Semi-parametric Model for Exchangeable Clustered Binary Outcomes

Xinran Qi* Aniko Szabo†

Abstract

Dependent or correlated binary data occur in experimental studies such as teratological risk assessment. Groups of correlated responses are often called clusters, and the response of interest is the number of affected units in a cluster. The simplest statistical model for binary outcomes is binomial distribution, which assumes individuals to be independent and identically distributed. However, the assumption of the binomial distribution is often violated. Both parametric (such as Beta-Binomial, q-power) and non-parametric (exchangeable binary) models have been proposed to model distribution of the number of affected individuals over several treatment groups. We propose a semi-parametric model that combines a non-parametric baseline describing the within-cluster dependence structure with a parametric between-group effect. The proposed model avoids making parametric assumptions about higher-order dependence, but is more parsimonious than non-parametric models. We fit the semi-parametric model with an Expectation Minimization Minorize-Maximize algorithm to the boron acid mouse dataset, and compare the semi-parametric estimates of joint probabilities from different dose levels with corresponding generalized estimating equations and non-parametric estimates.

Key Words: Clustered binary data, exchangeability, semi-parametric, relative risk, teratological risk assessment

1. Introduction

1.1 Example of Clustered Binary Outcomes

The National Toxicology Program conducted a study on the developmental toxicity of boron acid by providing pregnant mice with feed containing boron acid (BA) [Heindel et al., 1994]. BA was provided in mice's feed at different dose levels throughout gestation to guarantee fetuses' steady-state exposure during growth development. The number of fetuses in the litter would not be affected by the dose exposure because the mice were fed with BA after they were pregnant. At the end of study, each mouse's uterus was examined to determine number of resorptions, deaths, or other abnormalities.

As shown in figure 1, four dose groups corresponding to exposure levels of $0, 0.1, 0.2$ and 0.4% BA were used. For each pregnant mouse, represented with the rectangle, the circles within the cluster, regardless of their colors, represent implantation sites; and out of which the red circles

*Division of Biostatistics, Institute for Health and Equity, Medical College of Wisconsin, Milwaukee, Wisconsin 53226, U.S.A.

†Division of Biostatistics, Institute for Health and Equity, Medical College of Wisconsin, Milwaukee, Wisconsin 53226, U.S.A.

