INVESTIGATION OF STATISTICAL METHODS FOR DETERMINATION OF BENCHMARK DOSE LIMITS FOR RETINOIC ACID-INDUCED FETAL FORELIMB MALFORMATION IN MICE

by

TING LI

(Under the Direction of JAXK REEVES)

ABSTRACT

The benchmark dose (BMD) is an estimated dose that induces an adverse response above the background level after a toxic exposure. In the present study, different statistical methods, especially those proposed for developmental toxicity studies, were evaluated by using selected toxicological endpoints (fetal malformation probability, fetal ulna length, and roundness) from a study in which pregnant mice were exposed to retinoic acid. The intra-litter correlation was accommodated either by assuming a beta-binomial distribution or by using generalized estimating equation (GEE) methods. The lower confidence limits (BMDLs) were estimated by the delta method, the likelihood-ratio based method, and the bootstrap method. The results indicated that the malformation probability was the most sensitive endpoint. The likelihood-ratio based method and bootstrap method tend to give tighter confidence limits compared to the delta method. While the likelihood-ratio based method could not be used in GEE models, the bootstrap method did not have this limitation.

INDEX WORDS: Benchmark Dose, Dose-Response, Retinoic Acid, Fetal Forelimb Malformation

INVESTIGATION OF STATISTICAL METHODS FOR DETERMINATION OF BENCHMARK DOSE LIMITS FOR RETINOIC ACID-INDUCED FETAL FORELIMB MALFORMATION IN MICE

by

TING LI

B.S., HuaZhong Agricultural University, China, 2002

Ph.D., The University of Georgia, 2007

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2008

INVESTIGATION OF STATISTICAL METHODS FOR DETERMINATION OF BENCHMARK DOSE LIMITS FOR RETINOIC ACID-INDUCED FETAL FORELIMB MALFORMATION IN MICE

by

TING LI

Major Professor: Jaxk Reeves

Committee: Abhyuday Mandal
Lynne Seymour

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
December 2008

DEDICATION

To my family and friends.

ACKNOWLEDGEMENTS

I would like to thank Dr. Jaxk Reeves for his guidance and encouragement throughout the study. I also would like to thank Dr. Lynne Seymour and Dr. Abhyuday Mandal for their invaluable support and assistance. Dr. Jerry Campbell, who generously provided the experimental data, also deserves my deepest gratitude.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

1.1 DOSE-RESPONSE ASSESSMENT

Dose-response assessment in toxicology is the study that links the hazard exposure (e.g. dosage) of a chemical with adverse effects. Historically, different dose-response assessment approaches have been applied for cancer and non-cancer health effects. For cancer health effects, dose-response models (i.e., linearized multistage model) are fitted to tumor incidence data and then extrapolated to estimate the risk at low doses assuming linearity (US EPA 2005). For non-cancer effects, a threshold method is used, since it is assumed that below a certain level of exposure, no adverse effects will be induced. In this case, the no-observed-adverse-effect level (NOAEL) or the lowest-observed-adverse-effect level (LOAEL) is determined by multiple comparisons among the control group and the treatment groups. NOAEL is the highest exposure level at which there is no statistically or biologically significant increase in the frequency or severity of adverse effect between the exposed population and its appropriate control. LOAEL is the lowest exposure level at which there is statistically or biologically significant increase in frequency or severity of adverse effects between the exposed population and its appropriate control group. However, the NOAEL/LOAEL method has many undesirable features. For example, the NOAEL/LOAEL must be one of the experimental dose levels, so it might vary from experiment to experiment and the uncertainty in its estimation cannot be assessed. The shape of the dose-response curve (e.g. slope and plateau) is also not fully utilized by NOAEL/LOAEL methods (US EPA, 2000).

As our understanding of the underlying biology of toxic effects has grown, however, the apparent differences between cancer and non-cancer effects have lessened to the point where it seems reasonable to develop quantitative methods based on similar considerations for all types of health effects, and to make approaches to dose-response assessment as consistent across endpoints as our current mechanistic understanding allows. The benchmark dose (BMD) approach has been proposed as an alternative dose-response assessment to both cancer and non-cancer effects (US EPA, 2000).

1.2 BENCHMARK DOSE APPROACH

According to US EPA (2000), the benchmark dose (BMD) is the dose that corresponds to a pre-determined change in an adverse response (benchmark response, BMR) compared to the response in untreated animals. It is determined by modeling a dose-response curve in the region of the dose-response relationship where biologically observable data are available. To account for the uncertainty of the estimation of BMD, the lower confidence limit (BMDL) is the dose of interest (US EPA, 2000). The adverse responses (i.e., endpoints or effects) produced in toxicological studies can be classified as dichotomous or continuous data. Due to the distinct characteristic of those two classes of data, different approaches are often used to define BMR, as well as to estimate BMD and BMDL.

1.2.1 Dichotomous Response Data

The simplest dichotomous data are also referred as 'quantal data'. Such data have only two values, response or not (e.g., dead or alive; normal or abnormal). Quantal data are the most reported form of data in toxicology studies. For a quantal response data, suppose that there are k dose groups ($d_1, d_2 \ldots\ldots d_k$) employed in the study and the total number of animals in the dose

groups are $n_1, n_2 \ldots n_k$, with corresponding responders $x_1, x_2 \ldots x_k$. The incidence, x_i is assumed to have probability

$$p_i = p(d_i;\Theta) \tag{1.1}$$

where Θ is a vector of parameters for the specified probability model. Since animals are considered to be independent from each other, and x_i is the number of animals with adverse effect in i th dose group, x_i is assumed to follow a binomial distribution:

$$\Pr(X_i = x_i \mid P = p_i) = \binom{n_i}{x_i} p_i^{x_i}(1-p_i)^{n_i - x_i} \tag{1.2}$$

with expectation mean $E(X_i) = n_i p_i$ and variance VAR(X)=$\sigma^2 = n_i p_i (1-p_i)$.

The BMR for quantal data is simply defined as the increase of the adverse response incidence probability from the background level. There are two commonly used methods to define this increase (additional risk and extra risk; Fillipsson et al., 2003), as shown in Table 1.1. The models used to model dichotomous data are probability density models such as logistic, probit, and Weibull-CDF, whose predictions lie between 0 and 1 for any possible dose (Table 1.3) (Fillipsson et al., 2003). The maximum likelihood method is the recommended method for dose-response model parameter estimation. The likelihood function, L, can be written as

$$L = \prod_{i=1}^{k} \binom{n_i}{x_i} p_i^{x_i}(1-p_i)^{n_i - x_i} \tag{1.3}$$

The log-transformed likelihood function is maximized in practice because it simplifies the mathematical calculations:

$$\log L = \sum_{i=1}^{k} \ln \binom{n_i}{x_i} + x_i \ln p_i + (n_i - x_i)\ln(1-p_i) \tag{1.4}$$

The model parameters that define p_i are the only unknowns and they are estimated by the values that maximize the value of L.

1.2.2 Continuous Response Data

1.2.2.1 Mean Value Method

Examples of continuous response data are the changes in bodyweight and height after a chemical exposure. For continuous data, the BMR is the minimal level of change in the endpoint that is generally considered to be biologically significant (US EPA, 2000). However, in many cases, the biologically significant cutoff values are unknown. Thus, an arbitrary cutoff value, such as 5% change or 2 standard deviations (SD) away from the mean response value in the control group, has often been used (US EPA, 2000). The definition of BMR for mean continuous response data is not as straight-forward as the BMR for quantal data. Table 1.2 presents five proposed possible definitions, among which the 'extra effect' and 'scaled effect' are the most commonly used (Table 1.2) (Fillipsson et al., 2003).

Dose-response models for continuous data include linear, polynomial, power models and other non-linear models such as Hill models (Table 1.4) (Fillipsson et al., 2003). If a normal distribution is assumed within each dose group, the likelihood function, L, and its log-transformed form $logL$ can be expressed as:

$$L = \prod_{i=1}^{k} \prod_{j=1}^{n_i} \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left(-\frac{(x_{ij} - \mu_i)^2}{2\pi\sigma_i^2}\right) \qquad (1.5)$$

$$\log L = -\frac{1}{2} \sum_{i=1}^{k} n_i \ln(2\pi) - \sum_{i=1}^{k} [n_i \ln \sigma_i + \frac{1}{2} \sum_{j=1}^{n_i} (\frac{x_{ij} - \mu_i}{\sigma_i})^2] \qquad (1.6)$$

where in a study containing k dose levels, x_{ij} is the response for the j th animal in the i th dose group. μ_i and σ_i^2 are the model predicted mean and variance for dose i, respectively.

1.2.2.2 Hybrid Method

The mean value methods of computing the BMD for continuous data consider only the change in the predicted mean values. Crump (1995) proposed a hybrid method which takes the variance of the continuous data into consideration. The hybrid method defines a model for continuous data that corresponds to a predetermined functional form for the probability of an adverse response, and consequently permits the same mathematical dose-response model to be applied to quantal and continuous data (Crump, 2000; Crump, 2002). Based on the assumption of a normal distribution at each dose, for the responses where smaller values are undesirable, the probability of an adverse response at dose d_i follows:

$$P_i = \Phi\left[\frac{c-\mu_i}{\sigma_i}\right] \tag{1.7}$$

where c is the cutoff value and Φ is the standard normal cumulative function. Thus, the BMR for continuous response data is harmonized with the BMR of the quantal data. For example, the additional risk follows:

$$BMR = P_{BMD} - P_0 = \Phi\left[\frac{c-\mu_{BMD}}{\sigma_{BMD}}\right] - \Phi\left[\frac{c-\mu_0}{\sigma_0}\right] \tag{1.8}$$

where P_0, μ_0 and σ_0 are the predicted probability of adverse response, mean value and SD in the control group, respectively. P_{BMD}, μ_{BMD}, and σ_{BMD} are the predicted probability of adverse response, mean value and SD at BMD. On the other hand, if P_0 (the adverse response in controlled animals) is specified, the cutoff value can be expressed as:

$$c = \Phi^{-1}(P_0) \times \sigma_0 + \mu_0 \tag{1.9}$$

Under this approach, the mean response value can be expressed as:

$$\mu_i = c - \sigma_i \Phi^{-1}[P_i] \tag{1.10}$$

P_i is defined using probability density models. The entire function can be fitted using the normal distribution maximum likelihood in equations (1.5) and (1.6).

1.2.2.3 Quantalization Method

A third approach of the continuous data BMD/BMDL estimation is to transform the continuous data into dichotomous format: the original continuous data are classified as 'normal' or 'abnormal' by comparison to a pre-determined cutoff and one then treats the 'quantalized' data as real quantal data. However, it has been found that substantial distributional information is lost during this transformation and the precision of the BMD estimation is seriously compromised. Therefore this method is not recommended (Gaylor, 1996).

1.2.3 Piecewise Linear-Link Model

In addition to the linear and non-linear dose-response models, a piecewise linear modeling approach has been adopted in a number of studies (Kim et al., 2004; Roberts et al., 2007). Steenland et al. (2001) and US EPA (2002) applied the piecewise linear method in the evaluation of dose response between 2,3,7,8-Tetrachlorodibenzo-*p*-Dioxin (TCDD) exposure and cancer mortality. Piecewise linear models typically are simply parameterized. Linearly connecting the response (or some function of the response) at the two doses that are adjacent to each other can usually provide adequate fit without assuming any consistent dose-response relationship across the range of the experimental doses.

1.2.4 Developmental Toxicology Data

The special character of developmental toxicology studies are that the adverse responses observed on fetuses are the consequence of maternal exposure. For example, the quantal endpoint in a developmental toxicology study is the proportion of the fetuses with malformation in a litter and the continuous endpoint is the individual fetal body weight. The fetuses from the same litter tend to react more similarly compared to the fetuses from other litters. The simplest method is to treat each fetus in a dose group as an independent observation, totally ignoring the

potential litter effects (Auton, 1994). However, this method is obviously an over-simplification that could not provide adequate estimates of the variance. More sophisticated statistical methods have been developed to account for the litter effect by estimating an intra-litter correlation (Ryan, 1992; Allen et al., 1994; Catalano et al., 1994).

1.2.4.1 Beta-Binomial for Quantal Data

The first method to characterize the intra-litter correlation is to model the proportions of offspring affected in a litter as beta-binomial distribution (Kupper et al., 1986). Assuming *k* dose groups ($d_1,d_2,....d_k$) , m_i is the total number of pregnant females in the *i* th dose group, m_1, $m_2....m_k$. The dichotomous random variable X_{ijg} (j=1,…, m_i and g=1,…, n_{ij}) takes the value 1 if the *g* th fetus within the *j* th litter in the *i* th dose group show adverse response (e.g., is malformed) and takes the value 0 otherwise. Hence, the random variable $x_{ij}= \sum_{k=1}^{n_{ij}} X_{ijg}$ is the number of affected fetuses within the *j* th litter at the *i* th dose. The litter is also assigned a litter-specific covariate, r_{ij} (this term is usually, but not necessarily, represented by the litter size). The probability that a fetus in the *j* th litter of the *i* th dose group is affected is represented by:

$$p_{ij} = p(d_i, r_{ij}, \Theta) \qquad (1.11)$$

where Θ is a vector of parameters for the specified probability model. Assuming the fetuses in the same litter are independent of each other ($\text{corr}(X_{ijg}, X_{ijg'} \mid P_i=P_{ij}) = 0$ for every $g \neq g'$), X_{ij} follows the binomial distribution.

$$\text{P}(X_{ij} = x_{ij} \mid P = p_{ij}) = \binom{n_{ij}}{x_{ij}} p_{ij}^{x_{ij}} (1-p_{ij})^{n_{ij}-x_{ij}} \qquad (1.12)$$

One logical way to introduce intra-litter correlation is to assume that P_i varies from litter to litter within the *i* th dose group and follows a beta distribution, the probability density function follows:

$$f(p_{ij}) = \frac{p_{ij}^{\alpha_i - 1}(1-p_{ij})^{\beta_i - 1}}{B(\alpha_i, \beta_i)} \qquad (0 < p_{ij} < 1,\ \alpha_i > 0, \beta_i > 0) \qquad (1.13)$$

where

$$B(\alpha_i, \beta_i) = \frac{\Gamma(\alpha_i)\Gamma(\beta_i)}{\Gamma(\alpha_i + \beta_i)} \qquad (1.14)$$

$\Gamma(\alpha_i)$ and $\Gamma(\beta_i)$ denote the gamma function. Under this assumption, the unconditional distribution of X_{ij} follows a beta-binomial distribution:

$$\mathrm{P}(X_{ij} = x_{ij}) = \binom{n_{ij}}{x_{ij}} \frac{B(\alpha_i + n_{ij}, \beta_i + n_{ij} - x_{ij})}{B(\alpha_i, \beta_i)} \qquad x_{ij} = 0,1,...,m_{ij} \qquad (1.15)$$

With expectation (mean):

$$E(X_{ij}) = n_{ij}E(P_i) = n_{ij}(\frac{\alpha_i}{\alpha_i + \beta_i}) = n_{ij}p_i \qquad (1.16)$$

$$Var(X_{ij}) = n_{ij}P_i(1-P_i)[1 + \rho_i(n_{ij} - 1)]$$

and for $g \neq g'$,

$$\mathrm{Corr}(X_{ijg}, X_{ijg'}) = \frac{1}{1 + \alpha_i + \beta_i} = \rho_i \qquad (1.17)$$

where ρ_i is the correlation between the individuals within a litter. A litter effect is reflected by a positive value of ρ, which is known as the over-dispersion parameter. The model parameters can be estimated by maximizing the log-likelihood function which is expressed as:

$$\log L = \sum_{i=1}^{k} \left\{ \sum_{j=1}^{m_i} \left[\sum_{g=1}^{x_{ij}} \ln(p_{ij} + (g-1)\psi_i) + \sum_{g=1}^{n_{ij} - x_{ij}} \ln(1 - p_{ij} + (g-1)\psi_i) - \sum_{g=1}^{n_{ij}} \ln(1 + (g-1)\psi_i) \right] \right\} \qquad (1.18)$$

where $\Psi_i = \frac{\rho_i}{1 - \rho_i}$.

1.2.4.2 GEE Method

Liang and Zeger (1986) and Zeger and Liang (1986) describe a general approach for the analysis of correlated data. This approach is referred to as Generalized Estimating Equations (GEE) and Ryan (1992) discusses the use of GEE method in developmental toxicology studies. The GEE approach requires specification of only the mean and variance functions of the data. An important addition in the GEE method is the inclusion of an empirical variance "fix-up" that relaxes the distributional assumptions so that the model parameters and their variances will be estimated correctly, even if the variance function is misspecified (Liang and Zeger, 1986). The GEE method works with individual fetal binary outcomes, instead of the litter counts. As defined earlier, assuming *k* dose groups ($d_1, d_2, \ldots . d_k$), X_{ijg} denotes the response of the *g* th fetus in the *j* th litter at the *i* th dose and x_{ij} is the vector of the outcomes for the *j* th litter in the *i* th dose group. Then, GEE estimating equations can be written in vector form as:

$$\sum_{i=1}^{k}\sum_{j=1}^{m_i} D_{ij}^{T} V_{ij}^{-1}(x_{ij} - u_i) = 0 \qquad (1.19)$$

where μ_i is the vector of means for the outcome x_{ij}. D_{ij} represents $\partial P_i / \partial \theta$, where θ represents the model parameters. V_{ij} is the covariance matrix of x_{ij}. V_{ij} then is written as:

$$V_{ij} = A_{ij}^{1/2} R A_{ij}^{1/2} \qquad (1.20)$$

where R is a suitable correlation matrix for x_{ij} and A_{ij} is the diagonal matrix

$$A_{ij} = diag[\mathrm{var}(x_{ijg})] = diag[P_i(1 - P_i)] \quad . \qquad (1.21)$$

If the assumed covariance of x_{ij} were correct, then the covariance of the estimated parameters could be estimated by:

$$\Sigma_N = (\sum_{i=1}^{k}\sum_{j=1}^{m_i} D_{ij}^{T} V_{ij}^{-1} D_{ij})^{-1} \qquad (1.22)$$

In practice, the correct covariance structure is uncertain. Therefore a robust empirical variance estimate uses the following expression:

$$\Sigma_E = (\sum_{i=1}^{k}\sum_{j=1}^{m_i} D_{ij}^T V_{ij}^{-1} D_{ij})^{-1} \times (\sum_{i=1}^{k}\sum_{j=1}^{m_i} D_{ij}^T V_{ij}^{-1} (x_{ij} - \mu_i)(x_{ij} - \mu_i)^T V_{ij}^{-1} D_{ij}) \times (\sum_{i=1}^{k}\sum_{j=1}^{m_i} D_{ij}^T V_{ij}^{-1} D_{ij})^{-1} \quad . \quad (1.23)$$

The intra-litter correlation, ρ_i is estimated using: $(x_{ij} - \mu_i)(x_{ij} - \mu_i)^T$.

1.2.5 BMDL Calculation

BMDL is the lower end of a one-side confidence limit for BMD (US EPA, 2000), with the most commonly used being the 95% confidence limits (α=0.05). Four methods widely used for BMDL computation are: delta method, likelihood-ratio related method, bootstrap method, and upper-band method (Crump and Howe, 1985; Moerbeek et al., 2004)

1.2.5.1 Delta Method

The delta method calculates confidence intervals assuming the asymptotic distribution of maximum likelihood estimates of BMD. The BMDL follows from:

$$BMDL = BMD - z_{1-\alpha}\sigma_{(BMD)} \quad (1.24)$$

where $\sigma^2_{(BMD)}$ is the estimated variance of BMD, $z_{1-\alpha}$ is the (1 - α)% percentile of the standard normal distribution. For 95% lower confidence interval of BMD, α = 0.05. Multiple studies have suggested that the delta method could not generate reliable confidence intervals for nonlinear dose-response models, either theoretically and practically. First, the estimator may not be asymptotically normal when the sample size is small. It was also observed that confidence intervals computed using the delta method appeared far too large and sometimes yield negative estimates of BMDL (Crump and Howe, 1985).

1.2.5.2 Likelihood-Ratio Based Method

The likelihood-ratio based method is recommended by US EPA and many researchers (Crump and Howe, 1985; US EPA, 2000; Moerbeek et al., 2004). In this method, the confidence limit is based on the asymptotic distribution of the likelihood ratio statistics. Briefly, one parameter is solved for explicitly and uniquely by BMD, BMR and other parameters. Then the expression for the parameter is substituted back into the model, and thus the newly parameterized model contains BMD as a parameter. Since the -2 log-likelihood ratio follows an approximate chi-square distribution if the assumed model is correct, the lower confidence limit for BMD is achieved by reducing the value of the BMD in the re-parameterized model until: $-2\ln(\frac{L_{\max}}{L}) = \chi^2_{(1-2\alpha,1)}$, where L_{max} is the maximum likelihood value of the 'fitted' model, L is the maximum likelihood value of the re-parameterized model, and $\chi^2_{(1-2\alpha,1)}$ is the chi-square distribution with one degree of freedom and an upper tail probability of 2α. The 1-2α is chosen because the confidence interval of interest is one sided (Crump et la., 2000; Wheeler, 2005).

1.2.5.3 Bootstrap Method

The bootstrap method is preferred by some researchers (Jacobson et al., 2002). The reference distributions are created by making a large number of simulations from the original data. The advantage of the bootstrap is that it does not require any statistical assumption for the BMD distributions. The BMDL then is then estimated using the percentile method.

1.2.5.4 Confidence Band Method (Delta Method in GEE)

Often seen in GEE estimation, a lower confidence limit on the BMD can be calculated as suggested by Kimmel and Gaylor (1988) based on the approximate normality of the estimated risk function. An upper confidence limit on the estimated risk function is computed and the dose

that corresponds to a 5% increased response above the background is determined form this upper limit. Take the quantal data for example. In the GEE model, the BMDL is the dose such that

$$g_{BMD} = g_{BMDL} + z_{1-\alpha}\sigma_{g_{BMDL}} \qquad (1.25)$$

where $g = \log(\frac{p}{1-p})$ and $\sigma^2_{g_{BMDL}}$ is the estimated variance of the logit function at BMDL, and $z_{1-\alpha}$ is the (1 - α)% percentile of the standard normal distribution.

1.3 DEVELOPMENTAL TOXICOLOGY OF RETINOIC ACID

All-trans retinoic acid (RA) is an endogenous metabolite of Vitamin A. It is required for normal pattern formation during embryogenesis and the developing limb is particularly susceptible to excessive or deficient amounts of RA (Kamm et al., 1984; Collins et l., 1999; Maden, 2000). Previous research suggested that all-trans RA produced limb defects and cleft palate in a dose-responsive manner. No fetal limb malformations were seen when a dose of 1 mg/kg RA or less was administered to pregnant mice on day 11 of the gestation (GD 11) or GD 13 (Kochhar et al., 1996; Hansen et al., 2001). Kochhar et al. (1996) reported that 10 mg/kg RA single administration to GD 11 mice produced mild limb reduction defects and cleft palate in 23% of the exposed fetuses (Kochhar et al., 1996). However, after a single dose of 10 mg/kg, Hansen et al. (2001) observed only 2.4% of the fetuses exhibiting cleft palate, while none of the fetuses had detectable forelimb malformations (Hansen et al., 2001). A single 100 mg/kg RA administration to GD 11 mice resulted in 90 -100% (Kochhar et al., 1996; Hansen et la., 2001) malformed fetuses.

The pattern of malformation in human fetuses closely resembled that produced in animal studies of retinoid teratogenesis (Lammer et al., 1985). Clinically, RA (brand name Vesanoid®) is used to treat acute promyelocytic leukemia. Although experience with pregnant women administered Vesanoid® is extremely limited, increased spontaneous abortions and major human

fetal abnormalities related to the use of other retinoids have been documented in humans (Vesanoid® Medication Guide, Roche Laboratories Inc., NJ, 1998). Accutane® (13-cis RA), a metabolites of all-trans RA, has been used clinically to treat nodular acne. Since there is an extremely high risk that severe birth defects will result if pregnancy occurs while taking Accutane® in any amount, even for short periods of time, strict contraindications are given to the use of Accutane® in female patients who are or may become pregnant (Accutane® Medication Guide, Roche Laboratories Inc., NJ, 2007). Other potential therapeutic uses of RA are currently under investigation such as the treatment of AIDS-related Kaposi's sarcoma (Cattelan et al., 2002).

