underlying distribution. Second, with the specification of only a joint working covariance function, GEEs will still yield consistent and asymptotically normal estimates under mild regulatory conditions although efficiency may decrease. Zhu et al. (1994)

illustrated that the parameter estimates from GEEs are nearly as efficient as the maximum likelihood estimators based on the Dirichlet-trinomial distribution, although those for the dispersion parameters are less efficient. Third, GEEs are often computationally simpler to implement than maximum likelihood estimation. We will use the generalized estimating equations method for the analysis of correlated data, both as binomial and trinomial counts.

Ryan et al. (1991) and Catalano and Ryan (1994) considered joint analysis of fetal malformation and fetal weight. Catalano et al. (1993) proposed a multivariate dose-response model of prenatal death, fetal weight and malformation. In this chapter we focus on both separate and joint modeling of fetal malformation and prenatal death.

3.2 Dose-Response Models

In this thesis the main objective is to compare benchmark doses from separate modeling of the endpoints malformation and prenatal death with joint modeling, and to compare these values to the traditional NOAEL. The benchmark doses are computed from the original data as well as the Rao-Scott transformed data. Mathematical dose-response models used to describe the dose effect on the probability of a developmental anomaly are of the form

$$\pi_i = F(a + b\eta(d_i)),$$

where F is a smooth cumulative distribution function, and $\eta(d)$ is a monotone function of dose d. Commonly used dose-response models include the probit model $F(x) = \Phi(x)$, linear logisitc model $F(x) = [1 + \exp(-x)]^{-1}$, and the extreme-value model

$F(x) = 1 - \exp[-\exp(x)]$. A more flexible class of models is represented when the function $\eta(d)$ is a power of d_i. For instance $\pi_i = F(a + bd_i^{\gamma})$, where the power γ is estimated along with the linear parameters. This class includes the Weibull model $F(d_i) = 1 - \exp[-(a + bd_i^{\gamma})]$. A special case of the Weibull is the exponential distribution $F(d_i) = 1 - \exp[-(a + bd_i)]$ when γ=1. In this thesis we present results based on the logistic model.

The probability of observing a malformed fetus among s_{ij} live fetuses in the j^{th} litter of the i^{th} dose is given by

$$\pi_{1i} = \pi_1(d_i) = \frac{1}{1 + \exp[-(a_1 + b_1 d_i)]}, \tag{1}$$

the probability of a prenatal death among m_{ij} implants is given by

$$\pi_{2i} = \pi_2(d_i) = \frac{1}{1 + \exp[-(a_2 + b_2 d_i)]}, \quad (i=1,...,t). \tag{2}$$

Thus the probability of observing an adverse effect (either a prenatal death or fetal malformation) is given by

$$\pi_3(d_i) = 1 - [1 - \pi_1(d_i)][1 - \pi_2(d_i)]. \tag{3}$$

This last probability, Eq. (3) gives a measure of overall toxicity. Note that when dose is equal to zero in (1) and (2), we get $\pi_k = [1 + \exp(-a_k)]^{-1}$ (k=1,2). These are the population average probabilities of an anomaly occuring spontaneously in the absence of exposure to the test agent. For model fitting purposes, the two sets of parameters $\theta_k=(a_k,b_k)^T$ (k=1,2) involved in π_k in joint modeling may be regarded as being independent of one another.

3.3 GEE for Separate Modeling

The generalized estimating equations (GEE) approach has gained widespread popularity and includes simpler models of Williams (1982) and McCullagh and Nelder (1989) as special cases. The GEE methodology requires specification of only the mean and the variance function of the data. In separate modeling we model the proportions of prenatal death r/m, fetal malformation y/s, and overall toxicity $(y+r)/m$, separately.

Let x_{ij} be the anomaly of interest (i.e. $x_{ij}=r_{ij}$ if we are interested in modeling prenatal death as an overdispersed binomial outcome, and $x_{ij}=y_{ij}+r_{ij}$ if we model overall toxicity as an overdispersed binomial outcome. Let π_i be the probability of a response occuring at dose level $i=1,...,t$. Assume that the mean and the variance for the number of toxic effects within the same litter are given by

$$E(x|\, m) = m\pi\,,$$

and

$$Var(x|\, m) = m[1 + (m - 1)\phi]\pi(1 - \pi)\,, \tag{4}$$

respectively. The subscripts are dropped for convenience of notation. The variance function given in (4) may be derived directly from the beta-binomial distribution and includes the generalized binomial models of Altham (1978), Kupper and Haseman (1978) and Paul (1987) under proper reparametrization of correlated binary data. The correlation coefficient is restricted to $\phi \geq -(m-1)^{-1}$. If the term $[1 + (m - 1)\phi]$ is replaced by one then the variance function in (4) corresponds to that of a binomial distribution. If the term is replaced by ρ then the variance function in (4) corresponds to the generalized linear model. The GEEs associated with prenatal death and overall toxicity are obtained from

$$\sum_{i=1}^{t}\sum_{j=1}^{n_i} D_i^T W_{ij}^{-1}(x_{ij} - m_{ij}\pi_i) = 0,$$

where $D_i = \partial m_{ij}\pi_i/\partial\theta$, and W_{ij} is the working variance function of x_{ij}. The working variance function is chosen to approximate the actual variance of x_{ij} in order to achieve high efficiency. When the variance function involves additional parameters such as the intralitter correlation coefficient $\{\phi_i\}$, these parameters need to be estimated using another set of equations. We let π_i be the logistic density with $\theta=(a,b)^T$, then the GEE given above reduces to solving the following two equations simultaneously

$$\sum_{i=1}^{t}\sum_{j=1}^{n_i} \frac{(x_{ij} - m_{ij}\pi_i)}{[1 + (m_{ij} - 1)\phi_i]} = 0,$$

and

$$\sum_{i=1}^{t}\sum_{j=1}^{n_i} \frac{d_i(x_{ij} - m_{ij}\pi_i)}{[1 + (m_{ij} - 1)\phi_i]} = 0. \qquad (5)$$

To estimate the dispersion parameters $\{\phi_i\}$ we use the moment based equations

$$\sum_{j=1}^{n_i} \frac{(x_{ij} - m_{ij}\pi_i)^2}{m_{ij}\pi_i(1 - \pi_i)[1 + (m_{ij} - 1)\phi_i]} = n_i\left(1 - \frac{p}{n}\right), \quad i=1,...,t \qquad (6)$$

where $n = \sum_{i=1}^{t} n_i$.

To model malformation as an extra-binomial outcome, assume that the mean and variance for the number of malformed fetuses within the same litter are given by

$$E(y|s) = s\pi$$

and

$$Var(y|s) = s[1+(s-1)\phi]\pi(1-\pi), \qquad (7)$$

respectively (Krewski and Zhu, 1994). The intralitter correlation among live fetuses is restricted to $\phi \geq -(s-1)^{-1}$ (Krewski and Zhu, 1994). Therefore the GEE given for

prenatal death and overall toxicity can be used for separate modeling of *y/s* with m_{ij} replaced by s_{ij} and $x_{ij}=y_{ij}$ to yield the following two equation to be solved simultaneously

$$\sum_{i=1}^{t}\sum_{j=1}^{n_i}\frac{(y_{ij}-s_{ij}\pi_i)}{[1+(s_{ij}-1)\phi_i]}=0,$$

and

$$\sum_{i=1}^{t}\sum_{j=1}^{n_i}\frac{d_i(y_{ij}-s_{ij}\pi_i)}{[1+(s_{ij}-1)\phi_i]}=0. \qquad (8)$$

To estimate the dispersion parameters $\{\phi_i\}$ for the anomaly of malformation we replace m_{ij} by s_{ij} and $x_{ij}=y_{ij}$ in Eq. (6), and thus the moment-based equations for the correlation coefficients among live animals are given by

$$\sum_{j=1}^{n_i}\frac{(y_{ij}-s_{ij}\pi_i)^2}{s_{ij}\pi_i(1-\pi_i)[1+(s_{ij}-1)\phi_i]}=n_i\left(1-\frac{p}{n}\right),\ i=1,\ldots,t. \qquad (9)$$

The estimates of $\theta=(a,\ b)^T$ and $\{\phi_i\}$ in the case of prenatal death or overall toxicity are obtained by iteratively solving equations (5) and (6) until convergence. At each iteration, equations (5) is solved for θ with $\{\phi_i\}$ fixed at $\{\hat{\phi}_i\}$; and equations (6) are solved for $\{\phi_i\}$ with θ fixed at $\hat{\theta}$. Similarily for the occurrence of fetal malformation the estimate of θ and $\{\phi_i\}$ are obtained by iteratively solving equations (8) and (9) until convergence.

Under mild regulatory conditions, parameter estimates of θ based on GEE are consistent and normally distributed as each $n_i \to \infty$. The "naive" covariance matrix of the estimates of θ, assuming that the working variance function is correctly specified, is given by

$$\Sigma_N = \left[\sum_{i=1}^{I} \sum_{j=1}^{n_i} D_i^T W_{ij}^{-1} D_i \right]^{-1}.$$

To guard against misspecification of the variance function of $(x_{ij}|m_{ij})$ or $(y_{ij}|s_{ij})$, robust empirical "sandwich" estimate

$$\Sigma_E = \Sigma_N \Sigma_M \Sigma_N , \tag{10}$$

proposed by Liang and Zeger (1986) are used where

$$\Sigma_M = \sum_{i=1}^{I} \sum_{j=1}^{n_i} D_i^T W_{ij}^{-1} (x_{ij} - m_{ij}\pi_i)(x_{ij} - m_{ij}\pi_i)^T W_{ij}^{-1} D_i .$$

Again, in separate modeling of fetal malformation one would replace m_{ij} by s_{ij}, and $x_{ij}=y_{ij}$.

3.4 GEE for Joint Modeling

The trinomial count of malformation and prenatal death (y,r) given litter size m is generally overdispersed relative to the standard trinomial distribution. It is very important to characterize this overdispersion in order to make any valid statistical inferences. The Dirichlet-multinomial distribution (cf. Johnson and Kotz, 1969) may be used to describe the extra-multinomial variation exhibited in the data from developmental toxicity studies. Again, the specification of a Dirichlet-multinomial distribution for fetal malformation and prenatal death given the litter size can be avoided if we use the generalized estimating equations (GEE) approach as suggested by Zhu et al. (1994) and Krewski and Zhu (1994). GEEs only require the first two moments in the data.

Let π_1 be the probability of a malformation in a live fetus, π_2 be the probability of a prenatal death, and ϕ be the intralitter correlation coefficient among implants. Letting μ_1

be the probability of a malformed fetus given the number of implants, we have that $\mu_1=\pi_1(1-\pi_2)$, and the probability of a prenatal death is $\mu_2=\pi_2$. Letting $z=(y,r)^T$ be the vector of responses, we have the mean and covariance matrix of z are:

$$E(z|m) = m\mu = m\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix} = m\begin{bmatrix}\pi_1(1-\pi_2)\\ \pi_2\end{bmatrix},$$

$$\begin{aligned} Cov(z|m) &= m[1+(m-1)\phi]\begin{bmatrix}\mu_1(1-\mu_1) & -\mu_1\mu_2\\ -\mu_1\mu_2 & \mu_2(1-\mu_2)\end{bmatrix} \\ &= m[1+(m-1)\phi](1-\pi_2)\begin{bmatrix}\pi_1[1-\pi_1(1-\pi_2)] & -\pi_1\pi_2\\ -\pi_1\pi_2 & \pi_2\end{bmatrix}, \end{aligned} \tag{11}$$

respectively (Zhu et al., 1994). The covariance function in (11) corresponds to that of a trinomial distribution when $\phi = 0$, and to a Dirichlet-trinomial distribution when $0 \le \phi \le 1$. The Dirichlet-trinomial covariance function may not be able to capture extreme overdispersion because of the upper bound $\phi \le 1$. Furthermore if we allow

$$0 > \phi \ge \max\left\{\frac{-\mu_i}{m-1-\mu_i}; i = 1,2,3; m \ge 2\right\}$$

where $\mu_3=1-\mu_1-\mu_2$ (probability of being alive and normal), the covariance function corresponds to that of an extended Dirichlet-trinomial distribution (Krewski and Zhu, 1994). We permit $-(m-1)^{-1} < \phi < 0$ in order to allow for moderate underdispersion, since this is the maximum lower value that the parameter can take on.

The overdispersion parameter is likely to increase with dose (Williams, 1988; Krewski and Zhu, 1994). Some authors (Kupper et al., 1986; Zhu et al., 1994; Krewski and Zhu, 1994) have suggested assigning a unique value for ϕ in each dose group, while others (Prentice, 1986; Moore, 1987; Williams, 1988) have indicated to possibly model

the intralitter correlation ϕ as a smooth function of dose. This latter choice seems difficult as the dependence of ϕ on dose differs substantially from one experiment to another. Chen and Li (1994), Zhu et al. (1994) and Krewski and Zhu (1994) use distinct values of ϕ for each dose group in order to allow for dose dependence.

If we allow $[1+(m-1)\phi]=\rho$, where ρ varies in each dose group and is independent of m then we have Ryan's (1993) model. The dispersion parameters $\{\phi_i ; i=1,2,...,t\}$ are estimated using Pearson's χ^2 squared residuals jointly with the GEEs for the regression parameters $\theta_k=(a_k, b_k)^T$ involved in π_k (k=1,2).

Let $z_{ij}=(y_{ij},r_{ij})^T$ denote the vector of outcomes for the j^{th} litter in the i^{th} dose (i=1,...,t; j=1,...,n_i). Given the parametric models for $\mu(d_i,\theta) = \mu_i = (\mu_{i1}, \mu_{i2})$, the GEE for estimating the unknown parameters $\theta = (\theta_1^T,\theta_2^T)^T$ are given by

$$\sum_{i=1}^{t}\sum_{j=1}^{n_i} m_{ij} D_i^T W_{ij}^{-1}(z_{ij} - m_{ij}\mu_i) = 0, \qquad (12)$$

where $D_i = \partial\mu_i/\partial\theta^T$ and W_{ij} is the weight or "working covariance" matrices. We let W_{ij} equal the Dirichlet-trinomial covariance function for the working covariance of z_{ij}, given in (11). Two separate sets of estimating equations result with respect to θ_1 and θ_2 from Eq. (12),

$$\sum_{i=1}^{t}\sum_{j=1}^{n_i} \frac{1}{1+(m_{ij}-1)\phi_i} \frac{1}{\pi_{1i}(1-\pi_{1i})} \frac{\partial\pi_{1i}}{\partial\theta_1}(y_{ij} - s_{ij}\pi_{1i}) = 0, \qquad (13)$$

and

$$\sum_{i=1}^{t}\sum_{j=1}^{n_i} \frac{1}{1+(m_{ij}-1)\phi_i} \frac{1}{\pi_{2i}(1-\pi_{2i})} \frac{\partial\pi_{2i}}{\partial\theta_2}(r_{ij} - m_{ij}\pi_{2i}) = 0. \qquad (14)$$

To estimate the intralitter correlation coefficients $\{\phi_i\}$, the multivariate Pearson squared residuals $(z_{ij} - m_{ij}\mu_i)^T W_{ij}^{-1}(z_{ij} - m_{ij}\mu_i)$ are used with the factor $1 - p(2n)^{-1}$, to adjust for the simultaneous estimation of p regression parameters in the dose response model (Krewski and Zhu, 1994)

$$\sum_{j=1}^{n_i}(z_{ij} - m_{ij}\mu_i)^T W_{ij}^{-1}(z_{ij} - m_{ij}\mu_i) = 2n_i\left(1 - \frac{p}{2n}\right), \quad i=1,...,t \tag{15}$$

The parameter estimates $\hat{\theta}_1$, $\hat{\theta}_2$ and $\{\hat{\phi}_i\}$ are obtained by iteratively solving equations (13), (14) and (15) until convergence. At each iteration, equations (13) and (14) are solved for θ_1 and θ_2 respectively with $\{\phi_i\}$ fixed at $\{\hat{\phi}_i\}$; and equations (15) are solved for $\{\phi_i\}$ with θ_1 and θ_2 fixed at $\hat{\theta}_1$ and $\hat{\theta}_2$, respectively. Note that the moment estimates of $\{\phi_i\}$ are biased under incorrect specification of the variance function. In this event, variance estimators for the estimates of the dose-response model parameters are inconsistent.

The robust "sandwich" estimate of the covariance matrix for the estimate of θ, is given by

$$\Sigma_E = \Sigma_N \Sigma_M \Sigma_N \tag{16}$$

(Liang and Zeger, 1986), where

$$\Sigma_N = \left[\sum_{i=1}^{t}\sum_{j=1}^{n_i} m_{ij}^2 D_i^T W_{ij}^{-1} D_i\right]^{-1},$$

and $$\Sigma_M = \sum_{i=1}^{t}\sum_{j=1}^{n_i} m_{ij}^2 D_i^T W_{ij}^{-1}(z_{ij} - m_{ij}\mu_i)(z_{ij} - m_{ij}\mu_i)^T W_{ij}^{-1} D_i .$$

Both Σ_N and Σ_M are evaluated at $\hat{\theta}$ and $\{\hat{\phi}_i\}$. Under the Dirichlet-trinomial covariance function and when the observed moments in Σ_M are replaced by their expectation, we have that $\Sigma_N^{-1} = \Sigma_M$ (Krewski et al., 1994). In this case the asymptotic covariance for $\hat{\theta}$ reduces to Σ_N, which is block diagonal with the two diagonal blocks corresponding to the variances of $\hat{\theta}_1$ and $\hat{\theta}_2$, respectively. As long as the working covariance matrix is reasonable for the data, the GEE approach will yield highly efficient estimates of the parameters characterizing the mean response rates, and use of the empirical formula given in (16), will ensure validity of resulting variance estimators.

3.5 Estimates of ED_α and BMD_α for Separate and Joint Models

Crump (1984) and Chen and Kodell (1989) discussed the BMD methodology based on dose-response models for a single endpoint. Allen et. al (1994b) estimated BMDs using several dose-response models fitted to data from a large database and found that BMDs at the 5% level are similar to NOAEL in magnitude on the average. Ryan (1992) and Krewski and Zhu (1994, 1995) used joint dose response models to estimate BMDs.

Under the logistic dose response model defined in Eq. (1) or (2), $\hat{ED}_\alpha$ for both joint and separate models, is given by

$$d_\alpha = \frac{-1}{b_k} \ln\left[\frac{1-\alpha}{1+\alpha e^{-a_k}}\right], \qquad (k=1,2,3) \qquad (17)$$

evaluated at the parameter estimates $(\hat{a}_k, \hat{b}_k)$. The subscript k refers to the endpoint of interest. When k=1 we use the estimate of $\theta_1=(a_1,b_1)^T$ from fitting the logistic dose-response model to the malformed data, evaluted from Eq. (8) for separate modeling and Eq. (13) in joint modeling. When k=2 we use the estimate of $\theta_2=(a_2,b_2)^T$ from fitting the logistic dose-response model to the prenatal death data, evaluted from Eq. (5) for separate modeling and Eq. (14) in joint modeling. Lastly, when k=3 we use the estimates of $\theta_3=(a_3,b_3)^T$ from fitting the logistic dose-response model to the overall toxicity data, evaluated from Eq. (5) in separate modeling. Estimates of $E\hat{D}_\alpha$ for overall toxicity in joint models will be discussed later.

To form a lower confidence limit, we need the variance of $E\hat{D}_\alpha$. The estimate of the variance of $E\hat{D}_\alpha$ calculated from Eq. (17) is based on the δ method and is given by

$$\text{var}(\hat{d}_\alpha) = \left(\frac{\partial d_\alpha}{\partial a_k}\right)^2 \text{var}(\hat{a}_k) + \left(\frac{\partial d_\alpha}{\partial b_k}\right)^2 \text{var}(\hat{b}_k) + 2\frac{\partial d_\alpha}{\partial a_k}\frac{\partial d_\alpha}{\partial b_k}\text{cov}(\hat{a}_k, \hat{b}_k)$$

$$= \left(\frac{\alpha e^{-\hat{a}_k}}{\hat{b}_k(1+\alpha e^{-\hat{a}_k})}\right)^2 \text{var}(\hat{a}_k) + \left(\frac{E\hat{D}_\alpha}{\hat{b}_k}\right)^2 \text{var}(\hat{b}_k) + 2\frac{E\hat{D}_\alpha}{\hat{b}_k}\frac{\alpha e^{-\hat{a}_k}}{\hat{b}_k(1+\alpha e^{-\hat{a}_k})}\text{cov}(\hat{a}_k, \hat{b}_k). \quad (18)$$

The ED_α for overall toxicity based on the trinomial model π_3=1-(1-π_1)(1-π_1) given in (3) is obtained as the solution to the equation

$$(1-\alpha)(1+e^{a_1+b_1 d_\alpha})(1+e^{a_2+b_2 d_\alpha}) = (1+e^{a_1})(1+e^{a_2}), \quad (19)$$

evaluated at the parameter estimates $(\hat{a}_1, \hat{b}_1, \hat{a}_2, \hat{b}_2)$, determined by equations (13) and (14). The variance of $E\hat{D}_\alpha$ for overall toxicity based on joint modeling is also found

using the δ method. We implicitly differentiate Eq. (19) with respect to each parameter (a_1,b_1,a_2,b_2) and then use the δ method to obtain

$$\begin{aligned}\mathrm{var}(\hat{d}_\alpha) = &\left(\frac{\partial d_\alpha}{\partial a_1}\right)^2 \mathrm{var}(\hat{a}_1) + \left(\frac{\partial d_\alpha}{\partial b_1}\right)^2 \mathrm{var}(\hat{b}_1) + \left(\frac{\partial d_\alpha}{\partial a_2}\right)^2 \mathrm{var}(\hat{a}_2) + \left(\frac{\partial d_\alpha}{\partial b_2}\right)^2 \mathrm{var}(\hat{b}_2) \\ &+2\frac{\partial d_\alpha}{\partial a_1}\frac{\partial d_\alpha}{\partial b_1}\mathrm{cov}(\hat{a}_1,\hat{b}_1) + 2\frac{\partial d_\alpha}{\partial a_1}\frac{\partial d_\alpha}{\partial a_2}\mathrm{cov}(\hat{a}_1,\hat{a}_2) + 2\frac{\partial d_\alpha}{\partial a_1}\frac{\partial d_\alpha}{\partial b_2}\mathrm{cov}(\hat{a}_1,\hat{b}_2) \\ &+2\frac{\partial d_\alpha}{\partial b_1}\frac{\partial d_\alpha}{\partial a_2}\mathrm{cov}(\hat{b}_1,\hat{a}_2) + 2\frac{\partial d_\alpha}{\partial b_1}\frac{\partial d_\alpha}{\partial b_2}\mathrm{cov}(\hat{b}_1,\hat{b}_2) + 2\frac{\partial d_\alpha}{\partial a_2}\frac{\partial d_\alpha}{\partial b_2}\mathrm{cov}(\hat{a}_2,\hat{b}_2),\end{aligned} \tag{20}$$

where $\dfrac{\partial d_\alpha}{\partial a_1} = \dfrac{e^{\hat{a}_1}(1+e^{\hat{a}_2}) - (1-\alpha)e^{\hat{\kappa}_1}(1+e^{\hat{\kappa}_2})}{c}$,

$$\frac{\partial d_\alpha}{\partial b_1} = \frac{-E\hat{D}_\alpha(1-\alpha)e^{\hat{\kappa}_1}(1+e^{\hat{\kappa}_2})}{c},$$

$$\frac{\partial d_\alpha}{\partial a_2} = \frac{(1+e^{\hat{a}_1})e^{\hat{a}_2} - (1-\alpha)(1+e^{\hat{\kappa}_1})e^{\hat{\kappa}_2}}{c},$$

$$\frac{\partial d_\alpha}{\partial b_2} = \frac{-E\hat{D}_\alpha(1-\alpha)(1+e^{\hat{\kappa}_1})e^{\hat{\kappa}_2}}{c},$$

$$c = (1-\alpha)[\hat{b}_1 e^{\hat{\kappa}_1}(1+e^{\hat{\kappa}_2}) + \hat{b}_2(1+e^{\hat{\kappa}_1})e^{\hat{\kappa}_2}],$$

$$\hat{\kappa}_k = \hat{a}_k + \hat{b}_k E\hat{D}_\alpha, \quad (k=1,2).$$

The benchmark dose is then computed as the lower 95% confidence limit of $E\hat{D}_\alpha$,

$$B\hat{M}D_\alpha = E\hat{D}_\alpha - 1.645\sqrt{\hat{\mathrm{var}}(E\hat{D}_\alpha)}, \tag{21}$$

where 1.645 corresponds to the upper 95 percentile of the standard normal distribution.

Chapter 4

THE RAO-SCOTT TRANSFORMATION

Rao and Scott (1992) introduced a transformation to eliminate the extra-binomial variation present in the data, hence permitting the use of methods for analysis for binomial data. The idea is to calculate a design effect (inflation factor) which is the ratio of the variance of the data under cluster sampling (which induces intra-cluster correlation) to the variance of the same statistic under simple random sampling where the observations are independent. The Rao-Scott transformation makes no specific assumptions about the distribution of the observations.

4.1 Rao-Scott Transformation for Extra-binomial Outcomes

Analyzing the data as binary responses, the following Rao-Scott transformation adopted from Fung et al. (1994) is used. Let p_i be the probability of a toxic response in the i^{th} treatment group and x_{ij} be the binary response (i.e., $x_{ij}=r_{ij}$ for prenatal death and $x_{ij}=y_{ij}+r_{ij}$ for overall toxicity). Then the design effect D_i is estimated by

$$\hat{D}_i = \hat{v}_i m_i [\hat{p}_i(1-\hat{p}_i)]^{-1}, \tag{22}$$

where $\hat{p}_i = x_{i.}/m_{i.}$, $x_{i.} = \sum_{j=1}^{n_i} x_{ij}$, $m_{i.} = \sum_{j=1}^{n_i} m_{ij}$ and

$$\hat{v}_i = \frac{n_i}{(n_i - 1)m_{i.}^2} \sum_{j=1}^{n_i} (x_{ij} - m_{ij}\hat{p}_i)^2 .$$

The probability of a fetal malformation is estimated by $\hat{p}_{i_y} = y_{i.}/s_{i.}$, where the subscript y refers to the fetal malformation outcome. The variance estimator is given by

$$\hat{v}_{i_y} = \frac{n_i}{(n_i - 1)s_{i.}^2} \sum_{j=1}^{n_i} (y_{ij} - s_{ij}\hat{p}_{i_y})^2 .$$

Hence the estimated design effect for fetal malformation is

$$\hat{D}_{i_y} = \hat{v}_{i_y} s_{i.} [\hat{p}_{i_y}(1 - \hat{p}_{i_y})]^{-1}, \tag{23}$$

where $s_{i.} = \sum_{j=1}^{n_i} s_{ij}$ and $y_{i.} = \sum_{j=1}^{n_i} y_{ij}$. The transformed data is then computed as $(x_{ij}, m_{ij}) \rightarrow (\tilde{x}_{ij}, \tilde{m}_{ij})$, where $\tilde{x}_{ij} = x_{ij}/\hat{D}_i$, and $\tilde{m}_{ij} = m_{ij}/\hat{D}_i$, for prenatal death or overall toxicity. The transformed data for malformation is $(y_{ij}/\hat{D}_{i_y}, s_{ij}/\hat{D}_{i_y}) = (\tilde{y}_{ij}, \tilde{s}_{ij})$. The transformed data are then treated as binomial random variables.