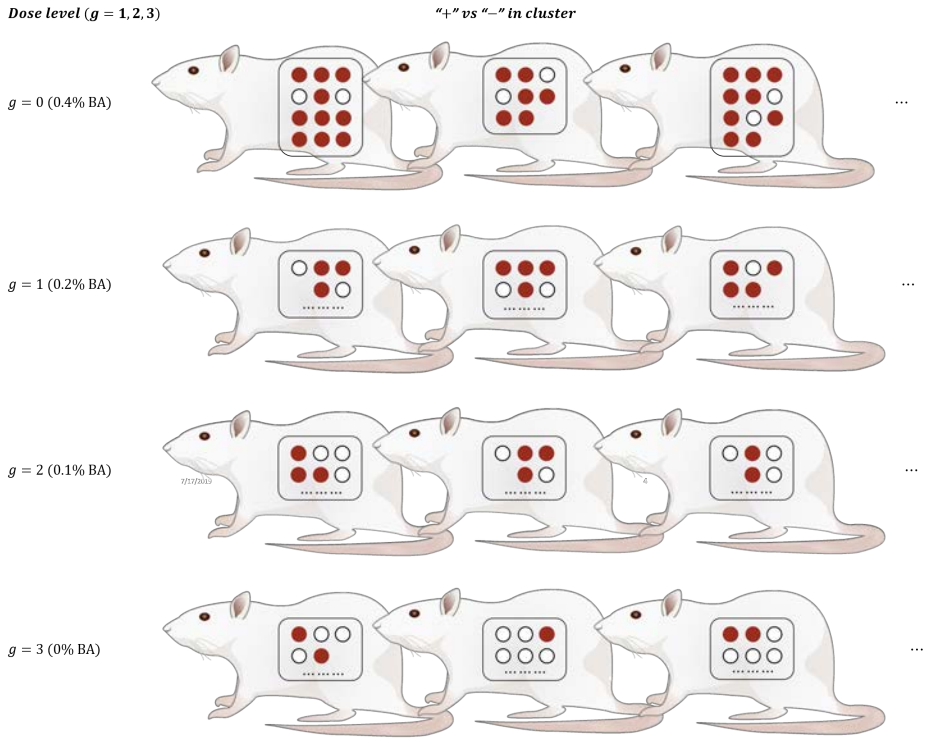

Figure 1: Developmental toxicity experiment of boron acid

represent the combined endpoints, either resorbed embryos, or dead fetuses, or fetuses with other abnormalities. The white circles represent healthy fetuses.

Figure 2 is a stacked bar plot illustrating responses from the mice's litters under different dose levels' exposure. We can see that doses of higher concentrations might increase the risk for individual due to the expanded tendency in the number of affected. Our primary parameter of interest is the entire distribution for the combined endpoints from varying cluster sizes and under different dose levels exposure. The fetuses are correlated not only because they gestate in the same litter, but also they are exposed to the same or different concentrations of boron acid. We want to model this within-cluster effects and exogenous risks and are going to use the second or even higher order correlation coefficients as parameterization in the proposed models.

1.2 Notations for Clustered Binary Outcomes

In many contexts with clustered binary outcomes, the cluster sizes are not fixed. For a sample of clusters $(i = 1, \ldots, I)$ of corresponding cluster sizes $(r_1, \ldots, r_i, \ldots, r_I)$, observations in the i^{th} cluster of cluster size r_i are denoted as $(X_{i1}, \ldots, X_{ij}, \ldots, X_{ir_i})$. Each observation within the cluster is a binary outcome with 1 indicating success and 0 indicating failure. Summing up