Campbell et al. (2004) reported a comprehensive dose-response study including 5 different doses RA and 110 pregnant mice (around 1000 fetuses). In addition to the traditional quantal response endpoints (incidence of defected forelimbs and cleft palate), the length, proximal width, distal width, area, perimeter, and roundness of the forelimb bones (humerus, radius, ulna and scapula) were measured for each fetus. The estimated NOAE levels for all the endpoints were not lower than 10 mg/kg. The estimated lowest BMDL was only 0.24 mg/kg for the fetal humerus roundness. The author compared BMD/BMDL of the quantal and continuous endpoints and concluded that the continuous endpoints were not more sensitive than the quantal endpoints.

1.4 OBJECTIVES AND SPECIFIC AIMS

The objective of this thesis is to compare the different statistical methods used in the BMD/BMDL analysis on selected endpoints reported by Campbell et al (2004). To achieve this objective, the following specific aims are to be accomplished:

1) To develop dose-response model for the selected quantal and continuous endpoints and assess the goodness of fit.
2) To estimate the BMD/BMDL.
3) To compare the BMD/BMDL estimated by different methods.

Table 1.1. Benchmark response for dichotomous data. P_{BMD} and P_0 are the probability of the adverse response at BMD and control.

Additional Risk	Extra Risk
$P_{BMD} - P_0$	$\frac{P_{BMD} - P_0}{1 - P_0}$

Table 1.2. Benchmark response for continuous data. μ_{BMD} and μ_0 are the mean value of the continuous response at BMD and control. Max(μ) is the maximum mean value estimated by dose-response model.

Point effect	Absolute effect	Relative effect	Extra effect	Scaled effect
μ_{BMD}	$\mu_{BMD} - \mu_0$	$\frac{\mu_{BMD} - \mu_0}{\mu_0}$	$\frac{\mu_{BMD} - \mu_0}{Max(\mu) - \mu_0}$	$\frac{\mu_{BMD} - \mu_0}{\sigma_0}$

Table 1.3. Examples of dose-response model for dichotomous data (Fillipsson et al., 2003; US EPA BMDS (version 1.4.1c) Help System).

Model Name	Model Equations ($0 \le \gamma < 1$)	No. Parameters
Weibull-CDF	$P(d) = \gamma + (1-\gamma)(1 - e^{-\beta \times d^{\alpha}})$ $\alpha \ge 1$	3
Logistic	$P(d) = \frac{1}{1 + e^{-(\alpha + \beta \times d)}}$	2
Log-logistic	$P(d) = \begin{cases} \gamma & d = 0 \\ \gamma + \frac{1-\gamma}{1 + e^{-(\alpha + \beta \times \ln(d))}} & d > 0 \end{cases}$ $\beta \ge 0$	3
Multistage	$P(d) = \gamma + (1-\gamma) \times (1 - e^{-\sum_{j=1}^{n} \beta_j d^j})$ $1 < n < 23$	n+1
Gamma	$P(d) = \gamma + (1-\gamma) \times \frac{1}{\Gamma(\alpha)} \int_0^{\beta \times d} t^{\alpha-1} e^{-t} dt$ $\Gamma(\alpha) = \int_0^{\infty} x^{(\alpha-1)} e^{-t} dx$ $\alpha > 0$ $\beta \ge 0$	3
Probit	$P(d) = \Phi(\alpha + \beta \times d) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\alpha + \beta \times d} e^{-\frac{x^2}{2}} dx$	2
Log-Probit	$P(d) = \begin{cases} \gamma & d = 0 \\ \gamma + (1-\gamma)\Phi(\alpha + \beta \times \ln(d)) = \gamma + \frac{1-\gamma}{\sqrt{2\pi}} \int_{-\infty}^{\alpha+\beta\times\ln(d)} e^{-\frac{x^2}{2}} dx & d > 0 \end{cases}$ $\beta \ge 0$	3

Table 1.4. Examples of dose-response model for continuous data (Fillipsson et al., 2003; US EPA BMDS (version 1.4.1c) Help System).

Model Name	Model Equations	No. Parameters
Polynomial	$\mu(d) = \beta_0 + \beta_1 \times d + \beta_2 \times d^2 + \ldots + \beta_n \times d^n$	n+1
Power	$\mu(d) = \gamma + \beta \times d^{\alpha}$ $\alpha \geq 1$	3
Hill	$\mu(d) = \gamma + \nu \frac{d^n}{\kappa^n + d^n}$	4

Table 1.5. Examples of dose-response model for the dichotomous response in developmental toxicology (Fillipsson et al., 2003; US EPA BMDS (version 1.4.1c) Help System).

Model Name	Model Equations (r_{ij} is the litter-specific covariate for the *j* th litter in the *i* th dose group)	No. Parameters
Rai and Van Ryzin	$P(d) = (1 - e^{-(\alpha + \beta \times d^{\rho})}) e^{-(\theta_1 + \theta_2 \times d) \times r_{ij}}$	5 + k
NCTR	$P(d) = 1 - e^{-(\alpha + \theta_1 (r_{ij} - \bar{r})) - (\beta + \theta_2 (r_{ij} - \bar{r})) \times d^{\rho}}$ $\alpha + \theta_1 (r_{ij} - \bar{r}) \geq 0$ $\beta + \theta_2 (r_{ij} - \bar{r}) \geq 0$ $\bar{r}$ is the overall mean for the litter-specific covariate	5 + k
Logistic Nested	$P(d) = \begin{cases} \alpha + \theta_1 \times r_{ij} & d = 0 \\ \alpha + \theta_1 \times r_{ij} + \frac{1 - \alpha - \theta_1 \times r_{ij}}{1 + e^{-\beta - \theta_2 \times r_{ij} - \rho \ln(d)}} & d > 0 \end{cases}$ $\alpha + \theta_1 \times r_{ij} \geq 0$	5 + k

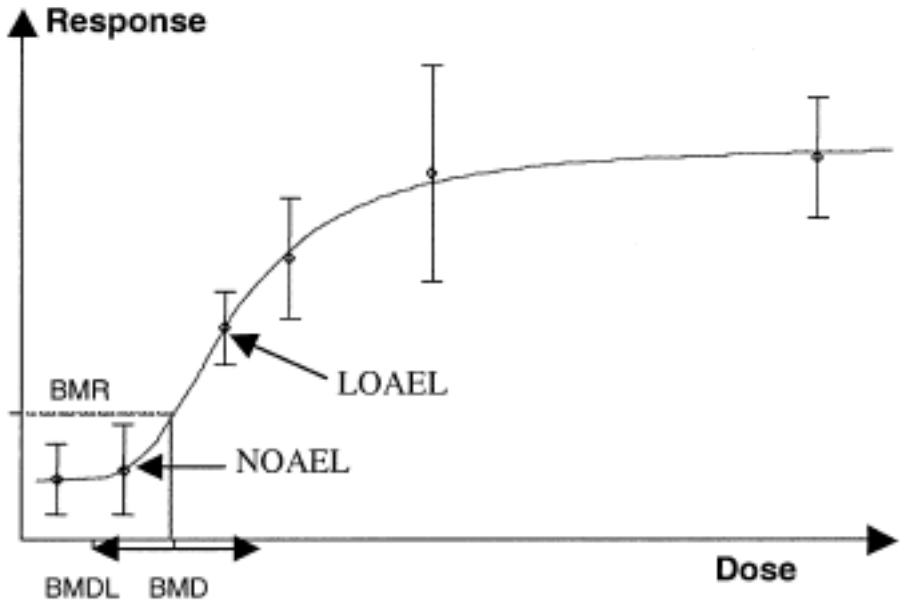

Figure 1.1. Illustration of dose-response assessment. The dot and vertical bar present the measured values and variation (e.g. +/- SD). The solid line presents the dose-response curve predicted by the model. NOAEL is the highest exposure level at which there is no statistically or biologically significant increases in the frequency or severity of adverse effect between the exposed population and its appropriate control. LOAEL is the lowest exposure level at which there is statistically or biologically significant increases in frequency or severity of adverse effects between the exposed population and its appropriate control group. BMR is a pre-determined increase in response (the benchmark response value). The BMD is the dose corresponds to BMR. The double-sided arrow indicated the confidence interval of the BMD. The lower confidence limit on the dose that would result in the defined response is the lower limit of BMD (BMDL). Adapted from Fillipsson et al., 2003.

CHAPTER 2

DATA AND METHODS

2.1 EXPERIMENTAL DATA

The experimental data were originally published by Campbell et al. (2004). Briefly, timed-pregnant CD-1 mice were obtained on gestation day 8 (GD8) and dosed with all-trans RA by oral gavage on GD11. Six dose groups (including control) were included: 0, 2.5, 10, 30, 60 and 100 mg/kg. The experiment was designed ideally to have 10 animals in the control group and approximately 25 animals in each of treatment groups. Because it was impossible to conduct the experiments on 135 animals simultaneously, the study was conducted in 5 sequential sub-experiments. Within each sub-experiment, at least 2 animals were randomized to the control group and 5 animals were randomized to each dose group. However because not all the animals were pregnant, the number of animals at each dose varied from 16 – 25. Mice remained in their home cage until GD18 when they were sacrificed by CO2 asphyxiation. The fetuses were removed from each mother's uterus and preserved for further analysis. All the animals were individually housed throughout the study.

As reported by Campbell et al. (2004), there were no maternal deaths in any of the treatment groups. The body weight of each fetus and the number of fetuses in each litter were recorded. The morphological evaluation was performed on the fetal forelimbs (humerus, radius, ulna and capula):

1) Incidences of fetal forelimb defect (humerus, radius, ulna and scapula) were obtained by visual inspection. The changes in overall size and shape of the ossified regions were

visually determined and recorded as normal (0) or abnormal (1). A fetus was given '1' when at least one malformed forelimb bone was observed.

2) Measurements of the fetal forelimbs were obtained by computerized image analysis (CIA) method. The maximal widths at both the proximal and distal ends of each bone were measured. Lengths were measured between the midpoints of the proximal and distal ends. The image analysis program calculated area, perimeter and roundness of each of the four bones.

2.2 SELECTION OF ENDPOINTS FOR ANALYSIS

Because it was of interest to compare the BMD/BMDL estimated from quantal response and continuous response data, the fetal forelimb defect (normal or abnormal) and fetal ulna length and roundness were selected for analysis. The morphological changes of ulna in the treatment groups were significant and easily observed by visual inspection. Both ulna length and roundness were analyzed, so that the sensitivity of different continuous responses on the same bone to all-trans RA could be compared. Meanwhile, the data were analyzed using litter-based and fetal-based piecewise-linear approaches. For the litter based approach, one litter was treated as an experimental unit and the proportion of defected fetus or the average fetal ulna length and roundness were analyzed. For the fetal-based piecewise-linear approach, each fetus was treated as a unit and the measurements for each fetus were analyzed directly.

2.3 NOAEL AND LOAEL CALCULATION

For comparison purposes, NOAEL and LOAEL were determined by statistically comparing each dose group against the control group. Dunnett's test (2-sided, α=0.05) was conducted on litter size (number of fetuses in each litter), litter average fetal body weight, litter

average fetal ulna length and roundness. Fisher's exact test was used to compare the proportion of defective fetuses.

2.4 DATA TRANSFORMATION

The distributions of the quantal and continuous responses were analyzed. Logarithmic transformations were applied to some continuous responses to ensure approximate normality at each dose level.

2.5 SELECTION OF BMR

For the quantal response, the BMR was set to 5% additional risk: BMR= $P_{BMD} - P_0 = 0.05$. This BMR was chosen because it was found to be similar to the statistically derived NOAEL in developmental toxicology (Allen et al., 1994; Kavlock et al., 1995). For the continuous response, the cutoffs for the abnormal ulna length and roundness were set to be 2SD smaller than the means in the control group, corresponding to a malformation probability of 0.023 in the control group (assuming a normal distribution for the control data).

2.6 LITTER-BASED APPROACH

The NLMIXED procedure in SAS® 9.1 was used to obtain the maximum likelihood parameter estimates.

2.6.1 Beta-Binomial Distribution for Incidence of Forelimb Defect

The probability of fetuses with forelimb malformations in a litter at each dose was assumed to follow a beta-binomial distribution. Because the beta-binomial distribution is not a built-in distribution in NLMIXED, it was specified through the general log-likelihood option (Nelson et al., 2006):

$$LL = \ln\left(\frac{\Gamma(x_{ij}+\alpha_i)\times\Gamma(n_{ij}-x_{ij}+\beta_i)}{\Gamma(\alpha_i+\beta_i+n_{ij})}\right) - \ln\left(\frac{\Gamma(\alpha_i)\times\Gamma(\beta_i)}{\Gamma(\alpha_i+\beta_i)}\right) \qquad (2.1)$$

where $\alpha_i = \frac{p_i(1-\rho_i)}{\rho_i}$, $\beta_i = \frac{(1-p_i)(1-\rho_i)}{\rho_i}$, n_{ij} is the total number of fetuses of *j* th litter of the *i* th dose group, x_{ij} is the number of fetuses with forelimb malformation in *j* th litter of the *i* th dose group, ρ_i is the intra-litter correlation parameter of *i* th dose group, p_i is probability of the fetus with forelimb malformation of the *i* th dose group.

2.6.2 Cutoff Values for Ulna Length and Roundness

Litter mean values of ulna length and roundness were modeled using the hybrid method. The cutoff value of malformation was set to be: $c = \mu_0 - 2\sigma_0$, where μ_0 and σ_0 are the model estimated mean and standard deviation of the ulna length or roundness in the control group. Since BMD falls in the low end of the range of doses and the standard deviation at BMD (σ_{BMD}) is expected to be close to σ_0, it was assumed that $\sigma_{BMD} = \sigma_0$.

$$\begin{aligned} P_{BMD} &= \Phi\left(\frac{c-\mu_{BMD}}{\sigma_{BMD}}\right) \\ &= P_0 + BMR = \Phi\left(\frac{c-\mu_0}{\sigma_0}\right) + BMR = \Phi(-2) + 0.05 = 0.073 \end{aligned} \qquad (2.2)$$

therefore,

$$\mu_{BMD} = c - \Phi^{-1}(0.073) \times \sigma_0 = \mu_0 - 0.545\sigma_0 . \qquad (2.3)$$

Since $P_0 = \Phi(-2) = 0.023$, using 2SD as cutoff value is equivalent to a background malformation probability of 0.023, assuming normality of the distribution of length or roundness.

2.6.3 Dose-Response Model and BMD Calculation

The log-logistic model was used to model the relationship between the endpoints and doses. The log-logistic model was chosen because it appears to be flexible to fit dose-response data from various studies (Rogers et al., 1993). Meanwhile, it has been suggested that different models would give similar estimates for BMD/BMDL as long as the model provides feasible fits

to the experimental data. Therefore, no effort was given to find the model providing the best fits. The probability of malformation at each dose level (P(d)) was modeled as:

$$P(d) = \begin{cases} \gamma + \theta_1 \times r_{ij} & d = 0 \\ \gamma + \theta_1 \times r_{ij} + \dfrac{1 - \gamma - \theta_1 \times r_{ij}}{1 + e^{-\alpha - \theta_2 \times r_{ij} - \beta \ln(d)}} & d > 0 \end{cases} \quad (2.4)$$

where γ, α and β are the logistic model parameters, θ_1 and θ_2 are the coefficient of the litter effect covariate and r_{ij} is the litter effect (litter size in this study). Therefore:

$$BMD = \exp\left(\frac{\alpha + \theta_2 \times r_{ij} + \log\left(\frac{1 - \gamma - \theta_1 \times r_{ij}}{0.05} - 1\right)}{-\beta}\right) \quad (2.5)$$

For the ulna length and roundness, the mean value at dose *i* was modeled as

$$\mu_i = \mu(d_i) = c - \sigma_i \Phi^{-1} P(d_i). \quad (2.6)$$

The μ_i were modeled using the normal distribution model in PROC NLMIXED.

2.6.4 BMDL Calculation

2.6.4.1 Delta Method

Delta method BMDLs were estimated via the PREDICT statement in the NLMIXED procedure. The lower end of the 2-sided 90% CI of the estimated BMD gave the 1-sided 95% BMDL.

2.6.4.2 Likelihood-Ratio Based Method

There were 3 major steps in likelihood ratio based BMDL calculation: 1) re-parameterizing the dose-response model so that the BMD was included as a parameter in the model, 2) running the maximum likelihood estimation processes multiple times. In each run, BMD was continually decreased by a small amount, until the new log-likelihood (LL_{BMDL}) was

reduced from LL_{max} to $\chi^2_{(1-2\alpha,1)}/2$ where $\chi^2_{(1-2\alpha,1)}$ is the chi-square distribution with one degree of freedom and an upper tail probability of 2α, and 3) BMDL was calculated for the final model.

A SAS MACRO reported by Wheeler (2005) was modified to perform the likelihood-ratio based BMDL calculation. The log-logistic model was refitted, in which β was substituted by an expression of BMD:

$$\beta = \frac{-\alpha - \log(\frac{1-\gamma-BMR}{BMR})}{\log(BMD)} \text{ , where BMR=0.05} \qquad (2.7)$$

The NLMIXED processes were run repeatedly with decreased BMD in each run (-2%), until the new log-likelihood (LL_{BMDL}) was reduced from original maximum log-likelihood (LL_{max}) to $\chi^2_{(0.9,1)}/2$. The BMDL is the BMD value in the model with LL_{BMDL}.

2.6.4.3 Bootstrap Method

The bootstrap method for BMDL calculation involved four steps: 1) new samples were created with the same number of observations in each dose group by random sampling with replacement from the original data, 2) dose-response models were re-fitted to the newly created data and the parameters and BMD for the new data were computed, 3) step1 and 2 were repeated thousands of times, and 4) the α-percentile of this bootstrap sample of BMDs served as an estimate of $BMDL_{1-\alpha}$.

The PROC SURVEYSELECT procedure was used to create 1000 bootstrap samples. The litters within each dose group were randomly sampled with replacement and the numbers of litters at each dose were kept the same as in the original experiment. The model was fitted to the 1000 samples, the BMD was calculated, and the BMDL was set to the 5th percentile of the bootstrap distribution.

2.6.5 Goodness of Fit Test

2.6.5.1 Pearson Chi-Square Test

The chi-square statistics for quantal endpoints model were calculated as:

$$\chi^2 = \sum_{i=1}^{N} \frac{(X_i - \widehat{X}_i)^2}{\widehat{X}_i} \qquad (2.8)$$

where N is the total number of observations, X_i and $\widehat{X}_i$ are the *i* th observation and prediction, respectively. The value of χ^2 was compared to the chi-square distribution at the degree of freedom = number of samples – 1. The χ^2 value, and its corresponding p-value are an indication of that "closeness". If the p-value is larger than 0.1 then the model was determined to adequately fit the data.

2.6.5.2 Log-likelihood Ratio Chi-Square Test

The chi-square statistics for continuous endpoints model were calculated as:

$$\chi^2 = 2(LL_{mean} - LL_{\mathrm{mod}\,el}) \qquad (2.9)$$

where LL_{mean} is the log-likelihood value for the mean model (full model with MLE mean and variance for each dose), $LL_{\mathrm{mod}\,el}$ is the log-likelihood value for the log-logistic model. The value of χ^2 was compared to the chi-square distribution with the degree of freedom equal to the difference between the number of parameters in the two models. If the p-value is larger than 0.1 then the log-logistic model was determined to adequately fit the data.

2.6.5.3 Pearson Residual Plots

The Pearson residuals of the quantal data model were plotted against the predicted values.

$$\text{Pearson Residuals} = \frac{X_i - \widehat{X}_i}{\sqrt{Var(\widehat{X}_i)}} \qquad (2.10)$$

where X_i is the *i* th observed value, $\widehat{X}_i$ is the predicted value, and $Var(\widehat{X}_i)$ is the estimate of the variance. Under the correct model assumptions, and assuming the X_i are not too small, Pearson residuals approximate a normal distribution and the residual plots are randomly distributed around 0.

2.6.5.4 Standardized Residual Plots

The standardized residuals of the continuous data model were plotted against the predicted values.

$$\text{Standardized Residuals} = \frac{X_i - \widehat{X}_i}{\sqrt{\widehat{Var}(\varepsilon_i)}} \qquad (2.11)$$

where X_i is the *i* th observed value, $\widehat{X}_i$ is the predicted value and $\varepsilon_i = X_i - \widehat{X}_i$. Under the correct model assumption, standard residuals approximate a normal distribution and the residual plots are randomly distributed around 0.

2.6.5.5 Normal Distribution of Residuals

Histogram and quantile-quantile (Q-Q) plots were used to examine the distribution of the Pearson residuals of the quantal endpoint model and the standardized residuals of the continuous endpoints model. Serious departure of the residuals from the normal distribution suggested mis-specification of the model, or very small m_i values.

2.7 FETAL-BASED PIECEWISE-LINEAR APPROCH

The PROC GENOMD procedure was used to fit the piecewise linear model using the data of each fetus. The correlation between the littermates was assumed exchangeable (Carr et al., 1993). The quantal responses were modeled as a binary distribution with logit as the link function. The continuous responses were modeled as normally distributed with the identity as the link function.

2.7.1 Piecewise Linear Model and BMD Calculation

The mean value at each dose group was linear linked and the intra-litter correlation was accounted for by the GEE method. The general form of the model equations are expressed as:

$$PV = \begin{cases} \beta_0 & d = 0 \\ \beta_0 + \beta_1 \times \ln d & 0 < d \le 2.5 \\ \beta_0 + \beta_1 \times \ln d + \beta_2 \times (\ln d - \ln(2.5)) & 2.5 < d \le 10 \\ \beta_0 + \beta_1 \times \ln d + \beta_2 \times (\ln d - \ln(2.5)) + \beta_3 \times (\ln d - \ln(10)) & 10 < d \le 30 \\ \beta_0 + \beta_1 \times \ln d + \beta_2 \times (\ln d - \ln(2.5)) + \beta_3 \times (\ln d - \ln(10)) & 30 < d \le 60 \\ \quad + \beta_4 \times (\ln d - \ln(30)) & \\ \beta_0 + \beta_1 \times \ln d + \beta_2 \times (\ln d - \ln(2.5)) + \beta_3 \times (\ln d - \ln(10)) & 60 < d \le 100 \\ \quad + \beta_4 \times (\ln d - \ln(30)) + \beta_5 \times (\ln d - \ln(60)) & \end{cases} \quad (2.12)$$

where PV is the predicted variable which presented the means for continuous data and the logistic form of the probability for the quantal data.