The design effect given in Eq. 22 (or 23 for malformation) represents the ratio of the estimated variance $\hat{v}_i$ (or $\hat{v}_{i_y}$) to the corresponding estimate of the variance for proportions under the assumption of no extra-binomial variation (i.e. $\hat{p}_i(1 - \hat{p}_i)/m_{i.}$ for prenatal death or overall toxicity, and $\hat{p}_{i_y}(1 - \hat{p}_{i_y})/s_{i.}$ for malformation). Correspondingly, values of $\hat{D}_i$ (or $\hat{D}_{i_y}$) less than or greater than unity indicate underdispersion and overdispersion, respectively, relative to the binomial variation.

4.2 Rao-Scott Transformation For Extra-Trinomial Counts

The Rao-Scott transformation procedure has been extended to trinomial counts (Krewski and Zhu, 1995; Zhu and Fung, 1996). Employing the same notation as before, let $z_{ij}=(y_{ij}, r_{ij})^T$, be the vector of responses given the number of implants m_{ij} from the j^{th} litter in the i^{th} dose group (i=1,2,...,t; j=1,2,...,n_i). Let μ_{1i} be the response rate of fetal malformation and μ_{2i} be the response rate for prenatal death. Then the expected number of malformations and prenatal deaths is given by

$$E(z_{ij}|m_{ij}) = m_{ij}(\mu_{1i}, \mu_{2i})^T = m_{ij}\mu_i .$$

In addition, let π_{1i} denote the conditional probability of observing a malformation in a live pup, and π_{2i} the probability of a death occuring. Then we have

$$\pi_{2i} = \mu_{2i},$$

and

$$\pi_{1i} = \mu_{1i}/(1-\pi_{2i}) .$$

Since π_{1i} is the conditional probability of a malformation occuring in a live pup, we have the following relation

$$E(y_{ij}|s_{ij}, m_{ij}) = s_{ij}\pi_{1i} .$$

The probabilities π_{1i} and π_{2i} model the incidence of fetal malformation (y/s) and prenatal death (r/m), respectively. The probability of either a death or a malformation

$$\pi_{3i} = 1-(1-\pi_{1i})(1-\pi_{2i}),$$

provides a measure of overall toxicity.

For the i^{th} dose group, consistent estimates of the response rates μ_{ki} (k=1,2,3) are given by the sample proportions

$$\hat{\mu}_{1i} = \frac{y_{i.}}{m_{i.}},$$

$$\hat{\mu}_{2i} = \frac{r_{i.}}{m_{i.}}$$

and $$\hat{\mu}_{3i} = 1 - \hat{\mu}_{1i} - \hat{\mu}_{2i},$$

where $y_{i.}$, $r_{i.}$ and $m_{i.}$ are defined in section 4.1. In the present notation we have $z_{i.}=(y_{i.},r_{i.})$, with $E(z_{i.}|m_{i.}) = m_{i.}\mu_i$, and $Var(z_{i.}) = m_{i.}V(\mu_i)$. The design effect is the average of the generalized design effect

$$D_i = \frac{1}{2} trace\left[V_0^{-1}(\mu_i)V(\mu_i)\right] = \frac{1}{2}\left[\frac{v_{i11}}{\mu_{1i}} + \frac{v_{i22}}{\mu_{2i}} + \frac{v_{i33}}{\mu_{3i}}\right], \tag{24}$$

where $V_0(\mu_i) = diag(\mu_i) - \mu_i\mu_i^T$, and v_{ikk} (k=1,2,3) are the sample variances for the components of the group total counts for malformed, dead or resorbed, and alive and normal ($y_{i.}$,$r_{i.}$,$s_{i.}$-$y_{i.}$) given the total number of implants $m_{i.}$ per dose group. To compute $\hat{D}_i$, we replace the parameters by their sample estimates, $\hat{\mu}_{1i}$, $\hat{\mu}_{2i}$ and $\hat{\mu}_{3i}$ along with

$$\hat{v}_{i11} = \frac{n_i}{(n_i - 1)m_{i.}} \sum_{j=1}^{n_i} (y_{ij} - m_{ij}\hat{\mu}_{1i})^2$$

$$\hat{v}_{i22} = \frac{n_i}{(n_i - 1)m_{i.}} \sum_{j=1}^{n_i} (r_{ij} - m_{ij}\hat{\mu}_{2i})^2$$

$$\hat{v}_{i33} = \frac{n_i}{(n_i - 1)m_{i.}} \sum_{j=1}^{n_i} ((s_{ij} - y_{ij}) - m_{ij}\hat{\mu}_{3i})^2 .$$

Note that in the estimated variance of being alive and normal, if we substitute s_{ij}=m_{ij}-r_{ij}, and $\hat{\mu}_{3i} = 1 - \hat{\mu}_{1i} - \hat{\mu}_{2i}$ we get

$$\hat{v}_{i33} = \frac{n_i}{(n_i - 1)m_{i.}} \sum_{j=1}^{n_i} ((r_{ij} + y_{ij}) - m_{ij}(\hat{\mu}_{1i} + \hat{\mu}_{2i}))^2$$

which is an estimate for the variance of overall toxicity. A consistent estimate of the average value of the generalized design effect is given by

$$\hat{D}_i = \frac{1}{2}\left[\frac{\hat{v}_{i11}}{\hat{\mu}_{1i}} + \frac{\hat{v}_{i22}}{\hat{\mu}_{2i}} + \frac{\hat{v}_{i33}}{\hat{\mu}_{3i}}\right], \quad (i=1,...,t). \tag{25}$$

The transformed data are then computed by dividing the trinomial responses in the i^{th} dose group by $\hat{D}_i$. The new data appears as follows

$$\tilde{z}_{ij} = z_{ij}/\hat{D}_i \text{, where } \tilde{z}_{ij} = (\tilde{y}_{ij}, \tilde{r}_{ij})^T = \left(\frac{y_{ij}}{\hat{D}_i}, \frac{r_{ij}}{\hat{D}_i}\right)^T,$$

$$\tilde{m}_{ij} = m_{ij}/\hat{D}_i$$

and $$\tilde{s}_{ij} = \tilde{m}_{ij} - \tilde{r}_{ij}, \qquad (i=1,...,t; j=1,...,n_i).$$

The design effect $\hat{D}_i$ in (25) measures the degree inflation of variation in the data relative to the trinomial variation. Correspondingly values of $\hat{D}_i$ less than or greater than unity indicate underdispersion and overdispersion, respectively relative to the trinomial variation. The Rao-Scott transformation has many advantages over direct analysis of the original correlatd data (Krewski, Zhu and Fung, 1994; Krewski and Zhu, 1995). Since the transformation does not require specification of the covariance function of the response (y_{ij}, r_{ij}) given m_{ij}, it may enjoy certain degrees of robustness. As well, after applying the transformation, standard procedures associated with the trinomial distribution can be utilized. Hence the analysis of the transformed data is computationally simpler than the original data.

We can fit any of the dose-response models mentioned in section 3.2, or any other π_1 and π_2 with independent parameters θ_1 and θ_2 to the transformed data, where π_1 measures the probability of a malformation in a live pup, π_2 measure the probability of a prenatal death, and $\pi_3=1-(1-\pi_1)(1-\pi_2)$ gives a measure of overall toxicity. For dose reponse model fitting, we set all intralitter correlation coefficients $\{\phi_i\}$ equal to zero in equations (5) and (8) for separate modeling, and equations (13) and (14) for joint modeling. To calculate the variance of the parameter estimates, we also set $\phi_i=0$ in equations (10) and (16). After the data are transformed and model fitting is performed on the transformed data, then equations (17) through to (21) can be used for estimation of ED_α and BMD_α. Krewski and Zhu (1995) illustrate that the dose response models fitted to the transformed data were practically equivalent to models fitted to the original data using the Dirichlet-trinomial covariance function.

Chapter 5

SIMULATION STUDY

5.1 The Simulated Population

Developmental toxicity bioassay data were simulated in a setting similar to the experiments conducted under the U.S. National Toxicology Program. The one we selected here was analyzed by Krewski et al.(1994) involving female rats exposed to Ethylene Glycol (EG), at dose levels 0, 1250, 2500, 5000 (mg/kg/day) (Table 1). Within each dose group, the number of prenatal death (r_{ij}), the number of occurrences of any type of malformation (y_{ij}), and the number of implants (m_{ij}) from each animal were available.

In our simulations, the counts $(y,r|m)$ are assumed to follow a Dirichlet-trinomial distribution which can be expressed as a product of two beta-binomial distributions

$$P(y,r|m;\pi_1,\pi_2,\phi) = P(r|m;\ \pi_2,\phi)\ P(y|s;\pi_1,\rho)$$

$$= BB(a_2,b_2)\ BB(a_1,b_1)$$

where $s=m-r$ is the number of live fetuses,

π_1 is the probability of malformation in a live fetus,

π_2 is the probability of a prenatal death,

a_1, b_1, a_2, b_2 are the parameters of the beta-binomial distributions,

ϕ is the intra-litter correlation coefficient among implants, and

$\rho=\phi(1-\pi_2(1-\phi))^{-1}$ is the intra-litter correlation among live fetuses.

The probability of the dose response is expressed as a function of dose. For the above data set, Krewski et al. (1994) fitted Weibull dose-response models. We fitted the logistic model with

$$\pi_{1i} = \frac{1}{1+\exp(-a_1 - b_1 d_i)},$$

$$\pi_{2i} = \frac{1}{1+\exp(-a_2 - b_2 d_i)}.$$

The subscript i refers to the dose group i=1,...,t. Chi-square goodness-of-fit tests were performed on these two models. Chi-square goodness of fit test for the malformed and prenatal death data, with expected value and variance given in (11) are

$$\sum_{i=1}^{t}\sum_{j=1}^{n_i} \frac{[y_{ij} - m_{ij}\pi_{1i}(1-\pi_{2i})]^2}{\pi_{1i}(1-\pi_{2i})(1-\pi_{1i}(1-\pi_{2i}))m_{ij}[1+(m_{ij}-1)\phi_i]},$$

and

$$\sum_{i=1}^{t}\sum_{j=1}^{n_i} \frac{[r_{ij} - m_{ij}\pi_{2i}]^2}{\pi_{2i}(1-\pi_{2i})m_{ij}[1+(m_{ij}-1)\phi_i]},$$

respectively. Each has a χ^2 distribution with degrees of freedom (n-p), where $n = \sum_{i=1}^{t} n_i$, and p is the number of mean parameters to be estimated in the above test statistics. Results show that the logistic model (χ^2_{108}=122.24, P-value=0.165) seems to fit the malformed data better than the Weibull model (χ^2_{106}=126.31, P-value=0.087), while the fit is about the same for the prenatal death data (for the logisitc model χ^2_{110}=89.93, P-

value=0.919, and for the Weibull model χ^2_{109}=84.42, P-value=0.961). Hence, we will choose the logistic model and use the estimates a_1= -3.87, b_1= 5.02, a_2= -3.12, b_2= 1.63 as the parameters in generating our simulated data. The corresponding values of $\{\phi_i\}$ for the four dose groups are (0.04590, 0.08112, 0.16420, 0.48756).

This data set was tested for trend using the following trend test based on generalized score functions (Zhu and Fung, 1996):

$$T_2 = \frac{\left\{\sum_{i=1}^{I}\sum_{j=1}^{n_i} h_{ij}(d_i-\bar{d})(y_{ij}-s_{ij}\hat{\pi}_{10})\right\}^2}{\sum_{i=1}^{I}\sum_{j=1}^{n_i} h_{ij}m_{ij}(d_i-\bar{d})^2\hat{\pi}_{10}(1-\hat{\pi}_{10})(1-\hat{\pi}_{20})} + \frac{\left\{\sum_{i=1}^{I}\sum_{j=1}^{n_i} h_{ij}(d_i-\bar{d})(r_{ij}-m_{ij}\pi_{20})\right\}^2}{\sum_{i=1}^{I}\sum_{j=1}^{n_i} h_{ij}m_{ij}(d_i-\bar{d})^2\hat{\pi}_{20}(1-\hat{\pi}_{20})}, \tag{26}$$

where $\bar{d} = \frac{\sum_{i=1}^{I}\sum_{j=1}^{n_i} m_{ij}h_{ij}d_i}{\sum_{i=1}^{I}\sum_{j=1}^{n_i} m_{ij}h_{ij}}$, $\hat{\pi}_{10} = \frac{\sum_{i=1}^{I}\sum_{j=1}^{n_i} h_{ij}y_{ij}}{\sum_{i=1}^{I}\sum_{j=1}^{n_i} h_{ij}s_{ij}}$ and $\hat{\pi}_{20} = \frac{\sum_{i=1}^{I}\sum_{j=1}^{n_i} h_{ij}r_{ij}}{\sum_{i=1}^{I}\sum_{j=1}^{n_i} h_{ij}m_{ij}}$.

The term h_{ij} is any function chosen to accomodate the different types of variation, we choose $h_{ij}=[1+(m_{ij}-1)\phi_i]^{-1}$. Estimates of the parameters (a_k, b_k) $(k=1,2)$ can be estimated using the GEE in equations (13) and (14) and dispersion parameters may be estimated by Pearsons χ^2 statistic given in (15). The trend statistic in (26) has a chi-square distribution with p=2 degrees of freedom. The test statistic was significant with a value of T_2=157.67 and corresponding P-value < 0.001, implying that there is a strong trend of response with respect to dose.

5.2 The Simulated Experiment

One thousand experiments were generated for each design below. Beta-binomial random variates were generated via IMSL subroutines RNBET and RNBIN. The number of implants m_{ij} were sampled randomly from the empirical distribution of the actual data set mentioned above. The number of prenatal deaths were generated from a beta-binomial distribution characterized by the parameters (a_2, b_2) given above. The number of malformed were then generated from another beta-binomial based on those survived with parameters (a_1, b_1).

The simulated data were fit using joint trinomial logistic models and separate binomial logistic dose response models. In joint modeling we used the following three methods to analyze the data: 1) the Dirichlet-trinomial model with corresponding covariance matrix given in (11) (DTC), 2) The GLM model (Ryan, 1993) model, where the term $1+(m_{ij}-1)\phi_i$ in the Dirichlet-multinomial covariance (Eq. 11) is replaced by ρ_i, and 3) the trinomial model after applying the Rao-Scott transformation (RST) to the data. When we analyzed the endpoints malformation, prenatal death and overall toxicity as overdispersed binomial outcomes, the beta-binomial variance (BB) is used, and the Rao-Scott transformation for overdispersed binary outcomes Eq. (22) is used for prenatal death or overall toxicity, and Eq. (23) for fetal malformation.

To solve for the parameters $\{\phi_i\}$ in DTC for joint modeling of the outcomes malformation and prenatal death equation (15) is used and reduces to

$$\sum_{j=1}^{n_i} \frac{\hat{\pi}_{1i}(1-\hat{\pi}_{1i})(r-m_{ij}\hat{\pi}_{2i})^2 + \hat{\pi}_{2i}(y_{ij}-s_{ij}\hat{\pi}_{1i})^2}{m_{ij}[1+(m_{ij}-1)\hat{\phi}_i]\hat{\pi}_{1i}(1-\hat{\pi}_{1i})\hat{\pi}_{2i}(1-\hat{\pi}_{2i})} = 2n_i\left(1-\frac{p}{2n}\right), \quad i=1,\ldots,t. \tag{27}$$

To estimate $\hat{\rho}_i$ in GLM for joint modeling we replace the term $1+(m_{ij}-1)\phi_i$ in equation (15) or in the above equation (27) by ρ_i and then solve for ρ_i to get,

$$\hat{\rho}_i = \frac{1}{2n_i\left(1-\frac{p}{2n}\right)} \sum_{j=1}^{n_i} \frac{\hat{\pi}_{1i}(1-\hat{\pi}_{1i})(r_{ij}-m_{ij}\hat{\pi}_{2i})^2 + \hat{\pi}_{2i}(y_{ij}-s_{ij}\hat{\pi}_{1i})^2}{m_{ij}\hat{\pi}_{1i}(1-\hat{\pi}_{1i})\hat{\pi}_{2i}(1-\hat{\pi}_{2i})}, \quad i=1,...,t \tag{28}$$

where the probabilities $\hat{\pi}_{1i}$ and $\hat{\pi}_{2i}$ are estimated by replacing the estimates $(\hat{a}_1, \hat{b}_1)$ in π_{1i} and $(\hat{a}_2, \hat{b}_2)$ in π_{2i}, computed by equations (13) and (14). Here p is the total number of regression parameters being estimated. Since we are fitting the logistic model, p is equal to four in joint trinomial modeling of the endpoints malformation and prenatal death.

In separate binomial models for prenatal death and overall toxicity the overdispersion parameters $\{\phi_i\}$ in the beta-binomial variance are solved iteratively using Eq. (6). For the endpoint fetal malformation, Eq. (9) is used to solve for the overdispersion parameter $\{\phi_i\}$. To estimate the parameter $\{\rho_i\}$ in GLM for separate modeling of prenatal death or overall toxicity the term $1+(m_{ij}-1)\phi_i$ in equation (6) is replaced by $\{\rho_i\}$, and then solve for ρ_i ($i=1,...,t$). For the endpoint of fetal malformation we replace the term $1+(s_{ij}-1)\phi_i$ in Eq. (9) by ρ_i. Solving for ρ_i in equation (9) yields

$$\hat{\rho}_i = \frac{1}{n_i\left(1-\frac{p}{n}\right)} \sum_{j=1}^{n_i} \frac{(y_{ij}-s_{ij}\hat{\pi}_i)^2}{s_{ij}\hat{\pi}_i(1-\hat{\pi}_i)}, \quad i=1,...,t. \tag{29}$$

Again, $\hat{\pi}_i$ is evaluated by replacing the estimates $(\hat{a}, \hat{b})$ computed by equation (5) (for prenatal death and overall toxicity) or computed by equation (8) (for malformation) into π_i. Here p would be equal to two when fitting the logistic model to one of the endpoints alone, and x_{ij} is the overdispersed binomial outcome (i.e., $x_{ij}=r_{ij}$ for prenatal death, and

$x_{ij}=y_{ij}+r_{ij}$ for overall toxicity). In both joint modeling and separate modeling, after applying the appropriate Rao-Scott transformation to the data, the overdispersion parameter in DTC or BB is set to zero.

Estimates of ED_{05} and benchmark dose BMD_{05} of 5% excess risk over the control rate (or relative risk) were calculated from the estimated models using equations (17) to (21) respectively. The determination of the traditional NOAEL was also made using Fisher's exact test.

5.3 Criteria for Evaluation

The mean, relative bias, and relative MSE for $\hat{ED}_{05}$ were calculated from the 1,000 experiments. Relative bias (RBIAS) and relative MSE (RMSE) were computed as follows

$$RBIAS = \frac{mean(\hat{ED}_{05}) - ED_{05}}{ED_{05}}$$

and,

$$RMSE = \frac{MSE}{ED_{05}^2} = \frac{\sum_{l=1}^{s} (\hat{ED}_{05,l} - ED_{05})^2}{sED_{05}^2},$$

where s is the number of simulated experiments that were used in the above calculations. The proportion of $\hat{ED}_{05}$ smaller than the actual 5% relative risk (ED_{05}) was also recorded. If this proportion is large, the estimate is conservative but safe. The mean of $\hat{BMD}_{05}$ and

NOAEL, along with the proportion that they are smaller than the actual ED_{05} were recorded.

In some of the simulated experiments, $\hat{ED}_{05}$ was greater than the highest dose in the experiment. This happened when the slope of the dose response curve was positive and small or close to zero, indicating that there may not be an increasing dose-response relationship. Another situation that occurred, suggesting that there is no increasing dose-response relationship, is when the slope of the curve ($\hat{b}_k$, k=1,2) was negative producing a negative point estimate of ED_{05}. When either of these situations arose, we did not include the estimates $\hat{ED}_{05}$ and $\hat{BMD}_{05}$ in the calculation of the mean, RBIAS, RMSE and P(<ED_{05}). This criteria was followed because in a true experiment the highest dose is assumed to produce 10% mortality (EPA, 1991), if this is not satisfied then the experiment is repeated. If the estimated variance of $\hat{ED}_{05}$ was large it would cause $\hat{BMD}_{05}$ to take on a negative value. In this instance we set $\hat{BMD}_{05}$ equal to zero and both estimates were included in further analysis. When calculating the NOAEL in some experiments the lowest dose group is significantly different from the control group, the NOAEL would not be assigned to zero. Usually, either a new experiment would be conducted or a lowest-observed-adverse-effect-level (LOAEL) would be set. For the purposes of this thesis, we follow Leisenring and Ryan (1992) and let NOAEL take the value of $d_2/10$, when this was the case. In some of the simulated experiments the proportion of responses for a particular anomaly in the highest dose level was not

significantly different from the control group, in this case NOAEL is assigned the highest dose level i.e., 1.0 in that experiment for that particular endpoint.

When we used the Rao-Scott transformation for overdispersed binomial outcomes in the simulated experiments, some of the response rate for the binary counts r_{ij}, y_{ij}, or $y_{ij}+r_{ij}$ are all equal to zero for dose level i causing the proportion $\hat{p}_i$ (or $\hat{p}_{i_v}$) to be zero, hence the design effect would be infinite. Following Fung et al., (1994) we set the design effect $\hat{D}_i$ (or $\hat{D}_{i_v}$) equal to 1 in this instance.

In the case of joint modeling the same situation may arise. In some of the simulated studies, one or more of the response rates $\hat{\mu}_{ki}$ (k=1,2,3; i=1,...t;) were equal to zero causing the design effect $\hat{D}_i$ to be infinite for one or more of the dose groups in a simulated experiment. The subscript k=1,2,3 corresponds to malformation $y_{i.}$, prenatal death $r_{i.}$, and alive and normal $s_{i.}$-$y_{i.}$, respectively at dose level i. Upon evaluation of the matrix $V_0^{-1}(\mu_i)V(\mu_i)$ in equation (24) for the Rao-Scott transformation, we find that the matrix can be expressed as

$$V_0^{-1}(\mu_i)V(\mu_i) = \begin{pmatrix} \dfrac{(1-\mu_{2i})v_{i11}}{\mu_{1i}\mu_{3i}} + \dfrac{v_{i12}}{\mu_{3i}} & \dfrac{(1-\mu_{2i})v_{i12}}{\mu_{1i}\mu_{3i}} + \dfrac{v_{i22}}{\mu_{3i}} \\ \dfrac{(1-\mu_{1i})v_{i12}}{\mu_{2i}\mu_{3i}} + \dfrac{v_{i11}}{\mu_{3i}} & \dfrac{(1-\mu_{1i})v_{i22}}{\mu_{2i}\mu_{3i}} + \dfrac{v_{i12}}{\mu_{3i}} \end{pmatrix}, \quad i=1,...,t$$

where v_{i11} and v_{i22} are the variances of malformation and prenatal death respectively as defined in section 4.2. The term v_{i12} is the covariance of y_{ij} and r_{ij} and is defined as

$$\hat{v}_{i12} = \frac{n_i}{m_{i.}(n_i-1)} \sum_{j=1}^{n_i} (r_{ij} - m_{ij}\hat{\mu}_{2i})(y_{ij} - m_{ij}\hat{\mu}_{1i}), \quad i=1,...,t.$$

It turns out that v_{i33}, the variance of being alive and normal (or the variance of overall toxicity) is equal to $(v_{i11}+ v_{i22}+ 2v_{i12})$. The design effect $\hat{D}_i$ is equal to one half the trace of the matrix $V_0^{-1}(\mu_i)V(\mu_i)$. Hence if the response rate for malformation $\hat{\mu}_{1i}$ is equal to zero then we set the term in the upper left hand corner of the matrix $V_0^{-1}(\mu_i)V(\mu_i)$ equal to one. If the prenatal death rate $\hat{\mu}_{2i}$ is equal to zero then we set the term in the lower right hand corner of the matrix $V_0^{-1}(\mu_i)V(\mu_i)$ equal to one. It can also happen that the estimated proportion of being alive and normal $\hat{\mu}_{3i}$ is equal to zero. In this case we set the ratio $\hat{v}_{i33}/\hat{\mu}_{3i}$ in (25) equal to one, hence

$$\hat{D}_i = \frac{1}{2}\left(\frac{\hat{v}_{i11}}{\hat{\mu}_{1i}} + \frac{\hat{v}_{i22}}{\hat{\mu}_{2i}} + 1\right).$$

By following this procedure, we are in effect saying that there is no overdispersion or underdispersion for the endpoint in which $\hat{\mu}_{ki} = 0$ for (k=1,2,3; i=1,...,t).

5.4 Effect of Study Design in Terms of Number of Litters Per Dose Group

Following the original study, we used four dose groups (0, 1250, 2500, 5000) (mg/kg/day) as our basic design. Without loss of generality, we scaled the doses in our analyses such that the highest dose is 1 (i.e., 0, 0.25, 0.5, 1.0). First, we explored the

effect of the number of animals in each group on the various estimates using group sizes (10, 10, 10, 10), (20, 20, 20, 20), and (30, 30, 30, 30).

As expected results in Table 2 show that as sample size increases, the relative bias (RBIAS) and relative MSE (RMSE) of $E\hat{D}_{05}$ decreases for the endpoints malformation and overall toxicity in all three methods DTC, GLM and RST. For the endpoint prenatal death, RMSE decreases as sample size increases in all three methods, however, the relative bias decreases when the sample size increases from 10 to 20, and then increases when sample size increases from 20 to 30 animals per dose group. This may be due to random fluctuation. In all anomalies concerned, the decrease in RBIAS and RMSE is more dramatic from experiments with sample size 10 to 20, but not as dramatic from 20 to 30. This indicates that designing a study with sample size 20 in each dose group is probably sufficient, and adding more animals to each group does not increase the accuracy. The number of times that $E\hat{D}_{05}$ is smaller than the true ED_{05} is generally less than 50% for experiments with 10 to 30 animals per dose group. The proportion increases as sample size increases from 10 to 30 animals per dose group for the endpoints malformation and overall toxicity. The estimate $E\hat{D}_{05}$ is closer to the true parameter as sample size increases.