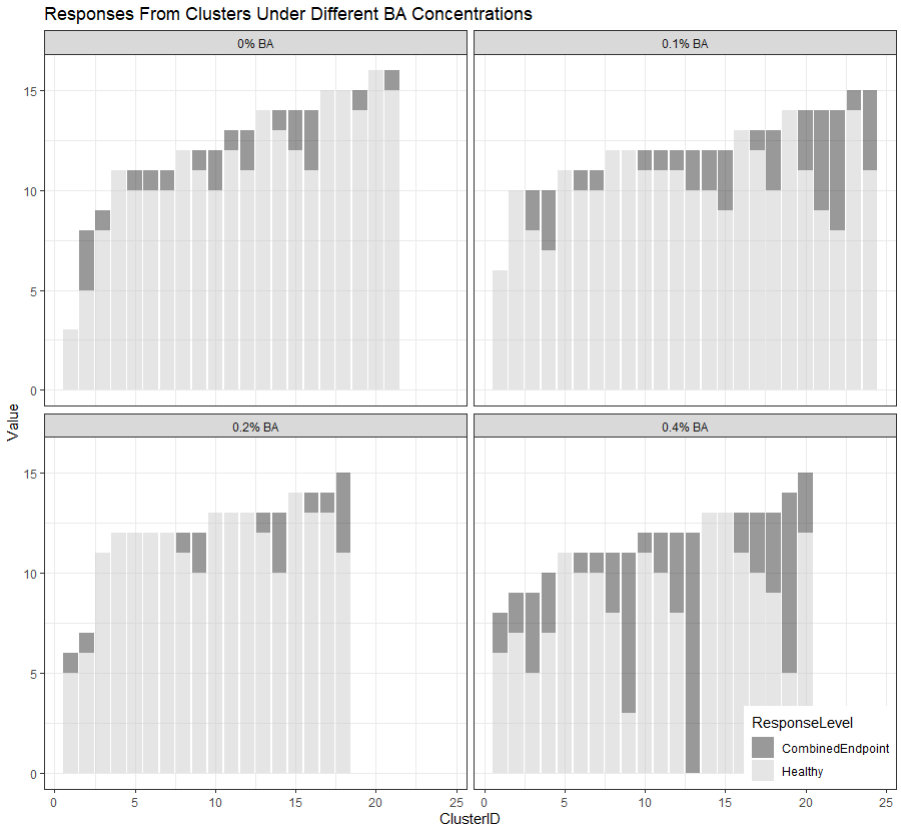

Figure 2: Bar plot for developmental toxicity experiment of boron acid

observations' responses in the i^{th} cluster of cluster size r_i gives the total number of successes within this cluster, denoted as $S_i = \sum_{j=1}^{r_i} X_{ij}$. The marginal probability of achieving s successes from a cluster of cluster size r is $q_{s,r} = \mathbf{Pr}_r(S = s)$, where $\mathbf{Pr}_r$ is the probability statement conditional on the cluster size being r. For a sample consisting of I clusters, the maximum cluster size is $R = \max(r_i)$, where $1 \leq i \leq I$; and the number of clusters of cluster size r and achieving s successes is denoted with n_{rs}.

1.3 Existing Methods to Model Clustered Binary Outcomes

In the study of clustered binary outcomes, the assumptions of multinomial distribution are violated in most situations. There are typically two ways to look at the clustering. One is the lack of independence among observations within the same cluster, and the other is the unobserved heterogeneity within a cluster.

The fundamental assumption lying behind all models constructed to handle the clustered binary

outcomes is that conditional on the probability of response, the observation is a random variable subject to Bernoulli distribution. Based on this assumption, there are mainly three types of models proposed to handle clustered binary outcomes.

The first type of models to handle clustered binary outcomes are the fixed or random effect models. They treat the correlation among observations within the same cluster as a fixed constant or a random variable. For fixed effect model, Tallis [1962] added an additional parameter ρ, which was the correlation coefficient for two variables being marginally distributed as multinomial variates with common means, to the probability generating functions and derived the joint probability distribution of correlated observations within the same cluster. Kupper and Haseman [1978] used the "correction factor", which was a function of pairwise, higher order correlations and binomial probability parameters, to multiply the standard binomial probability distributions to "correct for" the lack of mutual independence among observations. Rao and Scott [1992] applied the concepts of design effect and effective sample size from sample survey and constructed an adjusted multinomial distribution model with no assumption of intra-cluster correlations. For the random effect model, Williams [1975] proposed the beta-binomial model assuming that within each cluster, the binary responses formed a set of Bernoulli trial whose probability parameter varied among clusters in the same population according to a two-parameter Beta distribution.

The second type of models for clustered binary outcomes are the quasi-likelihood or generalized estimating equation (GEE) models. Depending on the research objectives, the expectation and covariance matrix of the objective can be obtained using the properties of generalized exponential family or direct derivation from the given distributions and definitions. The estimating equations consisting of the regression parameters can be solved using different numerical approaches such as Newton-Raphson algorithm. GEE model produces robust estimates of regression coefficients and other parameters even when the working correlation matrix characterizing the relation between observations within the cluster is misspecified. The research objectives can be marginal means and pairwise correlations [Prentice and Zhao, 1991], odds ratios [Lipsitz et al., 1991], or mean responses and intra-cluster correlations [Bowman et al., 1995].