2.7.1.1 BMD for Quantal Data

For the quantal data, instead of modeling the absolute proportion of abnormality at each dose, the logistic form of the probability at dose was linearly linked, PV= $\ln(\frac{P}{1-P})$, which yields Equation (2.13):

$$BMD = \begin{cases} \exp\left(\dfrac{\ln(\frac{P_{BMD}}{1-P_{BMD}}) - \beta_0}{\beta_1}\right) & 0 < BMD \le 2.5 \\ \exp\left(\dfrac{\ln(\frac{P_{BMD}}{1-P_{BMD}}) - \beta_0 + \ln(2.5) \times \beta_2}{\beta_1 + \beta_2}\right) & 2.5 < BMD \le 10 \\ \exp\left(\dfrac{\ln(\frac{P_{BMD}}{1-P_{BMD}}) - \beta_0 + \ln(2.5) \times \beta_2 + \ln(10) \times \beta_3}{\beta_1 + \beta_2 + \beta_3}\right) & 10 < BMD \le 30 \\ \exp\left(\dfrac{\ln(\frac{P_{BMD}}{1-P_{BMD}}) - \beta_0 + \ln(2.5) \times \beta_2 + \ln(10) \times \beta_3 + \ln(30) \times \beta_4}{\beta_1 + \beta_2 + \beta_3 + \beta_4}\right) & 30 < BMD \le 60 \\ \exp\left(\dfrac{\ln(\frac{P_{BMD}}{1-P_{BMD}}) - \beta_0 + \ln(2.5) \times \beta_2 + \ln(10) \times \beta_3 + \ln(30) \times \beta_4 + \ln(60) \times \beta_5}{\beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5}\right) & 60 < BMD \le 100 \end{cases}$$

2.7.1.2 BMD for Continuous Data

For the continuous data, PV equals to the fetal mean value at each dose level and $\mu(BMD) = 0.95\mu(0)$. Therefore, BMD could be expressed by Equation (2.14):

$$BMD = \begin{cases} \exp(\frac{\mu_{BMD} - \beta_0}{\beta_1}) & 0 < BMD \le 2.5 \\ \exp(\frac{\mu_{BMD} - \beta_0 + \ln(2.5) \times \beta_2}{\beta_1 + \beta_2}) & 2.5 < BMD \le 10 \\ \exp(\frac{\mu_{BMD} - \beta_0 + \ln(2.5) \times \beta_2 + \ln(10) \times \beta_3}{\beta_1 + \beta_2 + \beta_3}) & 10 < BMD \le 30 \\ \exp(\frac{\mu_{BMD} - \beta_0 + \ln(2.5) \times \beta_2 + \ln(10) \times \beta_3 + \ln(30) \times \beta_4}{\beta_1 + \beta_2 + \beta_3 + \beta_4}) & 30 < BMD \le 60 \\ \exp(\frac{\mu_{BMD} - \beta_0 + \ln(2.5) \times \beta_2 + \ln(10) \times \beta_3 + \ln(30) \times \beta_4 + \ln(60) \times \beta_5}{\beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5}) & 60 < BMD \le 100 \end{cases}$$

1) Mean Value Model

The mean value of the fetal ulna length and roundness were modeled by the piecewise linear model. Two BMD/BMDL values were estimated based on two different BMR definitions. The first BMR was defined by the changes of the mean value from the control group:

$$BMR = 0.05 = \frac{\mu_0 - \mu_{BMD}}{\mu_0} \Rightarrow \mu_{BMD} = 0.95\mu_0 \qquad (2.15)$$

The second BMR was defined to be the same as in the 'hybrid method' (Equation 2.3). Assuming the fetal data in the control group were normally distributed with the litter effects ignored temporarily, the cutoff for malformation was 2SD shorter than the mean in the control group. The mean value at BMD was: $\mu_{BMD} = \mu_0 - 0.545\sigma_0$.

2) Quantalized Data Model

Although the quantalization method for continuous data is not recommended, it has been utilized in developmental toxicology studies (Kavlock et al., 1995). The cutoff value $c = \mu_0 - 2\sigma_0$ was used to dichotomize the continuous data to binary data. If the observed value

was larger than the cutoff, then the fetus was considered to be 'normal', if the observed value was smaller than the cutoff then the fetus was considered to be 'abnormal'. Finally, the binary data were modeled (Equation 2.15).

2.7.1.3 BMDL Calculation

The GEE method is different from maximum-likelihood based method and the likelihood ratio based method could not be used for BMDL estimation (Ryan, 1992).

2.7.1.4 Confidence Band Method (Delta method in GEE)

The upper-limit bound method was to find the BMDL in which the 95% upper limit of P(BMDL) was equal to P(BMD) for the quantal data or 95% lower limit of μ(BMDL) equal to μ(BMD) for the continuous data. A dose that was smaller than BMD was manually inserted into the experimental data and its corresponding responses were set to missing. GENMOD calculated the predicted values and CIs for those 'pseudo' doses. This process was repeated until the BMDL was obtained (SAS/STAT 9.1 User's Guide, 2004).

2.7.1.5 Bootstrap Method

The individual fetal data within each litter were randomly sampled with replacement and the numbers of fetuses in each litter were kept the same as in the original experiment. Since BMD was not expected to be larger than 30 mg/kg in this study, to reduce the unnecessary computation, the bootstrap was conducted only on the 0 – 30 dose groups. However, not all the bootstrap samples could be used in BMD estimation. The bootstrap samples in which none of the fetuses at a dose level were defective (observed probability of malformation (P) was 0) in the control group had to be dropped, because the logit form ($\ln(\frac{P}{1-P})$) could not be calculated in this case. The bootstrap was run for 1500 or more times and the first 1000 samples satisfying the requirements were kept. The BMDL was the 5% percentile of the bootstrap distribution.

Deleting the bootstrap samples with zero malformation incidence at any dose level, unavoidably introduced bias resulting in overestimated malformation probabilities in the remaining bootstrap samples. However, assuming positive responses in low dose groups were not unrealistic. This bias wasn't too extreme, since a potential bootstrap sample would be dropped from the analysis only if every litter in a dose group failed to produce a deformity (or if every litter produced all deformities). In 1131 bootstrap samples, 131 had to be dropped because of no simulated malformations among the low dose groups.

2.7.2 Goodness of Fit Test

2.7.2.1 GEE-R^2

Since GEE model did not specify a likelihood, an extension of the R^2 measurement was calculated as the proportion of variance in the outcome that was explained by the model (Hardin and Hilbe, 2003):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(X_i - \widehat{X}_i)^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2} \tag{2.16}$$

where n is the number of observations, X_i is the *i* th observed value, $\widehat{X}_i$ is the predicted value, and $\overline{X}$ is the overall mean calculated as $\sum_{i=1}^{n} X_i / n$. A R^2 value of 0.6 was considered a good-fit for clustered data (Ballinger, 2004; Soaita, 2006).

2.7.2.2 Pearson Residual Plots

The Pearson residuals were plotted against the predicted values.

$$\text{Pearson Residuals} = \frac{X_i - \widehat{X}_i}{\sqrt{Var(\widehat{X}_i)/w_i}} \tag{2.17}$$

where X_i is the *i* th observed value, $\widehat{X_i}$ is the predicted value, $Var(\widehat{X_i})$ is the variance function, and w_i is the weight. A random distribution of residuals around 0 suggests adequacy of the fitted model. The Pearson residuals of the quantal data were plotted against the predicted probability with the locally weighted least squares smooth (Lowess). The Lowess smooth was created by PROC LOESS in SAS and the smooth parameters were obtained by globally minimizing the AICC criterion (SAS/STAT 9.1 User's Guide, 2004). An adequate modeling fitting was confirmed if the Lowess smooth was approximately a straight line with zero intercept and zero slope (Kutner et al., 2005).

CHAPTER 3

RESULTS

3.1 LITTER-BASED APPROACH RESULTS

Under the litter-based approach, all the statistical testes were conducted on the mean responses of a littler (e.g., average fetal ulna length and roundness or proportion of defected fetuses in one litter). Accordingly, the sample size in each dose group is the number of dams (litters).

3.1.1 NOAEL and LOAEL for Selected Endpoints

The least-squared means, p-values, NOAEL and LOAEL are presented in Table 3.1. The litter size and fetal body weights in all the dose groups were not significantly different from the control group at α=0.05. Therefore the NOAEL and LOAEL for litter sizes and fetal body weights could not be determined in the present study. From a toxicology point of view, similar litter sizes and fetal body weights in all dose groups indicated that the all-trans RA did not affect the fetal fatality and body weight significantly after a single oral dose of 2.5 to100 mg/kg. The endpoints that were significantly affected by all-trans RA in the experimental dose range were further analyzed by BMD method. The box-plots and sample size (N), mean and SD of fetal ulna length, roundness, and proportion of defected fetuses (malformation quantal data) are presented in Figure 3.1.

Table 3.2 summarizes the data on malformations, ulna length, and ulna roundness in slightly more detail than does Table 3.1. A question that one must resolve in performing model fits is whether results within a dose-level should be averaged at the fetus level or at the litter

level. Of course, if every litter had the same number of fetuses, these would yield the same value. On the average, each litter has about 10 fetuses, but the number of fetuses is not constant, so the two averaging methods won't be identical. In Table 3.2, the 'pooled' columns mean that the average is calculated by pooling over all fetuses in the dose level, while the 'unpooled' column represents calculating the average separately for each litter and then averaging these values, with each litter being weighted equally, independent of sample size. For most of the analyses in this thesis, we are using the 'unpooled' results, as can be seen from matching the results in Tables 3.1 and 3.2. The reasoning is that we believe effects are more likely be felt at the maternal ('dam') level, with results from the same mother likely to be highly correlated. Since the litter sizes tend to be similar (around n=10), there is usually not much difference between the two methods. One exception occurs for P(malformation) at Dose=10 mg/kg, where the unpooled proportion (.2977) is much higher than the pooled proportion (.2171), a result which occurred, as can be noted from the box plots of Figure 3.1 and the histograms in Figure 3.2, because of many litters with small deformation probabilities and a few with all or almost all deformed. We feel that the process which we are modeling is more appropriately modeled by averaging the litter averages than by pooling over all the fetuses in the class.

3.1.2 Model for Malformation Probability

Figure 3.2 presents the histogram of the distribution of malformation probability in each litter and the estimated beta distributions at each dose. The distribution of the malformation probability did not match the beta distribution perfectly. The beta-binomial distribution is only a simplified approach to account for over-dispersion in order to use maximum-likelihood estimation and it appears to work well for clustered quantal data in developmental toxicology (Chen et al., 1989 and Fung et al., 1998).

3.1.2.1 Model Building

The log-logistic model is widely used to model quantal data response due to its flexibility. In addition, the parameters in the log-logistic model relate to certain characteristics of biological responses too. For example, p_0 represents the background malformation probability and β is the slope parameter controlling the intensity of the change (Equation 2.4). To avoid unnecessary complexity while providing adequate fits to the experimental data, three model structures were tested for the malformation probability quantal data:

1) Full model (Model I), containing parameters for log-logistic model (p_0, α, and β), parameters for covariate of litter size (θ_1 and θ_2) and parameters for intra-litter correlation at each dose (ρ_1 - ρ_6).
2) Reduced model without litter size covariate (Model II);
3) Reduced model without litter size covariate and one fixed intra-litter correlation across all doses (Model III).

The maximum-likelihood estimates for the parameters are presented in Table 3.3 and the fit statistics are in Table 3.4. Since the coefficient for litter size covariate (θ_1 and θ_2) were not significantly different from 0 (P-value > 0.1) and Model II had almost the same AIC as the full model, Model II was used in BMD calculation. The value of ρ_1 - ρ_6 were limited between 0.01 – 1. ρ_1, ρ_5, and ρ_6 reached the lower limit (0.01) suggesting the over-dispersion parameters were not necessary in those dose groups. Since such small value of ρ_1, ρ_5, and ρ_6 would have little impact on the model fittings, they were left in the model to be consistent with other dose groups.

3.1.2.2 Goodness of Fit Test Results

1) Pearson Chi-Square Test. The calculated Chi-square statistics was 77.61 with 109 degree of freedom. The corresponding P-value was 0.99, suggesting adequate fit of the beta-

binomial log-logistic model.

2) Diagnostic plots. The Pearson residuals were plotted against the predicted values. No pattern was identified for the residual plots. The histogram of the residuals and the Q-Q plot indicates that the residuals are approximately normally distributed (Figure 3.3). Randomly distributed residual plots and approximately normally distributed Pearson residuals confirmed that the beta-binomial model was appropriate for the quantal malformation endpoints.

3.1.2.3 Model Prediction and BMD/BMDL

The observed mean malformation probability and model predicted value at each dose is presented in Table 3.5. The estimated BMD was 3.42 mg/kg. The estimated BMDL was 1.87 mg/kg by delta method, 2.24 mg/kg by likelihood ratio based method, and 2.37 mg/kg by bootstrap method (Table 3.6). The model predictions and the experimental data along with the reduced likelihood model are presented in Figure 3.4. The distribution of bootstrap BMD for fetal malformation probability is presented in Figure 3.5.

3.1.3 Hybrid Method for Fetal Ulna Length

Different transformations were tested for litter average ulna length in each dose group and the original untransformed data was determined to be most normally distributed (Figure 3.6) across all the doses. Approximate normality of the continuous data ensured the reliable estimates of the CDF of the assumed normal distribution in the hybrid method.

3.1.3.1 Model Building

The normal distribution was assumed for the ulna length in each dose group. The cutoff value for the malformation was 2SD's shorter than the mean of the control group. The log-logistic model was used to simulate the CDF of malformation at each dose. Similar to the model for the quantal data, μ_0 (easily converted to p_0 assuming a normal distribution), α and β were the

log-logistic model parameters and the σ_1 - σ_6 are the estimated standard deviations for the dose groups. The maximum-likelihood estimates are presented in Table 3.7.

3.1.3.2 Goodness of Fit Test Results

1) Likelihood Ratio Chi-Square Test. The calculated Chi-square statistic was 5.4 with 3 degree of freedom (12 parameters in the mean model minus 9 parameters in the log-logistic hybrid model). The corresponding P-value was 0.86, suggesting adequate fits of the log-logistic hybrid model.

2) Diagnostic plots. The standardized residuals were plotted against the predicted values. No pattern was identified for the residual plots. The histogram of the residuals and the Q-Q plot indicates that the residuals are approximately normally distributed (Figure 3.7). Random and approximately normally distributed residuals confirmed that the relationship between fetal ulna lengths and doses were appropriately modeled by the log-logistic hybrid model.

3.1.3.3 Model Prediction and BMD/BMDL

The observed mean fetal ulna length, the model predicted value and the malformation probability assuming normal distribution at each dose are presented in Table 3.8. The estimated BMD was 8.00 mg/kg. The BMDL was estimated as 3.26 mg/kg by delta method, 3.64 mg/kg by likelihood ratio based method and 3.81 mg/kg by bootstrap method (Table 3.9). The model predictions and the experimental data along with the reduced likelihood model are presented in Figure 3.8. The distribution of bootstrap BMD for fetal ulna length is presented in Figure 3.9.

3.1.4 Hybrid Method for Fetal Ulna Roundness

After a power transformation (power = 2), the distribution of ulna- roundness-square appeared to be approximately normally distributed at all the doses (Figure 3.10).

3.1.4.1 Model Building

The normal distribution was assumed for the ulna roundness-square in each dose group and ulna roundness-square was the dependent variable in the dose-response model. The log-logistic model was used to simulate the CDF of malformation at each dose. Similar to the model for the ulna length, the maximum-likelihood estimates for μ_0, α, β, and σ_1 - σ_6 are presented in Table 3.10.

3.1.4.2 Goodness of Fit Test Results

1) Likelihood Ratio Chi-Square Test. The calculated Chi-square statistic was 2.3 with 3 degree of freedom (12 parameters in the mean model minus 9 parameters in the log-logistic hybrid model). The corresponding P-value was 0.49, suggesting adequate fit of the log-logistic hybrid model.

2) Diagnostic plots. The standardized residuals were plotted against the predicted values. No pattern was identified for the residual plots. The histogram of the residuals and the Q-Q plot indicates that the residuals are approximately normally distributed (Figure 3.11). Random and approximately normally distributed residuals confirmed that the relationship between fetal ulna roundness and doses were appropriately modeled by the log-logistic hybrid model.

3.1.4.3 Model Prediction and BMD/BMDL

The observed mean fetal ulna roundness-square, the model predicted value, and the malformation probability (assuming a normal distribution at each dose) are presented in Table 3.11. The estimated BMD was 5.36 mg/kg. The BMDL estimated was 3.02 mg/kg by delta method, 2.92 mg/kg by likelihood ratio based method, and 1.16 mg/kg by bootstrap method (Table 3.12). The model predictions and the experimental data along with the reduced likelihood

model are presented in Figure 3.12. The distribution of the bootstrap BMD for fetal ulna roundness is presented in Figure 3.13.

3.2 FETAL-BASED PIECEWISE-LINEAR APPROACH RESULTS

The GEE method was applied to the dose-response analysis based on fetal data to account the intra-litter correlation. SAS PROC GENMOD was used to conduct the GEE estimation and binary distribution and normal distribution was assumed for the quantal and continuous data.

3.2.1 Model for Malformation Probability

3.2.1.1 Model Building

The estimated model parameters were β_0, β_1, β_2, β_3, β_4 and β_5 in equation (2.12). The GEE estimates for the parameters and the correlation coefficient are presented in Table 3.13.

3.2.1.2 Goodness of Fit Results

1) The calculated R^2 was 0.73, suggesting 73% of the variation of the outcome was explained by the model. 2) Pearson Residual plots. The Pearson residuals were plotted against the predicted values (Figure 3.14). The Lowess smooth approximated a horizontal line with zero intercept, suggesting that the model was adequate.

3.2.1.3 Model Prediction and BMD/BMDL

The observed mean malformation probability and model predicted value at each dose is presented in Table 3.14. The estimated BMD was 2.92 mg/kg. The delta method and bootstrap method estimated BMDL were 1.91 mg/kg and 2.21 mg/kg, respectively (Table 3.15). The model prediction, observed litter mean probability and the joint 90% CI are presented graphically in Figure 3.15. The histogram of the BMD distribution for the 1000 bootstrap samples is presented in Figure 3.16.

3.2.2 Model for Fetal Ulna Length

3.2.2.1 Fetal Ulna Length Treated as Continuous Data

The histogram of the distribution of the fetal ulna length in each dose group is presented in Figure 3.17. The distributions in the control group (0 mg/kg) and 2.5, 10 and 30 mg/kg dose groups approximate normal distributions. However, larger variances were observed in 60 and 100 mg/kg dose groups. The variances were weighted to squared means at each dose in the GEE model. The GEE estimates for the parameters and the correlation coefficient are presented in Table 3.16.

The calculated R^2 was 0.59, suggesting 59 % of the variation of the outcome was explained by the model. The Pearson residuals appear to be randomly distributed (Figure 3.18). The observed fetal mean and predicted ulna length at each dose are presented in Table 3.18. The model prediction and experimental data, along with the joint 90% CI are presented in Figure 3.19. By the definition of equation 2.15, the mean ulna length at BMD was $0.95\mu_0$. The estimated BMD was 13.30 mg/kg. The delta method and bootstrap method estimated BMDL were 11.3 and 12.14 mg/kg, respectively (Table 3.19). According to the hybrid BMR definition, the mean ulna length at BMD was μ_0-$0.545\sigma_0$, which yielded 13.26 mg/kg as the estimated BMD. The delta method and bootstrap method estimated BMDL were 11.3 and 11.15 mg/kg, respectively (Table 3.19). The distributions of BMD, by the two BMR definitions, based on 1000 bootstrap samples are presented in Figure 3.22 a) and b).

3.2.2.2 Quantalized Fetal Ulna Length

Although quantalization of continuous data method is not recommended, it has been reported in some developmental studies (Kavlock et al., 1995). For comparison purposes, the fetal ulna lengths and roundness-square were converted to binary data using (μ0 -2SD) as the

cutoff and then modeled as binary data. The GEE estimates of the model parameters and intra-litter correlation coefficient for the quantalized data are presented in Table 3.17.

The calculated R^2 was 0.44, suggesting 44 % of the variation of the outcome was explained by the model. The Pearson residuals were plotted against the predicted values (Figure 3.20). The Lowess smooth approximated a horizontal line with zero intercept, indicating that the model was adequate.

The predicted malformation probability at each dose is presented in Table 3.18. The model prediction, quantalized experimental data, and the joint 90% CI are presented in Figure 3.21. As the cutoff was 2.227 cm, less than 1% of the data at 10 mg/kg were below the cutoff. However, due to relatively larger variance, more than 1% of the data at 2.5 mg/kg were below the cutoff. Although, the estimated mean ulna length at 2.5 mg/kg was greater than at 10 mg/kg (Table 3.18), the malformation probability based on the quantalized data at 2.5 mg/kg was larger than at 10 mg/kg. This conflict suggested that the information from the continuous data was not accurately kept during the quantalization. The BMD was 15.33 kg/mg and the delta method BMDL was as low as 2.31 mg/kg. The bootstrap samples were directly extracted from the quantalized data and the BMDL was estimated as 13.00 mg/kg (Table 3.19). The BMD distribution for the 1000 bootstrap samples is presented in Figure 3.22 c).

3.2.3 Model for Fetal Ulna Roundness

3.2.3.1 Fetal Ulna Roundness Treated as Continuous Data

The histogram of the distribution of the fetal ulna roundness-square in each dose group is presented in Figure 3.23. The distribution in each dose group visually approximated a normal distribution. The fetal ulna roundness-square for FetusID 12 in DAM 235 at 2.5 mg/kg was 19.8 cm, as compared to shorter than 8.7 cm for all other fetuses in the same litter. Therefore,

FetusID 12 in DAM 235 was identified as an outlier and excluded from the analysis. The GEE estimates for the parameters and the correlation coefficient are presented in Table 3.20.

The calculated R^2 was 0.74, suggesting 74 % of the variation of the outcome was explained by the model. No trend was identified from the Pearson residual plots, indicating that the normal distribution assumption was appropriate and the model described the data adequately (Figure 3.24).

The observed fetal mean and predicted ulna roundness-square at each dose level are presented in Table 3.22. The model prediction and experimental data, along with the joint 90% CI are presented in Figure 3.25. Similar to the ulna length, two BMR definitions (mean changes and hybrid) were evaluated. The ulna roundness-square at BMD was 7.28 mg/kg, which was equal to $0.9025\mu_0$. Accordingly, the BMD was 4.41 mg/kg. The BMDL was 3.30 mg/kg by delta method and 3.72 mg/kg by bootstrap method (Table 3.23). By the hybrid BMR definition, the ulna roundness-square was 7.47mg/kg at BMD (calculated as $\mu_0 - 0.545\sigma_0$). The estimated BMD was 3.61 mg/kg. The BMDL was 2.65 mg/kg by delta method and 3.00 mg/kg by bootstrap method (Table 3.23). The distributions of BMD, by the two BMR definitions, based on 1000 bootstrap samples are presented in Figure 3.28 a) and b).

3.2.3.2 Quantalized Fetal Ulna Roundness

By using $(\mu_0 - 2\sigma)$ as the cutoff, the fetal ulna roundness-square data were dichotomized into binary data. The GEE estimates of the model parameters and intra-litter correlation coefficient for the quantalized data are presented in Table 3.21.

The calculated R^2 was 0.74, suggesting 74 % of the variation of the outcome was explained by the model. The Pearson residuals were plotted against the predicted values (Figure

3.26). The Lowess smooth approximated a horizontal line with zero intercept, indicating that the model was adequate.

The predicted malformation probability at each dose level is presented in Table 3.22. The model prediction and quantalized experimental data, along with the joint 90% CI are presented in Figure 3.27. The BMD was 4.89 kg/mg and the delta method BMDL was 2.65 mg/kg. The bootstrap samples were directly extracted from the dichotomized data and BMDL was estimated as 3.64 mg/kg (Table 3.23). The BMD distribution for the 1000 bootstrap samples is presented in Figure 3.28 c).

Table 3.1. NOAEL and LOAEL determination for litter mean values of number of fetuses, fetal body weight, forelimb malformation probability, fetal ulna length and roundness.

	Number of Fetuses[1]		Fetal Body Weight[1]		Fetal Ulna Length		Fetal Ulna Roundness		Malformation Probability[2]	
Dose (mg/kg)	LSMean	H0:LSMean =Control Pr > |t|	LSMean	H0:LSMean =Control Pr > |t|	LSMean	H0:LSMean =Control Pr > |t|	LSMean	H0:LSMean =Control Pr > |t|	LSMean	H0:LSMean =Control Pr > |Chi|
0	10.20		1.16		2.73		2.84		0.023	
2.5	8.43	0.438	1.21	0.5884	2.74	1	2.78	0.91	0.055	0.1497
10	7.17	0.0828	1.24	0.2362	2.67	0.9769	2.54	0.0034*	0.298	<.0001*
30	8.48	0.45	1.18	0.9573	2.34	0.0029*	2.15	<.0001*	0.906	<.0001*
60	9.63	0.9855	1.21	0.706	2.23	0.0002*	1.95	<.0001*	0.961	<.0001*
100	10.89	0.9663	1.13	0.9222	1.49	<.0001*	1.55	<.0001*	0.980	<.0001*
NOAEL		NA		NA		10		2.5		2.5
LOAEL		NA		NA		30		10		10

[1] Number of fetuses and fetal body weight were not significant different between control and treatment groups at α=0.05. NA = Not Available.
[2] Fisher's exact test (2-tail).
* Significantly different from control group ($p < 0.05$).

Table 3.2. Summary of selected endpoints: forelimb malformation, fetal ulna length, and roundness.