Table 3 presents the mean of $B\hat{M}D_{05}$ under the logistic model for the three methods DTC, GLM and RST with dose levels (0,0.25,0.5,1), and also the NOAEL. Results show that the mean of NOAEL is generally smaller and more conservative than

$\hat{BMD}_{05}$ except in two cases for prenatal death. For the endpoints prenatal death and overall toxicity in the design with 10 litters per dose group the proportion of NOAEL less than the true ED_{05} falls below the 50% mark, showing that smaller experiments lead to higher NOAELs than larger studies, and this is not desirable. The proportion of NOAEL less than the true ED_{05} is always smaller than the proportion of $\hat{BMD}_{05}$ less than the targeted value (i.e, ED_{05}) for the endpoints prenatal death and overall toxicity in the designs with 10 to 30 litters per dose group. Figures 2*a*, 2*b* and 2*c* present histograms of the ratio NOAEL/BMD for the endpoint overall toxicity in each experiment of the simulation. The proportion of ratios that are within a factor of two i.e., the probability that the ratio falls between 1/2 and 2, decreases as sample size increases from 10 to 30 litters per dose group. For example, 52% of the ratios in the simulation with 10 litters per dose group, 28% of the ratios in the simulation with 20 litters per dose group and 15% of the ratios in the simulation with 30 litters per dose group are within a factor of two. The mean of the ratios decreases as sample size increases. As sample size increases from 10 to 30 litters per dose group, NOAEL decreases (see Table 3), causing the proportion of ratios NOAEL/BMD less than 1 to increase. This indicates the proportion of NOAEL within a factor of two of $\hat{BMD}_{05}$ decreases as sample size increases. Thus, while the accuracy of $\hat{BMD}_{05}$ increases as sample size increases (smaller relative bias and larger proportion less than the true ED_{05}) that of NOAEL decreases (larger relative bias and smaller proportion less than the true ED_{05}). The chance that $\hat{BMD}_{05}$ is less than the true ED_{05} is over 80% in most cases, except when the number of animals is as small as 10 in

each dose group. The proportion of $\hat{BMD}_{05}$ less than the true ED_{05} for prenatal death is always the highest and close to 1. The proportion of NOAEL less than the true ED_{05} increases as sample size increases per dose group. The average estimated risk associated with the NOAEL was also computed by substituting each estimated NOAEL level into the estimated logistic dose-response curve to obtain an estimated risk, and then an average of these estimated risks was calculated. The average estimated risk associated with the NOAEL is largest for the endpoint overall toxicity, and is greater than 0.05, (which is the α-percent that refers to relative risk). The estimated risk associated with NOAEL for overall toxicity ranges from 9% for 10 litters per dose group to 7% for 30 litters per dose group, for prenatal death it ranges from 6.5% for 10 litters per dose group to 5.7% for 30 litters per dose group, and for malformation it ranges from 4% for 10 litters per dose group to 2.7% for 30 litters per dose group.

We also considered unbalanced designs with the total number of litters fixed at 80: (30,30,10,10), (30,10,10,30) and (10,30,30,10). Comparing these three designs the RBIAS and RMSE of $\hat{ED}_{05}$ appear to be similar. There is no one unbalanced design that does consistently better for all three endpoints (malformation, prenatal death and overall toxicity) than the other unbalanced designs. We compare these unbalanced designs to the design with equal sample size of 20 litters per dose group (since the total sample size here is also 80). The design with equal sample size per dose group is as good as, if not better than the unbalanced designs. Thus it seems that a design with equal sample size in each

dose group is appropriate, in order to fit the model adequately to the data for best results of estimates $\hat{ED}_{05}$.

Table 3 presents the BMDs and NOAEL levels for the unbalanced designs. The proportion of $\hat{BMD}_{05}$ less than the true ED_{05} is greater than the proportion of NOAEL being less than ED_{05}, yet the mean of NOAEL is usually smaller than that of $\hat{BMD}_{05}$, for prenatal death and overall toxicity. Looking at the histograms for these three unbalanced designs in Figures 2*d*, 2*e* and 2*f*, we see that the proportion of NOAEL less than the estimated benchmark dose (i.e., proportion of ratios less than one) is greater than the proportion of ratios greater than one, and is around 70%. Comparing these histograms to the histogram with equal sample size of 20 litters per dose group (Fig. 2*b*), we see that the proportion of ratios less than one in the equal sample size design is similar to the unbalanced design. The average estimated risk corresponding to NOAEL in the unbalanced designs, for overall toxicity ranges from 7-8%, as compared to 7.6% in the balanced design with 20 litters per dose group. There does not seem to be anything gained in the unbalanced design as opposed to the balanced design with respect to the estimates of ED_{05}, BMD_{05} and NOAEL.

Similar results concerning the estimate of ED_{05} and BMD_{05} in separate modeling are found in Tables 8 and 9, respectively, for study design in terms of number of litters per dose group.

5.5 Effect of Study Design in Terms of Number of Doses and Placement of Doses

Our aim is not to find the optimal design, but to see how the number of doses in an experiment and the dose placement affects the estimates of ED_{05} and BMD_{05}. Fixing the total number of animals at 80, we considered designs of three doses with (27,26,27) litters per dose group, four doses with (20,20,20,20) and five doses with (16,16,16,16,16) litters per dose group. The placement of the doses considered were (0,0.5,1) and (0,0.3,1) for three dose groups, (0,0.25,0.5,1) and (0,0.15,0.35,1) for four doses, and (0,0.125,0.25,0.5,1) and (0,0.125,0.4,0.75,1) for five doses. The ϕ_i's used to generate the data for each dose group were chosen in an increasing linear fashion from 0.1 to 0.5, depending on the number of doses in the design. For example, in the three dose design, $\{\phi_i\}$ are chosen as {0.1, 0.2, 0.3}, in the four dose design $\{\phi_i\}$ are chosen as {0.1, 0.2, 0.3, 0.4}, and in the five dose design $\{\phi_i\}$ are chosen as {0.1, 0.2, 0.3, 0.4, 0.5}, for each dose group.

The simulated results are presented in Tables 4 and 5. For three and four dose groups, the estimate of ED_{05} was more accurate (smaller RBIAS, smaller RMSE and larger proportion of $\hat{ED}_{05}$ smaller than the true value of ED_{05}) in experimental designs that included a dose near the targeted effect level, i.e., near the true ED_{05}. For instance in the three dose design (0, 0.5, 1), and four dose design (0, 0.25, 0.5, 1), the relative bias and relative MSE were smaller than in their respective three and four dose designs.

However, for the five dose group designs, the estimate of ED_{05} seem to be less dependent on the true ED_{05}.

Results for $\hat{BMD}_{05}$ and NOAEL are shown in Table 5. The mean value of NOAEL is always smaller than $\hat{BMD}_{05}$, and it also has a larger relative bias as compared to that of $\hat{BMD}_{05}$, in all designs being presently considered. The proportion of NOAEL less than the true ED_{05} is usually larger than the proportion of $\hat{BMD}_{05}$ less than the true ED_{05}. Although in some of the cases the proportion of NOAEL less than the true ED_{05} is less than 70%. Figures 2*g* through to 2*l* present the histograms of the ratio NOAEL/BMD for the endpoint overall toxicity in the simulations mentioned. The proportion of ratios within a factor of two is smallest in the designs (0, 0.5, 1), (0, 0.25, 0.5, 1) and (0, 0.125, 0.4, 0.75, 1). This indicates that the NOAEL is very different from the estimated BMDs in these designs. In these three designs we see from the histograms that the proportion of NOAEL less than $\hat{BMD}_{05}$ (i.e., proportion of ratios that are less than 1) is larger as compared to their respective three, four and five dose designs. In histogram 2*g* for the design with doses at (0, 0.5, 1) it is interesting to note that all the ratios are less than 1, implying that all NOAELs are less than the corresponding BMDs. We note that the true ED_{05} is 0.201 for the endpoint overall toxicity. Since the NOAEL must choose one of the existing dose levels and 0.5 is very high for this endpoint, the NOAEL is usually assigned to 0.05 ($=d_2/10$), causing all the ratios NOAEL/BMD to be very small. The average estimated risk associated with the NOAEL in all of these designs ranges from 6-7% for the endpoint overall toxicity.

In the experimental designs that we have investigated, we have found that in the three and four dose designs, estimates of ED_{05} did relatively better when there was one dose that was near the targetted effect level. It is difficult to know the targeted effect level prior to the design of the experiment, though possible information may be gained in pilot studies. More research in the area of number of doses and dose placement in an experiment should be investigated, because in this setting we have only looked at a few selected designs and explained the effects of these designs on the estimates of ED_{05} and BMD_{05}. Comparable results for separate modeling are presented in Tables 10 and 11.

5.6 Effect of Intra-Litter Correlation

The effect of changing the size of the intra-litter correlation among implants to the various estimates was investigated. Results are in Tables 6 and 7. Values for $\{\phi_i\}$ used to generate the data, ranging from 0 to 0.4, were considered. In the generation of the data a common ϕ for all dose groups as well as varying ϕ_i for each dose group was used: (0,0,0,0), (0.1,0.1,0.1,0.1), (0.4,0.4,0.4,0.4), and (0.1,0.2,0.3,0.4). Results in Table 6 show that for constant intra-litter correlation in each dose group, the RBIAS and RMSE of $\hat{ED}_{05}$ increases as ϕ increases. The number of times that $\hat{ED}_{05}$ is less than the true parameter ED_{05}, decreases as ϕ increases. The mean of the estimates $\hat{BMD}_{05}$ and the proportion of $\hat{BMD}_{05}$ less than the true ED_{05} decreases as the constant ϕ increases. All these estimates are sensitive to intra-litter correlation. The mean of NOAEL is always

lower than the mean of $\hat{BMD}_{05}$, but the proportion of NOAEL less than the true effect level ED_{05}, is smaller than the proportion of $\hat{BMD}_{05}$ less than the true effect level ED_{05}, for the endpoints prenatal death and overall toxicity, indicating that although $\hat{BMD}_{05}$ has a higher mean value it is still a more accurate measure of developmental toxicity than the traditional NOAEL. Figures 2*m* to 2*p* present the histograms of the ratio NOAEL/BMD for the endpoint overall toxicity. The proportion of ratios within a factor of two and the mean of the ratio increases as ϕ increases. The average risk associated with the NOAEL is around 0.08 for overall toxicity. Corresponding results for separate modeling are presented in Tables 12 and 13.

5.7 Comparison of the Three Methods: DTC, GLM and RST

In all the experimental designs discussed above we analyzed the simulated data with the Dirichlet-trinomial covariance given in (11), generalized linear model and with the Rao-Scott transformation. Above we discussed the results concerning the estimate of ED_{05} between the different experimental designs. Now we will discuss how these results compare between the three different methods of analysis.

Generally estimates of ED_{05} derived from using the GLM method did not differ from and are very comparable to those derived by using the Dirichlet-trinomial covariance. For the estimate $\hat{BMD}_{05}$, again the mean and $P(\hat{BMD}_{05} < ED_{05})$ were

comparable and in some cases the proportion of $\hat{BMD}_{05}$ less than the true effect level ED_{05} was greater for GLM than DTC. Since the two methods give similar results, it is simpler to estimate the parameter $\{\rho_i\}$ using GLM.

Results attained from applying the Rao-Scott transformation to the trinomial counts are agreeable and favourable. In all simulation studies the RBIAS and RMSE are smaller for the estimate $\hat{ED}_{05}$ after transforming the data as compared to the RBIAS and RMSE of the estimate $\hat{ED}_{05}$ for the original overdispersed data using either DTC or GLM. Also, the proportion of $\hat{ED}_{05}$ less than the true ED_{05} and the proportion of $\hat{BMD}_{05}$ less than the true ED_{05} are larger for the transformed data. The transformation is simple, does not require specification of an underlying distribution of the data, and the analysis of the transformed data is computationally easier. Hence the Rao-Scott transformation is recommended for use in analyzing developmental toxicity data.

5.8 Overdispersed Binomial Versus Overdispersed Trinomial Models

Estimates of ED_{05} for the three endpoints based on separate binomial modeling are compared with estimates of ED_{05} based on joint trinomial modeling. For overall toxicity as a binary outcome, we combine the incidence of prenatal death and malformation i.e., $(y+r)/m$. From Tables 2, 4, 6, 8, 10 and 12, we see that the mean, RBIAS and RMSE of $\hat{ED}_{05}$ for malformation are smaller in joint modeling for DTC, GLM and RST as compared to separate modeling. Also the proportion of $\hat{ED}_{05}$ less than the true ED_{05} is

also larger in joint modeling for this endpoint. With respect to the endpoint prenatal death the results for $E\hat{D}_{05}$ are very similar, yet leaning in favour towards joint modeling. In contrast, for the incidence of overall toxicity, separate modeling seems to do better (smaller RBIAS and RMSE and larger proportion of $E\hat{D}_{05}$ less than the target value ED_{05}) than in joint modeling, using BB and GLM variance functions. The Rao-Scott transformation does much more favourably in joint modeling than in separate modeling for overall toxicity. It was observed that when the Rao-Scott transformation was applied to trinomial counts, the RMSE of $E\hat{D}_{05}$ is generally smaller than applying the transformation to binary counts, in all endpoints concerned. The mean of the estimates $E\hat{D}_{05}$ based on joint models are always lower than those based on binomial models, for prenatal death or malformation alone. The mean of $E\hat{D}_{05}$ for overall toxicity based on joint modeling is a more sensitive indicator of developmental toxicity than those for fetal malformation and prenatal death alone, in that the former is always below the minimum of the latter two, as indicated by Krewski and Zhu (1994), and Krewski et al. (1994).

To compare the results for the estimates of BMD_{05} from joint trinomial models and separate binomial models we look at the results in Tables 3, 5, 7, 9, 11 and 13. In all simulation studies considered, the mean of $B\hat{M}D_{05}$ from joint models is generally lower than the mean of $B\hat{M}D_{05}$ from binomial models for the endpoints malformation and prenatal death in all methods of analysis considered. For overall toxicity the mean of $B\hat{M}D_{05}$ from separate modeling is smaller than its corresponding counterpart in all

methods of analysis considered. Comparing the proportion of $\hat{BMD}_{05}$ less than the true ED_{05} from joint trinomial models with separate binomial models we find that the proportion of $\hat{BMD}_{05}$ less than the true ED_{05}, is larger for estimates from joint models for the endpoints malformation and prenatal death, whereas the proportion of $\hat{BMD}_{05}$ less than the true ED_{05}, is larger in separate modeling in the case of overall toxicity.

Chapter 6

APPLICATION TO DEVELOPMENTAL TOXICITY DATA

6.1 Developmental Toxicity Data

The modeling techniques discussed in sections 3.3 and 3.4 were applied to developmental toxicity bioassay data. The database consists of developmental toxicity evaluations of 246 experiments described in detail by Faustman, et al. (1994). Each experiment has several subsets, with each data subset defined as the dose-response data for one endpoint from one experiment. There are up to nine endpoints measured for each experiment with a total of 1825 specific endpoints. For purposes of this thesis, we are only interested in the endpoints prenatal death (r), fetal malformation (y), and overall toxicity ($r+y$). The intention of this thesis is to compare joint modeling with separate modeling of these endpoints, to do so we had to combine data files where possible. To merge files we only considered those data files corresponding to the endpoints prenatal death or resorption (DR) and total affected (DM) which include dead plus total malformed. If an experiment is missing one of these endpoints, then that experiment would not be considered further in our analysis. The two files DR and DM were then compared to see if: 1) both files had

the same sequence number at the beginning of the file, 2) if each litter had the same number of pups, 3) if each dose had the same number of litters, and 4) if the total number affected was greater than the total number of dead in each litter. If all four conditions were satisfied, then the two data files were combined for that experiment, and the same was done for all 246 experiments. In the usual notation the DR file gives (r,m) and the DM file gives $(y+r,m)$.

There were a total of 78 experiments used for joint modeling, with only 28 of these experiments exhibiting a significant increase in trend based on the trinomial count prenatal death and fetal malformation given litter size, after applying the trend test Eq. (26) with $h_i=\rho_i^{-1}$. The 28 developmental toxicity experiments are presented in Table 14. Of these 28 laboratory experiments, 16 of them were conducted under the U.S. National Toxicology Program (NTP), and 12 were conducted under the WIL Laboratory. Compounds used in the assay at the WIL Laboratories were not identified.

These teratological studies involved groups as small as 5 to as large as 35 female animals exposed to one of three to five doses, including an unexposed control group. For both joint modeling and separate modeling, we fit the logistic model to the endpoints of interest,

$$\pi_{1i} = \frac{1}{1+\exp(-a_1 - b_1 d_i)},$$

$$\pi_{2i} = \frac{1}{1+\exp(-a_2 - b_2 d_i)},$$

π_{1i} is the probability of malformation in a live fetus at dose level $i=1,...,t$

π_{2i} is the probability of a prenatal death at dose level $i=1,...,t$

a_1, b_1, a_2, b_2 are the parameters to be estimated in the logistic model.

The GEE for separate and joint modeling presented in sections 3.3 and 3.4 respectively, were fit to the 28 data sets, with the highest dose scaled to unity. In both separate and joint modeling we used the generalized linear model (GLM) variance-covariance function, i.e., $[1+(m_{ij}-1)\phi_i]=\rho_i$. We used the GLM rather than the Dirichlet-trinomial (DTC) or beta-binomial (BB) variance-covariance functions because it was seen in our simulation study that estimates of ED_α and BMD_α derived from GLM were very comparable to those derived from DTC and BB variance-covariance functions, and also because the estimate of the overdispersion parameter in GLM is simpler to compute. The determination of the traditional NOAEL was also made for the 28 data sets.

6.2 Evaluation of $\hat{ED}_{05}$, $\hat{BMD}_{05}$ and NOAEL

For some of the endpoints there was no increase in dose response. To determine this we fit the logistic model to the data for both overdispersed trinomial and binomial outcomes and then extrapolated to find the 5% relative risk level. If $\hat{ED}_{05}$ is greater than the highest dose for a particular endpoint in an experiment then we investigate the data and find that the slope $\hat{b}_k$, (k=1,2) is very small or close to zero for that particular endpoint. After plotting the response of interest along with the logistic model previously fit, it is clear that there is no dose response relationship, for example the curve is very flat. Plots, similar to Figure 6 for BPA in rats (case #22) were constructed, and it is seen that in this experiment there is no increasing dose-response relationship for fetal malformation. This

occured for one or both of the endpoints prenatal death or malformation in case studies 1, 4, 5, 8, 10, 11, 12, 13, 14, 16, 17, 19, 22, 23, 24, 25 and 26. In case study 14 $\hat{BMD}_{05}$ was negative due to a large standard error, in this situation we set $\hat{BMD}_{05}$ to the lowest dose level i.e., $\hat{BMD}_{05}$=0.0.

The Rao-Scott transformation was also applied to the 28 data sets. Joint models were fit to the transformed data. The same criteria were used here to evaluate $\hat{BMD}_{05}$ as in the analysis for the original data. When the design effect $\hat{D}_i$ was infinite as was the case in studies # 4, 14, 17, 23 and 26, we followed the same criteria as in our simulation study (section 5.3).

We also computed the NOAEL for the 28 assays. The same criteria to evaluate the NOAEL was used as in our simulation study. When the NOAEL was determined to be zero, we followed Leisenring and Ryan, (1992) and set the NOAEL to $d_2/10$ in case studies 1, 3, 5, 6, 10, 13, 14, 17, 19, 20 and 28. If the highest dose was not significantly different than the control group then the NOAEL is set to 1.0, as in case studies 6, 8, 10, 11, 12, 13, 14, 16, 17, 22, 23, 25 and 26.

6.3 Overdispersed Binomial Versus Overdispersed Trinomial Models

Point estimates of ED_α are not presented here. We only present BMD_α, because we compare benchmark doses with NOAEL levels. Benchmark doses corresponding to a 5% relative risk level for malformation, prenatal death and overall toxicity based on both the trinomial and the binomial models are presented in Table 15. In all the experiments with

the exception of study 23, the BMDs for overall toxicity from joint modeling are lower than the minimum of the BMDs of the single endpoints malformation and prenatal death, fit with the overdispersed trinomial model. Similarly modeling (y+r)/m as a binomial outcome gives smaller BMDs than modeling the malformation rate (*y/s*) or embryolethality rate (*r/m*) separately. The BMDs for overall toxicity are always lower than the BMDs for fetal malformation or prenatal death. This concurs with the results from our simulation. Even when the BMDs could not be approximated for both of the endpoints malformation and prenatal death, we were still able to approximate the BMD for the endpoint overall toxicity. For instance, there was no dose-response relationship for either of the endpoints malformation and prenatal death alone in studies 12, 24 and 26, but the BMD could be estimated for overall toxicity. This demonstrates that the endpoint overall toxicity is a more sensitive measurement for teratogenicity.

In Figures 3a, 3b and 3c, the estimate $\hat{BMD}_{05}$ from joint and separate models are plotted against one another for the 28 experiments. If the estimates of the endpoints of interest are similar for the two models, most of the observations should lie on the straight line $\hat{BMD}_{05,t} = \hat{BMD}_{05,s}$. Here the subscript s or t indicates whether the estimate is from separate modeling or trinomial modeling. For the endpoint malformation, (in Figure 3a) most of the points lie above the line $\hat{BMD}_{05,s} = \hat{BMD}_{05,t}$ indicating that the BMDs from separate modeling are larger than those from joint modeling.

Comparing the two BMDs for prenatal death (Fig. 3b), most of the points lie above the line of interest, giving the same conclusion as for malformation. There is one

noticeable outlier in Figure 3b. This point corresponds to case 9 with $\hat{BMD}_{05,I}$=0.5240 and $\hat{BMD}_{05,S}$=0.1852. We find that there is an outlier in the prenatal death data in the first dose group causing a large variation among responses between the different litters within the group. After removing this one outlier, the BMD in joint modeling for prenatal death decreased slightly from 0.5240 to 0.5236, whereas the BMD in separate modeling for the same endpoint increased from 0.1852 to 0.3128 with a corresponding decrease in its standard error, from 0.4116 to 0.2845. This example illustrates how sensitive separate modeling is to outliers in comparison to joint modeling.

For the case of overall toxicity (Fig. 3c) most of the points lie below the line $\hat{BMD}_{05,S} = \hat{BMD}_{05,I}$. This implies that estimates of BMD for overall toxicity from separate modeling are smaller than those from joint modeling, giving the opposite conclusion as compared to the endpoints malformation and prenatal death, where estimates from joint models were smaller than those from separate models. Note that the intra-litter correlation in the two models are determined differently. In separate modeling each anomaly has its own set of $\{\phi_i\}$ to be determined for each dose group, whereas in joint modeling, the intra-litter correlation must take into account the overdispersion from the malformation and the prenatal death data (see Eq. 27).

For all anomalies being considered, the standard errors of $\hat{ED}_{05}$ used for estimation of $\hat{BMD}_{05}$ are smaller in joint modeling than in separate modeling, though not uniformly. This result is due to the fact that the trinomial model utilizes all of the available information, it is expected to yield estimates of BMD with smaller variability

than the binomial model (Krewski and Zhu, 1994). Hence, estimates of BMD_{05} from joint modeling for overall toxicity are higher than those from separate modeling, and they have smaller standard errors.

The NOAEL for each of the 28 experiments are also presented in Table 15. The NOAEL estimates do not in general agree with the benchmark dose estimates, for all three endpoints. For the anomaly fetal malformation, when there was no dose-response relationship present in the data the NOAEL, as expected, was usually estimated at 1.0 (the highest dose). Although in case study 4 the estimates of BMD from both joint and separate models are 0.1606 and 0.2741, respectively, the NOAEL was still calculated as 1.0. On the other hand, cases such as # 28 the respective BMDs from joint and separate models for malformation are 0.2141 and 0.2235, and the NOAEL is estimated at 0.0250 ($=d_2/10$). Similar examples can be found in the prenatal death and overall toxicity results (e.g., case # 1, 3, 5, 6, 10, 12, 13, 17, 19, 21, 24). When a dose-response relationship did not exist for both the endpoints malformation and prenatal death (case study 12 and 26), the BMDs for overall toxicity was high, as expected, yet the estimate of NOAEL is small (e.g., in case #12 the BMDs from joint and separate models are 0.6071 and 0.5636, respectively, and the NOAEL is $0.0625=d_2/10$). All these examples illustrate that the NOAEL is very sensitive to dose placement, as the NOAEL must choose one of the doses in the experiment.

In Figure 4 we compare the BMDs from trinomial models along with NOAEL levels for overall toxicity. More points lie above the line NOAEL= $\hat{BMD}_{05,I}$ than below.

This demonstrates that the BMD levels from joint modeling are lower than the traditional NOAELs.

6.4 Comparison of BMDs Calculated From The Original and Rao-Scott Transformed Data

The Rao-Scott transformation was applied to the 28 data sets. Table 16 reports the BMDs from joint models for the transformed data. We duplicated the results of the BMDs from joint models of the original data in this table for ease of comparison. The results are very comparable and can be seen in Figures 5a, 5b and 5c. The subscript (RST) refers to BMDs computed from the Rao-Scott transformed data and (GLM) refers to BMDs computed from joint modeling using the generalized linear model variance-covariance function. Most of the BMDs lie on the straight line $\hat{BMD}_{05,RST} = \hat{BMD}_{05,GLM}$ or just below it, indicating that BMDs computed from applying the Rao-Scott transformation to the data may be a more conservative measure of estimating safe dose levels. There is one case that appears as an outlier and is clearly seen in figures 5a, 5b and 5c. This corresponds to experiment 9, the same case aforementioned when comparing the BMDs from joint and separate models. After deleting the outlier from the data set the BMDs tabulated from the transformed data have increased from 0.2916 to 0.3005 for malformation, from 0.1534 to 0.2641 for prenatal death, and from 0.2242 to 0.2307 for overall toxicity. For all three endpoints there was also a decrease in the standard error (from 0.0448, 0.4403, 0.0485 to 0.0022, 0.1103, 0.0023, for malformation, prenatal death and overall toxicity, respectively). In this particular case after removing the outlier from

the data set, the BMDs from the transformed data not only increased, but were still smaller than the BMDs from the original data and have correspondingly smaller standard errors than the estimates of BMD_{05} from the original data. In experiment 23 there is no BMD level reported for the malformation data after the Rao-Scott transformation, because the $\hat{ED}_{05}$ level after fitting the logistic model to this endpoint was slightly greater than the highest dose level. Thus we followed our usual criteria and dismissed this endpoint as having no dose response relationship.