The third type of model for correlated binary data is using higher order moments or correlations as parameterization to handle probability distribution under the exchangeability assumption. A sequence of binary random variables $X_1, X_2, \ldots$ is exchangeable if for any n and any vector $(x_1, x_2, \ldots, x_n)'$, the joint probability $\mathbf{Pr}(X_{\pi(1)} = x_1, \ldots, X_{\pi(n)} = x_n) = \mathbf{Pr}(X_1 = x_1, \ldots, X_n = x_n)$ for any permutation of indexes $1, 2, \ldots, n$. Based on the exchangeability assumption, Bowman and George [1995] proposed to use the joint probability $\lambda_k = \mathbf{Pr}(X_{j_1} = X_{j_2} = \cdots = X_{j_k} = 1)$, where $\{j_1, j_2, \ldots, j_k\}$ was a subset of $\{1, 2, \ldots, n\}$, $k = 1, \ldots, n$ and $\lambda_0 = 1$, to reparametrize the likelihood function. Estimators of $\{\lambda_k\}$ under equal cluster size were derived using the maximum likelihood estimation. Stefanescu and Turnbull [2003] proposed an additional assumption "marginal compatibility" to estimate $\{\lambda_k\}$ under varying cluster sizes. They assumed that clusters of different sizes might be viewed as coming from a sample of equal-size clusters, with some observations missing completely at random. The marginal compatibility assumption is that the probability distribution of $X_{1,k}, X_{2,k}, \ldots, X_{k,k}$ in a cluster of size $k < n$ should be the same as that of $X_{1,n}, X_{2,n}, \ldots, X_{k,n}$ in a cluster of size n. For $1 \leq k \leq n$,

$\lambda_{k,n} = \mathbf{Pr}(X_{1,n} = \cdots = X_{k,n} = 1) = \mathbf{Pr}(X_{1,k} = \cdots = X_{k,k} = 1) = \lambda_{k,k}$, which no longer depends on the cluster size. This assumption links up the distributions for different cluster sizes so that estimation can be based on the combined data from all cluster sizes. The researchers estimated the $\{\lambda_k\}$ under varying cluster sizes using the EM algorithm. Pang and Kuk [2007] proposed a penalized kernel smoothing method which performed smoothing in both covariate and response space. They also proposed a generalized Armitage's trend test to test for the marginal compatibility assumption.

Many parametric models have been proposed to model the completely monotonic sequence $\{\lambda_k\}$. Kuk [2004] proposed a more manageable two-parameter family of distribution that could be parameterized in terms of marginal response probability and the intra-cluster association under the exchangeability assumption. This was achieved by applying power transformation directly to the response probability. Swapping the 1's with 0's for observations gave the q-power model, which fit real data better than p-power model in some toxicological experiments. The risk of at least one adverse response within a cluster took on a simple form under the q-power distribution and could be reduced further to a generalized linear model if a complementary log-log link function was used.

In addition to using $\{\lambda_k\}$ to parameterize the likelihood function, there are other higher order moments and parameters proposed for reparameterization of likelihood. Ekholm et al. [2003] proposed to use the dependence ratios of all orders together with the first-order moments to represent the path probability. Lovison [2015] proposed the dependence ratio (DR)-binomial distribution under exchangeability assumption. The likelihood function was reparametrized using marginal probability and dependence ratio. The DR-binomial distribution is the reparameterization of Altham's additive-binomial and Kupper-Haseman's correlated-binomial distribution.

The sum of exchangeable Bernoulli random variables can be represented as continuous-time Markov processes via a technique called "probabilistic embedding". Crawford and Zelterman [2015] proposed a Markov counting model for correlated binary responses. By introducing an auxiliary variable, the binary responses are made to depend on the arrival times of points in a Markov counting process. This formulation provided a flexible way to parameterize and fit models for correlated binary outcomes, and accommodated different cluster size ascertainment schemes.

Going into more details about clustered exchangeable outcomes, multiple assumptions have been proposed in order to construct parametric, non-parametric or semi-parametric models to incorporate correlations among observations within the same cluster. The semi-parametric model proposed in our study is based on two significant assumptions. One is the exchangeability assumption, which is stated as given the cluster size being r, the joint probability of a sequence of binary random variables remains the same for any permutation. This assumption is essentially embedded in many parametric and non-parametric models. For example, models using the sum of successes from the cluster $S = \sum_{j=1}^{r} X_j$ as the response variable are all based on the exchangeability assumption. Otherwise, the sum is not sufficient to model multiple measurements of the cluster. Based on the exchangeability assumption, the joint probability $\lambda_{k,r} = \mathbf{Pr}_r(X_1 = X_2 = \cdots = X_k = 1)$ was proposed by Bowman and George [1995] and could be estimated using the maximum likelihood estimator under equal cluster size. Stefanescu and Turnbull [2003] proposed an additional assumption, "marginal compatibility", to estimate $\{\lambda_{k,r}\}$ under varying cluster sizes using the EM

algorithm. This assumption links up values of $\{\lambda_{k,r}\}$ for different cluster sizes so that the parameter estimation is based on the combined data with all cluster sizes.