Dose (mg/kg)	No. litters	No. live fetuses	Malformed fetuses	Probability of Malformation		Fetal Ulna Length		Fetal Ulna Roundness	
				Pooled	Unpooled	Pooled	Unpooled	Pooled	Unpooled
0	10	102	2	0.0196	0.0225	2.712	2.728	2.837	2.839
2.5	23	194	12	0.0619	0.0552	2.727	2.739	2.805	2.783
10	18	129	28	0.2171	0.2977	2.698	2.673	2.580	2.540
30	25	212	191	0.9009	0.9057	2.345	2.343	2.145	2.151
60	16	154	148	0.9610	0.9614	2.210	2.227	1.950	1.954
100	18	196	192	0.9796	0.9802	1.491	1.488	1.588	1.545
Total	110	987	573	0.5806	NA	2.313	NA	2.262	NA

NA = Not applicable.

Table 3.3. Log-logistic model parameters for malformation probability. The beta-binomial distribution was assumed for each dose group and the parameters were estimated using maximum-likelihood methods.

Parameters[§]	Model I (Full Model)			Model II (Reduced Model without litter size covariate)			Model III (Reduced Model with fixed rho)		
	Estimate	Standard Error	Pr > |t|	Estimate	Standard Error	Pr > |t|	Estimate	Standard Error	Pr > |t|
p_0	0.05034	0.01777	0.0055	0.02595	0.01561	0.0993	0.04393	0.02071	0.0361
α	-7.6052	1.7989	<.0001	-5.5045	0.8446	<.0001	-5.8439	0.9782	<.0001
β	3.3633	0.8381	0.0001	2.1032	0.2404	<.0001	2.1697	0.3070	<.0001
θ_1	0.001695	0.001153	0.1443	NA	NA	NA	NA	NA	NA
θ_2	-0.1749	0.1110	0.1178	NA	NA	NA	NA	NA	NA
ρ_1	0.2274	0.2666	0.3955	0.01000	NA	NA	NA	NA	NA
ρ_2	0.02778	0.04372	0.5265	0.02750	0.04385	0.5318	NA	NA	NA
ρ_3	0.2339	0.1194	0.0526	0.3643	0.1174	0.0024	NA	NA	NA
ρ_4	0.3291	0.1375	0.0183	0.3904	0.1329	0.0040	NA	NA	NA
ρ_5	0.01000	NA	NA	0.01000	NA	NA	NA	NA	NA
ρ_6	0.01000	NA	NA	0.01000	NA	NA	NA	NA	NA
ρ	NA	NA	NA	NA	NA	NA	0.1714	0.05954	0.0048

NA: Not Available.

[§] p_0, α and β are the log-logistic model parameters. θ_1 and θ_2 are the litter size (number of fetuses per litter) covariate coefficients. $\rho_1 - \rho_6$ are the estimated correlation coefficients of beta-binomial distribution at each dose and ρ is the correlation coefficient for all the dose groups.

Table 3.4. Fit statistics of the three log-logistic models for malformation probability.

	Model I	Model II (Final Model)	Model III
-2 Log Likelihood	429.3	433.2	459.7
AIC (smaller is better)	451.3	451.2	467.7
AICC (smaller is better)	454.0	453.0	468.1
BIC (smaller is better)	481.0	475.5	478.5

Table 3.5. Log-logistic model estimated malformation probability at each dose.

	Dose (mg/kg)					
Malformation Probability	0	2.5	10	30	60	100
Observed Mean	0.023	0.055	0.298	0.906	0.981	0.980
Model Predicted	0.026	0.052	0.357	0.843	0.958	0.985

Table 3.6. Estimated BMD and BMDL for malformation probability by log-logistic model.

	BMR = $P_{BMD} - P_0 = 0.05$	
P(BMD)		0.076
BMD (mg/kg)		3.42
BMDL (mg/kg)	Delta Method	1.87
	Likelihood Ratio Based Method	2.24
	Bootstrap Method	2.37

Table 3.7. Log-logistic model parameters for fetal ulna length. The normal distribution was assumed for ulna length at each dose and the parameters were estimated using maximum-likelihood methods.

Parameter[§]	Estimate	Standard Error	Pr > |t|
α	-7.8108	1.6941	<.0001
β	2.3520	0.4390	<.0001
μ_0	2.7462	0.04503	<.0001
σ_1	0.1572	0.02505	<.0001
σ_2	0.1614	0.01957	<.0001
σ_3	0.1658	0.02611	<.0001
σ_4	0.2295	0.03309	<.0001
σ_5	0.2521	0.03942	<.0001
σ_6	0.5522	0.07443	<.0001

[§] α is the background parameter, β is the slope parameter in log-logistic model, μ_0 is the estimated mean value in the control group and σ_1- σ_6 are the standard deviation at each dose.

Table 3.8. Log-logistic model estimated fetal ulna length and malformation probability based on normal distribution at each dose.

	Dose (mg/kg)					
Fetal Ulna Length	0	2.5	10	30	60	100
Observed Mean (cm)	2.73	2.74	2.67	2.34	2.23	1.49
Model Predicted Mean (cm)	2.75	2.75	2.64	2.40	2.16	1.50
Model Predicted Malformation Probability	0.023	0.026	0.104	0.537	0.864	0.955

Table 3.9. Estimated BMD and BMDL for fetal ulna length by log-logistic model.

$$BMR = \Phi\left[\frac{c-\mu_{BMD}}{\sigma_0}\right] - \Phi\left[\frac{c-\mu_0}{\sigma_0}\right] = 0.05, \text{ where } c = \mu_0 - 2\sigma_0$$

P(BMD)		0.073
BMD (mg/kg)		8.00
BMDL (mg/kg)	Delta Method	3.26
	Likelihood Ratio Based Method	3.64
	Bootstrap Method	3.81

Table 3.10. Log-logistic model parameters for fetal ulna roundness. The normal distribution was assumed for fetal ulna roundness-square at each dose and the parameters were estimated using maximum-likelihood methods.

Parameter[§]	Estimate	Standard Error	Pr > |t|
α	-9.1717	1.7250	<.0001
β	3.7242	0.5787	<.0001
μ_0	7.9628	0.2302	<.0001
σ_1	0.8617	0.1460	<.0001
σ_2	0.8029	0.1060	<.0001
σ_3	0.6730	0.1122	<.0001
σ_4	0.8349	0.09734	<.0001
σ_5	0.8114	0.09486	<.0001
σ_6	1.1115	0.1153	<.0001

[§] α is the background parameter, β is the slope parameter in log-logistic model, μ_0 is the estimated mean value in the control group and σ_1- σ_6 are the standard deviation at each dose.

Table 3.11. Log-logistic model estimated fetal ulna roundness-square and malformation probability based on normal distribution at each dose.

	Dose (mg/kg)					
Fetal Ulan Roundness Square	0	2.5	10	30	60	100
Observed Mean	8.08	7.77	6.47	4.61	3.85	2.53
Model Predicted Mean	7.96	7.80	6.46	4.65	3.93	2.46
Malformation Probability Assuming Normal Distribution	0.023	0.026	0.370	0.971	0.998	1.00

Table 3.12. Estimated BMD and BMDL for ulna roundness by log-logistic model.

$$BMR = \Phi\left[\frac{c-\mu_{BMD}}{\sigma_0}\right] - \Phi\left[\frac{c-\mu_0}{\sigma_0}\right] = 0.05, \text{ where } c = \mu_0 - 2\sigma_0$$

P(BMD)		0.073
BMD (mg/kg)		5.36
BMDL (mg/kg)	Delta Method	3.02
	Likelihood Ratio Based Method	2.92
	Bootstrap Method	1.16

Table 3.13. Piecewise linear model parameters for forelimb malformation probabilities by GEE method.

Exchangeable Working Correlation						
Correlation	0.0777867553					
Analysis Of GEE Parameter Estimates						
Empirical Standard Error Estimates						
Parameter	Estimate	Standard Error	90% Confidence Limits		Z	Pr > |Z|
β_0	-3.8299	0.6579	-4.9120	-2.7477	-5.82	<.0001
β_1	1.1828	0.8028	-0.1376	2.5032	1.47	0.1406
β_2	-0.0262	0.9672	-1.6170	1.5646	-0.03	0.9784
β_3	1.8983	0.7334	0.6920	3.1045	2.59	0.0096
β_4	-1.6025	1.1665	-3.5212	0.3163	-1.37	0.1695
β_5	-0.1518	1.5725	-2.7383	2.4347	-0.10	0.9231

Table 3.14. Piecewise linear GEE model estimated malformation probability at each dose.

	Dose (mg/kg)					
Malformation Probability	0	2.5	10	30	60	100
Observed Fetal Mean	0.020	0.062	0.217	0.901	0.961	0.980
Predicted Value	0.021	0.060	0.242	0.901	0.962	0.980

Table 3.15. Estimated BMD and BMDL of fetal malformation probability by piecewise linear GEE model.

	$BMR = P_{BMD} - P_0 = 0.05$	
P(BMD)		0.071
BMD (mg/kg)		2.92
BMDL (mg/kg)	Delta Method	1.91
	Bootstrap Method	2.21

Table 3.16. Piecewise linear model parameters for fetal ulna length by GEE method.

Exchangeable Working Correlation						
Correlation	0.572520267					
Analysis Of GEE Parameter Estimates						
Empirical Standard Error Estimates						
Parameter	Estimate	Standard Error	90% Confidence Limits		Z	Pr > |Z|
β_0	2.7263	0.0548	2.6361	2.8165	49.71	<.0001
β_1	0.0130	0.0703	-0.1026	0.1287	0.18	0.8533
β_2	-0.0571	0.0888	-0.2031	0.0890	-0.64	0.5204
β_3	-0.2615	0.0708	-0.3780	-0.1451	-3.69	0.0002
β_4	0.1383	0.1480	-0.1052	0.3818	0.93	0.3501
β_5	-1.2691	0.3513	-1.8469	-0.6914	-3.61	0.0003

Table 3.17. Piecewise linear model parameters for quantalized fetal ulna length by GEE method. The malformation was defined as ulna length that was 2SD shorter than the mean in the control group.

Exchangeable Working Correlation						
Correlation	0.3674216349					
Analysis Of GEE Parameter Estimates						
Empirical Standard Error Estimates						
Parameter	Estimate	Standard Error	90% Confidence Limits		Z	Pr > |Z|
β_0	-4.4939	0.9600	-6.0730	-2.9148	-4.68	<.0001
β_1	1.0518	1.2240	-0.9614	3.0651	0.86	0.3901
β_2	-1.5147	1.6493	-4.2275	1.1981	-0.92	0.3584
β_3	3.8297	1.7041	1.0267	6.6326	2.25	0.0246
β_4	-2.5900	1.2708	-4.6804	-0.4997	-2.04	0.0415
β_5	2.0003	1.6282	-0.6778	4.6783	1.23	0.2192

Table 3.18. Piecewise linear GEE model estimated fetal ulna length and malformation probabilities based on quantalized data at each dose.

Fetal Ulna Length (cm)	Dose (mg/kg)					
	0	2.5	10	30	60	100
Observed Mean	2.71	2.73	2.70	2.35	2.21	1.49
Model Predicted	2.73	2.74	2.68	2.34	2.22	1.49
Observed Malformation Probability Mean	0.010	0.021	0.008	0.387	0.506	0.821
Predicted Malformation Probability	0.011	0.028	0.015	0.384	0.516	0.815

Table 3.19. Estimated BMD and BMDL for fetal ulna length by piecewise linear GEE model.

BMR = 0.05		Continuous Data		Quantalized Data
		$\frac{\mu_0 - \mu_{BMD}}{\mu_0}$	$\Phi\left[\frac{c-\mu_{BMD}}{\sigma_0}\right] - \Phi\left[\frac{c-\mu_0}{\sigma_0}\right]$ where $c = \mu_0 - 2\sigma_0$	$P_{BMD} - P_0$
μ_{BMD} or P_{BMD}		2.59	2.59	0.061
BMD (mg/kg)		13.30	13.26	15.33
BMDL (mg/kg)	Delta Method	11.3	11.3	2.31
	Bootstrap Method	12.14	11.15	13.00

Table 3.20. Piecewise linear model parameters for fetal ulna roundness by GEE method. Fetal ulna roundness-square was used as predictive variable and Fetus 12 in dam 235 was excluded.

Exchangeable Working Correlation						
Correlation	0.5379068779					
Analysis Of GEE Parameter Estimates						
Empirical Standard Error Estimates						
Parameter	Estimate	Standard Error	90% Confidence Limits		Z	Pr > |Z|
β_0	8.0691	0.2312	7.6888	8.4494	34.90	<.0001
β_1	-0.2802	0.3207	-0.8077	0.2472	-0.87	0.3822
β_2	-0.6518	0.4293	-1.3580	0.0544	-1.52	0.1290
β_3	-0.7523	0.3251	-1.2870	-0.2176	-2.31	0.0207
β_4	0.5791	0.4878	-0.2232	1.3814	1.19	0.2352
β_5	-1.3124	0.8283	-2.6748	0.0501	-1.58	0.1131

Table 3.21. Piecewise linear GEE model parameters for quantalized fetal ulna roundness. The malformation was defined as 2SD shorter than the mean in the control group.

Exchangeable Working Correlation						
Correlation	0.203392089					
Analysis Of GEE Parameter Estimates						
Empirical Standard Error Estimates						
Parameter	Estimate	Standard Error	90% Confidence Limits		Z	Pr > |Z|
β_0	-4.5046	0.9615	-6.0861	-2.9232	-4.69	<.0001
β_1	0.9092	1.2050	-1.0728	2.8912	0.75	0.4505
β_2	0.4855	1.4703	-1.9330	2.9039	0.33	0.7413
β_3	1.7615	0.8552	0.3549	3.1681	2.06	0.0394
β_4	-0.8992	1.0369	-2.6048	0.8064	-0.87	0.3858
β_5	-3.4773	1.9151	-6.6273	-0.3273	-1.82	0.0694

Table 3.22. Piecewise linear GEE model estimated fetal ulna roundness-square and malformation probabilities based on quantalized data at each dose.

Ulna Roundness Square	Dose (mg/kg)					
	0	2.5	10	30	60	100
Observed Mean	8.09	7.92	6.69	4.66	3.88	2.74
Model Predicted	8.07	7.81	6.52	4.67	3.90	2.67
Observed Malformation Probability Mean	0.010	0.021	0.116	0.854	0.961	0.934
Predicted Malformation Probability	0.011	0.025	0.150	0.849	0.964	0.935

Table 3.23. Estimated BMD and BMDL for fetal ulna roundness by piecewise linear GEE model.

BMR = 0.05		Continuous Data		Quantalized Data
		$\frac{\mu_0 - \mu_{BMD}}{\mu_0}$	$\Phi\left[\frac{c-\mu_{BMD}}{\sigma_0}\right] - \Phi\left[\frac{c-\mu_0}{\sigma_0}\right]$ where $c = \mu_0 - 2\sigma_0$	$P_{BMD} - P_0$
μ_{BMD} or P_{BMD} *		7.28	7.47	0.061
BMD (mg/kg)		4.41	3.61	4.89
BMDL (mg/kg)	Delta Method	3.30	2.65	2.65
	Bootstrap Method	3.72	3.00	3.64

* Fetal ulna roundness-square was the predictive variable. All the mean and SD value were based on ulna roundness-square. (0.95×ulna roundness mean at control) was equivalent to (0.95^2×ulna roundness-square mean at control) in the model.

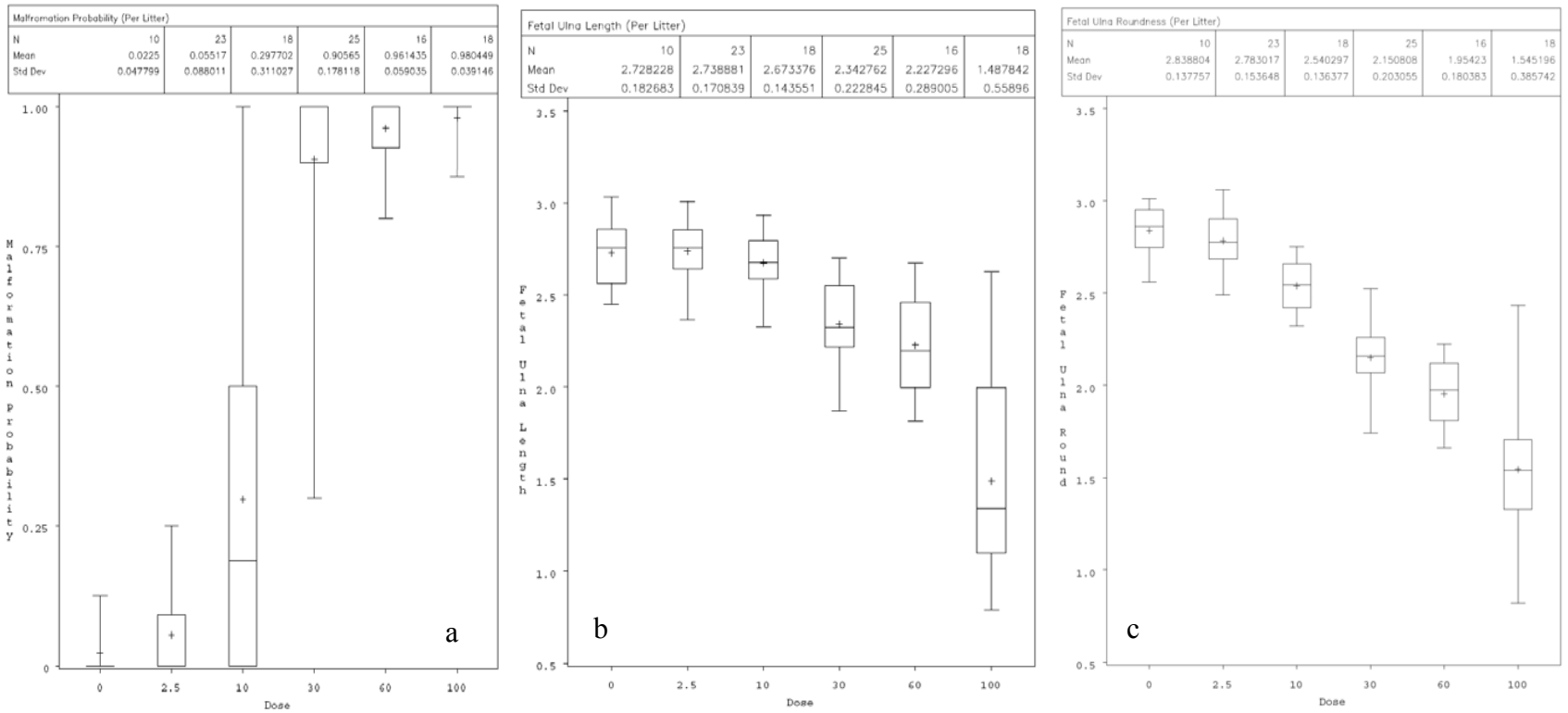

Figure 3.1. Box-plot and summary statistics of the litter means by dose. a) Malformation probability. b) Fetal ulna length. c) Fetal ulna roundness.

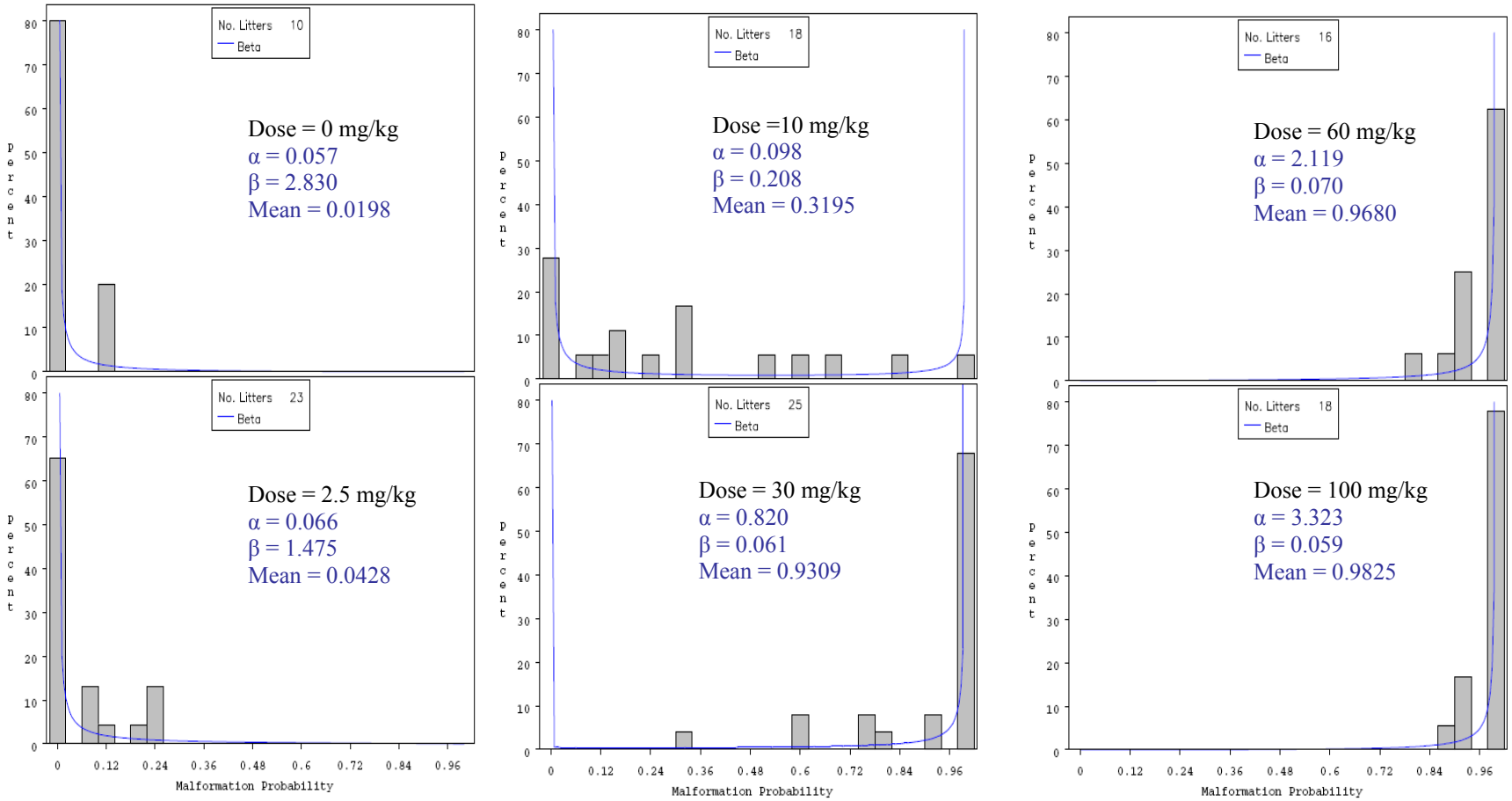

Figure 3.2. Histograms of forelimb malformation probability (proportion of defected fetuses per litter) and fitted beta distribution at each dose.