Chapter 7

SUMMARY of RESULTS and DISCUSSION

There is a need to move from using NOAEL to assess risk in developmental toxicity studies as the NOAEL method has many drawbacks. In this thesis we demonstrate again that the benchmark dose (BMD) approach is better than the NOAEL method. We have also compared BMDs obtained from using joint endpoints to using separate endpoints. In both joint and separate modeling we have also evaluated estimates of ED_{05} and BMD_{05} obtained from the Rao-Scott transformed data as opposed to the original data with DTC (or BB) and GLM covariance functions. These comparisons were all carried out first in a simulation setting and then to real-life developmental toxicity data. Some of the main results are summarized here.

Results from our simulation study indicate that a study design of four dose group with roughly 20 animals in each dose group provide good estimates of $\hat{ED}_{05}$ and $\hat{BMD}_{05}$ in terms of small RMSE and small RBIAS for $\hat{ED}_{05}$ and a large proportion of $\hat{BMD}_{05}$ less than the true ED_{05}. When we studied dose placement and number of doses in an experiment with fixed total sample size, results suggest that designs with three or four

doses with doses near the targeted effect value gave better estimates for $E\hat{D}_{05}$ and $B\hat{M}D_{05}$, i.e. smaller RBIAS and RMSE. The estimates of $E\hat{D}_{05}$ and $B\hat{M}D_{05}$ are sensitive to intra-litter correlation, thus intra-litter correlation should not be ignored in dose-response modeling.

It was observed that BMDs for overall toxicity are always below the minimum of BMDs for malformation or prenatal death for both joint trinomial and separate binomial models. Thus BMDs for the endpoint overall toxicity give lower safe dose levels than for malformation or prenatal death alone.

We have also compared the BMDs from joint trinomial and separate binomial models. Comparable results are obtained from our simulation study and application to developmental toxicity data. For the endpoints of malformation and prenatal death, BMDs from joint models are lower than those from separate binomial models. In contrast, BMDs from separate binomial models are smaller than the BMDs from joint models for overall toxicity. The BMDs from joint models have smaller standard errors, indicating that estimates of BMD_{05} from joint modeling are more precise because it makes optimum use of the available data (Krewski and Zhu, 1994).

The Dirichlet-trinomial (DTC) (or beta-binomial, BB) and generalized linear model (GLM) variance-covariance functions were investigated in model fitting for both joint and separate models. Results showed that they give comparable estimates of ED_{α} and BMD_{α}. This implies that these estimates of risk are quite insensitive to the variance structure as long as overdispersion is somehow characterized. In most instances,

estimates from using GLM had smaller RBIAS and RMSE, and larger P(<ED_{05}) than those derived from DTC or BB.

The Rao-Scott transformation was applied to developmental toxicity bioassay data and the simulated data. After correcting for overdispersion by this transformation, estimates of BMD_{05} seem to be very comparable and in some instances lower than the $\hat{BMD}_{05}$ computed from the original data. In our simulation study, the Rao-Scott transformation always produced lower relative bias and relative MSE, and larger P(<ED_{05}) for estimates of ED_{05}. Similarily for BMDs, the means were smaller and the proportion of $\hat{BMD}_{05}$ less than the true ED_{05} was larger for the RST data as compared to those from the original data. The Rao-Scott transformation, as applied to trinomial counts and then fitting joint models to the transformed data, makes use of all the available data while producing the most accurate point estimates ($\hat{ED}_{05}$) and P($\hat{BMD}_{05}$ <ED_{05}).

NOAELs do not provide a reliable means for arriving at a safe dose level. The NOAEL must choose one of the existing dose levels in the experiment. NOAEL levels are sensitive to small studies and ignore the shape of the dose response curve. NOAELs do not take into consideration the correlation among implants in the same litter. In our simulation study we saw that in most cases for the endpoint overall toxicity, the proportion of times that $\hat{BMD}_{05}$ (either from joint or separate models) was less than the true ED_{05} was greater than the proportion that NOAEL was less than the true ED_{05}, indicating that the benchmark dose methodology is a more conservative measure for arriving at a safe dose level than the traditional NOAEL method. In the application to

developmental toxicity data, we found that for overall toxicity, the BMD method from joint modeling reported lower safe dose levels than NOAEL in 16 out of 28 case studies (Fig.4).

There still remain areas in risk assessment of developmental toxicity studies to be explored. Research in the area of incorporating litter effect due to the litter size into the dose-response model should be considered, since response may depend on the litter size. Fitting models with the actual level of the teratogenic substance that is absorbed by the fetus is another area of interest. Perhaps some pharmacokinetic models have to be used, and the ultimate dose-response relationship may be different from those presented here. The robustness in the choice of a dose-response model (e.g., Weibull, probit, logistic etc.) in estimating risk can also be investigated. A simulation study can be conducted to compare the estimates of ED_α and BMD_α from different dose-response models and see how close they are. In our study we considered dose-response models with independent parameters i.e., (a_1, b_1) were independent of (a_2, b_2) in joint models. An area of investigation may include models where the parameters are not necessarily independent. Mathematical extrapolation methods do not presuppose any definte threshold. It would also be of interest to compare, in a simulation setting, estimates of ED_α and BMD_α computed from data generated with and without threshold. Finally, Kimmel and Gaylor (1988) suggest that the interaction of developmental toxicologists, pharmacokineticists and modelers, will be the development of more biologically-based models that will move us beyond the BMD or NOAEL paradigm and allow us to consider the continuum of

developmental effects ranging from functional deficits to growth alteration to malformation and death.

Bibliography

Aitkin, M. (1987). Modeling variance heterogeneity in normal regression using GLIM. Applied Statistics, 36, 332-339.

Allen, B.C., Kavlock, R.J., Kimmel, C.A., Faustman, E.M. (1994a). Dose-response assessment for developmental toxicity, comparison of generic benchmark dose estimates with no observed adverse effect levels. Fundamental and Applied Toxicology, 23, 487-495.

Allen, B.C., Kavlock, R.J., Kimmel, C.A., Faustman, E.M. (1994b). Dose-response assessment for developmental toxicity, statistical models. Fundamental and Applied Toxicology, 23, 496-509.

Altham, P.M.E. (1978). Two generalizations of the binomial distribution. Appl. Stat., 27, 162-167.

Anderson, E., and Carcinogen Assessment Group of the U.S. Environmental Protection Agency (1983). Quantitative approaches in use to assess cancer risk. Risk Anal., 3,277-295.

Barnes, D.G., Daston, G.P., Evans, J.S., Jarabek, A.M., Kavlock, R.J., Kimmel, C.A., Park, C.P., Spitzer, H.L. (1995). Benchmark dose workshop: Criteria for use of a benchmark dose to estimate a reference dose. Regulatory Toxicology and

Pharmacology, 21, 296-306.

Barnes, D.G., Dourson, M. (1988). Reference dose (RfD): Description and use in health risk assessments. Reg. Toxicol. Pharmacol., 8, 471-486.

Catalano, P.J., Ryan, L.M. (1994). Statistical methods in developmental toxicology. In Handbook of Statistics Vol. 12, (G.P. Patil, C.R. Rao, ed.), North-Holland, Amsterdam, New York, 507-532.

Catalano, P.J., Scharfstein, D.O., Ryan, L., Kimmel, C., Kimmel, G. (1993). Statistical model for fetal death, fetal weight, and malformation in developmental toxicity studies. Teratology, 47, 281-290.

Chen, J.J., Kodell, R.L. (1989). Quantitative risk assessment for teratological effects. JASA, 408, 966-971.

Chen, J.J., Kodell, R.L., Howe, R.B., Gaylor, D.W. (1991). Analysis of trinomial responses from reproductive and developmental toxicity experiments. Biometrics 47, 1049-1058.

Chen, J.J., Li, L.A. (1994). Dose-response modeling of trinomial responses from developmental experiments. Stat. Sin., 4, 265-274.

Crowder, M.J. (1978). Beta-binomial ANOVA for proportions. Applied Statistics, 27, 34-37.

Crump, K.S. (1984). A new method for determining allowable daily intakes. Fund. Appl. Toxicol., 4, 854-871.

Crump, K.S., Howe, R.B. (1983). A review of methods for calculating confidence limits in low dose extrapolation. In Toxicological Risk Assessment, eds. D.B. Clayson,

D.R. Krewski and I.C. Munroe. Boca Raton, FL:CRC Press, pp.187-203.

Dourson, M.L. (1986). New approaches in the derivation of acceptable daily intake (ADI). Comments Toxicology, 1, 35-48.

Efron, B. (1986). Double exponential families and their use in generalized linear regression. JASA, 81, 709-721.

EPA (1986). U.S. Environmental Protection Agency. Guidelines for the health assessment of suspect developmental toxicants. Federal Register 51, 34028-34040.

EPA (1989). U.S. Environmental Protection Agency. Proposed amendments to the guidelines for the health assessment of suspect developmental toxicants. Federal Register, 54, 9386-9403.

EPA (1991). U.S. Environmental Protection Agency. Guidelines for developmental toxicity and risk assessment. Federal Register 56, 63798-63826.

Faustman, E.M., Wellington, D.G., Smith, W.P., Kimmel, C.A. (1989). Characterization of a developmental toxicity dose-response model. Environ. Health Perspect., 79, 229-241.

Faustman, E.M., Allen, B.C., Kavlock, R.J., Kimmel, C.A. (1994). Dose-response assessment for developmental toxicity, characterization of database and determination of no observed adverse effect levels. Fund. Appl. Toxicol., 23, 478-486

Frankos, V.H. (1985). FDA perspectives on the use of teratology data for human risk assessment. Fund. Appl. Toxicol., 5, 616-625.

Fung, K.Y., Krewski, D., Rao, J.N.K., Scott, A.J. (1994). Tests for trend in developmental toxicity experiments with correlated binary data. Risk Analysis, 14,

639-647.

Gaylor, D.W. (1983). The use of safety factors for controlling health risks. J. Toxicol. Environ. Health, 11, 329-336.

Gaylor, D.W. (1988). Applicability of cancer risk assessment techniques to other effects. Toxicol. Ind. Health, 4, 453-459.

Gaylor, D.W. (1989). Quantitative risk analysis for quantal reproductive and developmental effects. Environ. Health Perspect., 79, 243-246.

Gaylor, D.W., Razzaghi, M. (1992). Process of building biologically based dose-response models for developmental defects. Teratology, 46, 573-581.

Gart, J.J., Krewski, D., Lee, P.N., Tarone, R.E., Wahrendorf,J. (1986). Statistical methods in cancer research. Vol.III, The Design and Analysis of Long-Term Animal Experiments. Oxford Univ. Press, Oxford, UK.

Haseman, J.K., Kupper, L.L. (1979). Analysis of dichotomous response data from certain toxicological experiments. Biometrics, 35, 281-293.

Holmes, L.B. (1992). Fetal environmental toxins. Pediatrics in Review, 13, 10.

International Life Sciences Institute (1989). Interpretation and extrapolation of reproductive data to establish human safety standards. In Current Issues in Toxicology. Springer-Verlog: New York, 1-133.

IRLG (1986). Interagency Regulatory Liaison Group (eds. Kimmel, C.A., Kimmel, G.L., Frankos, V.H.). Interagency Regulatory Liaison Group Workshop on Reproductive Toxicity Risk Assessment. Environ. Health Perspect., 66, 193-221.

Johnson, N.L., Kotz, S. (1969). Discrete Distributions. New York: Wiley.

Kaplan, N., Hoel, D., Portier, C., Hogan, M. (1987). An evaluation of the safety factor approach in risk assessment. In Banbury Report 26: Developmental Toxicology: Mechanisms and Risk. Cold Spring Harbor Laboratory. Cold Spring Harbor, NY.

Kimmel, C.A. (1990). Quantitative approaches to human risk assessment for noncancer health effects. Neurotoxicology, 11, 201-210.

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

Kleinman, J.C. (1973). Proportions with extraneous variance: single and independent samples. JASA, 68, 46-54.

Kodell, R.L., Howe, R.B., Chen, J.J., Gaylor, D.W. (1991). Mathematical modeling of reproductive and developmental toxic effects for quantitative risk assessment. Risk Analysis, 11, 583-590.

Krewski, D., Zhu, Y. (1994). Applications of multinomial dose-response models in developmental toxicity risk assessment. Risk Analysis, 14, 613-627.

Krewski, D., Zhu, Y. (1995). A simple data transformation for estimating benchmark doses in developmental toxicity experiments. Risk Analysis, 15, 29-39.

Krewski, D., Zhu, Y., Fung, K.Y. (1994). Statistical Analysis of overdispersed multinomial data from developmental toxicity studies. In Statistics in Toxicity, ed. by B. Morgan, Oxford Univ. Press.

Kupper, L.L., Haseman, J.K. (1978). The use of correlated binomial model for the analysis of certain toxicological experiments. Biometrics, 34, 69-76.

Kupper, L.L., Portier, C., Hogan, M.D., Yamamoto, E. (1986). The impact of litter effects

on dose response modeling in teratology. Biometrics 42, 85-98.

Leisenring, W., Ryan, L.M. (1992). Statistical properties of the NOAEL. Regulatory Toxicology and Pharmacology, 15, 161-171.

Liang, K.Y., Henfalt, S.L. (1993). On the use of the quasi-likelihood method in teratology experiments. Technical Report, Dept. of Biostatistics, John Hopkins University, Baltimore, Maryland.

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

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

Moore, D.F. (1987). Modeling the extraneous variation in the presence of extra-binomial variation. Appl. Stat., 36, 8-14.

Nelder, J.A., Pregibon, D. (1987). An extended quasi-likelihood function. Biometrika, 74, 221-232.

Pack, S.E. (1986). Hypothesis testing for proportions with overdispersion. Biometrics, 42, 967-972.

Paul, S.R. (1982). Analysis of proportions of effected fetuses in teratological experiments. Biometrics, 38, 361-370.

Paul, S.R. (1987). On the beta correlated binomial (bcb). Distribution - a three parameter generalization of the binomial distribution. Commun. Stat. Theory Methods, 16, 1473-1478.

Prentice, R.L. (1986). Binary regression using an extended beta-binomial distribution,

with discussion of correlation induced by covariate measurement errors. J. Am. Stat. Assoc., 81, 321-327.

Rai, K., Van Ryzin, J. (1985). A generalized multihit dose-response model for low dose extrapolation. Biometrics, 39, 341-352.

Rao, J.N.K., Scott, A. (1992). A simple method for analysis of clustered binary data. Biometrics, 48, 577-585.

Rodricks, J.V., Tardiff, R.G., Brett, S.M., Putzrath, R.H., St. Hilaire, C. (1986). Elements of toxicology and chemical risk assessment. Environ. Corporation. Washington, D.C.

Ryan, L. (1992). Quantitative risk assessment for developmental toxicity. Biometrics, 48, 163-174.

Ryan, L. (1993). Using historical controls in the analysis of developmental toxicity data. Biometrics, 49, 1126-1135.

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

Smyth, G.K. (1989). Generalized linear models with varying dispersion. JRSS, B, 51, 47-60.

WHO (1984). World Health Organization. Principles for evaluating health risks to progeny associated with exposure to chemicals during pregnancy. Environmental Health Criteria 30. World Health Organization, Geneva.

Williams, D.A. (1975). The analysis of binary responses from toxicological experiments involving reproduction and teratogenicity. Biometrics, 31, 949-952.

Williams, D.A. (1982). Extra-binomial variation in logistic linear models. Appl. Stat., 31,

144-148.

Williams, D.A. (1987). Dose response models for teratological experiments. Biometrics, 43, 1013-1016.

Williams, D.A. (1988). Estimation bias using the beta-binomial distribution in teratology. Biometrics, 44, 305-309.

Zeger, S., Liang, K.Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. Biometrics, 42, 121-130.

Zhu, Y., Krewski, D., Ross, W.H. (1994). Dose-response models for correlated multinomial data from developmental toxicity studies. Appl. Statist., 43, 583-598.

Zhu, Y., Fung, K.Y. (1996). Statistical methods in developmental toxicity risk assessment. In Toxicology and Risk Assessment, (A. Fan, L.W. Chang, ed.), Marcel Dekker, Inc., New York, 413-445.

LIST OF TABLES AND FIGURES

toxicity.

Figure 5a BMDs from GLM and RST for the endpoint fetal malformation.

Figure 5b BMDs from GLM and RST for the endpoint prenatal death.

Figure 5c BMDs from GLM and RST for the endpoint overall toxicity.

Figure 6 Logistic Model fit to Bisphenol A (BPA) in Rats with Joint (___) and Separate Modeling (----) to the Three Endpoints. Diameter of circles are proportional to the squareroot of the number of replicates.

Table 1

Developmental Toxicity Data in Rats Exposed to Ethylene Glycol (EG)

Dose (mg/kg/day)			
0	**1250**	**2500**	**5000**
$(0,0,12)_2$	(1,0,11)	(0,0,1)	(4,2,6)
(0,2,12)	(2,1,11)	(3,0,6)	(5,1,6)
$(0,0,13)_4$	$(0,0,12)_4$	(2,1,9)	(2,0,10)
(0,2,13)	(0,1,12)	(1,0,10)	(9,2,11)
(2,0,13)	(1,0,12)	(4,1,11)	(0,12,12)
$(0,0,14)_4$	(1,1,12)	(0,0,12)	(1,11,12)
$(0,2,14)_2$	(4,1,12)	$(0,1,12)_2$	(6,4,12)
(3,1,14)	(0,0,13)	(2,3,12)	(8,1,12)
$(0,0,15)_4$	(0,1,13)	(4,1,12)	$(10,0,12)_2$
$(0,1,15)_2$	(0,2,13)	(7,0,12)	(10,3,13)
$(0,2,15)_2$	$(0,0,14)_2$	(9,2,12)	(3,2,14)
(0,0,16)	(0,1,14)	(0,0,13)	(9,3,14)
(0,1,16)	(1,2,14)	(0,1,13)	(10,3,14)
(0,0,17)	(2,0,14)	(1,0,13)	(11,3,14)
(0,3,17)	(0,0,15)	(1,1,13)	(12,2,14)
	$(0,1,15)_2$	(5,0,13)	(0,0,15)
	(0,3,15)	(7,0,13)	(3,0,15)
	(1,0,15)	(2,1,14)	(12,0,15)
	(2,0,15)	(3,3,14)	(13,1,15)
	(3,2,15)	(9,0,14)	(13,2,15)
	(3,7,15)	(1,0,15)	(1,2,16)
	(0,1,17)	(3,1,15)	(8,3,16)
	(0,0,18)	(9,3,15)	(10,2,16)
		(0,2,16)	(7,2,17)
		(4,0,16)	(17,0,17)
		$(0,1,17)_2$	(3,15,18)
		(9,0,17)	

Data $(y,r,m)_k$ represent y malforamtions and r deaths among m implants in a single litter; the subscript k denotes the multiplicity of litters producing the same response.

Table 2

Mean, Relative Bias (RBIAS), Relative MSE (RMSE) and P(<ED_{05}) of $E\hat{D}_{05}$ Based on Logistic Dose-Response Model with Dirichlet-Trinomial Covariance (DTC), Generalized Linear Model (GLM) and the Rao-Scott Transformation (RST) (The true ED_{05} for malformed is 0.254, for death is 0.497, overall toxicity is 0.201).

Number of Litters per Dose Group		MALFORMATION			PRENATAL DEATH			OVERALL TOXICITY		
		DTC	GLM	RST	DTC	GLM	RST	DTC	GLM	RST
(10,10,10,10)	MEAN	0.283	0.284	0.270	0.551	0.549	0.544	0.236	0.238	0.223
	RBIAS	0.116	0.120	0.067	0.110	0.105	0.095	0.177	0.182	0.109
	RMSE	0.0548	0.0562	0.0325	0.0972	0.0982	0.0979	0.0926	0.0951	0.0506
	P(<ED_{05})	0.312	0.298	0.367	0.436	0.430	0.456	0.245	0.244	0.323
(20,20,20,20)	MEAN	0.271	0.271	0.266	0.545	0.544	0.538	0.220	0.219	0.215
	RBIAS	0.071	0.067	0.049	0.098	0.095	0.083	0.097	0.089	0.068
	RMSE	0.0224	0.0204	0.0170	0.0753	0.0777	0.0695	0.0328	0.0291	0.0225
	P(<ED_{05})	0.308	0.315	0.357	0.418	0.437	0.442	0.259	0.286	0.318
(30,30,30,30)	MEAN	0.264	0.265	0.261	0.551	0.545	0.548	0.213	0.214	0.210
	RBIAS	0.041	0.045	0.028	0.110	0.097	0.103	0.062	0.065	0.046
	RMSE	0.0110	0.0117	0.0095	0.0696	0.0613	0.0673	0.0143	0.0160	0.0119
	P(<ED_{05})	0.349	0.340	0.407	0.389	0.405	0.394	0.292	0.292	0.334
(30,30,10,10)	MEAN	0.268	0.267	0.264	0.542	0.547	0.535	0.222	0.221	0.216
	RBIAS	0.057	0.052	0.042	0.091	0.101	0.077	0.104	0.100	0.074
	RMSE	0.0186	0.0176	0.0151	0.0971	0.1052	0.0943	0.0389	0.0360	0.0263
	P(<ED_{05})	0.344	0.345	0.375	0.472	0.462	0.498	0.271	0.281	0.322
(30,10,10,30)	MEAN	0.274	0.274	0.267	0.554	0.552	0.546	0.221	0.221	0.214
	RBIAS	0.082	0.083	0.052	0.115	0.112	0.100	0.100	0.100	0.066
	RMSE	0.0321	0.0319	0.0225	0.0761	0.0749	0.0727	0.0358	0.0352	0.0240
	P(<ED_{05})	0.336	0.333	0.392	0.375	0.391	0.422	0.292	0.293	0.352
(10,30,30,10)	MEAN	0.269	0.269	0.264	0.532	0.536	0.532	0.222	0.222	0.216
	RBIAS	0.063	0.062	0.043	0.072	0.080	0.072	0.104	0.106	0.074
	RMSE	0.0159	0.0151	0.0121	0.0828	0.0872	0.0852	0.0336	0.0354	0.0239
	P(<ED_{05})	0.294	0.281	0.355	0.479	0.488	0.494	0.257	0.248	0.302

Table 3

Mean and P(<ED_{05}) of $\hat{BMD}_{05}$ Under the Logistic Dose-Response Model with DTC, GLM and RST Compared to the NOAEL.

(The true ED_{05} for malformed is 0.254, for death is 0.497, overall toxicity is 0.201).

Number of Litters per Dose Group		MALFORMATION				PRENATAL DEATH				OVERALL TOXICITY			
		DTC	GLM	RST	NOAEL[a]	DTC	GLM	RST	NOAEL[a]	DTC	GLM	RST	NOAEL[a]
(10,10,10,10)	MEAN	0.220	0.195	0.210	0.151	0.242	0.212	0.228	0.396	0.177	0.150	0.167	0.148
	P(<ED_{05})	0.782	0.931	0.875	0.989	0.978	0.998	0.996	0.478	0.779	0.946	0.882	0.479
(20,20,20,20)	MEAN	0.226	0.217	0.221	0.086	0.301	0.278	0.295	0.285	0.181	0.172	0.178	0.088
	P(<ED_{05})	0.822	0.946	0.878	1.000	0.999	1.000	1.000	0.685	0.833	0.955	0.885	0.719
(30,30,30,30)	MEAN	0.227	0.226	0.224	0.056	0.343	0.333	0.338	0.223	0.182	0.180	0.179	0.059
	P(<ED_{05})	0.881	0.911	0.907	1.000	1.000	0.999	1.000	0.796	0.864	0.889	0.894	0.847
(30,30,10,10)	MEAN	0.221	0.219	0.218	0.059	0.232	0.220	0.230	0.301	0.176	0.173	0.172	0.061
	P(<ED_{05})	0.891	0.916	0.921	0.997	0.996	0.998	0.999	0.625	0.866	0.898	0.926	0.844
(30,10,10,30)	MEAN	0.222	0.222	0.216	0.110	0.332	0.330	0.324	0.284	0.179	0.179	0.173	0.115
	P(<ED_{05})	0.831	0.824	0.867	0.998	0.995	0.995	0.999	0.632	0.819	0.826	0.888	0.602
(10,30,30,10)	MEAN	0.229	0.222	0.224	0.109	0.256	0.233	0.247	0.378	0.179	0.171	0.175	0.110
	P(<ED_{05})	0.827	0.914	0.892	1.000	0.999	1.000	1.000	0.563	0.857	0.927	0.916	0.625

[a] NOAEL=$d_2/10$ when NOAEL is equal to zero

Table 4

Mean, Relative Bias (RBIAS), Relative MSE (RMSE) and P(<ED_{05}) of $\hat{ED}_{05}$ Based on Logistic Dose-Response Model with Dirichlet-Trinomial Covariance (DTC), Generalized Linear Model (GLM) and the Rao-Scott Transformation (RST) (The true ED_{05} for malformed is 0.254, for death is 0.497, overall toxicity is 0.201).

Number of litters per dose group and placement of dose		MALFORMATION			PRENATAL DEATH			OVERALL TOXICITY		
		DTC	GLM	RST	DTC	GLM	RST	DTC	GLM	RST
(27,26,27)	MEAN	0.270	0.271	0.266	0.543	0.544	0.541	0.218	0.218	0.214
(0,0.5,1)	RBIAS	0.067	0.068	0.050	0.093	0.095	0.089	0.083	0.084	0.065
	RMSE	0.0266	0.0270	0.0224	0.0565	0.0583	0.0568	0.0290	0.0296	0.0243
	P(<ED_{05})	0.374	0.359	0.382	0.402	0.400	0.417	0.308	0.296	0.330
(27,26,27)	MEAN	0.274	0.275	0.268	0.549	0.550	0.547	0.221	0.221	0.216
(0,0.3,1)	RBIAS	0.082	0.083	0.058	0.105	0.108	0.102	0.099	0.100	0.074
	RMSE	0.0360	0.0369	0.0276	0.0568	0.0602	0.0594	0.0378	0.0386	0.0282
	P(<ED_{05})	0.338	0.345	0.385	0.346	0.338	0.354	0.294	0.292	0.342
(20,20,20,20)	MEAN	0.279	0.279	0.269	0.547	0.547	0.540	0.229	0.226	0.217
(0,0.25,0.5,1)	RBIAS	0.100	0.100	0.060	0.102	0.103	0.088	0.124	0.124	0.079
	RMSE	0.0409	0.0415	0.0270	0.0670	0.0668	0.0647	0.0502	0.0507	0.0312
	P(<ED_{05})	0.303	0.304	0.372	0.377	0.384	0.412	0.269	0.267	0.324
(20,20,20,20)	MEAN	0.282	0.283	0.271	0.552	0.551	0.547	0.229	0.229	0.219
(0,0.15,0.35,1)	RBIAS	0.114	0.116	0.069	0.111	0.110	0.102	0.138	0.141	0.090
	RMSE	0.0508	0.0517	0.0314	0.0840	0.0816	0.0845	0.0581	0.0589	0.0360
	P(<ED_{05})	0.283	0.278	0.356	0.416	0.407	0.426	0.249	0.238	0.316
(16,16,16,16,16)	MEAN	0.293	0.294	0.276	0.562	0.565	0.551	0.241	0.242	0.225
(0,0.125,0.25,0.5,1)	RBIAS	0.156	0.158	0.087	0.131	0.138	0.109	0.198	0.202	0.118
	RMSE	0.0696	0.0715	0.0349	0.0914	0.0985	0.0892	0.0929	0.0959	0.0452
	P(<ED_{05})	0.233	0.228	0.317	0.366	0.366	0.409	0.183	0.186	0.259
(16,16,16,16,16)	MEAN	0.288	0.288	0.272	0.557	0.555	0.542	0.235	0.235	0.220
(0,0.125,0.4,0.75,1)	RBIAS	0.134	0.136	0.072	0.121	0.117	0.091	0.167	0.169	0.093
	RMSE	0.0637	0.0634	0.0329	0.0833	0.0799	0.0730	0.0794	0.0792	0.0375
	P(<ED_{05})	0.281	0.277	0.355	0.365	0.378	0.424	0.219	0.218	0.298

Table 5

Mean and P(<ED_{05}) of $\hat{BMD}_{05}$ Under the Logistic Dose-Response Model with DTC, GLM and RST Compared to the NOAEL.
(The true ED_{05} for malformed is 0.254, for death is 0.497, overall toxicity is 0.201).