The parameter $\{\lambda_k\}$ has a one-to-one mapping to the marginal probabilities $\{q_s\}$. For $s = 0, \ldots, r$, the relation between $\{q_s\}$ and $\{\lambda_k\}$ is:

$$q_s = \binom{r}{s} \sum_{j=0}^{r-s} (-1)^j \binom{r-s}{j} \lambda_{s+j}, \tag{1}$$

$$\lambda_k = \sum_{j=0}^{r-k} \frac{\binom{r-k}{j} q_{r-j}}{\binom{r}{r-j}}, \tag{2}$$

as proved by Bowman and George [1995]. The marginal probabilities $\{q_0, q_1, \ldots, q_r\}$ are subject to the constraint $\sum_{s=0}^{r} q_s = 1$ to be a probability mass function. Since there is a one-to-one mapping from $\{q_s\}$ to $\{\lambda_k\}$, $\{\lambda_k\}$ should also be subject to certain constraint to be a new set of reparametrization for the marginal probabilities $\{q_s\}$ with respect to a distribution. The constraint is that $\{\lambda_k\}$ should be completely monotone, defined as follows:

$$\Delta^{r-k}(\lambda_k) \geq 0, k = 0, \ldots, r, \tag{3}$$

where $\Delta(\lambda_k) = \lambda_k - \lambda_{k+1}, \Delta^2(\lambda_k) = \Delta(\lambda_k) - \Delta(\lambda_{k+1})$, etc, are the forward differences. The complete monotonicity property guarantees that the parameters $\{\lambda_k\}$ define a distribution.

2. Semi-parametric model for exchangeable clustered binary outcomes

Based on the exchangeability and marginal compatibility assumptions, we propose the following semi-parametric model to describe relationship among $\left\{\lambda_k^{(g)}\right\}$ from different dose levels ($g = 1, \ldots, G$):

$$\lambda_k^{(g)} = \lambda_k^{(0)} \times \theta_g^k, k = 1, \ldots, r, \tag{4}$$

where θ_g is a parameter with range $[0, 1]$; $\left\{\lambda_k^{(0)}\right\}$ is the joint probability for the baseline dose level; and $\left\{\lambda_k^{(g)}\right\}$ is the joint probability for the g^{th} dose level.

When we set $k = 1$, $\lambda_1^{(g)} = \lambda_1^{(0)} \times \theta_g$ according to the proposed semi-parametric model (4). Therefore, $\theta_g = \dfrac{\lambda_1^{(g)}}{\lambda_1^{(0)}}$ and is essentially the relative risk of the g^{th} dose level compared with the baseline.

Our model incorporates a flexible baseline $\left\{\lambda_k^{(0)}\right\}$ in order to relax the restrictions of parametric model such as Beta-Binomial, and assigns a parameter θ_g to each dose level's $\left\{\lambda_k^{(g)}\right\}$ so that the model is much more parsimonious than fully non-parametric models. Our semi-parametric

model also guarantees the complete monotonicity property of $\{\lambda_k\}$ because the baseline $\left\{\lambda_k^{(0)}\right\}$ is completely monotone, and the exponential function θ_g^k is also completely monotone. The product of these two terms yields a completely monotone sequence. In addition, if we use another type of reparameterization, the dependence ratio [Lovison, 2015], defined under the exchangeability assumption as follows:

$$\begin{aligned}\tau_k &= \frac{\mathbf{Pr}(X_1 = X_2 = \cdots = X_k = 1)}{\mathbf{Pr}(X_1 = 1) \times \mathbf{Pr}(X_2 = 1) \times \cdots \times \mathbf{Pr}(X_k = 1)} \\ &= \frac{\pi_{1_1,1_2,\ldots,1_k}}{\pi_{1_1}\pi_{1_2}\ldots\pi_{1_k}}.\end{aligned} \tag{5}$$