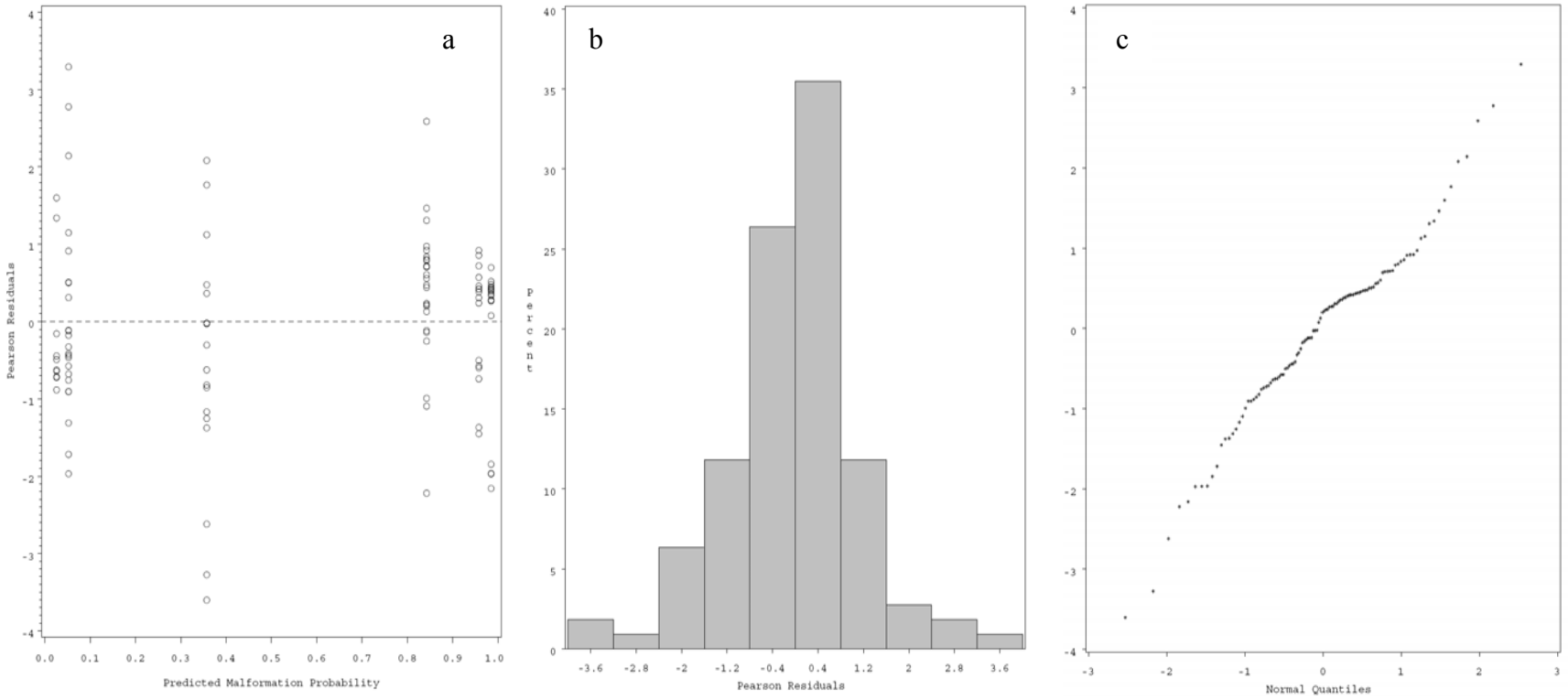

Figure 3.3. Diagnostic plots of the log-logistic model for malformation probabilities. a) Pearson residuals vs. predicted values. b) Histogram of Pearson residuals. c) Q-Q plot of Pearson residuals.

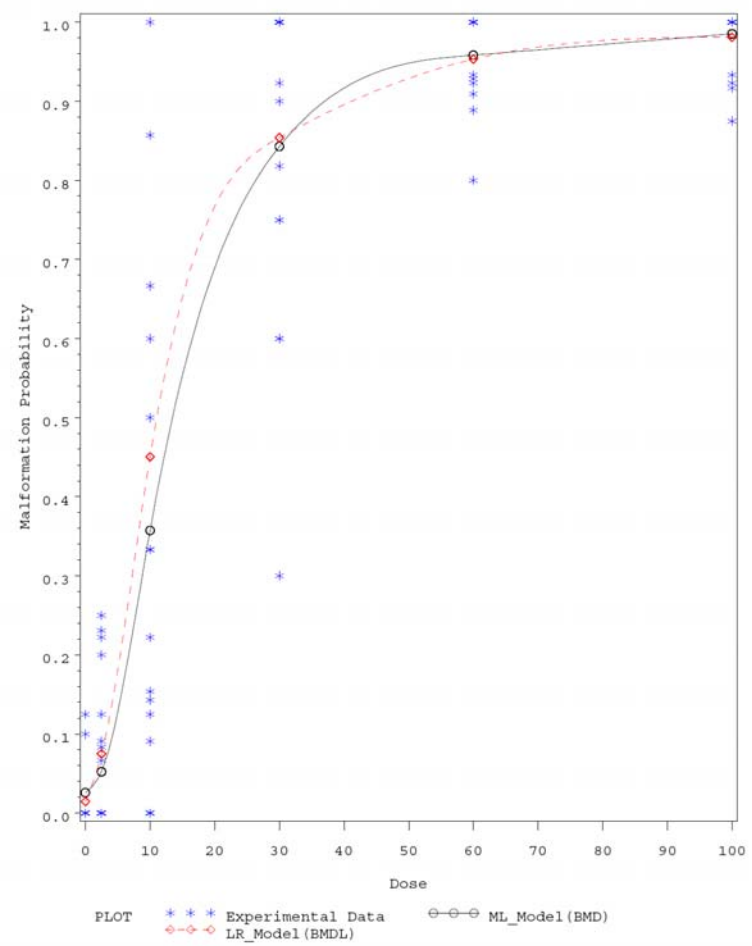

Figure 3.4. Log-logistic model predictions for malformation probability. The solid line is the maximum-likelihood model prediction, the dashed line is the reduced likelihood model for BMDL, and the starts are the experimental data.

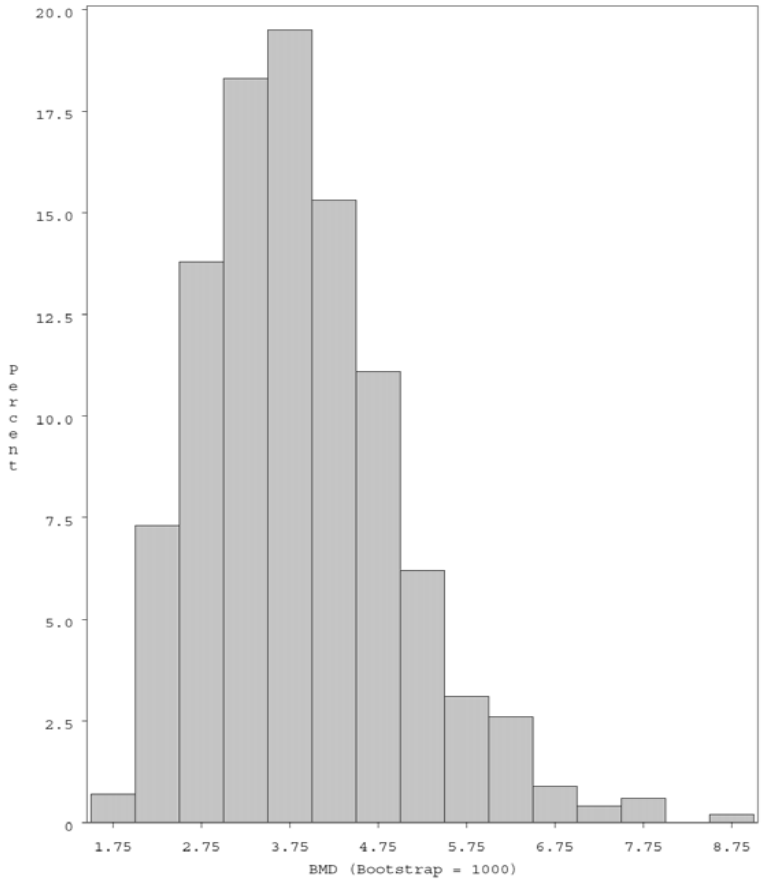

Figure 3.5. Histogram of BMD for malformation probability by log-logistic beta-binomial model based on 1000 bootstrap samples.

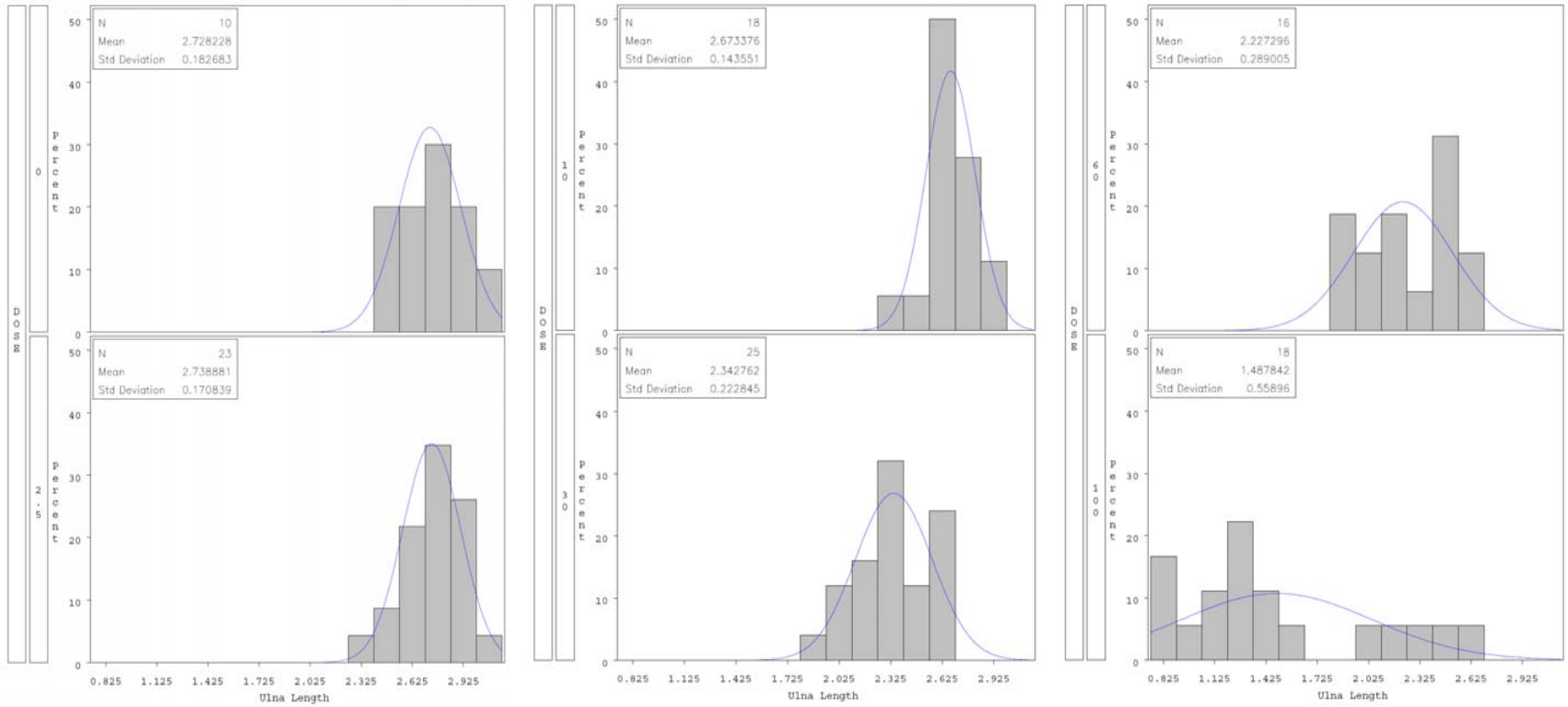

Figure 3.6. Histograms of litter mean values of fetal ulna length and fitted normal distribution at each dose.

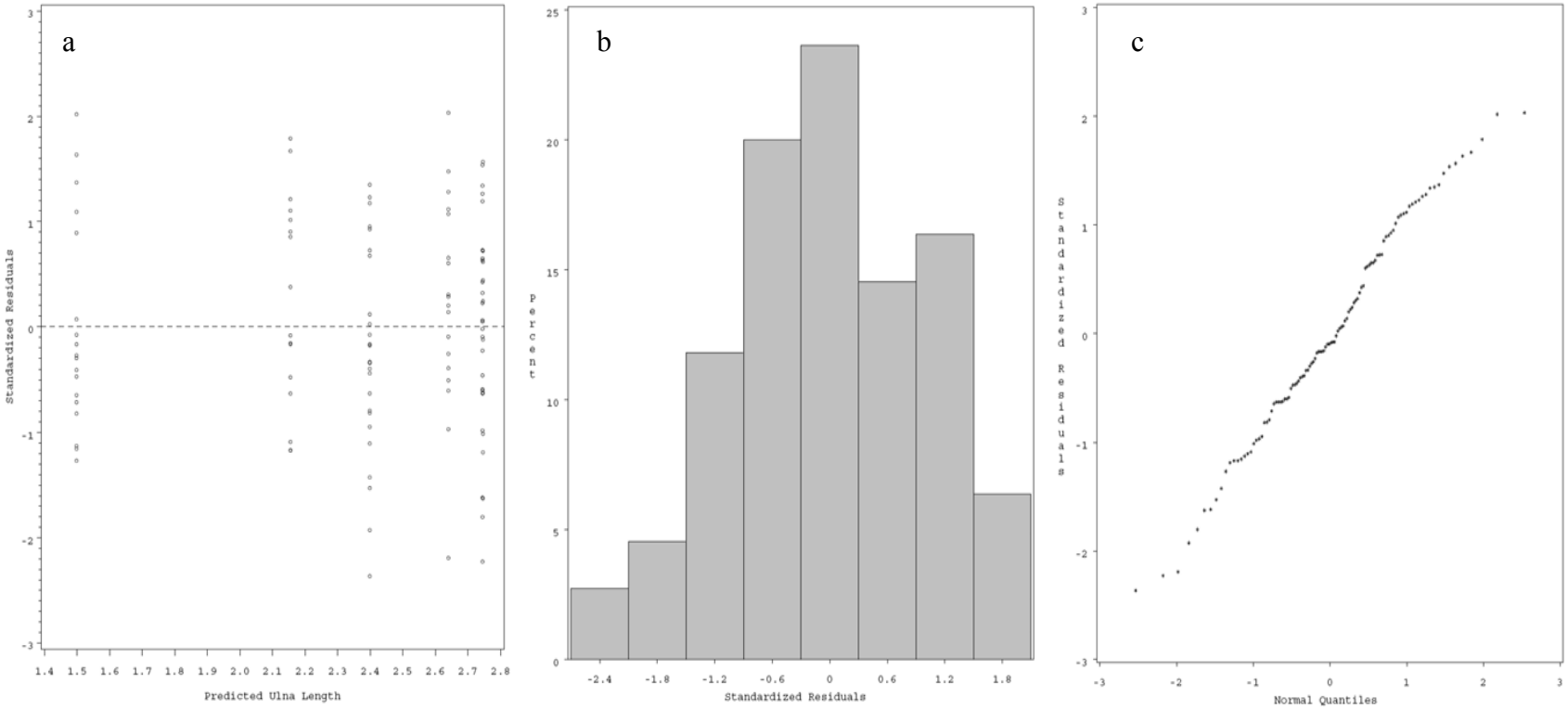

Figure 3.7. Diagnostic plots of the log-logistic model for fetal ulna length. a) Standardized residuals vs. predicted values. b) Histogram of standardized residuals. c) Q-Q plot of standardized residuals.

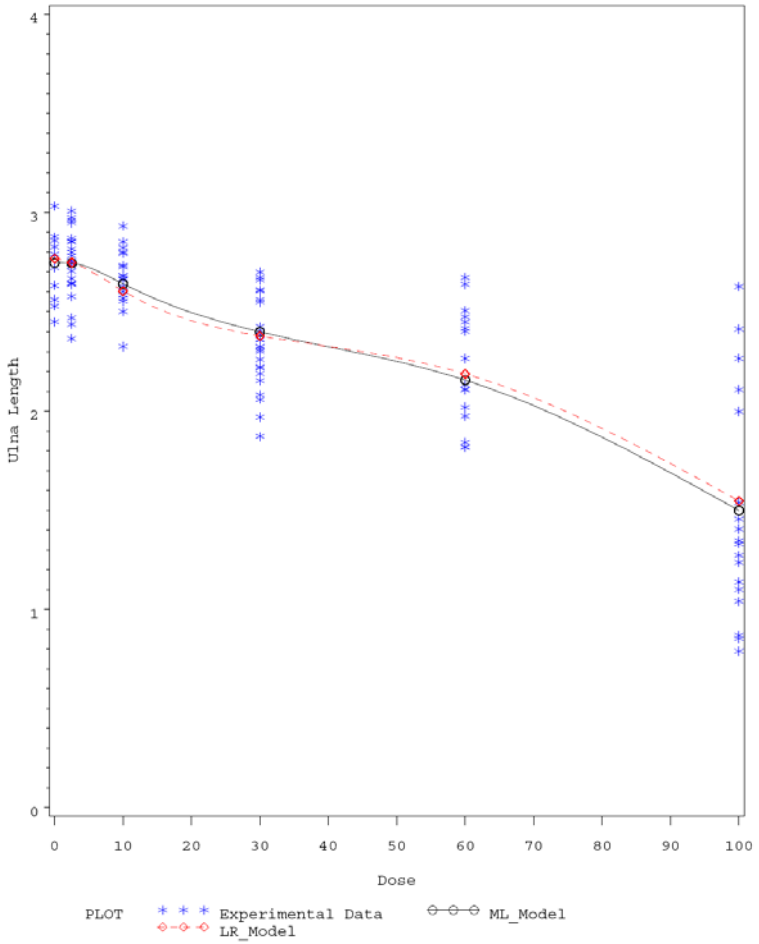

Figure 3.8. Log-logistic model predictions for fetal ulna length. The solid line is the maximum-likelihood model prediction, the dashed line is the reduced likelihood model for BMDL, and the starts are the experimental data.

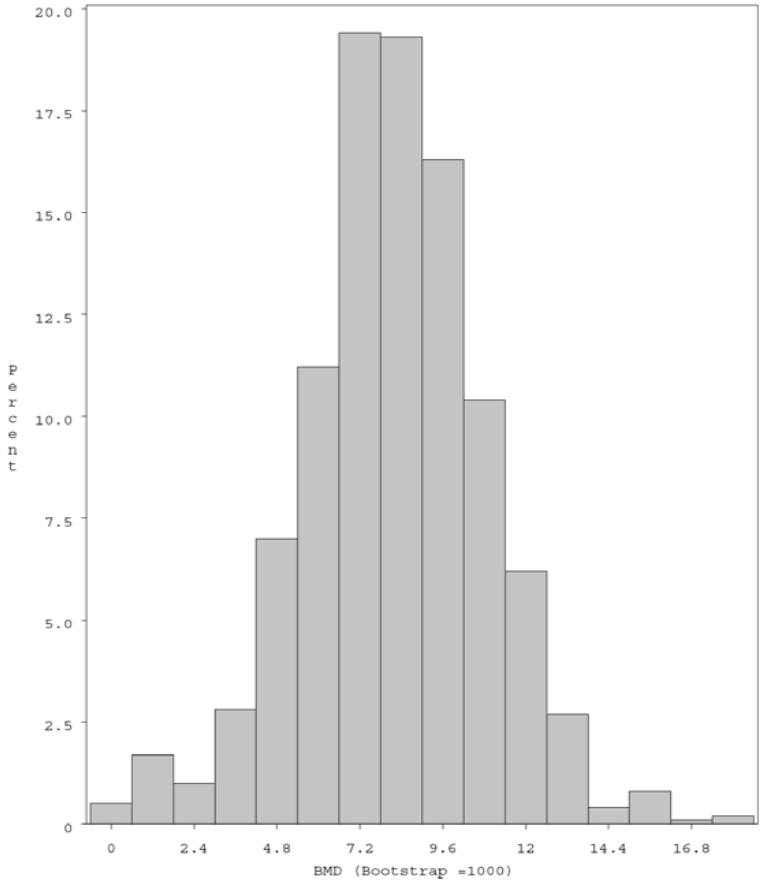

Figure 3.9. Histogram of BMD for fetal ulna length by hybrid log-logistic model based on 1000 bootstrap samples.

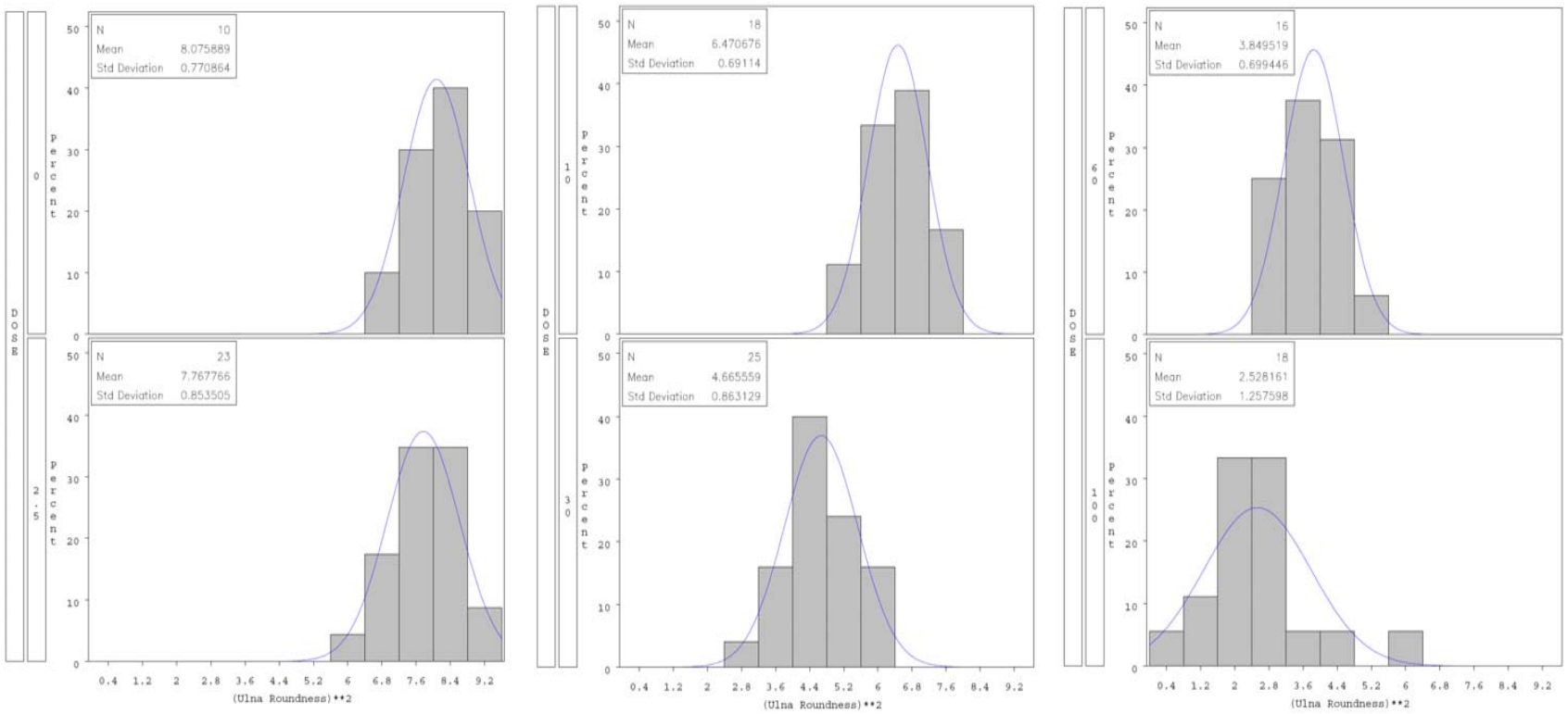

Figure 3.10. Histograms of litter mean values of transformed fetal ulna roundness by dose.