Number of Litters per Dose Group and Placement of Dose		MALFORMATION				PRENATAL DEATH				OVERALL TOXICITY			
		DTC	GLM	RST	NOAEL[a]	DTC	GLM	RST	NOAEL[a]	DTC	GLM	RST	NOAEL[a]
(27,26,27)	MEAN	0.215	0.215	0.211	0.050	0.359	0.358	0.354	0.208	0.174	0.174	0.171	0.050
(0,0.5,1)	P(<ED_{05})	0.862	0.856	0.891	1.000	0.988	0.987	0.990	0.660	0.863	0.863	0.897	1.000
(27,26,27)	MEAN	0.216	0.216	0.211	0.063	0.358	0.356	0.351	0.184	0.175	0.175	0.171	0.066
(0,0.3,1)	P(<ED_{05})	0.841	0.835	0.886	0.877	0.987	0.986	0.989	0.995	0.831	0.823	0.871	0.866
(20,20,20,20)	MEAN	0.222	0.222	0.214	0.098	0.321	0.320	0.309	0.257	0.179	0.179	0.171	0.100
(0,0.25,0.5,1)	P(<ED_{05})	0.795	0.792	0.871	0.998	0.980	0.981	0.992	0.690	0.802	0.803	0.874	0.669
(20,20,20,20)	MEAN	0.222	0.222	0.212	0.108	0.313	0.312	0.299	0.207	0.179	0.179	0.171	0.106
(0,0.15,0.35,1)	P(<ED_{05})	0.806	0.805	0.880	0.927	0.990	0.990	0.996	0.972	0.797	0.785	0.876	0.922
(16,16,16,16,16)	MEAN	0.231	0.232	0.216	0.131	0.291	0.288	0.264	0.214	0.187	0.187	0.173	0.123
(0,0.125,0.25,0.5,1)	P(<ED_{05})	0.733	0.722	0.848	0.986	0.973	0.972	0.998	0.787	0.714	0.709	0.849	0.712
(16,16,16,16,16)	MEAN	0.226	0.227	0.212	0.103	0.316	0.317	0.301	0.245	0.183	0.183	0.170	0.105
(0,0.125,0.4,0.75,1)	P(<ED_{05})	0.756	0.741	0.856	0.944	0.974	0.973	0.992	0.887	0.744	0.735	0.862	0.936

[a] NOAEL=$d_2/10$ when NOAEL is equal to zero

Table 6

Mean, Relative Bias (RBIAS), Relative MSE (RMSE) and P(<ED_{05}) of $\hat{ED}_{05}$ Based on Logistic Dose-Response Model with Dirichlet-Trinomial Covariance (DTC), Generalized Linear Model (GLM) and the Rao-Scott Transformation (RST) (The true ED_{05} for malformed is 0.254, for death is 0.497, overall toxicity is 0.201).

ϕ_i per Dose Group		MALFORMATION			PRENATAL DEATH			OVERALL TOXICITY		
		DTC	GLM	RST	DTC	GLM	RST	DTC	GLM	RST
(0,0,0,0)	MEAN	0.262	0.262	0.259	0.513	0.521	0.513	0.208	0.209	0.206
	RBIAS	0.032	0.032	0.021	0.033	0.050	0.032	0.035	0.040	0.023
	RMSE	0.0071	0.0070	0.0059	0.0127	0.0147	0.0129	0.0068	0.0072	0.0056
	P(<ED_{05})	0.361	0.364	0.412	0.422	0.358	0.428	0.343	0.323	0.392
(0.1,0.1,0.1,0.1)	MEAN	0.269	0.269	0.263	0.536	0.536	0.534	0.216	0.216	0.211
	RBIAS	0.061	0.061	0.039	0.079	0.080	0.075	0.074	0.075	0.051
	RMSE	0.0201	0.0200	0.0154	0.0380	0.0395	0.0388	0.0216	0.0216	0.0161
	P(<ED_{05})	0.335	0.322	0.380	0.365	0.375	0.386	0.295	0.287	0.346
(0.4,0.4,0.4,0.4)	MEAN	0.294	0.294	0.278	0.565	0.565	0.553	0.242	0.243	0.227
	RBIAS	0.158	0.161	0.097	0.137	0.137	0.114	0.203	0.207	0.130
	RMSE	0.0842	0.0860	0.0497	0.0844	0.0827	0.0796	0.1111	0.1139	0.0623
	P(<ED_{05})	0.281	0.289	0.358	0.322	0.330	0.384	0.248	0.245	0.303
(0.1,0.2,0.3,0.4)	MEAN	0.276	0.276	0.267	0.551	0.551	0.543	0.224	0.224	0.219
	RBIAS	0.090	0.090	0.055	0.109	0.110	0.093	0.114	0.115	0.074
	RMSE	0.0362	0.0366	0.0251	0.0754	0.0772	0.0716	0.0424	0.0434	0.0286
	P(<ED_{05})	0.328	0.326	0.398	0.381	0.387	0.416	0.267	0.268	0.343

Table 7

Mean and P(<ED_{05}) of $\hat{BMD}_{05}$ Under the Logistic Dose-Response Model with DTC, GLM and RST Compared to the NOAEL.
(The true ED_{05} for malformed is 0.254, for death is 0.497, overall toxicity is 0.201).

ϕ_i per Dose Group		MALFORMATION				PRENATAL DEATH				OVERALL TOXICITY			
		DTC	GLM	RST	NOAEL[a]	DTC	GLM	RST	NOAEL[a]	DTC	GLM	RST	NOAEL[a]
(0,0,0,0)	MEAN	0.231	0.232	0.183	0.075	0.431	0.436	0.430	0.276	0.185	0.186	0.183	0.077
	P(<ED_{05})	0.880	0.880	0.911	1.000	0.974	0.969	0.976	0.696	0.881	0.861	0.911	0.767
(0.1,0.1,0.1,0.1)	MEAN	0.225	0.225	0.219	0.091	0.395	0.395	0.389	0.257	0.181	0.181	0.177	0.096
	P(<ED_{05})	0.842	0.839	0.882	1.000	0.968	0.966	0.980	0.689	0.826	0.829	0.883	0.686
(0.4,0.4,0.4,0.4)	MEAN	0.222	0.223	0.209	0.094	0.330	0.334	0.305	0.206	0.182	0.182	0.169	0.102
	P(<ED_{05})	0.739	0.734	0.831	0.995	0.941	0.930	0.974	0.766	0.719	0.710	0.812	0.669
(0.1,0.2,0.3,0.4)	MEAN	0.220	0.220	0.212	0.096	0.318	0.317	0.307	0.254	0.177	0.177	0.171	0.097
	P(<ED_{05})	0.813	0.816	0.881	0.996	0.992	0.991	0.998	0.707	0.813	0.812	0.886	0.686

[a] NOAEL=$d_2/10$ when NOAEL is equal to zero

Table 8

Mean, Relative Bias (RBIAS), Relative MSE (RMSE) and P(<ED_{05}) of $\hat{ED}_{05}$ Based on Separate Modeling with Logistic Dose-Response Model, with the Beta-Binomial variance (BB), the Generalized Linear Model (GLM) and Rao-Scott Transformation (RST) for Binary Outcomes

(The true ED_{05} for malformed is 0.254, for death is 0.497, overall toxicity is 0.201).

Number of Litters per Dose Group		MALFORMATION			PRENATAL DEATH			OVERALL TOXICITY		
		BB	GLM	RST	BB	GLM	RST	BB	GLM	RST
(10,10,10,10)	MEAN	0.284	0.286	0.273	0.559	0.559	0.549	0.193	0.194	0.182
	RBIAS	0.120	0.129	0.079	0.126	0.126	0.106	-0.041	-0.036	-0.093
	RMSE	0.0574	0.0583	0.0356	0.0953	0.0996	0.1023	0.0397	0.0404	0.0304
	P(<ED_{05})	0.290	0.277	0.352	0.385	0.399	0.438	0.659	0.646	0.765
(20,20,20,20)	MEAN	0.277	0.277	0.268	0.556	0.553	0.544	0.183	0.183	0.178
	RBIAS	0.092	0.094	0.057	0.121	0.113	0.095	-0.089	-0.090	-0.114
	RMSE	0.0277	0.0282	0.0184	0.0827	0.0811	0.0774	0.0198	0.0200	0.0224
	P(<ED_{05})	0.267	0.254	0.345	0.388	0.403	0.434	0.816	0.809	0.893
(30,30,30,30)	MEAN	0.269	0.270	0.262	0.561	0.554	0.551	0.179	0.179	0.175
	RBIAS	0.060	0.063	0.034	0.129	0.117	0.109	-0.109	-0.108	-0.127
	RMSE	0.0149	0.0154	0.0100	0.0782	0.0696	0.0702	0.0189	0.0187	0.0218
	P(<ED_{05})	0.299	0.290	0.386	0.355	0.373	0.388	0.898	0.906	0.940
(30,30,10,10)	MEAN	0.273	0.272	0.265	0.557	0.558	0.550	0.178	0.177	0.278
	RBIAS	0.075	0.074	0.047	0.123	0.123	0.109	-0.113	-0.118	0.383
	RMSE	0.0230	0.0254	0.0163	0.1117	0.1116	0.1112	0.0235	0.0240	0.3790
	P(<ED_{05})	0.308	0.305	0.370	0.438	0.430	0.472	0.877	0.887	0.096
(30,10,10,30)	MEAN	0.285	0.285	0.269	0.566	0.566	0.552	0.175	0.176	0.195
	RBIAS	0.125	0.126	0.063	0.140	0.139	0.111	-0.127	-0.126	-0.029
	RMSE	0.0504	0.0506	0.0249	0.0833	0.0833	0.0752	0.0337	0.0332	0.0183
	P(<ED_{05})	0.272	0.267	0.367	0.334	0.333	0.397	0.864	0.862	0.622
(10,30,30,10)	MEAN	0.271	0.272	0.266	0.548	0.547	0.532	0.189	0.189	0.271
	RBIAS	0.069	0.071	0.048	0.103	0.101	0.071	-0.059	-0.061	0.346
	RMSE	0.0176	0.0175	0.0128	0.1011	0.0990	0.0833	0.0148	0.0140	0.4012
	P(<ED_{05})	0.282	0.267	0.338	0.451	0.467	0.492	0.754	0.741	0.147

Table 9

Mean and P(<ED_{05}) of $\hat{BMD}_{05}$, Under Separate Modeling with the Logistic Dose-Response Model with BB variance, GLM and RST for Binary outcomes.
(The true ED_{05} for malformed is 0.254, for death is 0.497, overall toxicity is 0.201).

Number of Litters per Dose Group		MALFORMATION			PRENATAL DEATH			OVERALL TOXICITY		
		BB	GLM	RST	BB	GLM	RST	BB	GLM	RST
(10,10,10,10)	MEAN	0.221	0.197	0.214	0.263	0.220	0.235	0.151	0.136	0.143
	P(<ED_{05})	0.794	0.932	0.844	0.969	0.995	0.996	0.935	0.986	0.977
(20,20,20,20)	MEAN	0.231	0.218	0.224	0.307	0.279	0.297	0.154	0.149	0.150
	P(<ED_{05})	0.761	0.918	0.845	0.996	0.999	1.000	0.984	0.999	0.995
(30,30,30,30)	MEAN	0.232	0.228	0.226	0.347	0.333	0.340	0.155	0.154	0.153
	P(<ED_{05})	0.825	0.885	0.891	0.998	0.998	1.000	0.995	0.998	0.999
(30,30,10,10)	MEAN	0.227	0.223	0.220	0.231	0.220	0.230	0.149	0.147	0.124
	P(<ED_{05})	0.840	0.864	0.902	0.997	0.999	1.000	0.992	0.996	1.000
(30,10,10,30)	MEAN	0.232	0.232	0.219	0.343	0.342	0.328	0.145	0.145	0.160
	P(<ED_{05})	0.732	0.742	0.848	0.978	0.982	0.998	0.980	0.983	0.970
(10,30,30,10)	MEAN	0.230	0.223	0.226	0.260	0.232	0.255	0.161	0.156	0.119
	P(<ED_{05})	0.838	0.906	0.875	1.000	1.000	1.000	0.980	0.993	0.994

Table 10

Mean, Relative Bias (RBIAS), Relative MSE (RMSE) and P(<ED_{05}) of $\hat{ED}_{05}$ Based on Separate Modeling with Logistic Dose-Response Model, with the Beta-Binomial variance (BB), the Generalized Linear Model (GLM) and Rao-Scott Transformation (RST) for Binary Outcomes

(The true ED_{05} for malformed is 0.254, for death is 0.497, overall toxicity is 0.201).

Number of litters per dose group and placement of dose		MALFORMATION			PRENATAL DEATH			OVERALL TOXICITY		
		BB	GLM	RST	BB	GLM	RST	BB	GLM	RST
(27,26,27)	MEAN	0.275	0.276	0.269	0.548	0.549	0.543	0.181	0.182	0.174
(0,0.5,1)	RBIAS	0.085	0.087	0.060	0.104	0.105	0.094	-0.097	-0.095	-0.133
	RMSE	0.0313	0.0319	0.0247	0.0585	0.0596	0.0583	0.0295	0.0294	0.0311
	P(<ED_{05})	0.329	0.320	0.362	0.382	0.378	0.414	0.786	0.792	0.880
(27,26,27)	MEAN	0.282	0.282	0.272	0.551	0.552	0.549	0.177	0.177	0.171
(0,0.3,1)	RBIAS	0.111	0.114	0.072	0.110	0.112	0.105	-0.121	-0.120	-0.151
	RMSE	0.0459	0.0475	0.0309	0.0572	0.0579	0.0594	0.0326	0.0327	0.0365
	P(<ED_{05})	0.286	0.288	0.353	0.330	0.318	0.349	0.840	0.837	0.898
(20,20,20,20)	MEAN	0.288	0.289	0.274	0.560	0.559	0.547	0.187	0.188	0.182
(0,0.25,0.5,1)	RBIAS	0.134	0.138	0.080	0.127	0.126	0.102	-0.068	-0.066	-0.092
	RMSE	0.0548	0.0569	0.0323	0.0706	0.0716	0.0697	0.0293	0.0304	0.0330
	P(<ED_{05})	0.259	0.243	0.334	0.331	0.329	0.403	0.727	0.726	0.779
(20,20,20,20)	MEAN	0.296	0.297	0.278	0.564	0.562	0.549	0.183	0.183	0.189
(0,0.15,0.35,1)	RBIAS	0.167	0.172	0.098	0.135	0.131	0.106	-0.090	-0.088	-0.061
	RMSE	0.0713	0.0738	0.0396	0.0883	0.0832	0.0801	0.0329	0.0329	0.0387
	P(<ED_{05})	0.222	0.219	0.300	0.365	0.348	0.400	0.765	0.754	0.688
(16,16,16,16,16)	MEAN	0.311	0.312	0.286	0.580	0.577	0.564	0.195	0.195	0.217
(0,0.125,0.25,0.5,1)	RBIAS	0.226	0.229	0.127	0.167	0.162	0.136	-0.030	-0.028	0.078
	RMSE	0.1071	0.1098	0.0478	0.1048	0.0998	0.1023	0.0431	0.0432	0.1102
	P(<ED_{05})	0.166	0.163	0.244	0.317	0.319	0.367	0.657	0.655	0.464
(16,16,16,16,16)	MEAN	0.298	0.299	0.280	0.578	0.578	0.552	0.189	0.189	0.181
(0,0.125,0.4,0.75,1)	RBIAS	0.175	0.180	0.105	0.164	0.163	0.112	-0.060	-0.059	-0.097
	RMSE	0.0771	0.0808	0.0423	0.0990	0.0987	0.0809	0.0366	0.0374	0.0360
	P(<ED_{05})	0.216	0.220	0.292	0.293	0.297	0.400	0.712	0.715	0.766

Table 11

Mean and P(<ED_{05}) of $\hat{BMD}_{05}$, Under Separate Modeling with the Logistic Dose-Response Model with BB variance, GLM and RST for Binary outcomes.
(The true ED_{05} for malformed is 0.254, for death is 0.497, overall toxicity is 0.201).

Number of Litters per Dose Group and Placement of Dose		MALFORMATION			PRENATAL DEATH			OVERALL TOXICITY		
		BB	GLM	RST	BB	GLM	RST	BB	GLM	RST
(27,26,27)	MEAN	0.220	0.220	0.215	0.366	0.366	0.357	0.144	0.144	0.139
(0,0.5,1)	P(<ED_{05})	0.822	0.818	0.869	0.974	0.974	0.988	0.983	0.981	0.997
(27,26,27)	MEAN	0.223	0.223	0.215	0.368	0.367	0.355	0.142	0.142	0.137
(0,0.3,1)	P(<ED_{05})	0.779	0.768	0.846	0.969	0.969	0.988	0.986	0.985	0.997
(20,20,20,20)	MEAN	0.231	0.232	0.220	0.342	0.340	0.315	0.150	0.151	0.141
(0,0.25,0.5,1)	P(<ED_{05})	0.718	0.716	0.809	0.953	0.958	0.989	0.960	0.956	0.993
(20,20,20,20)	MEAN	0.234	0.235	0.221	0.336	0.337	0.310	0.146	0.146	0.142
(0,0.15,0.35,1)	P(<ED_{05})	0.688	0.677	0.816	0.968	0.968	0.997	0.962	0.962	0.991
(16,16,16,16,16)	MEAN	0.248	0.249	0.228	0.320	0.323	0.277	0.155	0.155	0.134
(0,0.125,0.25,0.5,1)	P(<ED_{05})	0.585	0.584	0.751	0.941	0.941	0.996	0.914	0.913	0.983
(16,16,16,16,16)	MEAN	0.237	0.237	0.222	0.345	0.345	0.315	0.149	0.149	0.139
(0,0.125,0.4,0.75,1)	P(<ED_{05})	0.671	0.656	0.772	0.941	0.941	0.990	0.945	0.944	0.989

Table 12

Mean, Relative Bias (RBIAS), Relative MSE (RMSE) and P(<ED_{05}) of $\hat{ED}_{05}$ Based on Separate Modeling with Logistic Dose-Response Model, with the Beta-Binomial variance (BB), the Generalized Linear Model (GLM) and Rao-Scott Transformation (RST) for Binary Outcomes

(The true ED_{05} for malformed is 0.254, for death is 0.497, overall toxicity is 0.201).

ϕ_i per Dose Group		MALFORMATION			PRENATAL DEATH			OVERALL TOXICITY		
		BB	GLM	RST	BB	GLM	RST	BB	GLM	RST
(0,0,0,0)	MEAN	0.264	0.264	0.259	0.514	0.522	0.513	0.179	0.180	0.155
	RBIAS	0.041	0.042	0.021	0.036	0.052	0.033	-0.108	-0.103	-0.227
	RMSE	0.0083	0.0083	0.0060	0.0126	0.0147	0.0130	0.0166	0.0155	0.103
	P(<ED_{05})	0.324	0.321	0.407	0.416	0.342	0.431	0.938	0.925	0.955
(0.1,0.1,0.1,0.1)	MEAN	0.278	0.276	0.266	0.539	0.539	0.535	0.185	0.186	0.161
	RBIAS	0.088	0.090	0.050	0.085	0.085	0.077	-0.077	-0.077	-0.199
	RMSE	0.0266	0.0269	0.0170	0.0379	0.0380	0.0395	0.0187	0.0186	0.0812
	P(<ED_{05})	0.278	0.266	0.347	0.351	0.352	0.381	0.779	0.774	0.926
(0.4,0.4,0.4,0.4)	MEAN	0.298	0.303	0.293	0.576	0.578	0.561	0.203	0.204	0.169
	RBIAS	0.174	0.193	0.157	0.160	0.165	0.130	0.011	0.013	-0.159
	RMSE	0.0840	0.0961	0.0735	0.0861	0.0891	0.0799	0.0519	0.0524	0.117
	P(<ED_{05})	0.252	0.240	0.268	0.274	0.271	0.340	0.553	0.541	0.844
(0.1,0.2,0.3,0.4)	MEAN	0.286	0.286	0.272	0.561	0.559	0.550	0.186	0.186	0.181
	RBIAS	0.126	0.129	0.075	0.131	0.125	0.108	-0.077	-0.075	-0.099
	RMSE	0.0507	0.0515	0.0298	0.0808	0.0759	0.0779	0.0261	0.0270	0.0404
	P(<ED_{05})	0.272	0.265	0.352	0.332	0.346	0.400	0.726	0.720	0.790

Table 13

Mean and P(<ED_{05}) of $\hat{BMD}_{05}$, Under Separate Modeling with the Logistic Dose-Response Model with BB variance, GLM and RST for Binary outcomes.
(The true ED_{05} for malformed is 0.254, for death is 0.497, overall toxicity is 0.201).

ϕ_i per Dose Group		MALFORMATION			PRENATAL DEATH			OVERALL TOXICITY		
		BB	GLM	RST	BB	GLM	RST	BB	GLM	RST
(0,0,0,0)	MEAN	0.234	0.234	0.229	0.434	0.439	0.430	0.159	0.160	0.121
	P(<ED_{05})	0.851	0.850	0.909	0.968	0.960	0.977	0.995	0.995	1.000
(0.1,0.1,0.1,0.1)	MEAN	0.232	0.232	0.223	0.405	0.405	0.393	0.155	0.156	0.119
	P(<ED_{05})	0.761	0.754	0.857	0.954	0.950	0.974	0.982	0.980	1.000
(0.4,0.4,0.4,0.4)	MEAN	0.227	0.232	0.228	0.360	0.360	0.329	0.15	0.152	0.118
	P(<ED_{05})	0.678	0.649	0.695	0.866	0.863	0.947	0.898	0.893	0.988
(0.1,0.2,0.3,0.4)	MEAN	0.229	0.230	0.219	0.338	0.339	0.316	0.149	0.149	0.141
	P(<ED_{05})	0.740	0.723	0.832	0.962	0.964	0.993	0.973	0.970	0.994

Table 14

Overview of Experimental Design for 28 Developmental Toxicity Studies

Study No. and Lab	Compound	Species	Dose units	Dose levels (dams per dose group)
1 NTP	Codeine (COD)	hamster	mg/kg	0,10,50,150 (19,23,26,22)
2 NTP	Chlorpromazine (CPZ)	Mice	mg/kg	0,2.5,5,15,30 (27,25,24,29,24)
3 NTP	Diethylhexalphthalate (DEHP)	Mice	% in diet	0,0.03,0.05,0.1,0.15 (30,26,26,24,25)
4 NTP	Ethylene glycol diethyl ether (EGDE)	Mice	mg/kg	0,50,150,500,1000 (23,24,22,23,23)
5 NTP	Theophylline (THEO)	Mice	% in diet	0,0.08,0.15,0.20 (26,26,33,23)
6 NTP	Carbon disulfide (CS)	Rabbits	mg/kg	0,25,75,150 (27,23,28,25)
7 NTP	EGDE	Rabbits	mg/kg	0,25,50,100 (28,32,26,31)
8 NTP	Gentian violet (GV)	Rabbits	mg/kg	0,0.5,1,2 (26,19,21,23)
9 WIL	W023	Rabbits	mg/kg	0,1,5,8 (17,18,17,20)
10 WIL	W025	Rabbits	mg/kg	0,100,300,500 (14,16,15,9)
11 WIL	W060	Rabbits	mg/kg	0,0.2,0.5,1.5 (15,14,14,14)
12 WIL	W062	Rabbits	mg/kg	0,2.5,10,40 (15,16,17,17)
13 WIL	W073	Rabbits	mg/kg	0,12,25,50 (11,14,14,12)
14 WIL	W094	Rabbits	mg/kg	0,15,35,75 (16,14,12,5)

Table 14 (continued)

Overview of Experimental Design for 28 Developmental Toxicity Studies

Study No. and Labs	Compound	Species	Dose units	Dose levels (dams per dose group)
15 WIL	W095	Rabbits	mg/kg	0,15,35,75 (16,12,12,6)
16 WIL	W101	Rabbits	mg/kg	0,0.25,1,4 (17,20,18,10)
17 WIL	W105	Rabbits	mg/kg	0,5,25,70 (19,17,12,7)
18 NTP	Diethylene glycol dimethyl ether (DYME)	Rabbits	mg/kg	0,25,50,100,175 (22,17,15,21,17)
19 NTP	Nitrofurazone (NF)	Rabbits	mg/kg	0,5,10,15,20 (25,23,27,22,24)
20 NTP	Triethylene glycol dimethyl ether (TGDM)	Rabbits	mg/kg	0,75,125,175,250 (25,22,25,23,23)
21 WIL	W103	Rabbits	mg/kg	0,5,10,30,60 (13,13,7,13,11)
22 NTP	Bisphenol A (BPA)	Rats	mg/kg	0,160,320,640,1280 (23,26,24,29,18)
23 NTP	DEHP	Rats	% in diet	0,0.5,1,1.5,2 (24,23,22,24,25)
24 NTP	L-5-hydroxytryptophan (HTP)	Rats	mg/kg	0,50,100,150,300 (35,33,34,35,34)
25 NTP	Sulfamethezine (SM)	Rats	mg/kg	0,545,685,865 (26,22,24,24)
26 WIL	W037	Rats	mg/kg	0,50,150,450 (22,20,21,20)
27 WIL	W092	Rats	mg/kg	0,10,50,175 (22,18,22,21)
28 NTP	Ethylene glycol (EG)	Rats	mg/kg	0,1250,2500,5000 (28,28,29,27)