In our semi-parametric model setting, we can come to the conclusion that the model is equivalent to assuming that the dependence ratio is the same across all dose levels. The g^{th} dose level's k^{th} dependence ratio in our semi-parametric model setting is:

$$\begin{aligned}\tau_k^{(g)} &= \frac{\mathbf{Pr}^{(g)}(X_1 = X_2 = \cdots = X_k = 1)}{\mathbf{Pr}^{(g)}(X_1 = 1) \times \mathbf{Pr}^{(g)}(X_2 = 1) \times \cdots \times \mathbf{Pr}^{(g)}(X_k = 1)} \\ &= \frac{\lambda_k^{(g)}}{\left(\lambda_1^{(g)}\right)^k} = \frac{\lambda_k^{(0)} \times \theta_g^k}{\left(\lambda_1^{(0)} \times \theta_g^1\right)^k} = \frac{\lambda_k^{(0)}}{\left(\lambda_1^{(0)}\right)^k},\end{aligned} \tag{6}$$

which means that the k^{th} order dependence ratio is a function of the baseline joint probabilities $\boldsymbol{\lambda}^{(0)}$, and is independent of the variation of dose level g.

3. Parameter estimation using the EM MM algorithm

Parameter estimation is based on the maximization likelihood estimator with respect to likelihood function of observed data. Under the marginal compatibility assumption, the parameters $\boldsymbol{\lambda}$ are independent of cluster sizes. The likelihood function based on the observed data is written as follows:

$$L = \prod_{g=0}^{G}\prod_{r=1}^{R}\prod_{s=0}^{r}\left\{\binom{r}{s}\sum_{j=0}^{r-s}(-1)^j\binom{r-s}{j}\lambda_{s+j}^{(g)}\right\}^{n_{rs}^{(g)}}. \tag{7}$$

We implement an algorithm called the Expectation Maximization Minorize-Maximize, abbreviated as EM MM, to estimate the parameters. For dose level g, let $\boldsymbol{\phi}_g = (\boldsymbol{\lambda}^{(0)}, \theta_g)$ denote the parameters in the semi-parametric model. Stefanescu and Turnbull [2003] have proved that the marginal compatibility assumption is equivalent to assuming that clusters are from a sample of clusters sharing the same cluster size (the maximum cluster size R is a good choice and is used

in our study), but some observations are completely missing at random, abbreviated as MCAR. The expectation step in the EM MM algorithm can be performed based on the MCAR assumption. For complete data in which all clusters have R observations, the marginal probability of achieving t successes out of R observations within a cluster is denoted with parameter $q_t^{(g)}$. Estimation of $\boldsymbol{\lambda}^{(g)}$ is much more demanding than that of $\boldsymbol{q}^{(g)}$ because the complete monotonicity property of $\boldsymbol{\lambda}^{(g)}$ is difficult to control during each EM MM updating iteration. We hence focus on estimating the parameters $\boldsymbol{q}^{(g)}$ instead of $\boldsymbol{\lambda}^{(g)}$ because of its robustness during interation. Therefore, we let $\boldsymbol{\phi}_g = (\boldsymbol{q}^{(0)}, \theta_g)$ to denote the set of parameters in our semi-parametric model. Based on the semi-parametric model (4), and one-to-one mapping formulas (1) and (2), the following equation holds for the baseline dose level marginal probabilities $\boldsymbol{q}^{(0)}$ and $\boldsymbol{q}^{(g)}$ for dose level g;