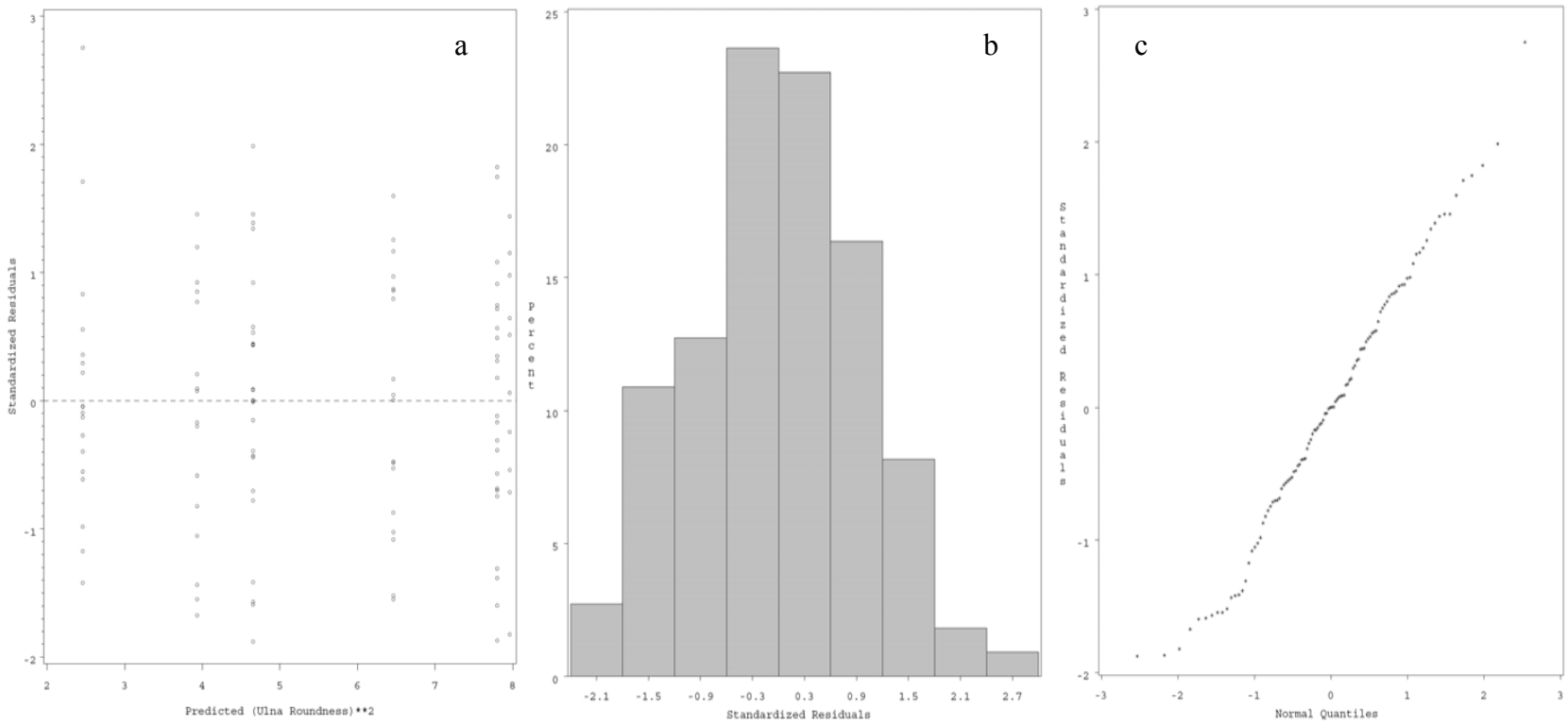

Figure 3.11. Diagnostic plots of the log-logistic model for fetal ulna roundness-square. a) Standardized residuals vs. predicted values. b) Histogram of standardized residuals. c) Q-Q plot of standardized residuals.

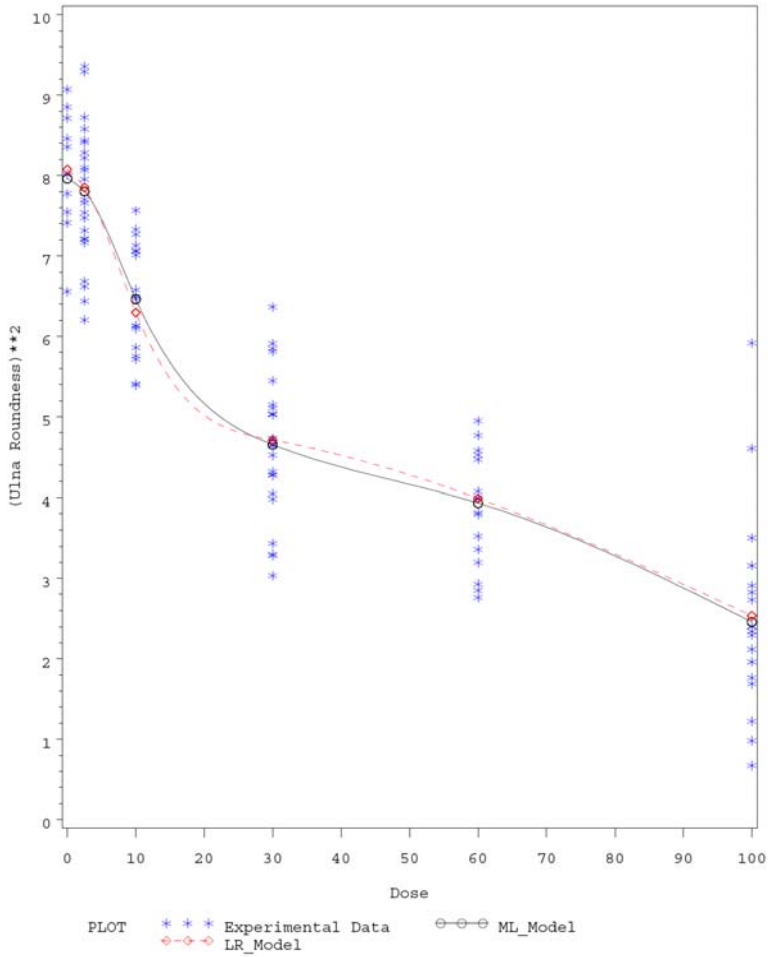

Figure 3.12. Log-logistic model predictions for fetal ulna roundness-square. The solid line is the maximum-likelihood model prediction, the dashed line is the reduced likelihood model for BMDL, and the starts are the experimental data.

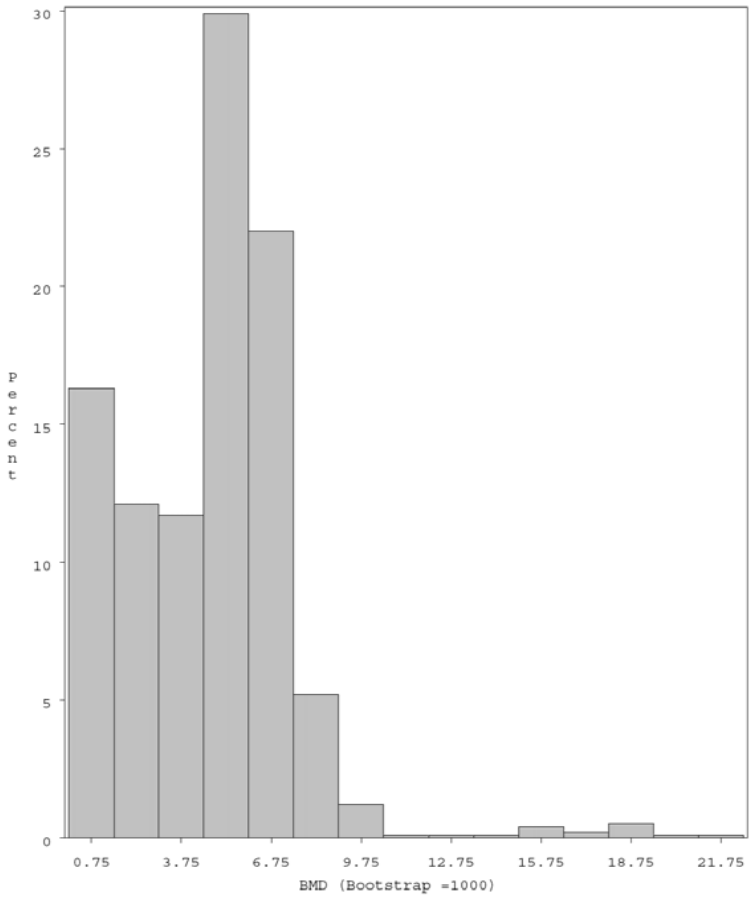

Figure 3.13. Histogram of BMD for fetal ulna roundness by hybrid log-logistic model based on 1000 bootstrap samples.

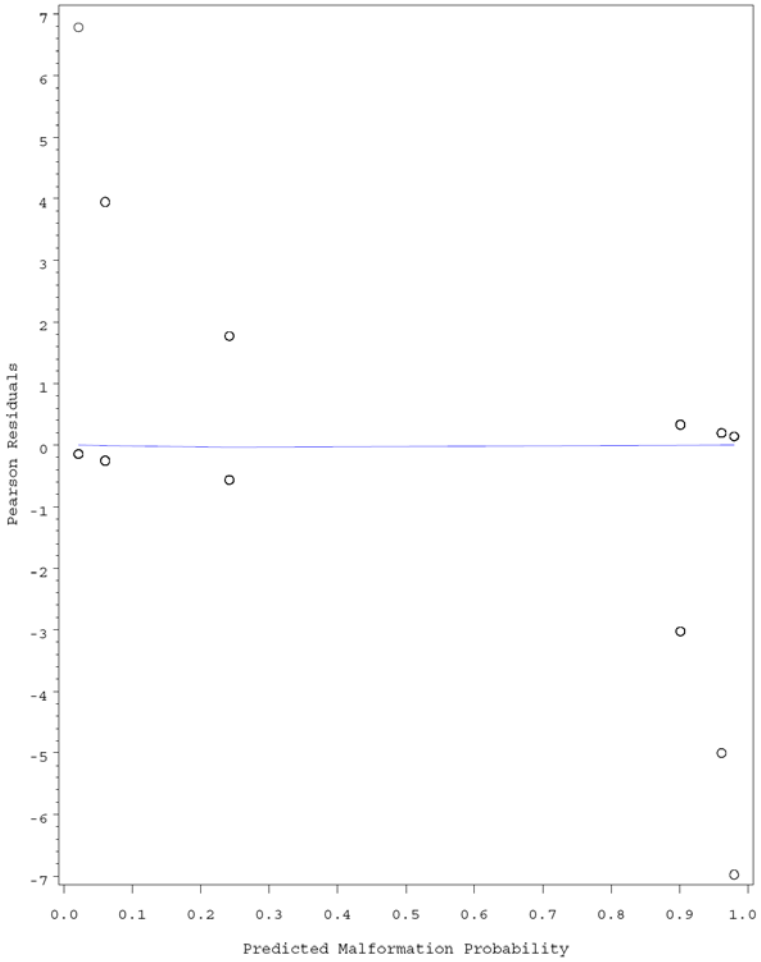

Figure 3.14. Pearson residuals vs. predicted malformation probability by piecewise linear GEE model with Lowess

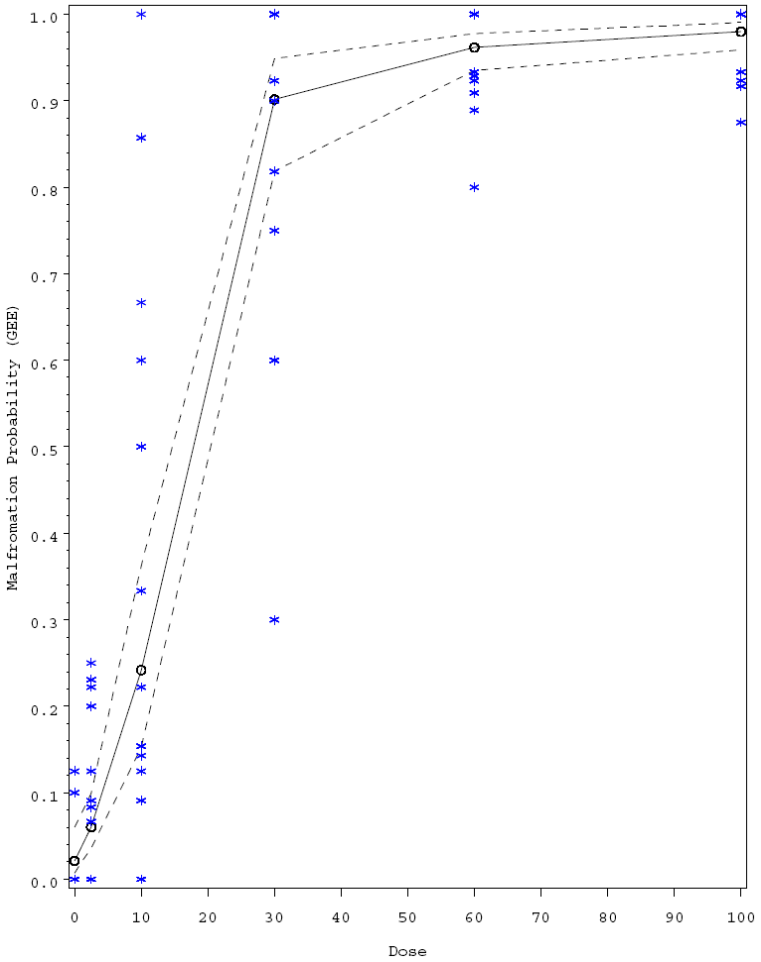

Figure 3.15. Model prediction of the malformation probabilities by GEE method. Solid line is the piecewise linear model prediction. Dashed lines are the joint 90% confidence interval. The stars are the experimental malformation probability per litter.

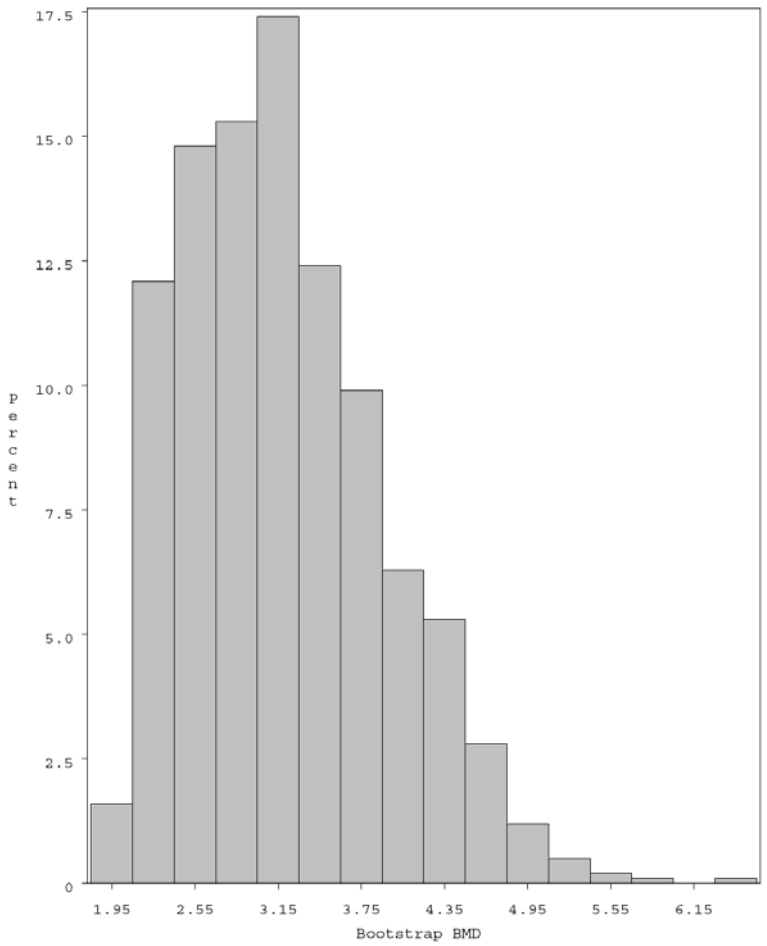

Figure 3.16. Histogram of BMD for malformation probability by GEE method based on 1000 bootstrap samples.

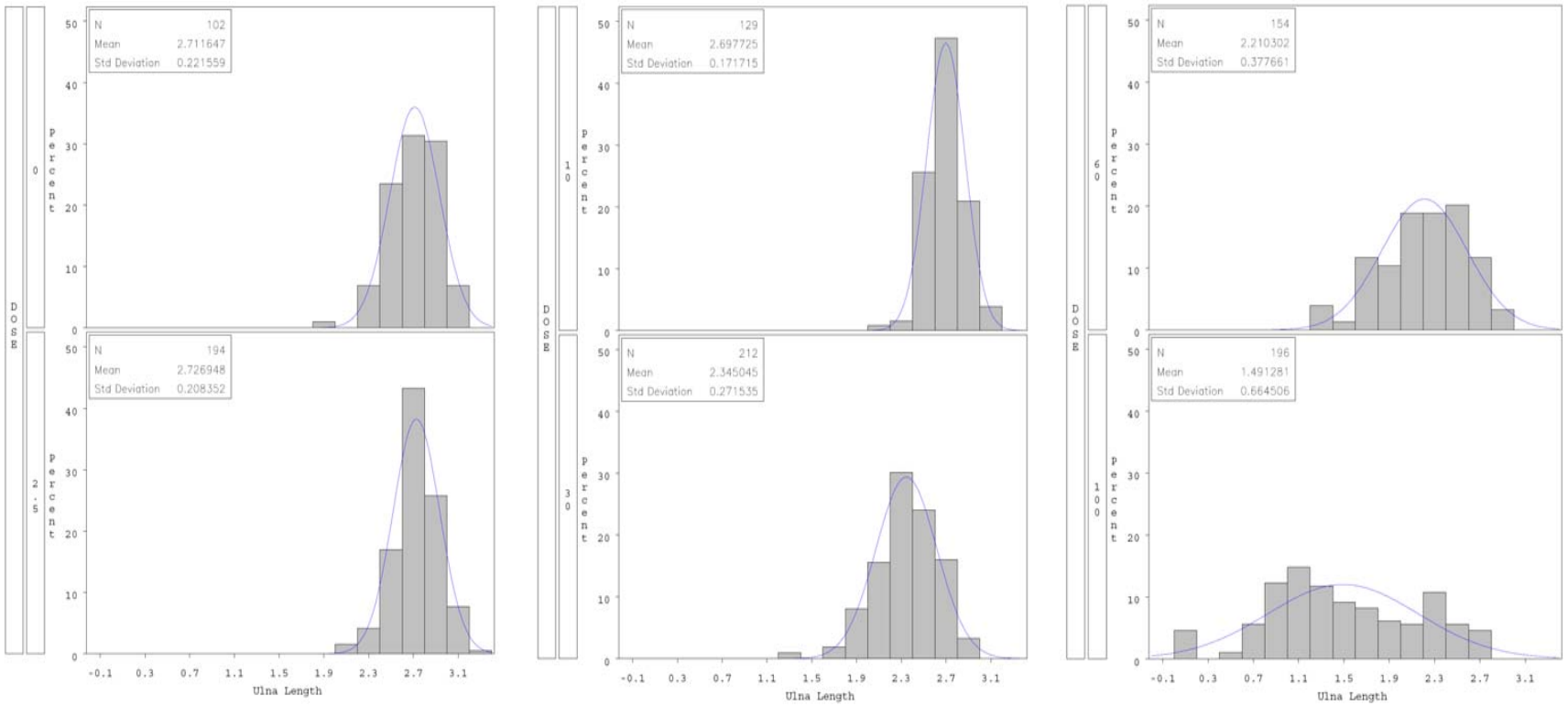

Figure 3.17. Histograms of fetal ulna length at each dose.

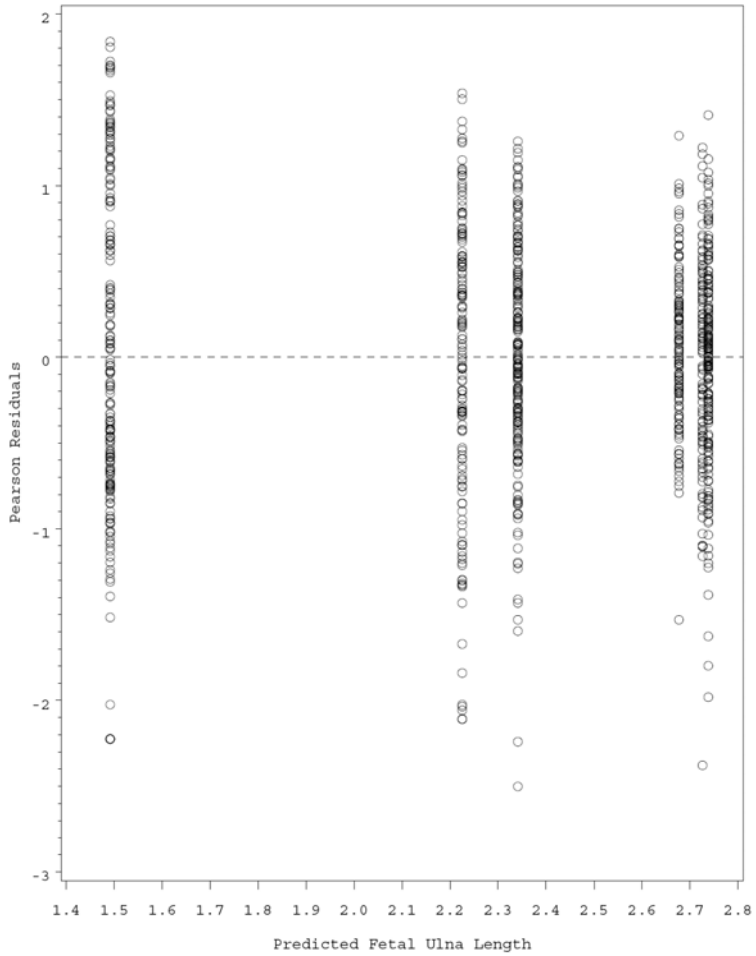

Figure 3.18. Pearson residuals vs. predicted fetal ulna length by piecewise linear GEE model with Lowess.

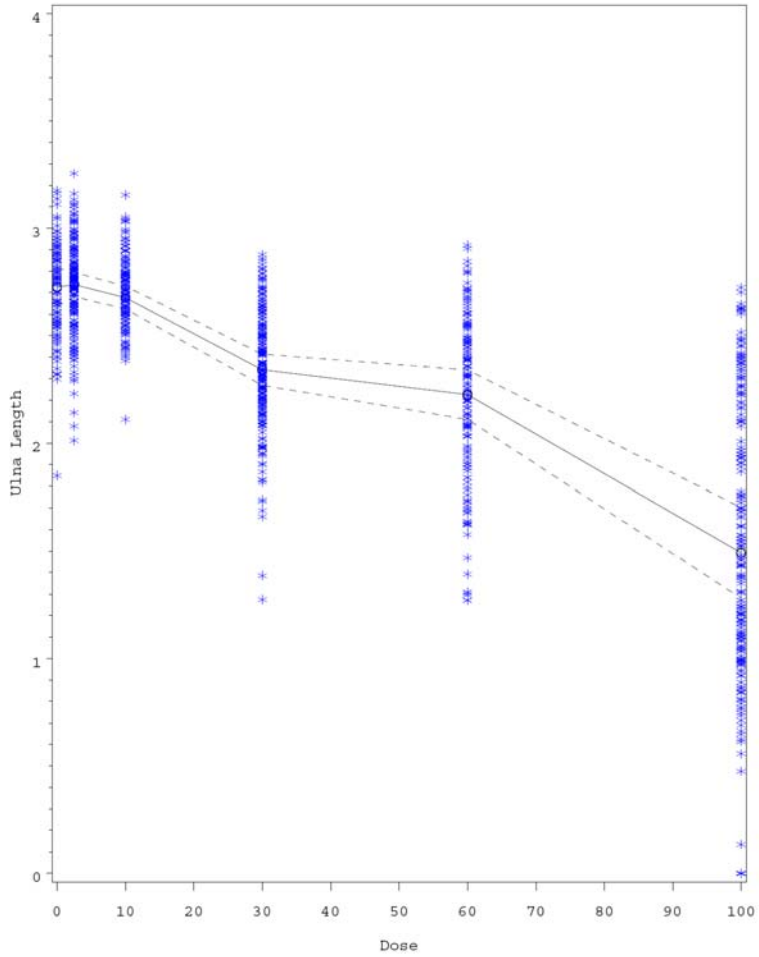

Figure 3.19. Model prediction of the fetal ulna length by GEE method. Solid line is the piecewise linear model prediction. Dashed lines are the joint 90% confidence interval. The stars are the experimental data.