Table 15

$\hat{BMD}_{05}$ Based on Trinomial and Binomial Dose Response Models Using GLM, and Estimated NOAEL
(Standard Error in Parentheses)

Study No.	Unit	MALFORMATION			PRENATAL DEATH			OVERALL TOXICITY		
		Trinomial	Binomial	NOAEL	Trinomial	Binomial	NOAEL	Trinomial	Binomial	NOAEL
1	mg/kg	-----[b]	-----[b]	0.3333	0.2493 (0.0412)	0.2571 (0.0425)	0.0067[a]	0.2356 (0.0361)	0.2335 (0.0356)	0.0670
2	mg/kg	0.7014 (0.0828)	0.6897 (0.0871)	0.5000	0.2256 (0.0782)	0.2149 (0.1152)	0.5000	0.2191 (0.0592)	0.2207 (0.0615)	0.5000
3	% in diet	0.2556 (0.0237)	0.2750 (0.0243)	0.2000	0.1514 (0.0126)	0.1663 (0.0126)	0.0200[a]	0.1221 (0.0105)	0.1197 (0.0110)	0.0200[a]
4	mg/kg	0.3091 (0.0302)	-----[b]	0.0500	0.2472 (0.0436)	0.2513 (0.0438)	0.1500	0.1911 (0.0221)	0.1204 (0.0130)	0.1500
5	% in diet	-----[b]	-----[b]	0.4000	0.2477 (0.0559)	0.2475 (0.0550)	0.0400[a]	0.2389 (0.0490)	0.2377 (0.0466)	0.0400[a]
6	mg/kg	0.1606 (0.4132)	0.2741 (0.3278)	1.0000	0.1031 (0.0173)	0.1057 (0.0173)	0.0167[a]	0.0924 (0.0161)	0.0864 (0.0148)	0.0167[a]
7	mg/kg	0.2527 (0.0279)	0.2562 (0.0290)	0.2500	0.2111 (0.0418)	0.2060 (0.0439)	0.5000	0.1386 (0.0173)	0.1218 (0.0164)	0.2500
8	mg/kg	-----[b]	-----[b]	1.0000	0.1837 (0.1262)	0.1857 (0.1240)	0.5000	0.0940 (0.1937)	0.0906 (0.1767)	0.5000
9	mg/kg	0.5978 (0.0396)	0.5999 (0.0400)	0.6250	0.5240 (0.1017)	0.1852 (0.4116)	0.6250	0.4580 (0.0620)	0.4668 (0.0570)	0.6250
10	mg/kg	-----[b]	-----[b]	1.0000	0.2346 (0.1334)	0.2340 (0.1403)	0.0200[a]	0.1220 (0.2185)	0.1527 (0.1529)	0.0200[a]

Table 15 (continued)

$B\hat{M}D_{05}$ Based on Trinomial and Binomial Dose Response Models Using GLM, and Estimated NOAEL
(Standard Error in Parentheses)

Study No.	Unit	MALFORMATION			PRENATAL DEATH			OVERALL TOXICITY		
		Trinomial	Binomial	NOAEL	Trinomial	Binomial	NOAEL	Trinomial	Binomial	NOAEL
11	mg/kg	-----[b]	-----[b]	1.0000	0.2600 (0.2015)	0.2100 (0.2696)	0.3333	0.1831 (0.1625)	0.1540 (0.1778)	0.3333
12	mg/kg	-----[b]	-----[b]	1.0000	-----[b]	-----[b]	0.0625	0.6071 (0.2197)	0.5636 (0.2400)	0.0625
13	mg/kg	-----[b]	-----[b]	1.0000	0.2844 (0.1020)	0.2865 (0.0987)	0.0240[a]	0.2548 (0.0931)	0.2491 (0.0837)	0.0240[a]
14	mg/kg	-----[b]	-----[b]	1.0000	0.0769 (0.0839)	0.0000[c] (0.2198)	0.0200[a]	0.0000[c] (0.1789)	0.0000[c] (0.2192)	0.0200[a]
15	mg/kg	0.1764 (0.2481)	0.1935 (0.2800)	0.4670	0.1922 (0.1728)	0.0623 (0.3959)	0.2000	0.1596 (0.0979)	0.1209 (0.0822)	0.2000
16	mg/kg	-----[b]	-----[b]	1.0000	0.1683 (0.0777)	0.1886 (0.0890)	0.2500	0.1509 (0.0898)	0.1289 (0.0881)	0.2500
17	mg/kg	-----[b]	-----[b]	1.0000	0.1341 (0.1392)	0.1582 (0.1173)	0.0071[a]	0.0772 (0.2406)	0.0876 (0.1621)	0.0071[a]
18	mg/kg	0.2152 (0.0297)	0.2261 (0.0324)	0.1430	0.2069 (0.0318)	0.2061 (0.0311)	0.2860	0.1303 (0.0158)	0.1227 (0.0148)	0.1430
19	mg/kg	-----[b]	0.7835 (0.0521)	0.7500	0.2934 (0.0780)	0.3331 (0.0554)	0.0250[a]	0.2367 (0.0876)	0.2331 (0.0838)	0.7500
20	mg/kg	0.1524 (0.0184)	0.1555 (0.0190)	0.0300[a]	0.1763 (0.0230)	0.1794 (0.0237)	0.3000	0.0953 (0.0100)	0.0854 (0.0100)	0.0300[a]

Table 15 (continued)

$B\hat{M}D_{05}$ Based on Trinomial and Binomial Dose Response Models Using GLM, and Estimated NOAEL
(Standard Error in Parentheses)

Study No.	Unit	MALFORMATION			PRENATAL DEATH			OVERALL TOXICITY		
		Trinomial	Binomial	NOAEL	Trinomial	Binomial	NOAEL	Trinomial	Binomial	NOAEL
21	mg/kg	0.2034 (0.0924)	0.1730 (0.1190)	0.5000	0.1054 (0.1242)	0.1350 (0.1008)	0.5000	0.0975 (0.0508)	0.1004 (0.0454)	0.5000
22	mg/kg	-----[b]	-----[b]	1.0000	0.3312 (0.1101)	0.3377 (0.1015)	0.5000	0.2844 (0.1783)	0.1228 (0.2674)	0.5000
23	% in diet	0.7384 (0.0581)	-----[b]	1.0000	0.3661 (0.2904)	-----[b]	0.7500	0.4909 (0.0894)	0.4301 (0.1815)	0.7500
24	mg/kg	-----[b]	-----[b]	0.5000	-----[b]	-----[b]	0.5000	0.3538 (0.2186)	0.3395 (0.2425)	0.5000
25	mg/kg	0.4778 (0.0837)	0.6323 (0.0406)	0.6300	-----[b]	-----[b]	1.0000	0.4050 (0.1153)	0.3529 (0.1171)	0.7920
26	mg/kg	-----[b]	-----[b]	1.0000	-----[b]	-----[b]	0.3333	0.5202 (0.2535)	0.5167 (0.2488)	0.3333
27	mg/kg	0.4785 (0.0553)	0.4854 (0.0527)	0.2857	0.2854 (0.0286)	0.3045 (0.0321)	0.2857	0.2441 (0.0265)	0.2462 (0.0283)	0.2857
28	mg/kg	0.2141 (0.0261)	0.2235 (0.0264)	0.0250[a]	0.3488 (0.0926)	0.3641 (0.1272)	0.5000	0.1698 (0.0202)	0.1450 (0.0157)	0.0250[a]

[a] When the NOAEL was equal to d_1=0.0 then we set the NOAEL = $d_2/10$

[b] The trinomial or binomial model could not be fit due to the lack of a dose response relationship

[c] $B\hat{M}D_{05}$ was set to zero if $B\hat{M}D_{05}$ was negative due to a large variance

Table 16

$\hat{BMD}_{05}$ Based on GLM and Rao-Scott Transformation (RST)
(Standard Error in Parentheses)

Study No.	Unit	MALFORMATION		PRENATAL DEATH		OVERALL TOXICITY	
		GLM	RST	GLM	RST	GLM	RST
1	mg/kg	-----[b]	-----[b]	0.2493 (0.0412)	0.2338 (0.0410)	0.2356 (0.0361)	0.2225 (0.0358)
2	mg/kg	0.7014 (0.0828)	0.7018 (0.0867)	0.2256 (0.0782)	0.2158 (0.0785)	0.2191 (0.0592)	0.2101 (0.0597)
3	% in diet	0.2556 (0.0237)	0.2282 (0.0232)	0.1514 (0.0126)	0.1405 (0.0118)	0.1221 (0.0105)	0.1087 (0.0095)
4	mg/kg	0.3091 (0.0302)	0.2529 (0.0351)	0.2472 (0.0436)	0.2519 (0.0429)	0.1911 (0.0221)	0.1674 (0.0224)
5	% in diet	-----[b]	-----[b]	0.2477 (0.0559)	0.2477 (0.0532)	0.2389 (0.0490)	0.2384 (0.0468)
6	mg/kg	0.1606 (0.4132)	0.1936 (0.3853)	0.1031 (0.0173)	0.1020 (0.0176)	0.0924 (0.0161)	0.0913 (0.0161)
7	mg/kg	0.2527 (0.0279)	0.2284 (0.0279)	0.2111 (0.0418)	0.2051 (0.0455)	0.1386 (0.0173)	0.1317 (0.0176)
8	mg/kg	-----[b]	-----[b]	0.1837 (0.1262)	0.1840 (0.1263)	0.0940 (0.1937)	0.0943 (0.1922)
9	mg/kg	0.5978 (0.0396)	0.2916 (0.0448)	0.5240 (0.1017)	0.1534 (0.4403)	0.4580 (0.0620)	0.2242 (0.0485)
10	mg/kg	-----[b]	-----[b]	0.2346 (0.1334)	0.2234 (0.1261)	0.1220 (0.2185)	0.1258 (0.2010)

Table 16 (continued)

$\hat{BMD}_{05}$ Based on GLM and Rao-Scott Transformation (RST)
(Standard Error in Parentheses)

Study No.	Unit	MALFORMATION		PRENATAL DEATH		OVERALL TOXICITY	
		GLM	RST	GLM	RST	GLM	RST
11	mg/kg	-----[b]	-----[b]	0.2600 (0.2015)	0.2605 (0.1991)	0.1831 (0.1625)	0.1804 (0.1657)
12	mg/kg	-----[b]	-----[b]	-----[b]	-----[b]	0.6071 (0.2197)	-----[b]
13	mg/kg	-----[b]	-----[b]	0.2844 (0.1020)	0.2700 (0.1057)	0.2548 (0.0931)	0.2397 (0.0975)
14	mg/kg	-----[b]	-----[b]	0.0769 (0.0839)	0.0168 (0.1463)	0.0000[c] (0.1789)	0.0000[c] (0.3086)
15	mg/kg	0.1764 (0.2481)	0.1952 (0.2335)	0.1922 (0.1728)	0.1657 (0.1229)	0.1596 (0.0979)	0.1532 (0.0726)
16	mg/kg	-----[b]	-----[b]	0.1683 (0.0777)	0.1630 (0.0853)	0.1509 (0.0898)	0.1420 (0.0971)
17	mg/kg	-----[b]	-----[b]	0.1341 (0.1392)	0.1183 (0.1511)	0.0772 (0.2406)	0.0544 (0.2398)
18	mg/kg	0.2152 (0.0297)	0.2066 (0.0285)	0.2069 (0.0318)	0.2044 (0.0308)	0.1303 (0.0158)	0.1262 (0.0152)
19	% in diet	-----[b]	-----[b]	0.2934 (0.0780)	0.2772 (0.0844)	0.2367 (0.0876)	0.2323 (0.0802)
20	mg/kg	0.1524 (0.0184)	0.1480 (0.0173)	0.1763 (0.0230)	0.1743 (0.0230)	0.0953 (0.0100)	0.0930 (0.0095)

Table 16 (continued)

$\hat{BMD}_{05}$ Based on GLM and Rao-Scott Transformation (RST)
(Standard Error in Parentheses)

Study No.	Unit	MALFORMATION		PRENATAL DEATH		OVERALL TOXICITY	
		GLM	RST	GLM	RST	GLM	RST
21	mg/kg	0.2034 (0.0924)	0.1959 (0.0988)	0.1054 (0.1242)	0.0950 (0.1333)	0.0975 (0.0508)	0.0936 (0.0541)
22	mg/kg	-----[b]	-----[b]	0.3312 (0.1101)	0.3171 (0.1194)	0.2844 (0.1783)	0.2768 (0.2050)
23	mg/kg	0.7384 (0.0581)	-----[b]	0.3661 (0.2904)	0.4089 (0.0402)	0.4909 (0.0894)	0.3735 (0.0321)
24	mg/kg	-----[b]	-----[b]	-----[b]	0.3692 (0.3324)	0.3538 (0.2186)	0.3700 (0.1697)
25	mg/kg	0.4778 (0.0837)	0.4701 (0.0773)	-----[b]	-----[b]	0.4050 (0.1153)	0.3955 (0.0952)
26	mg/kg	-----[b]	-----[b]	-----[b]	-----[b]	0.5202 (0.2535)	0.5215 (0.2514)
27	mg/kg	0.4785 (0.0553)	0.4782 (0.0506)	0.2854 (0.0286)	0.2390 (0.0245)	0.2441 (0.0265)	0.2082 (0.0217)
28	mg/kg	0.2141 (0.0261)	0.2105 (0.0259)	0.3488 (0.0926)	0.3511 (0.0903)	0.1698 (0.0202)	0.1677 (0.0200)

[b] GLM and/or the Rao-Scott transformed data could not be fit due to a lack of a dose response relationship

[c] $\hat{BMD}_{05}$ was set to zero if $\hat{BMD}_{05}$ was negative due to a large variance

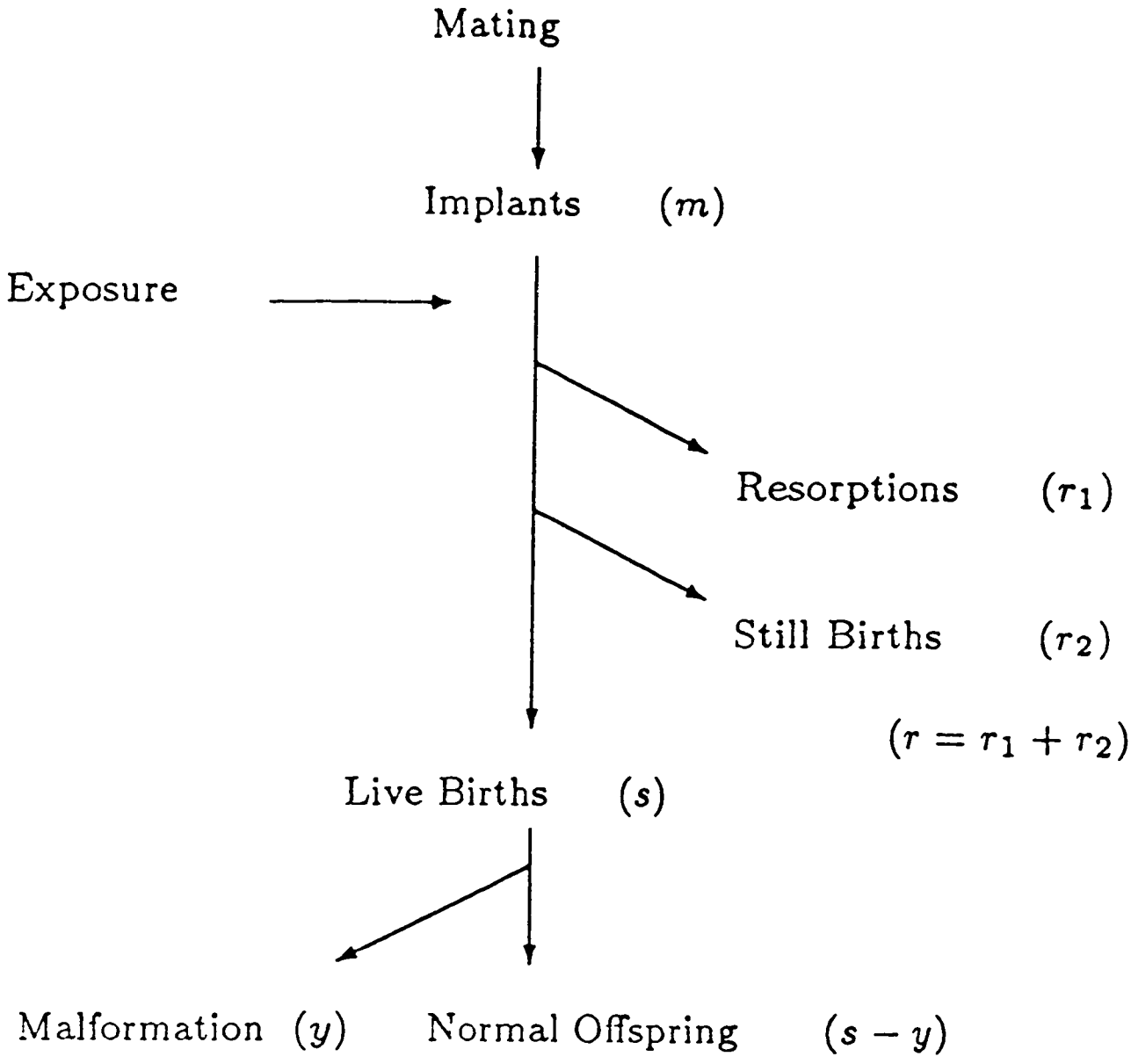


Figure 1: Schematic representation of a developmental toxicity experiment. Taken from Krewski et al. (1994)

FIG.2a: Histogram of the ratios NOAEL/BMD from joint models in simulation with 10 litters per dose group. There are 515/983 ratios within a factor of 2. The mean ($\pm$ SD) and median value for $NOAEL/\hat{BMD}_{05}$ are 0.633 ($\pm$ 0.017) and 0.754, respectively.

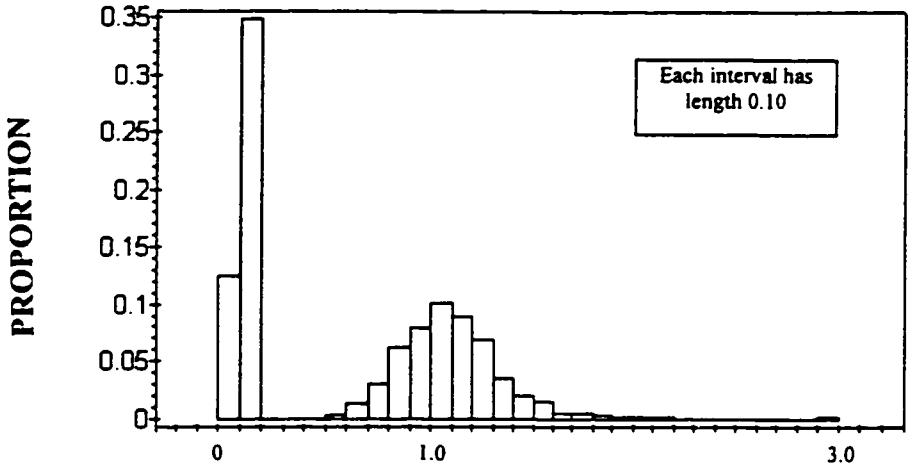


FIG.2b: Histogram of the ratios NOAEL/BMD from joint models in simulation with 20 litters per dose group. There are 280/994 ratios within a factor of 2. The mean ($\pm$ SD) and median value for $NOAEL/\hat{BMD}_{05}$ are 0.396 ($\pm$ 0.014) and 0.124, respectively.

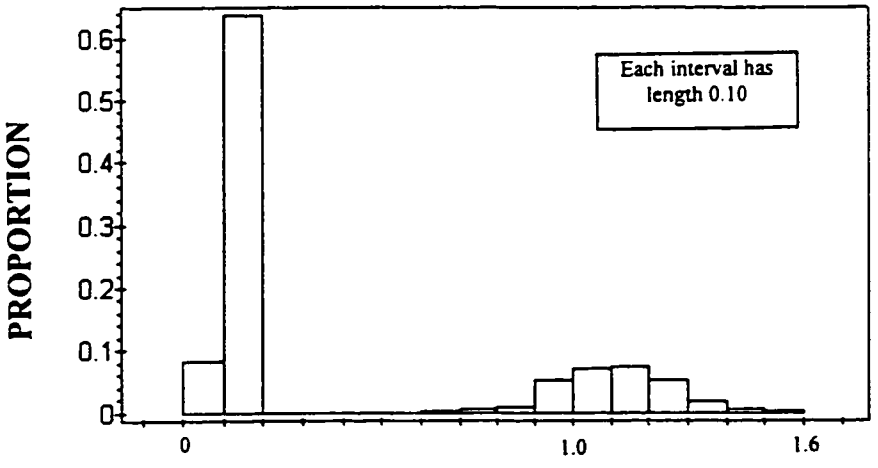


FIG.2c: Histogram of the ratios NOAEL/BMD from joint models in simulation with 30 litters per dose group. There are 152/999 ratios within a factor of 2. The mean ($\pm$ SD) and median value for $NOAEL/\hat{BMD}_{05}$ are 0.273($\pm$0.012) and 0.122, respectively.

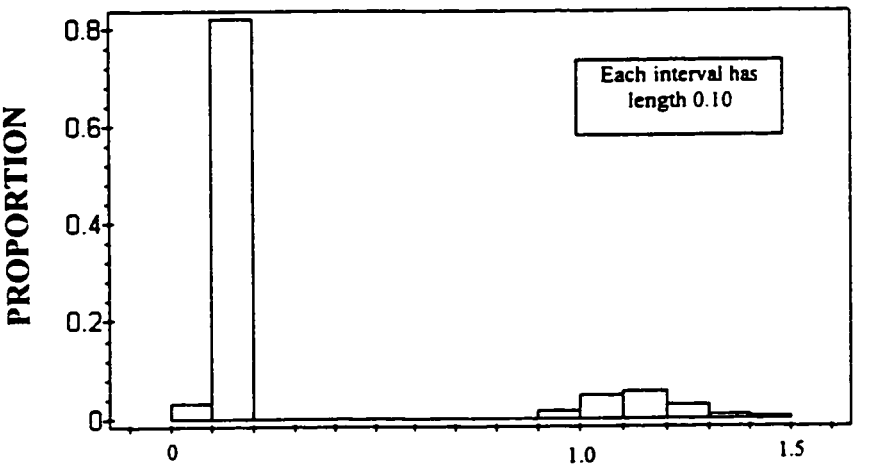


FIG.2d: Histogram of the ratios NOAEL/BMD from joint models in simulation with (30, 30, 10, 10) litters per dose group. There are 153/996 ratios within a factor of 2. The mean ($\pm$ SD) and median value for $NOAEL/\hat{BMD}_{05}$ are 0.267 ($\pm$0.011) and 0.121, respectively.

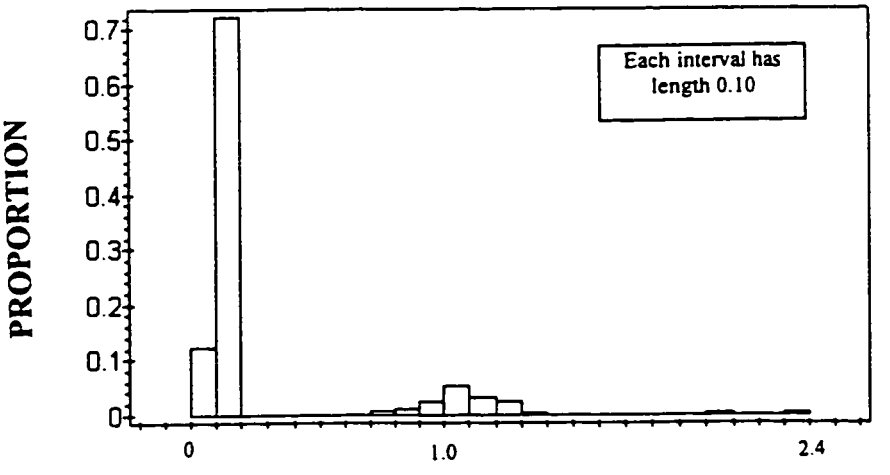


FIG.2e: Histogram of the ratios NOAEL/BMD from joint models in simulation with (30, 10, 10, 30) litters per dose group. There are 396/996 ratios within a factor of 2. The mean ($\pm$ SD) and median value for $NOAEL/\hat{BMD}_{05}$ are 0.514 ($\pm$ 0.016) and 0.133, respectively.

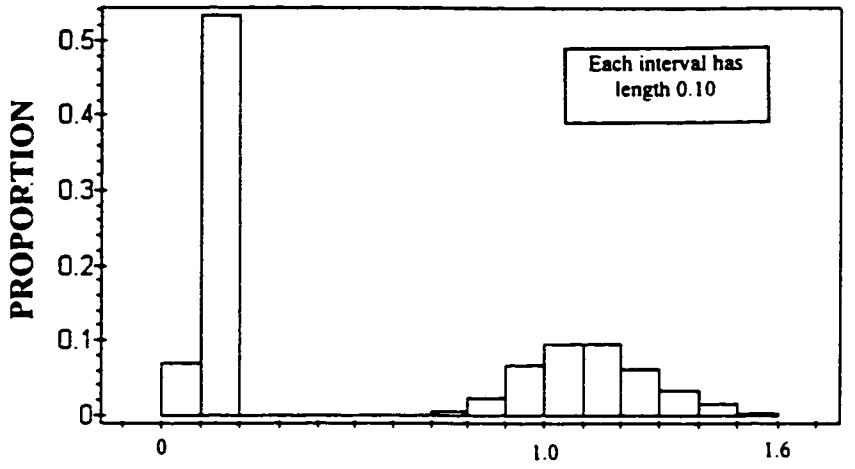


FIG.2f: Histogram of the ratios NOAEL/BMD from joint models in simulation with (10, 30, 30, 10) litters per dose group. There are 372/991 ratios within a factor of 2. The mean ($\pm$ SD) and median value for $NOAEL/\hat{BMD}_{05}$ are 0.495 ($\pm$ 0.016) and 0.128, respectively.

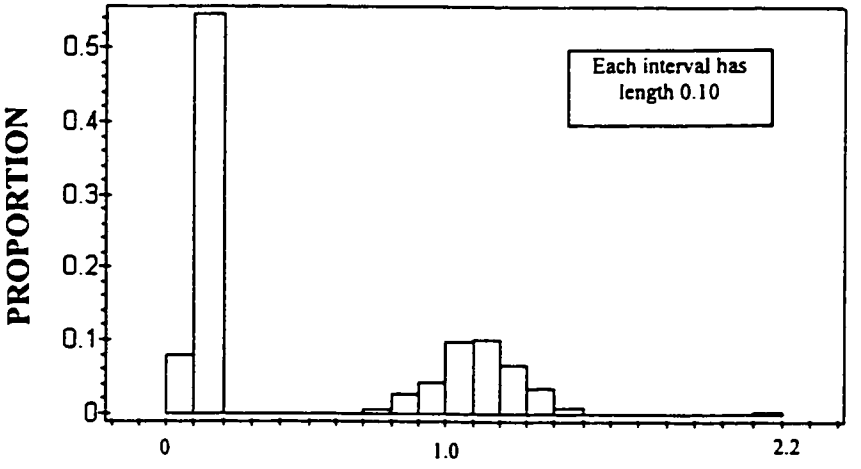


FIG.2g: Histogram of the ratios NOAEL/BMD for simulation with dose placements at (0, 0.5, 1). There are 0/998 ratios within a factor of 2. The mean ($\pm$ SD) and median value for $NOAEL/\hat{BMD}_{05}$ are 0.234 ($\pm$ 0.001) and 0.233, respectively.

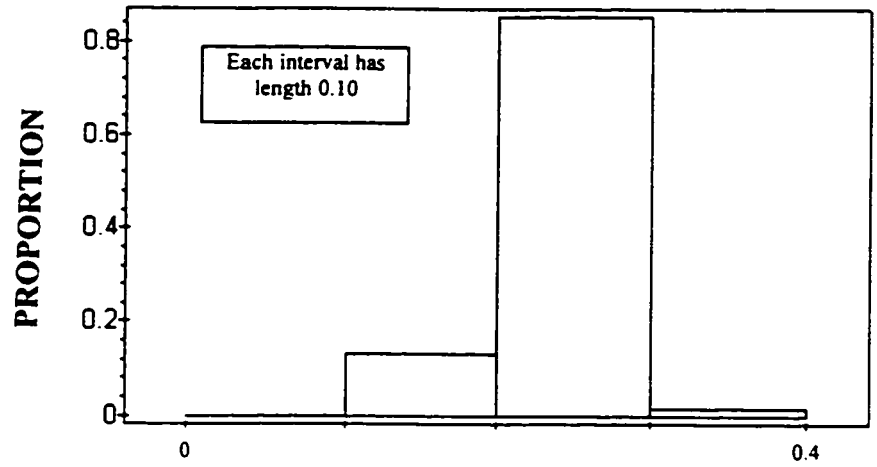


FIG.2h: Histogram of the ratios NOAEL/BMD for simulation with dose placements at (0, 0.3, 1). There are 133/997 ratios within a factor of 2. The mean ($\pm$ SD) and median value for $NOAEL/\hat{BMD}_{05}$ are 0.290 ($\pm$ 0.012) and 0.144, respectively.

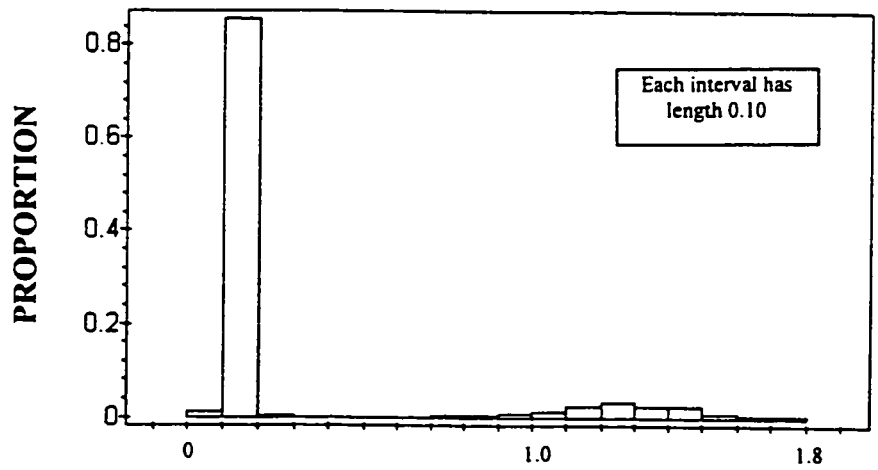


FIG.2i: Histogram of the ratios NOAEL/BMD for simulation with dose placements at (0, 0.25, 0.5, 1). There are 330/996 ratios within a factor of 2. The mean ($\pm$ SD) and median value for $NOAEL/\hat{BMD}_{05}$ are 0.444 ($\pm$0.015) and 0.126, respectively.

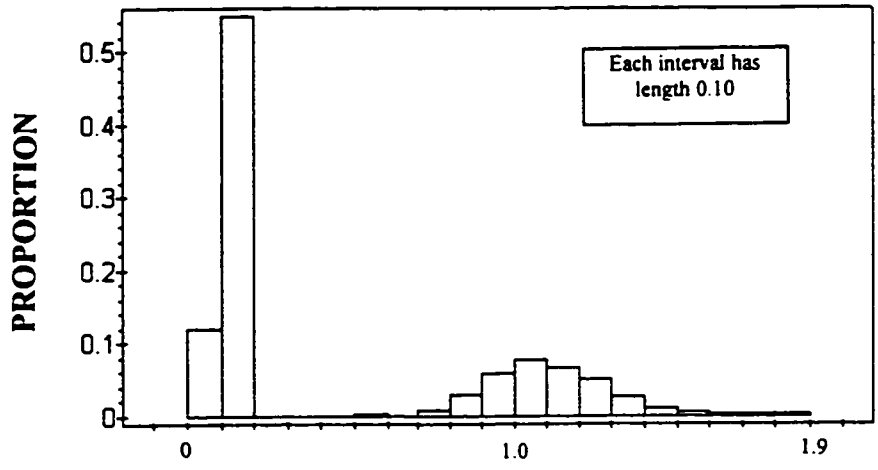


FIG.2j: Histogram of the ratios NOAEL/BMD for simulation with dose placements at (0, 0.15, 0.35, 1). There are 538/997 ratios within a factor of 2. The mean ($\pm$ SD) and median value for $NOAEL/\hat{BMD}_{05}$ are 0.465 ($\pm$0.013) and 0.566, respectively.

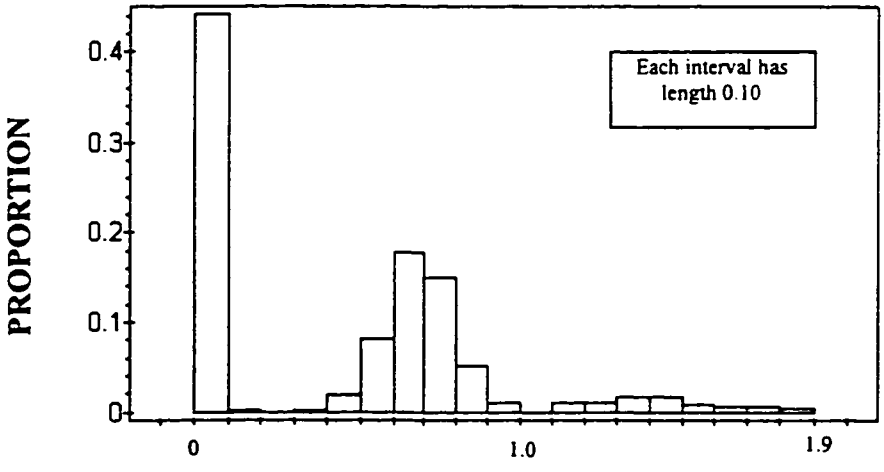


FIG.2k: Histogram of the ratios NOAEL/BMD for simulation with dose placements at (0, 0.125, 0.25, 0.5, 1). There are 531/993 ratios within a factor of 2. The mean ($\pm$ SD) and median value for $NOAEL/\hat{BMD}_{05}$ are 0.523 ($\pm$ 0.014) and 0.530, respectively.

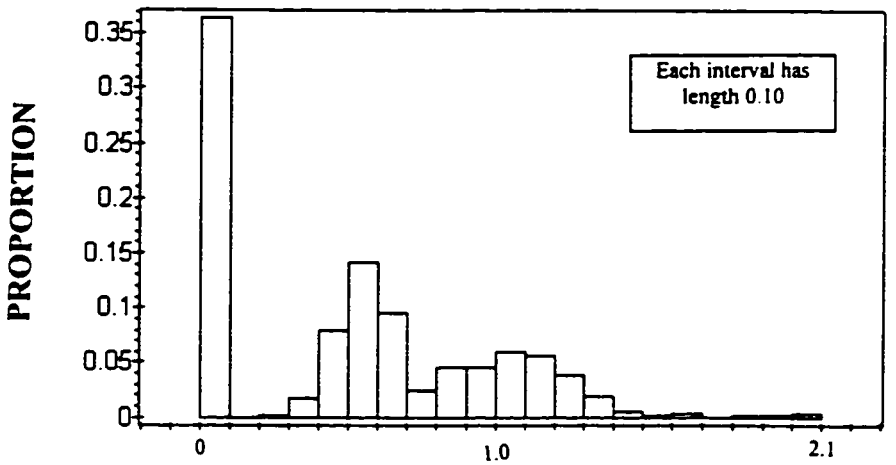


FIG.2l: Histogram of the ratios NOAEL/BMD for simulation with dose placements at (0, 0.125, 0.4, 0.75, 1). There are 502/994 ratios within a factor of 2. The mean ($\pm$ SD) and median value for $NOAEL/\hat{BMD}_{05}$ are 0.455 ($\pm$ 0.012) and 0.502, respectively.

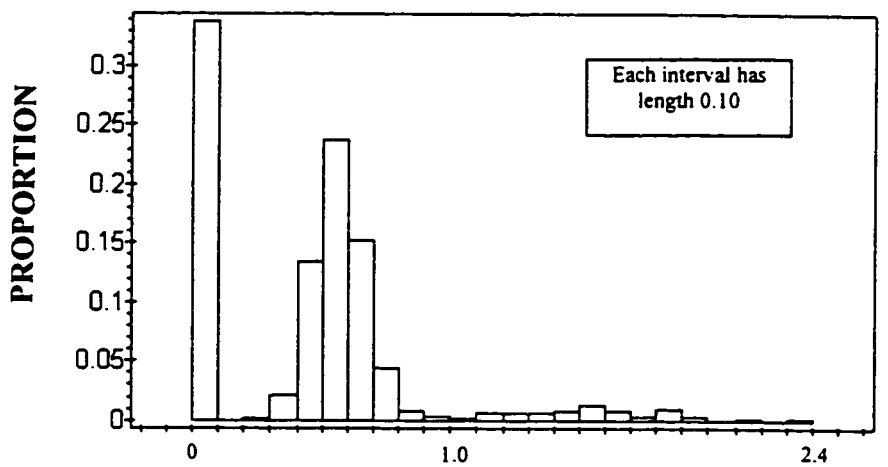


FIG.2m: Histogram of the ratios NOAEL/BMD for simulation generated by correlation coefficients $\{\phi_i\}=\{0, 0, 0, 0\}$. There are 233/1000 ratios within a factor of 2. The mean ($\pm$ SD) and median value for $NOAEL/\hat{BMD}_{05}$ are 0.371 ($\pm$ 0.014) and 0.125, respectively.

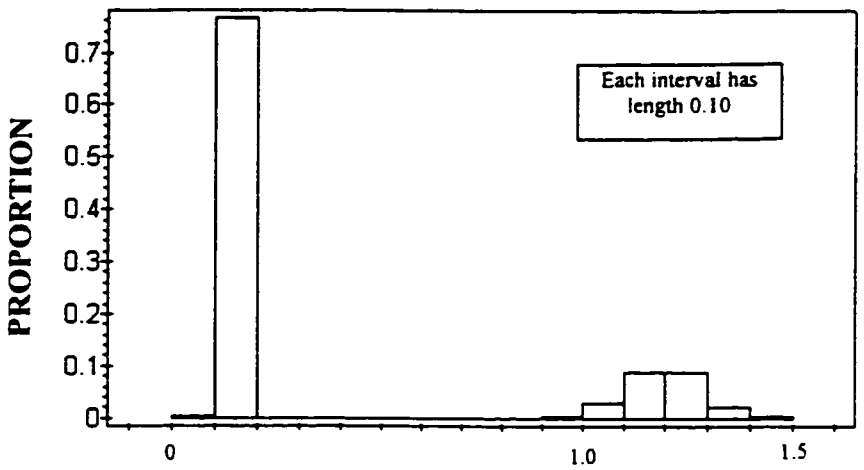


FIG.2n: Histogram of the ratios NOAEL/BMD for simulation generated by correlation coefficients $\{\phi_i\}=\{0.1, 0.1, 0.1, 0.1\}$. There are 313/996 ratios within a factor of 2. The mean ($\pm$ SD) and median value for $NOAEL/\hat{BMD}_{05}$ are 0.441 ($\pm$ 0.015) and 0.126, respectively.

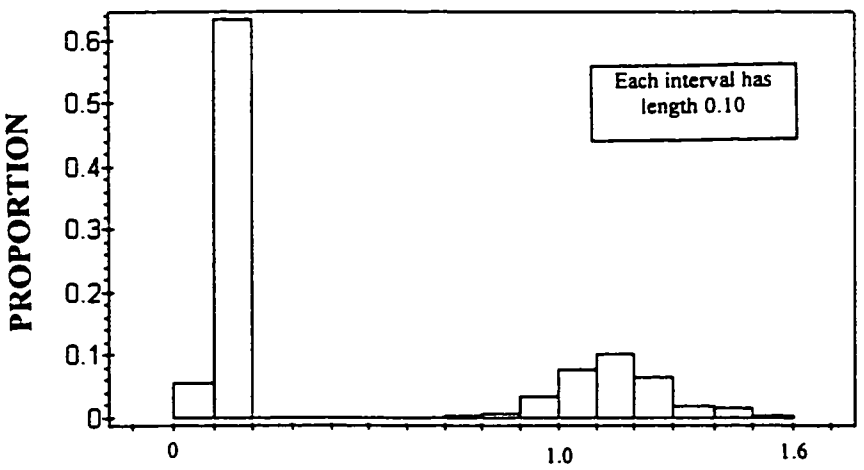


FIG.2o: Histogram of the ratios NOAEL/BMD for simulation generated by correlation coefficients $\{\phi_i\}=\{0.4, 0.4, 0.4, 0.4\}$. There are 327/990 ratios within a factor of 2. The mean ($\pm$ SD) and median value for $NOAEL/\hat{BMD}_{05}$ are 0.439 ($\pm$ 0.016) and 0.126, respectively.

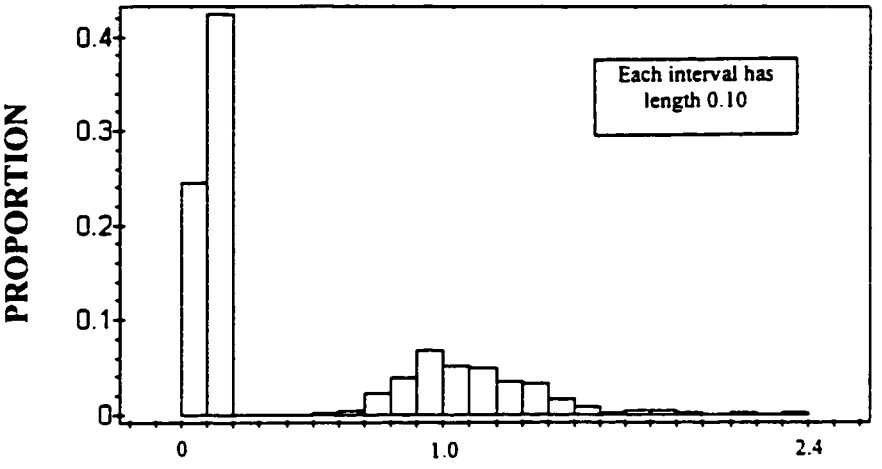
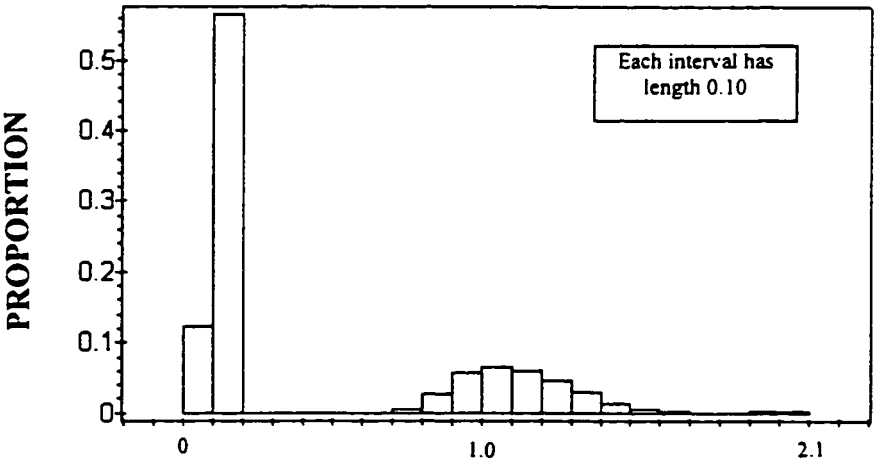


FIG.2p: Histogram of the ratios NOAEL/BMD for simulation generated by correlation coefficients $\{\phi_i\}=\{0.1, 0.2, 0.3, 0.4\}$. There are 313/997 ratios within a factor of 2. The mean ($\pm$ SD) and median value for $NOAEL/\hat{BMD}_{05}$ are 0.434 ($\pm$ 0.015) and 0.125, respectively.



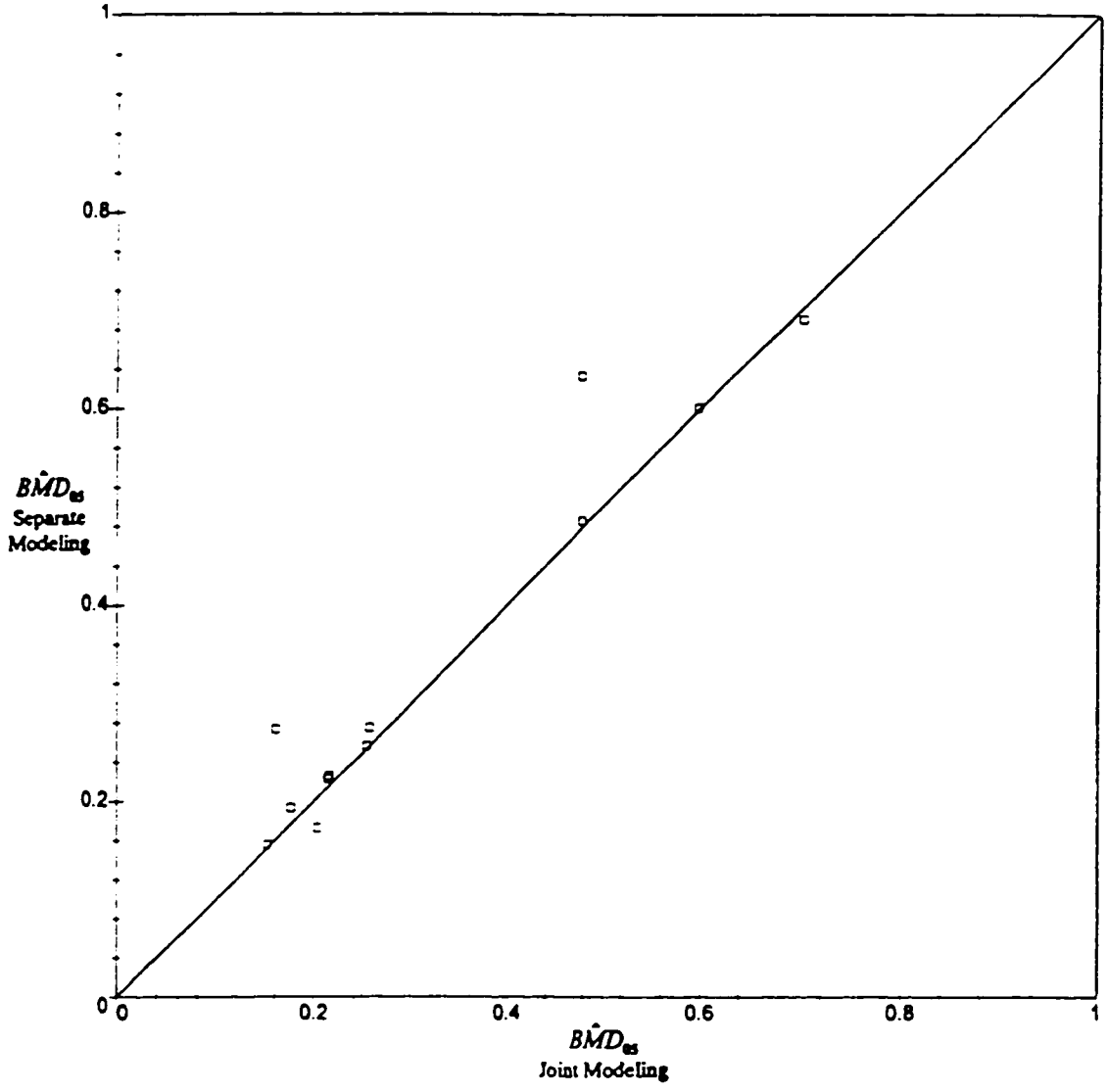


Figure 3a: BMDs from joint and separate modeling for the endpoint fetal malformation in the developmental toxicity data.

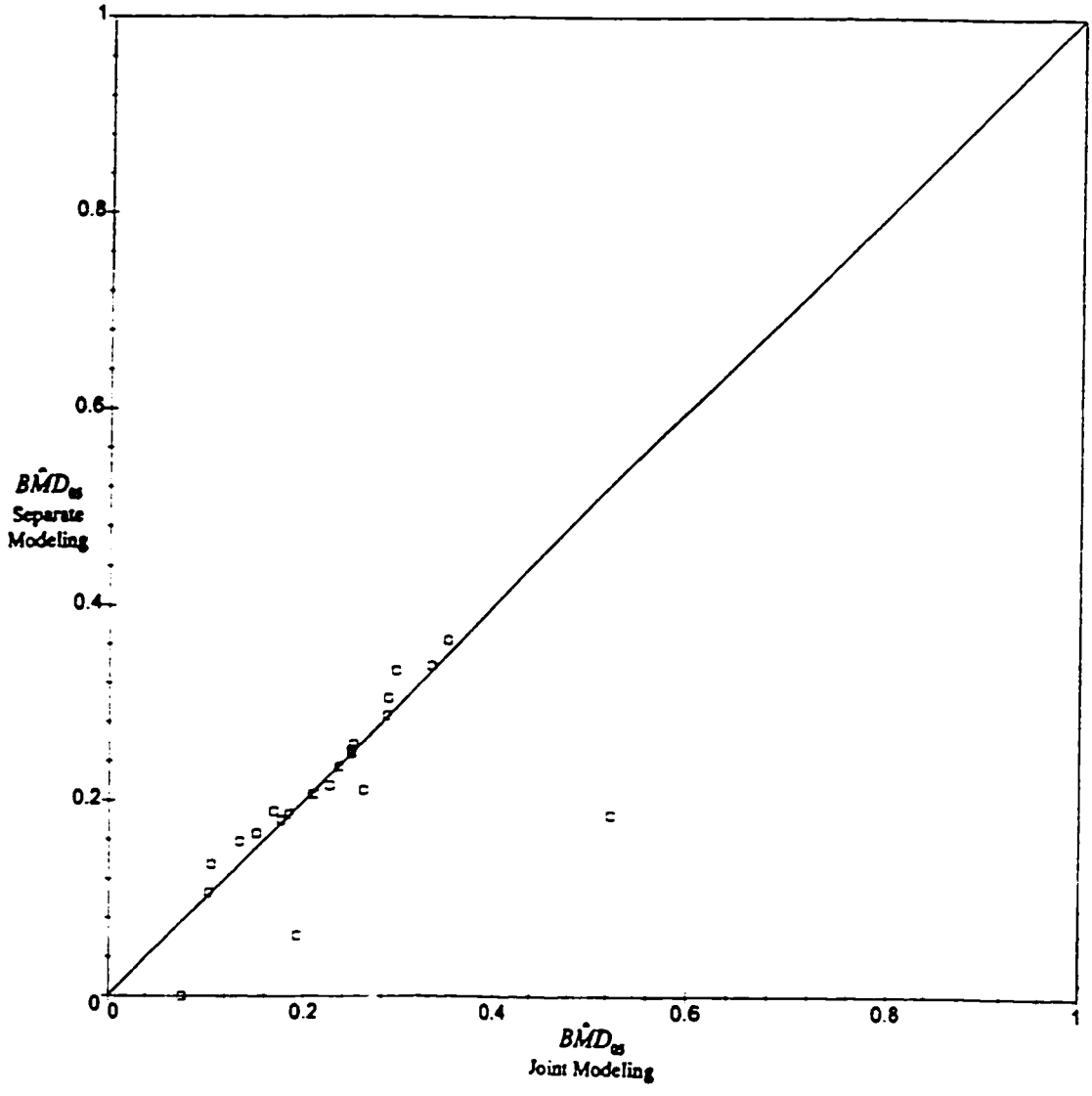


Figure 3b: BMDs from joint and separate modeling for the endpoint prenatal death in the developmental toxicity data.