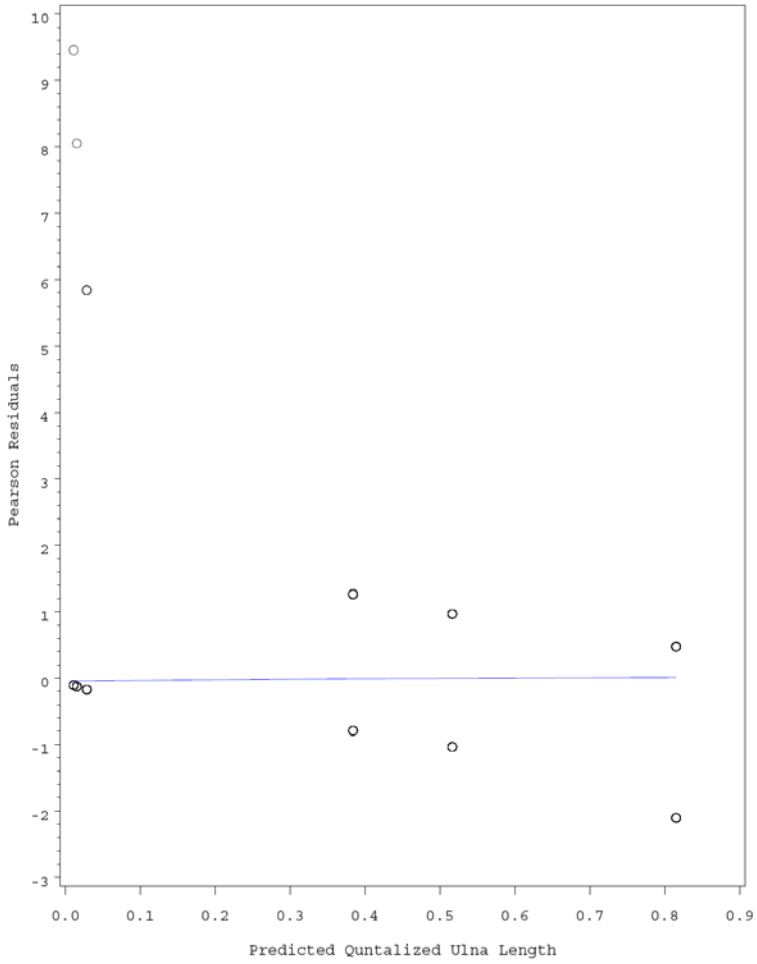

Figure 3.20. Pearson residuals vs. predicted quantalized fetal ulna length by piecewise linear GEE model with Lowess .

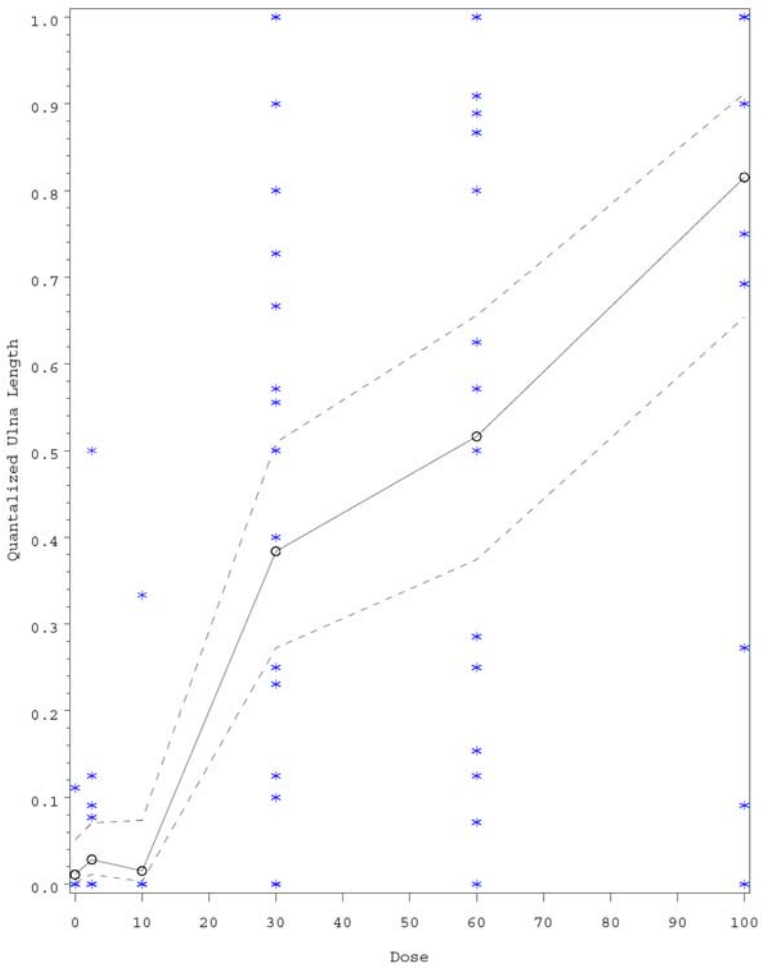

Figure 3.21. Model predictions of the quantalized fetal ulna length by GEE method. Solid line is the piecewise linear model prediction. Dashed lines are the joint 90% confidence interval. The stars are the experimental abnormal ulna length probability per litter.

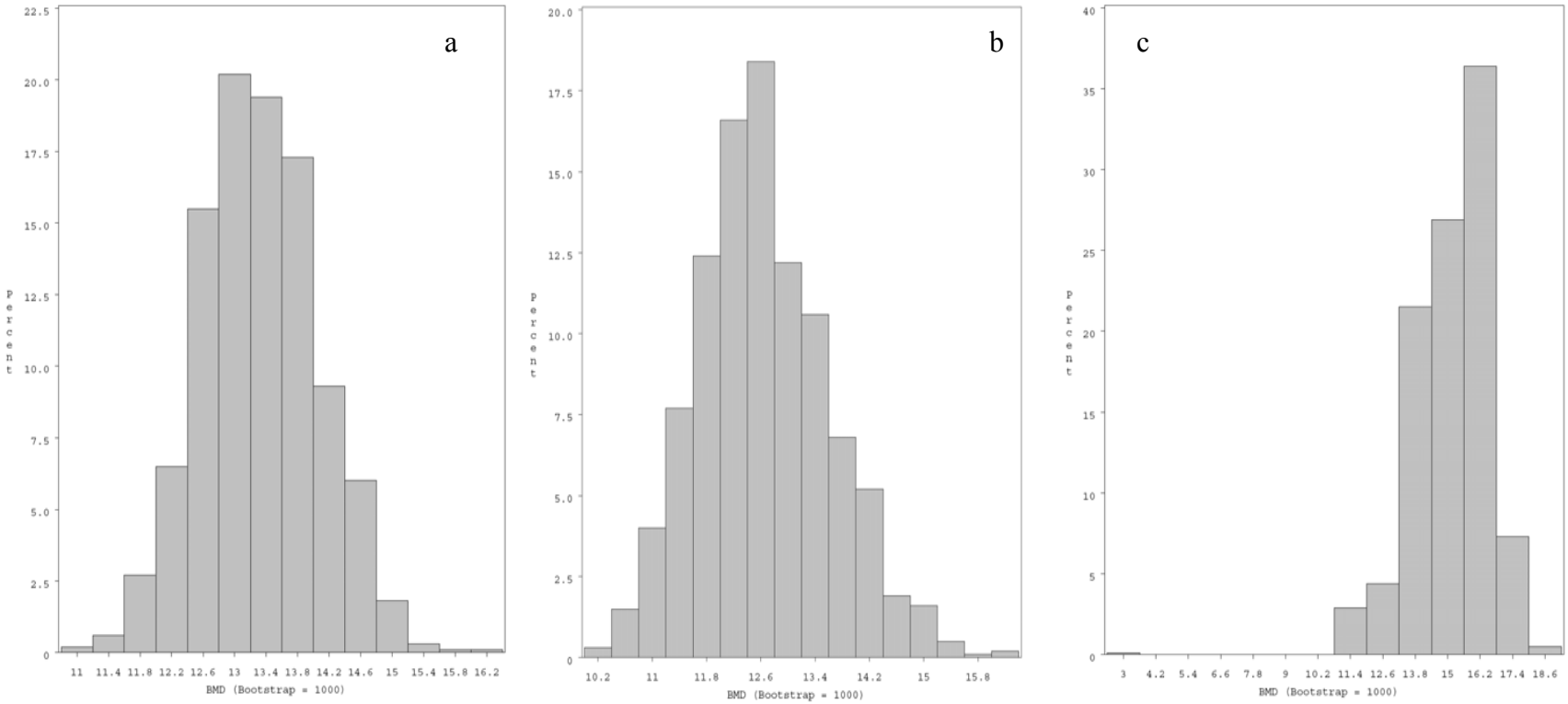

Figure 3.22. Histograms of BMD for fetal ulna length by GEE method based on 1000 bootstrap samples. a) $\mu_{BMD} = 0.95\mu_0$. b) $\mu_{BMD} = \mu_0 - 0.545\sigma_0$. c) Quantalized data, $P_{BMD} = P_0 + 0.05$.

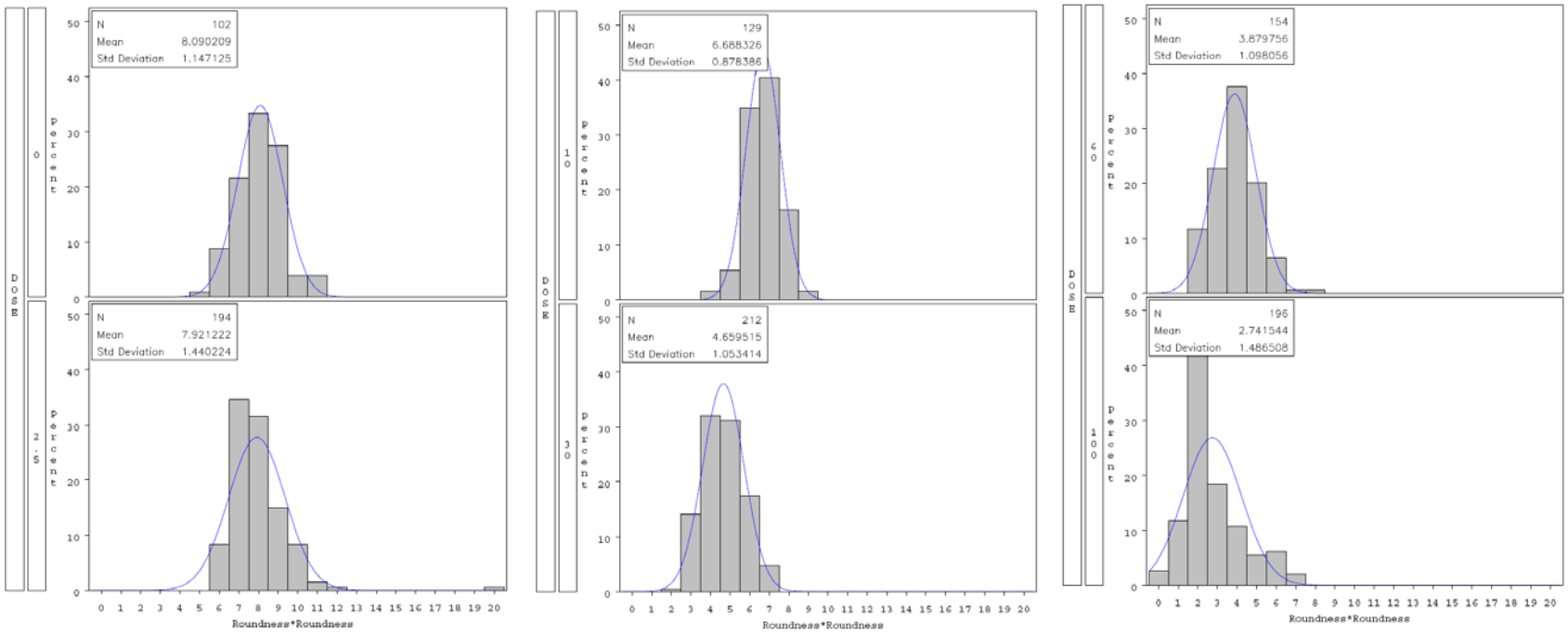

Figure 3.23. Histograms of fetal ulna roundness-square at each dose. Ulna roundness-square approximate normal distribution at each dose.

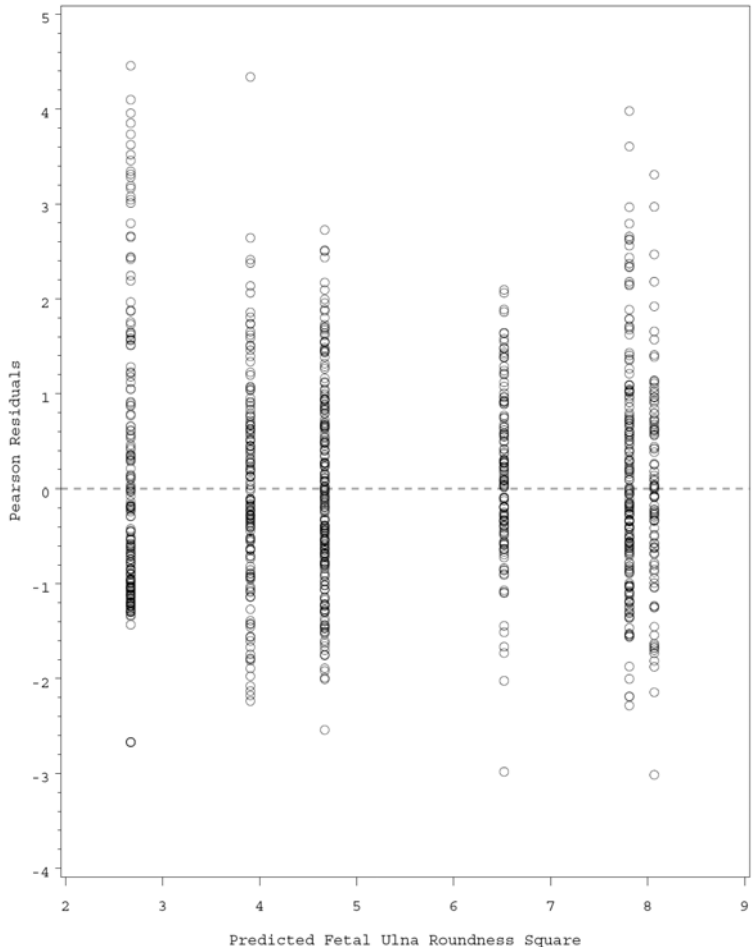

Figure 3.24. Pearson residuals vs. predicted fetal ulna roundness-square by piecewise linear GEE model with Lowess.

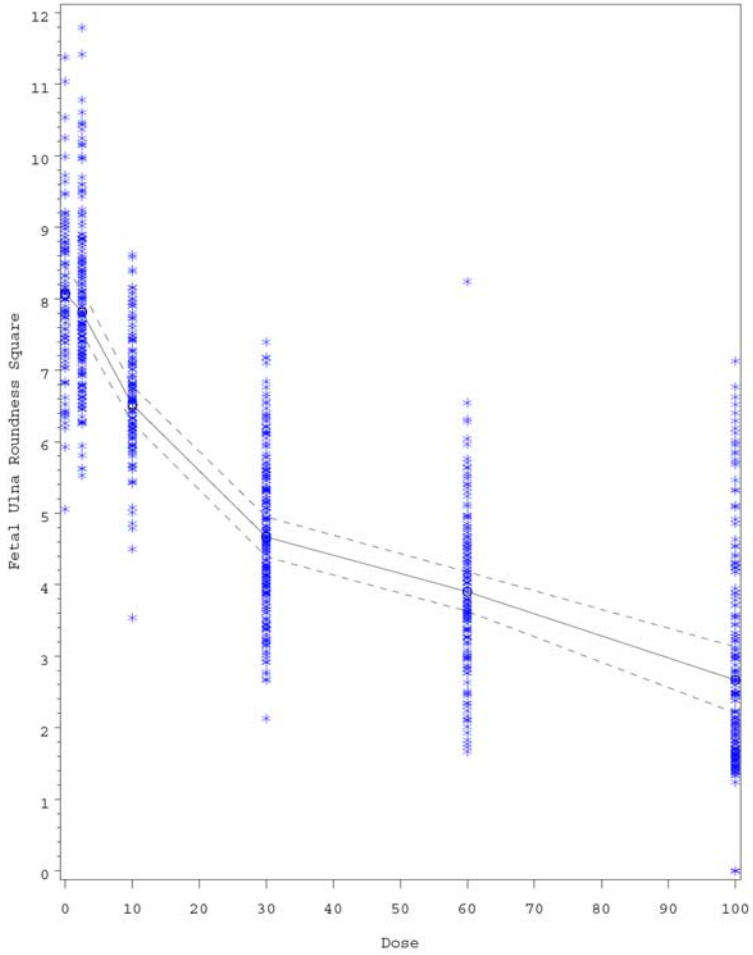

Figure 3.25. Model predictions of the fetal ulna roundness-square by GEE method. Solid line is the piecewise linear model prediction. Dashed lines are the joint 90% confidence interval. The stars are the experimental data.

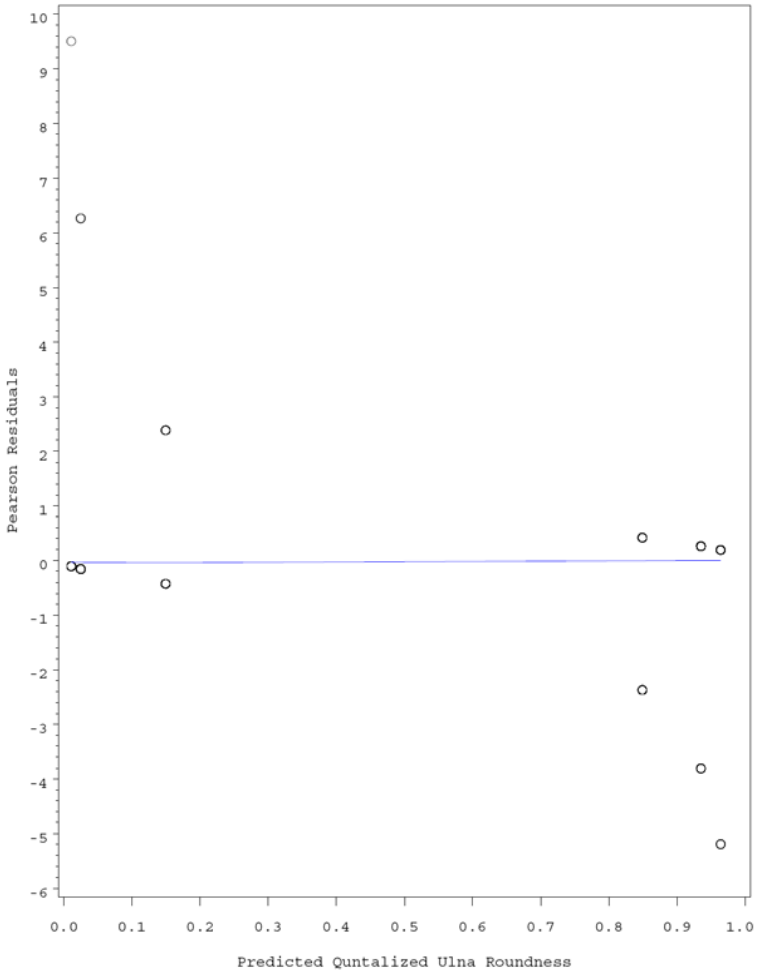

Figure 3.26. Pearson residuals vs. predicted quantalized fetal ulna roundness with Lowess.

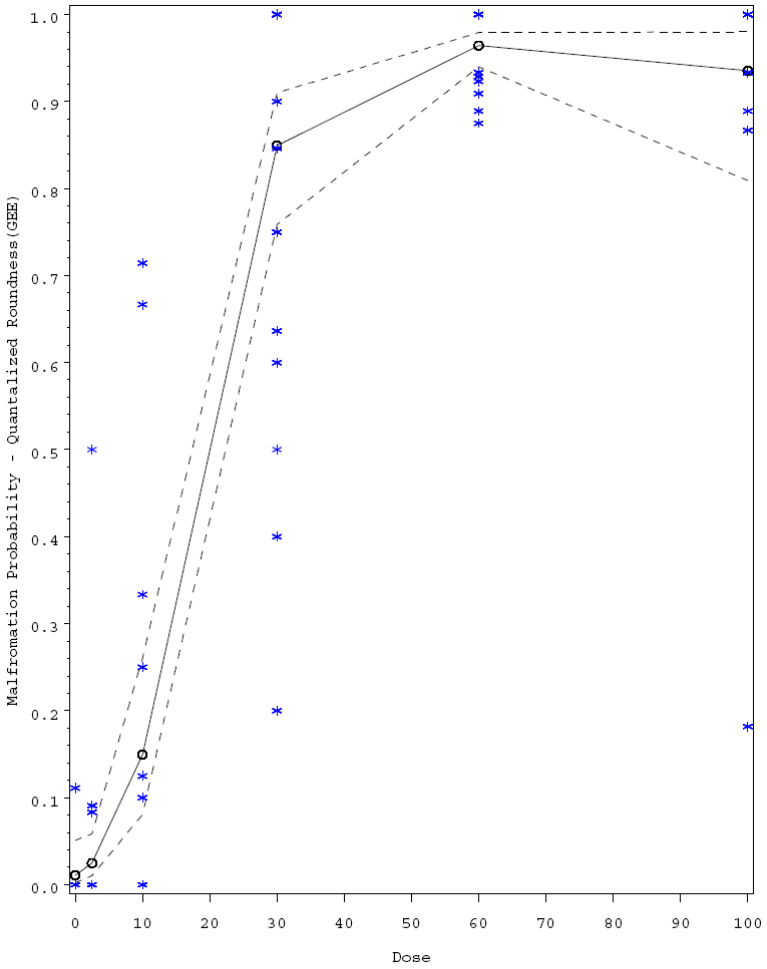

Figure 3.27. Model predictions of the quantalized fetal ulna roundness by GEE method. Solid line is the piecewise linear model prediction. Dashed lines are the joint 90% confidence interval. The stars are the experimental abnormal ulna length probability per litter.

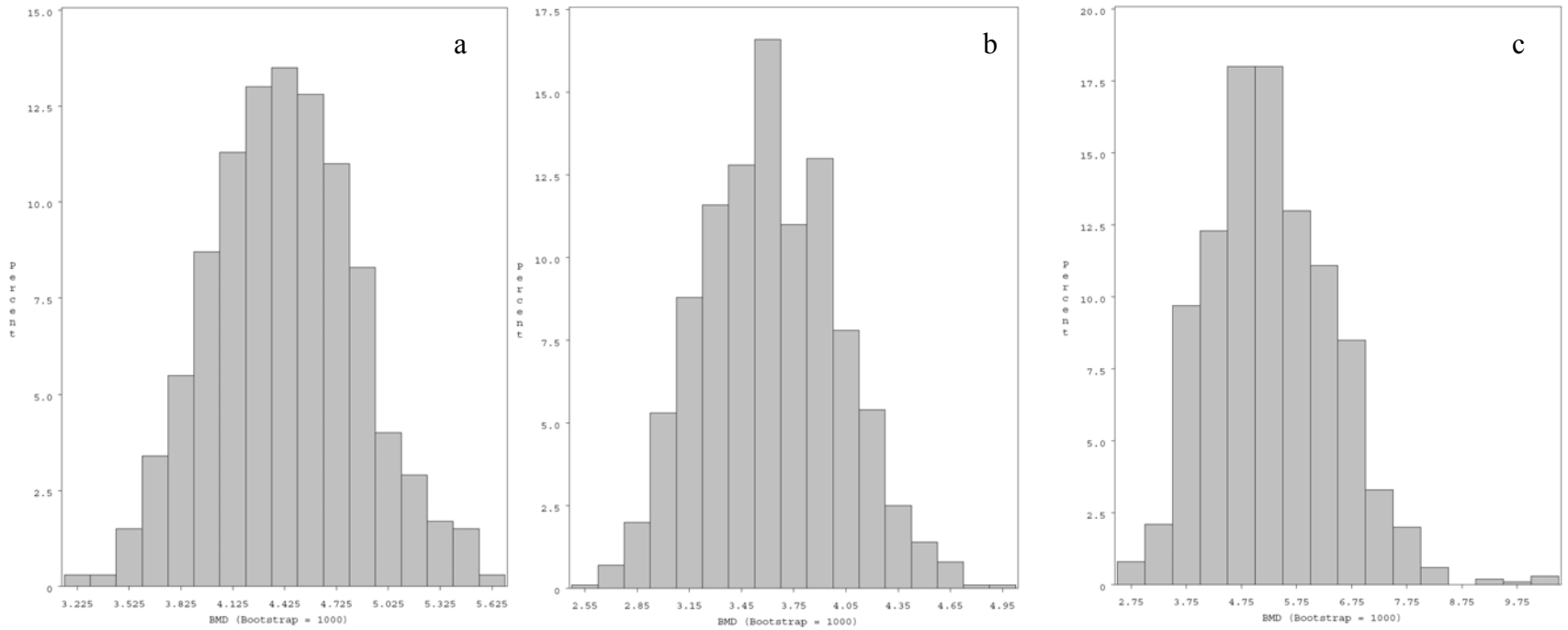

Figure 3.28. Histograms of BMD for ulna roundness by GEE method based on 1000 bootstrap samples. a) $\mu_{BMD} = 0.95\mu_0$. b) $\mu_{BMD} = \mu_0 - 0.545\sigma_0$. c) Quantalized data, $P_{BMD} = P_0 + 0.05$.