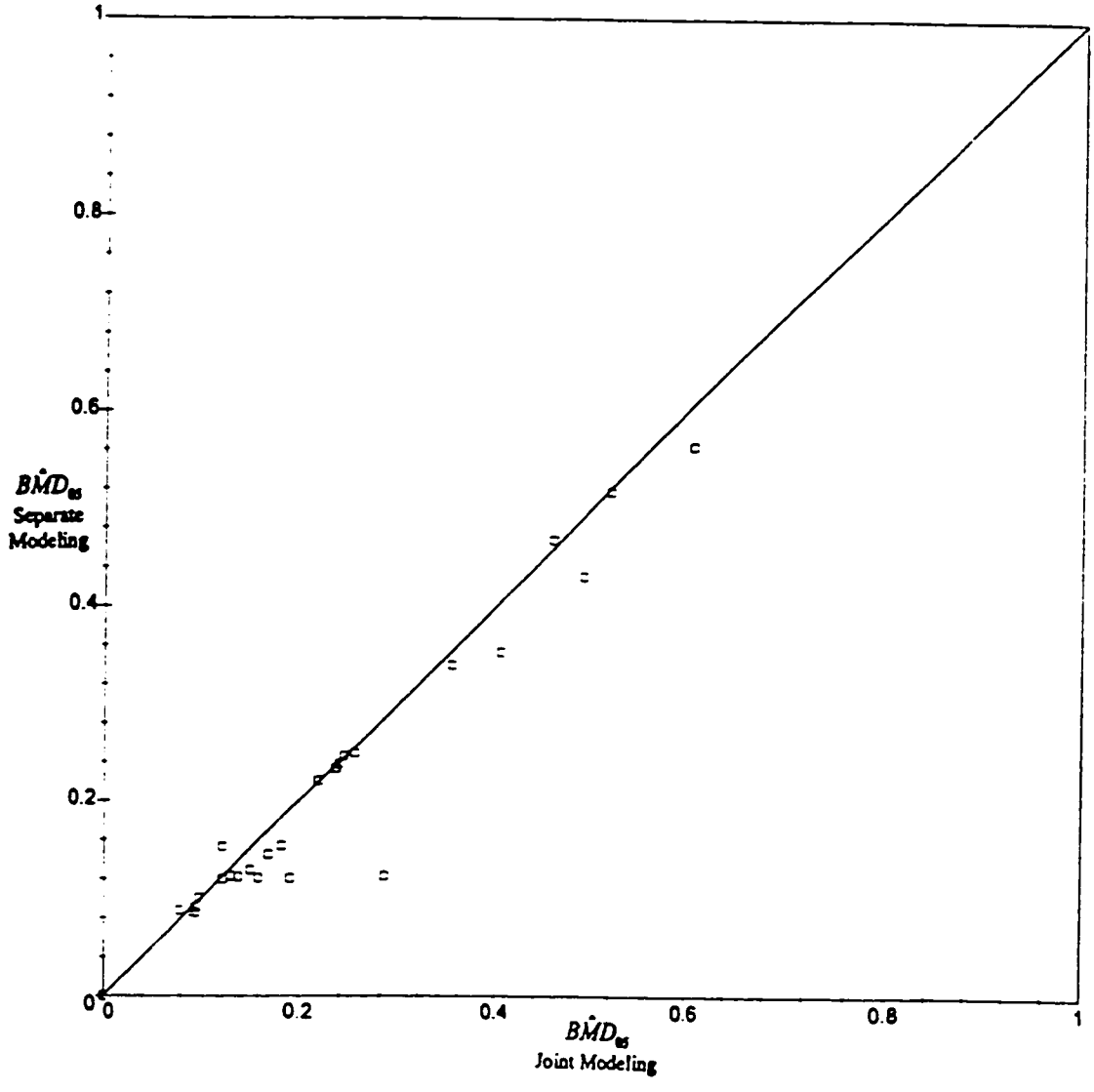


Figure 3c: BMDs from joint and separate modeling for the endpoint overall toxicity in the developmental toxicity data.

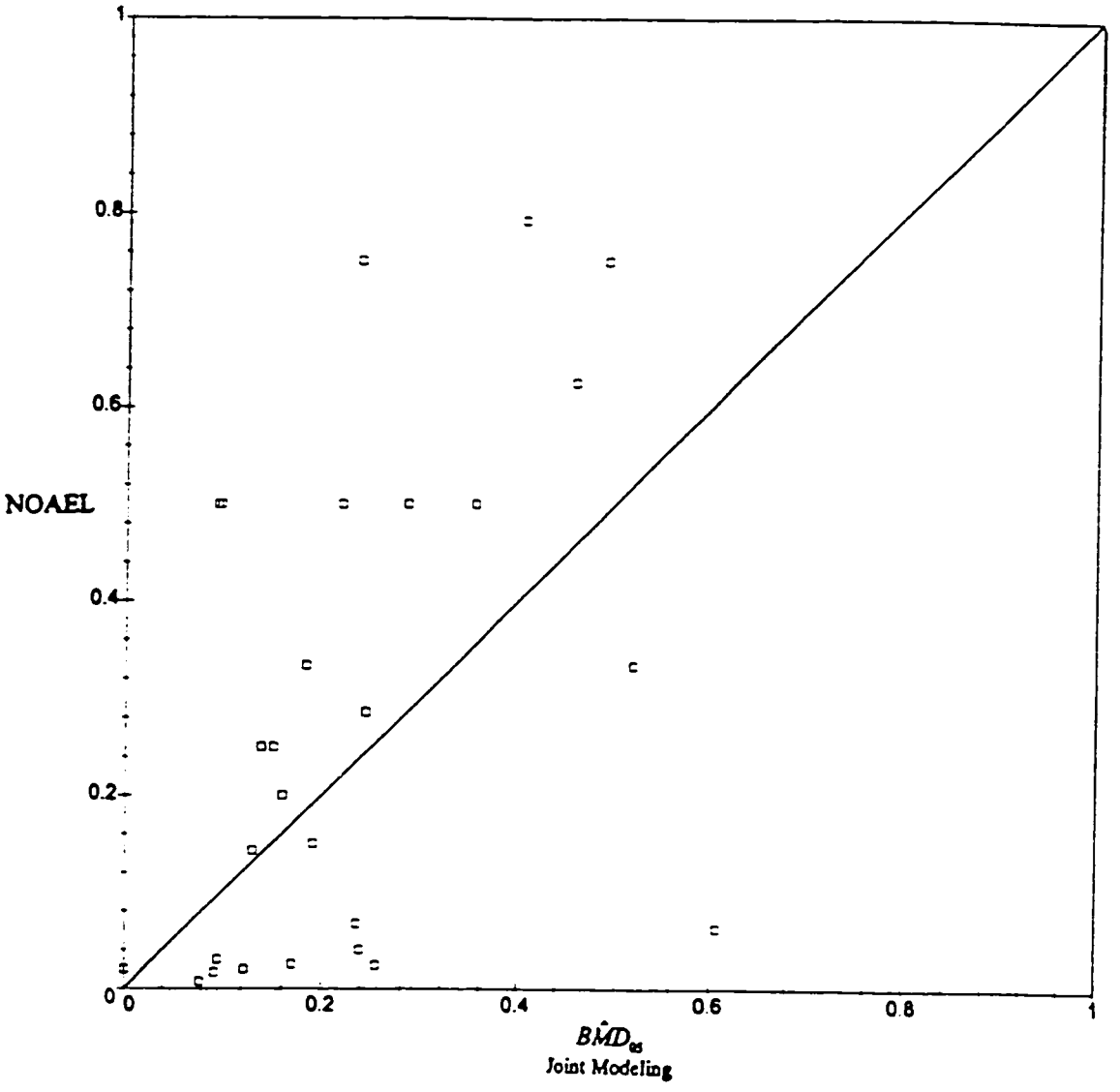


Figure 4: BMD from joint modeling and NOAEL levels for the endpoint overall toxicity.

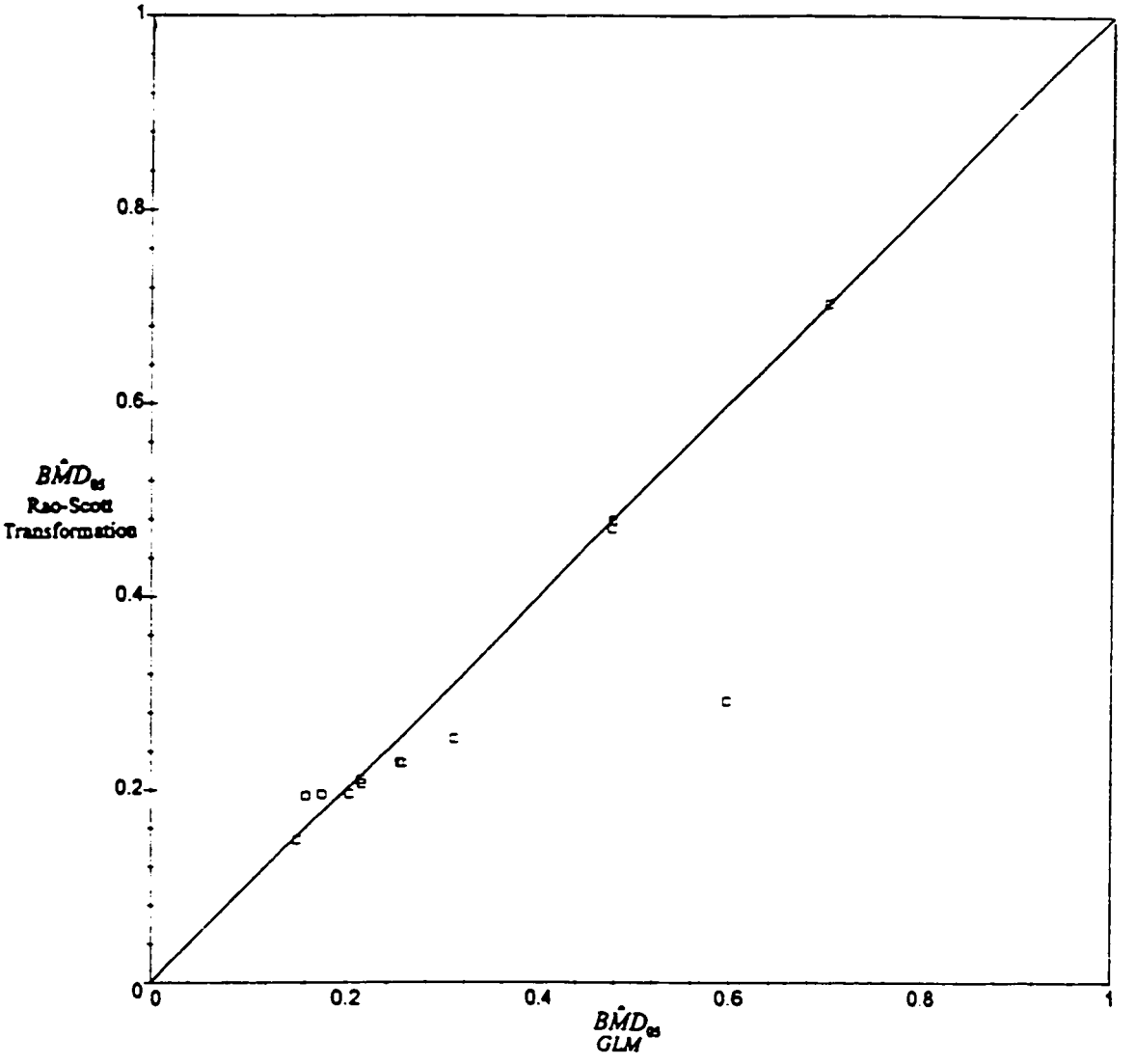


Figure 5a: BMDs from GLM and RST for the endpoint fetal malformation.

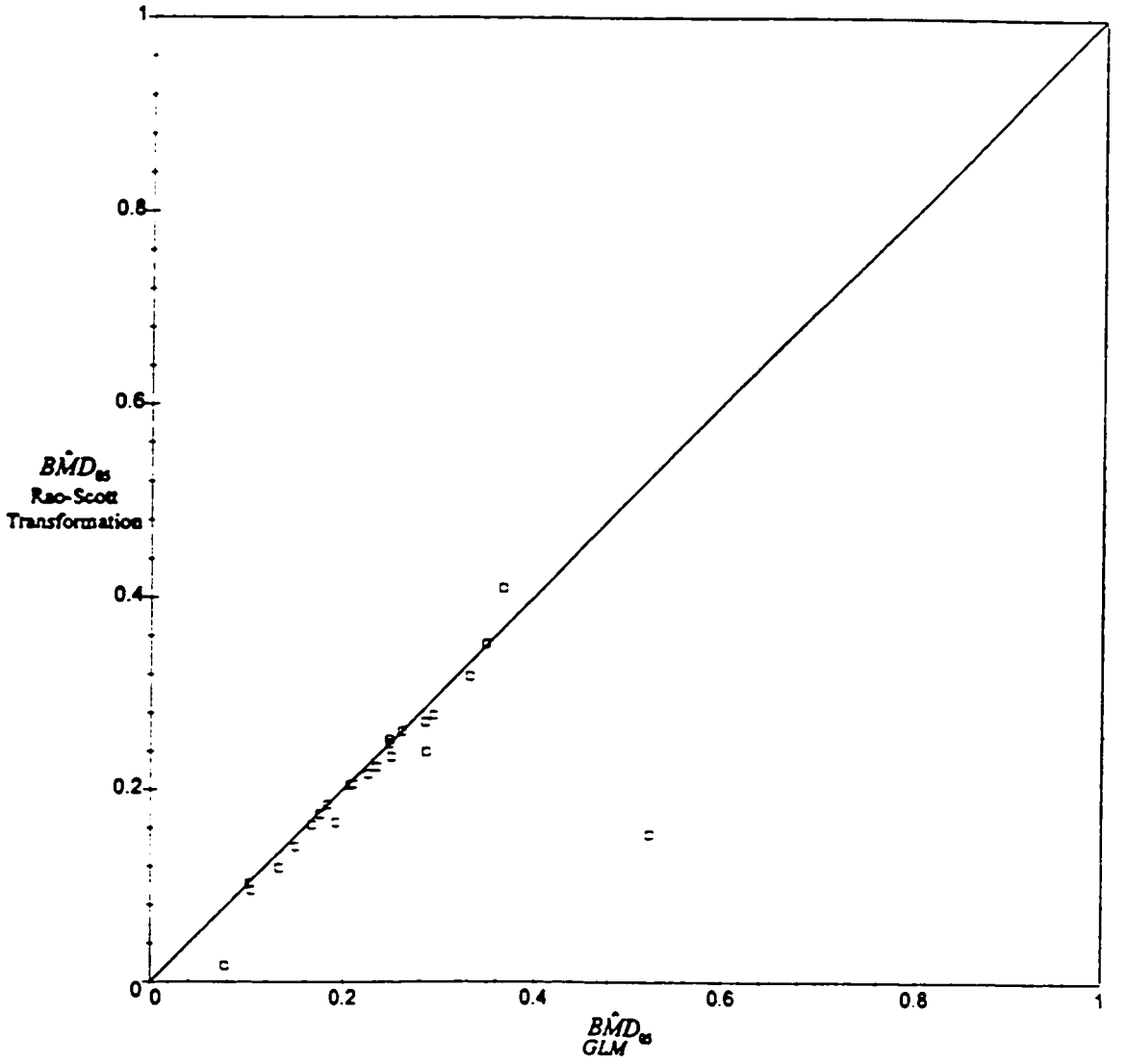


Figure 5b: BMDs from GLM and RST for the endpoint prenatal death.

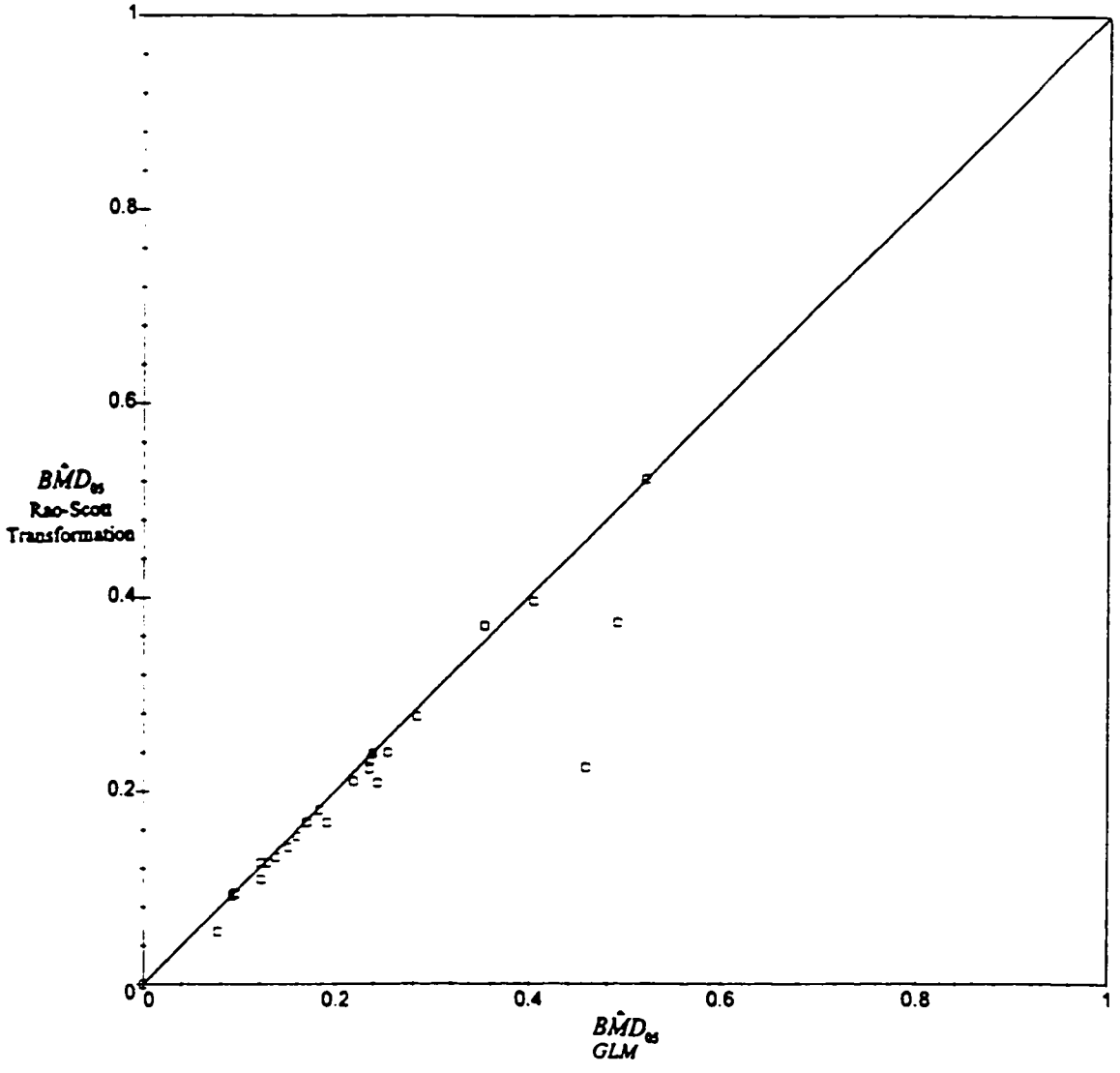


Figure 5c: BMDs from GLM and RST for the endpoint overall toxicity.

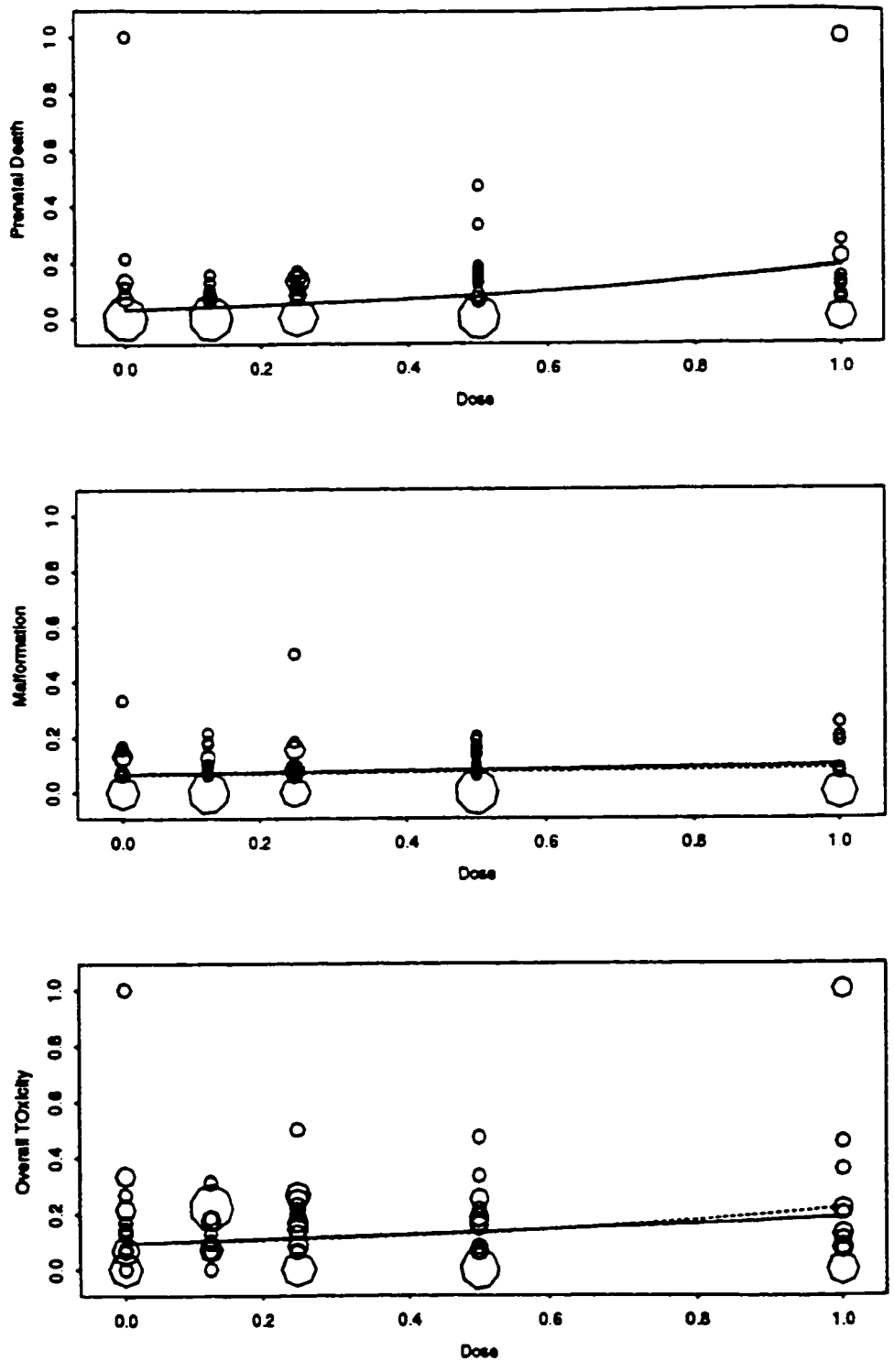
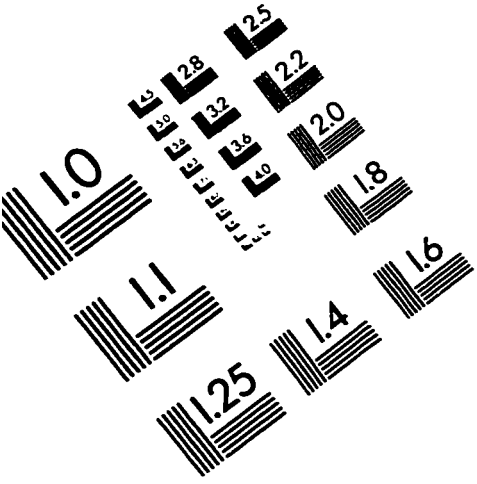


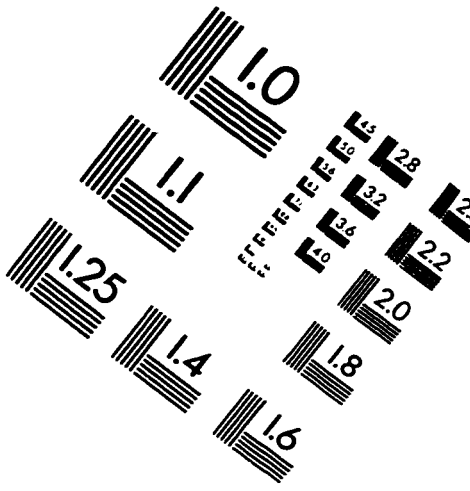
Figure 6: Logistic Model fit to Bisphenol A (BPA) in Rats with Joint (___) and Separate Modeling (----) to the Three Endpoints. Diameter of circles are proportional to the squareroot of the number of replicates.

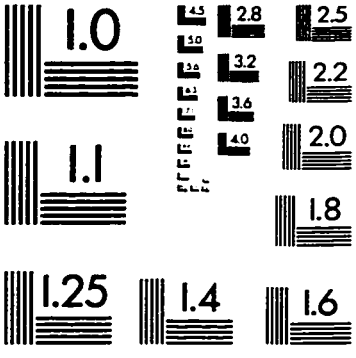
VITA AUCTORIS

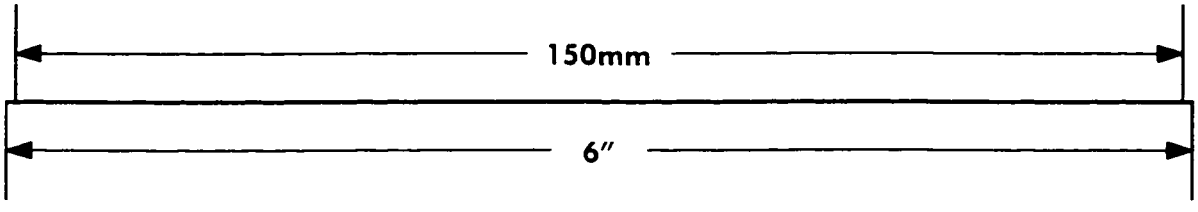
Leonora Marro was born in 1972 in Windsor Ontario, Canada. She graduated from high school in June 1991. In September 1991, she went to the University of Windsor where she obtained an Honours B.A. in Mathematics in 1995. She is now a student of the University of Windsor, and hopes to obtain a Master's degree in Statistics in Fall 1996.

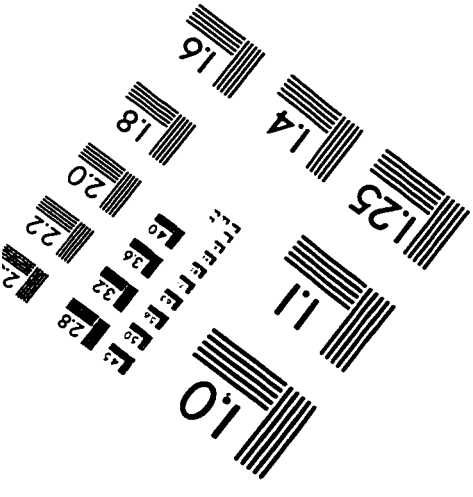
IMAGE EVALUATION TEST TARGET (QA-3)

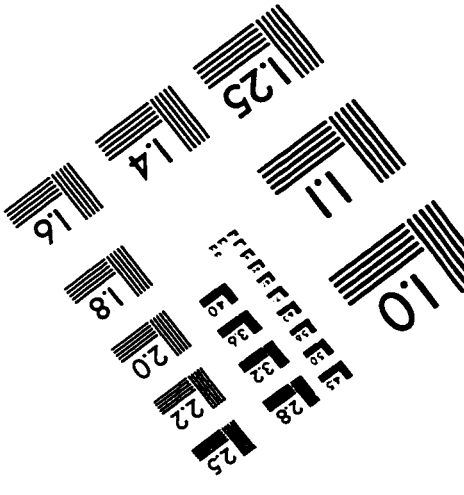












APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989