CHAPTER 4

CONCLUSION AND DISCUSSION

Rodents are widely used in developmental toxicology to identify chemicals that may pose danger to a developing fetus. Typically, pregnant dams are exposed to chemicals and the responses are collected from the fetuses, such as numbers of fetal deaths, fetal weights, and malformations. The most prominent feature of a developmental toxicity study is the presence of intra-litter correlations. Beta-binomial or GEE methods have been used to account for over-dispersion caused by variation in spread between different litters. The endpoints collected in the majority of developmental toxicity studies are quantal data, and limited attention has been paid to applying the methodology to continuous responses. The BMD results from quantal and continuous endpoints are often not comparable due to the different BMR definitions and dose-response models. Quantalization and hybrid methods were proposed so that the same BMR definition and model can be used for both types of endpoints. In the present study, the different statistical approaches were evaluated by using the data from the study of Campbell et al. (2004).

The log-logistic model was applied to malformation probability per litter, assuming a beta-binomial distribution at each dose. By applying the hybrid method, the log-logistic model provided adequate fits to the dose-responses of fetal ulna length and roundness. Thus, the BMDs/BMDLs for the quantal endpoints (malformation probability) and continuous endpoints (fetal ulna length and roundness) were comparable. During the process of this manuscript's preparation, US EPA released a new version of the BMD software (BMDS, version 1.4.1c)

which included a beta version of the hybrid method using the log-logistic model. However, the hybrid method has never been used in developmental toxicology studies.

The piecewise linear dose-response model was employed to analyze the individual fetal responses directly by GEE methods. GEE methods yield consistent estimates even when the variance and covariates are misspecified. In addition to the model parameters, a correlation parameter was estimated for the observations from the littermates. The piecewise linear model has its own advantages in dose-response modeling. First of all, the fits to the lower doses are not sacrificed for the sake of better overall fit, which is always problematic for studies containing large-range doses. Second, the piecewise linear model can be fit to monotonic or non-monotonic dose-response relationships. Linear extrapolation between dose groups is simple without assuming any consistent function across all the doses. Third, when the study contains too few doses to employ non-linear models or when generalized linear models fail to describe the data adequately, a piecewise linear model can always be used in conjunction with GEE methods.

4.1 LITTER-BASED APPROACH VS. FETAL-BASED PIECEWISE-LINEAR APPROACH

The BMDs estimated by litter-based and fetal-based piecewise-linear approaches are in fairly good agreement with each other (Table 4.1). When the exact numbers were compared, slightly larger BMDs for malformation probability and ulna roundness were obtained from the litter-based approach. Larger BMDs for ulna length were obtained by the fetal-based piecewise-linear approach. The NOAEL was 2.5 mg/kg for malformation probability, 10 mg/kg for ulna length, and 2.5 mg/kg for ulna roundness based on the litter mean values. All the estimated BMDs for malformation probability and fetal ulna roundness were larger than the NOAELs but smaller than the LOAELs. The BMD for fetal ulna length by the litter-based approach was

slightly smaller than the NOAEL (8 mg/kg vs. 10 mg/kg), while the BMDs estimated by the fetal-based piecewise-linear approach were slightly larger (13.30 – 15.33 mg/kg). The ulna length BMDs by both approaches were well below the LOAELs (30 mg/kg).

4.2 METHODS FOR BMDL

The delta method, likelihood-ratio based method, and bootstrap method were used in calculating the BMDL under the litter-based approach, while the delta method and bootstrap method were used under the fetal-based piecewise-linear approach. The BMDLs estimated by the three methods are in agreement with each other in most cases. 13 comparisons were made between delta BMDLs and the likelihood-based or bootstrap BMDLs. Delta BMDLs were smaller in 10 out of the 13 comparisons, although the numbers were close in the majority of the cases (Table 4.1). In the litter-based approach, the likelihood-ratio based BMDLs were slightly smaller than bootstrap BMDLs for malformation probability and ulna length. However, the bootstrap BMDLs of ulna roundness was only 1.16 mg/kg in litter-based, which was smaller than BMDLs estimated by either delta or likelihood-based method (3.02 and 2.92 mg/kg, respectively). In the fetal-based piecewise-linear approach, when the target μ(BMD) was set to $\mu_0 - 0.545 \times SD_0$, the bootstrap method produced slightly smaller BMDLs than did the delta method for the ulna length (11.15 vs. 11.30 mg/kg).

Unlike the delta method, which is based on the asymptotic normal distribution of the estimator, the likelihood-ratio method and the bootstrap method provide consistent estimates when this assumption is not valid. Since likelihood is not available in GEE models, the bootstrap method should provide the most robust estimates of BMDLs over all model types.

4.3 SENSITIVITY OF THE SELECTED ENDPOINTS

The quantal response (malformation probability) had the smallest BMD/BMDL and therefore was the most sensitive endpoint. This result was expected because a fetus was determined to be 'malformed' if there was even one 'abnormal' characteristic observed on any of the forelimb bones, while the other two continuous endpoints, shorter fetal ulna length and smaller roundness, were just two independent measurements from one bone. When the two continuous responses were compared, the ulna roundness had smaller BMD/BMDL, indicating that roundness (the shape of the bone) was a more sensitive endpoint than length to *all-trans* RA exposure. Dichotomized continuous data didn't represent the original continuous data faithfully and produced larger BMDs for fetal ulna length.

In conclusion, different statistical methods gave consistent BMD/BMDL estimates. Ryan (1992) recommended GEE methods over the beta-binomial maximum likelihood methods because the beta-binomial was an overly simple assumption. According to the results in this study, beta-binomial maximum likelihood models produced similar BMD/BMDLs as those estimated by GEE methods. However, the goodness of fit statistics of GEE models are not well established, and therefore are not included in the output of standard statistical software such as SAS. The delta method for BMDL was the easiest to apply because it is included in the default SAS output. Both the likelihood-ratio based method and the bootstrap method required extra coding efforts. The bootstrap method provides additional information on the distribution of BMDs in the re-sampled populations. One may find that the non-linear models fail to converge for some samples, but this problem did not occur in the present study (Zhu et al., 2007). Because piecewise linear GEE models paired with bootstrap methods always gave adequate estimates of BMD/BMDL without making any extra assumptions, we recommend piecewise linear GEE

models for BMD and bootstrap method for BMDL estimation in developmental toxicology studies.

In addition to the proper statistical models and methods, an optimal experimental design can increase the accuracy and precision of the BMDs (Kavlock et al., 1996). However, such 'optimal design' can typically be employed only after a pilot study has already been conducted. For example, for the Campbell (2004) study which has been the main focus of this manuscript, 'after-the-fact' examination of the data would suggest that rather than having 'wasted' efforts sampling at high dose levels such as 60 mg/kg or 100 mg/kg, more precise estimates of BMD and BMDL might have been obtained if measurements had instead been taken at dose levels of 1.0 mg/kg and 5.0 mg/kg, for example.

Table 4.1. Summary of BMD/BMDL by different methods for malformation probability, fetal ulna length, and roundness.

Approach	Endpoints	NOAEL	LOAEL	μ(BMD) or P(BMD)	BMD	BMDL		
						Delta or Confidence Band*	LRB	Bootstrap
Litter-Based	Malformation Probability	2.5	10	$P_0 + 0.05$	3.42	1.87	2.24	2.37
	Ulna Length	10	30	$P_0 + 0.05$	8.00	3.26	3.64	3.81
	Ulna Roundness	2.5	10	$P_0 + 0.05$	5.36	3.02	2.92	1.16
Fetal-Based	Malformation Probability	2.5	10	$P_0 + 0.05$	2.92	1.91	NC	2.21
	Ulna Length	10	30	$\mu_0 - 0.05\mu_0$	13.30	11.3	NC	12.14
				$\mu_0 - 0.545\sigma_0$	13.26	11.3	NC	11.15
				$P_0 + 0.05$	15.33	2.31	NC	13.00
	Ulna Roundness	2.5	10	$\mu_0 - 0.05\mu_0$	4.41	3.30	NC	3.72
				$\mu_0 - 0.545\ \sigma_0$	3.61	2.65	NC	3.00
				$P_0 + 0.05$	4.89	2.65	NC	3.64

NC = Not Calculated. LRB = Likelihood Ratio Based. Units for NOAEL, LOAEL, BMD, and BMDL are mg/kg.

* Confidence band for fetal-based GEE Piecewise Linear Model.

REFERENCES

Accutane® Medication Guide (2007). Roche Laboratories Inc., NJ.

Allen, B. C., Kavlock, R. J., Kimmel, C. A., and Faustman, E. M. (1994). Dose-response assessment for developmental toxicity .2. Comparison of generic benchmark dose estimates with no observed adverse effect levels. *Fundamental and Applied Toxicology* **23**, 487-495.

Auton, T. R. (1994). Calculation of benchmark doses from teratology data. *Regulatory Toxicology and Pharmacology* **19**, 152-167.

Ballinger G.A. (2004). Using generalized estimating equations for longitudinal data analysis. *Organizational Research Methods* **7**, 127-150.

Campbell, J. L., Smith, M. A., Fisher, J. W., and Warren, D. A. (2004). Dose-response for retinoic acid-induced forelimb malformations and cleft palate: A comparison of computerized image analysis and visual inspection. *Birth Defects Research Part B-Developmental and Reproductive Toxicology* **71**, 289-295.

Carr, G. J., and Portier, C. J. (1993). An evaluation of some methods for fitting dose-response models to quantal-response developmental toxicology data. *Biometrics* **49**, 779-791.

Catalano, P., Ryan, L., and Scharfstein, D. (1994). Modeling fetal death and malformation in developmental toxicity studies. *Risk Analysis* **14**, 629-637.

Cattelan AM, Trevenzoli M, and Aversa SM (2002). Recent advances in the treatment of AIDS-related Kaposi's sarcoma. *American Journal of Clinical Dermatology* **3**, 451-462.

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

Collins M.D. and Mao G.E. (1999). Teratology of retinoids. *Annual Review of Pharmacology and Toxicology* **39**, 399-430.

Crump, K. (2002). Critical issues in benchmark calculations from continuous data. *Critical Reviews in Toxicology* **32**, 133-153.

Crump, K. S. (1995). Calculation of benchmark doses from continuous data. *Risk Analysis* **15**, 79-89.

Crump, K. S.; R. Howe. (1985). Chapter 9. in *Toxicological Risk Assessment.* CRC Press, Inc.

Crump, K. S., Van Landingham, C., Shamlaye, C.*, et al.* (2000). Benchmark concentrations for methylmercury obtained from the Seychelles Child Development study. *Environmental Health Perspectives* **108**, 257-263.

Filipsson, A. F., Sand, S., Nilsson, J., and Victorin, K. (2003). The benchmark dose method - Review of available models, and recommendations for application in health risk assessment. *Critical Reviews in Toxicology* **33**, 505-542.

Fung, K. Y., Marro L., and Krewski D. (1998). A comparison of methods for estimating the benchmark dose based on overdispersed data from developmental toxicity studies. *Risk Analysis* **18**, 329-342.

Gaylor, D. W. (1996). Quantalization of continuous data for benchmark dose estimation. *Regulatory Toxicology and Pharmacology* **24**, 246-250.

Jacobson, J. L., Janisse, J., Banerjee, M., Jester, J., Jacobson, S. W., and Ager, J. W. (2002). A benchmark dose analysis of prenatal exposure to polychlorinated biphenyls. *Environmental Health Perspectives* **110**, 393-398.

Hansen D.K., Laborde J.B., Wall K.S., Hinson W.G., Pipkin J.L., Shaddock J., Lyn-Cook L., and Young J.F. (2001). Dose-response of retinoic acid induced stress protein synthesis and teratogenesis in mice. *Reproductive Toxicology* **15**:31-41.

Hardin J.W. and Hibe J.M. (2001) Generalized Estimating Equations. CRC Press, Inc.

Kamm J.J., Ashenfelter K.O., and Ehmann C.W. (1984). Preclinical and clinical toxicology of selected retinoids. In: Sporn MB, Roberts AB, Goodman DS, editors. *The Retinoids*. 2nd ed. Orlando: Academic Press. p. 287-326.

Kavlock, R. J., Allen, B. C., Faustman, E. M., and Kimmel, C. A. (1995). Dose-response assessments for developmental toxicity .4. Benchmark doses for fetal weight changes. *Fundamental and Applied Toxicology* **26**, 211-222.

Kavlock, R. J., Schmid, J.E., and Setzer, R.W. (1996). A simulation study of the influence of study design on the estimation of benchmark doses for developmental toxicity. *Risk Analysis* **16**, 399-410.

Kim, S. Y., Lee, J. T., Hong, Y. C., Ahn, K. J., and Kim, H. (2004). Determining the threshold effect of ozone on daily mortality: an analysis of ozone and mortality in Seoul, Korea, 1995-1999. *Environmental Research* **94**, 113-119.

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

Kochhar DM, Hiang H, Penner JD, Beard RL, and Chandranratna RAS (1996) Differential teratogenic response of mouse embryos to receptor selective analogs of retinoic acid. *Chemico-Biological Interactions* **100**, 1-12.

Kupper, L. L., Portier, C., Hogan, M. D., and Yamamoto, E. (1986). The impact of litter effects on dose-response modeling in teratology. *Biometrics* **42**, 85-98.

Kutner M.H., Nachtsheim C.J., Neter J., and Li W. (2005). Applied Linear Statistical Models. McGraw-Hill/Irwin.

Lammer E.J., Chen D.T., Hoar R.M., Agnish N.D., Benke P.J., Braun J.T., Curry C.J., Fernhoff P.M., Grix A.W., Lott I.T., Richard J.M., and Sun S.C. (1985). Retinoic acid embryopathy. *New England Journal of Medicine* **313**, 837-841.

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

Maden M. (2000). The role of retinoic acid in embryonic and post-embryonic development. *The Proceedings of the Nutrition Society* **59**, 65-73.

Moerbeek, M., Piersma, A. H., and Slob, W. (2004). A comparison of three methods for calculating confidence intervals for the benchmark dose. *Risk Analysis* **24**, 31-40.

Nelson, K. P., Lipsitz, S. R., Fitzmaurice, G. M., Ibrahim, J., Parzen, M., and Strawderman, R. (2006). Use of the probability integral transformation to fit nonlinear mixed-effects models with nonnormal random effects. *Journal of Computational and Graphical Statistics* **15**, 39-57.

Roberts, S., and Martin, M. A. (2007). A distributed lag approach to fitting non-linear dose-response models in particulate matter air pollution time series investigations. *Environmental Research* **104**, 193-200.

Rogers, J. M., Mole, M. L., Chernoff, N.*, et al.* (1993). The developmental toxicity of inhaled methanol in the cd-1 mouse, with quantitative dose-response modeling for estimation of benchmark doses. *Teratology* **47**, 175-188.

Ryan, L. (1992). The use of generalized estimating equations for risk assessment in developmental toxicity. *Risk Analysis* **12**, 439-447.

SAS/STAT 9.1 User's Guide (2004). SAS Institute Inc., Cary, NC.

Soaita, A.I. (2006). GEE Models for the longitudinal analysis of the effects of occupational radiation exposure on lymphocyte counts in Russian nuclear works. Ph.D. Dissertation, School of Public Health, University of Pittsburgh.

Steenland, K., Deddens, J., and Piacitelli, L. (2001). Risk assessment for 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) based on an epidemiologic study. *American Journal of Epidemiology* **154**, 451-458.

U.S. Environmental Protection Agency (US EPA) (2000). Benchmark Dose Technical Guidance Document. Risk Assessment Forum, Washington, DC: EPA/630/R-00/001.

U.S. Environmental Protection Agency (US EPA) (2005). Guidelines for Carcinogen Risk Assessment. Risk Assessment Forum, Washington, DC: EPA/630/P-03/001F.

Vesanoid® Medication Guide (1998). Roche Laboratories Inc., NJ.

Wheeler M. W. (2005). Benchmark Dose Estimation Using SAS®. SAS® Users Group International (SUGI) 30, Paper 201-30.

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

Zhu, Y., Wang, T., and Jelsovsky, J.Z.H. (2007). Bootstrap Estimation of Benchmark Doses and Confidence Limits with Clustered Quantal Data. *Risk Analysis* **27**, 447-465.

APPENDIX A

EXAMPLE SAS CODE

```
/*-------------------------------------------------------------------------*
| Purpose: Beta-Binomial log-logistic model for                            |
|          quantal malformation endpoint;                                  |
| Model Parameters: p0, alpha, beta, and rho1-rho6                         |
| Defect = Defected fetuses in a Litter                                    |
| Size = Number of fetuses in a litter                                     |
*-------------------------------------------------------------------------*/
data One;
X1 = 0; X2 =0; X3 =0; X4=0; X5=0; X6=0;
IF DOSE = 0 THEN X1 = 1;
IF DOSE = 2.5 THEN X2 =1;
IF DOSE = 10 THEN X3=1;
IF DOSE = 30 THEN X4 =1;
IF DOSE = 60 THEN X5 = 1;
IF DOSE =100 THEN X6 = 1;
proc sort data = one; by dose damid;

Proc nlmixed data=One;
parms p0=0.04 alpha=-5 beta=1.8
rho1=0.3 rho2=0.4 rho3=0.8 rho4=0.5 rho5=0.3 rho6=0.2;
bounds alpha < 0, beta > 1, 0.0 < p0 < 1, 0.01<=rho1<=1, 0.01<=rho2<=1,
0.01<=rho3<=1, 0.01 <=rho4<=1, 0.01<=rho5<=1, 0.01<=rho6<=1;

BMR = 0.05;
P = p0;
rho = x1*rho1+x2*rho2+x3*rho3+x4*rho4+x5*rho5+x6*rho6;
      if (dose > 0) then do;
        eta = -(alpha+beta*log(dose));
        expeta = exp(eta);
        P = p0+ (1-p0)/(1+ expeta);
      end;

   Dalpha=P*(1-rho)/rho;
   Dbeta=(1-P)*(1-rho)/rho;
link = log ((1-p0-BMR)/BMR);
BMD = exp ((alpha + link)/(-beta));

like1 = gamma(defect+Dalpha)*gamma(size-
defect+Dbeta)/gamma(Dalpha+Dbeta+size);
like2 = gamma(Dalpha)*gamma(Dbeta)/gamma(Dalpha+Dbeta);
ll = log(like1)-log(like2);

model Defect ~ general(ll);
predict p out = Pred1 alpha=0.1;
predict BMD out=BMD alpha=0.1;
run;
```

```
/*-------------------------------------------------------------------------*
| Purpose:  Macro for Maximum likelihood ratio based BMDL                  |
*-------------------------------------------------------------------------*/

%macro boundBMD(BMDL, MLIKE, CL);
%let BMR=0.05;
data _temp_;
cval = cinv(1-2*(1-&CL),1);
call symput("CRITVAL", cval);
run;

%Let CRITLIKE = &MLIKE;
%do %while(%SYSEVALF((&CRITLIKE - &MLIKE) <= &CRITVAL));
%Let BMDL = %SYSEVALF(0.98*&BMDL);
*set up the initial parameters for the new likelihood;
*%LET FIRSTTOKEN=1;

*fit this new "constrained" likelihood;
data Pe; set pe; If (Parameter = 'beta') then delete; run;

proc nlmixed data=One method = GAUSS NOAD;
ods output fitstatistics =FitR ParameterEstimates=PeR;
parms/ data=pe;
bounds p0 >= 0.0, p0 <= 1, 0.01<=rho1<=1, 0.01<=rho2<=1, 0.01<=rho3<=1,
0.01<=rho4<=1, 0.01<=rho5<=1, 0.01 <=rho6<=1;

link = log ((1-p0-&BMR)/&BMR);
beta = -(alpha +link)/log(&BMDL);
rho = x1*rho1+x2*rho2+x3*rho3+x4*rho4+x5*rho5+x6*rho6;

if (dose = 0) then do;
p = p0;
end;

if (dose >0) then do;
eta = -alpha-beta*log(dose);
expeta = exp(eta);
p = p0+ (1-p0)/(1+ expeta);
end;

DAlpha = p*(1-rho)/rho;
DBeta = (1-p)*(1-rho)/rho;

like1 = gamma(defect+Dalpha)*gamma(size-
defect+Dbeta)/gamma(Dalpha+Dbeta+size);
like2 = gamma(Dalpha)*gamma(Dbeta)/gamma(Dalpha+Dbeta);
ll = log(like1)-log(like2);

model Defect ~ general(ll);
predict p out=Pred2;
predict &BMDL out=BMDL;
run;
ods listing;
```

```
*obtain the Fit statistics to determine if the algorithm has bounded the
BMDL;
data _temp_;
set fitR;
format value best16.;
informat value best16.;
if (Descr = "-2 Log Likelihood");
keep value;
call symput("CRITLIKE", value);
run;
%END;
%put &bmdl;
%mend boundBMD;
%boundBMD(BMDL=3.42261, MLIKE=433.2, CL=0.95); run;


/*-------------------------------------------------------------------------*
| Purpose:  Hybrid Log-logistic Model for Continuous endpoint              |
| Model Parameters: Mu0, alpha, beta, and s1-s6                            |
| Length = Fetal Ulna Length                                               |
*-------------------------------------------------------------------------*/

proc nlmixed data=One;

parameters alpha=-5.8 beta=2.0 mu0=2.73 s1=0.18 s2=0.17 s3=0.14 s4=0.22
s5=0.29 s6=0.56;

sigma = x1*s1+ x2*s2+x3*s3+x4*s4+x5*s5+x6*s6;
var = sigma*sigma;
BMR = 0.05;

pred = Mu0;
X0 = Mu0-2*S1;
P0 = probnorm((X0-Mu0)/S1);
P = P0;

If (dose>0) then do;
eta = -alpha-beta*log(dose);
expeta = exp(eta);
P = P0 + (1-p0)/(1+ expeta);
Pred = X0- sigma*Probit(P);
end;

link = log ((1-p0-BMR)/BMR);
BMD = exp ((alpha + link)/(-beta));

model Length ~ normal(pred, var);
predict pred out=Pred1;
predict BMD out=BMD alpha=0.1;
predict P out=Percent1 alpha=0.1;
ods output ParameterEstimates=Pe;
ods output FitStatistics=fit;
run;
```

```
/*-------------------------------------------------------------------------*
| Purpose: Piecewise Linear GEE Model for                                  |
|          Qunatal Malformation endpoints                                  |
| LimbDefect = Binary endpoint: 1= Abnormal and 0 = Normal                 |
*-------------------------------------------------------------------------*/

Data GEEData;
set one;
p2=0; p3=0; p4=0; p5=0;
if dose > 2.5 then P2=1;
if dose > 10 then P3=1;
if dose > 30 then P4=1;
if dose > 60 then P5=1;
if dose = 0 then lgd= dose;
if dose > 0 then lgd = log(dose);
Td2 = Lgd-log(2.5);
Td3 = Lgd-log(10);
Td4 = Lgd-log(30);
Td5 = Lgd-log(60);
proc sort; by dose damid fetusid; run;

proc genmod data=GEEData descending;
class damid dose;
model LimbDefect = lgd P2*Td2 P3*Td3 P4*Td4 P5*Td5
/AGGREGATE= Damid link=logit dist=bin alpha=0.1 SCALE=PEARSON;
repeated subject = damid /type = cs;
ods output GEEEmpPEst=parms;
output out = Residuals
Pred = Pred;
run;
```