$$
\begin{aligned}
q_t^{(g)} &= \binom{R}{t} \sum_{j=0}^{R-t} (-1)^j \binom{R-t}{j} \lambda_{t+j}^{(0)} \theta_g^{t+j} \\
&= \binom{R}{t} \sum_{j=0}^{R-t} (-1)^j \binom{R-t}{j} \theta_g^{t+j} \left\{ \sum_{l=0}^{R-t-j} \frac{\binom{R-t-j}{l} q_{R-l}^{(0)}}{\binom{R}{R-l}} \right\} \\
&= \left(\frac{\theta_g}{1-\theta_g} \right)^t \times \sum_{l=0}^{R-t} \binom{R-l}{t} q_{R-l}^{(0)} (1-\theta_g)^{R-l} \\
&= \left(\frac{\theta_g}{1-\theta_g} \right)^t \times \sum_{\alpha=t}^{R} \binom{\alpha}{t} q_{\alpha}^{(0)} (1-\theta_g)^{\alpha} \\
&= \sum_{\alpha=t}^{R} \binom{\alpha}{t} \theta_g^t (1-\theta_g)^{\alpha-t} q_{\alpha}^{(0)}.
\end{aligned}
\tag{8}
$$

Let the parameter $p_{rst}^{(g)} = \mathbf{Pr}_R(T = t | S = s, r)$ denote the conditional probability of achieving t successes out of complete cluster size R for the g^{th} dose level, given that s successes have been achieved out of the original observed cluster size r. Based on the MCAR assumption, it can be derived from $q_t^{(g)}$ [Stefanescu and Turnbull, 2003], shown as follows:

$$
\begin{aligned}
p_{rst}^{(g)} &= \mathbf{Pr}_R^{(g)}(T = t | S = s, r) \\
&= \frac{\mathbf{Pr}_r^{(g)}(S = s | T = t) \mathbf{Pr}_R^{(g)}(T = t)}{\sum_{t'=0}^{R} \mathbf{Pr}_r^{(g)}(S = s | T = t') \mathbf{Pr}_R^{(g)}(T = t')} \\
&= \frac{\binom{t}{s}\binom{R-t}{r-s} q_t^{(g)}}{\sum_{t'=s}^{R-r+s} \binom{t'}{s}\binom{R-t'}{r-s} q_{t'}^{(g)}}.
\end{aligned}
\tag{9}
$$

For dose level g, based on the conditional probability $p_{rst}^{(g)}$ and number of clusters achieving s

successes out of cluster size r, $n_{rs}^{(g)}$, from the observed data, the expectation of number of successes $n_t^{(g)\prime}$ achieved out of complete cluster size R given the k^{th} iteration's parameters $\boldsymbol{\phi}^{(k)}$ can be derived as follows [Stefanescu and Turnbull, 2003]:

$$n_t^{(g)\prime(k)} = E_Z[N_t^{(g)}|S_k, \boldsymbol{\phi}^{(k)}] = \sum_{r=1}^{R} \sum_{s=max(0,t+r-R)}^{min(t,r)} n_{rs}^{(g)} p_{rst}^{(g)(k)}. \tag{10}$$

The log likelihood function for the complete data given the k^{th} iteration's parameters $\boldsymbol{\phi}^{(k)}$ is:

$$\log(L(\boldsymbol{\phi}|\boldsymbol{\phi}^{(k)})) = \sum_{t=0}^{R} n_t^{(0)\prime(k)} \log(q_t^{(0)}) + \sum_{g=1}^{G}\sum_{t=0}^{R} n_t^{(g)\prime(k)} \log(q_t^{(g)}(\theta_g, \boldsymbol{q}^{(0)})). \tag{11}$$

The maximization step uses Minorize-Maximize algorithm to update the parameters $\boldsymbol{\phi}$. We apply the Jensen's inequality to lower bound the term $\log(q_t^{(g)}(\theta_g, \boldsymbol{q}^{(0)}))$ by assigning elementwise weights to each $q_\alpha^{(0)}$, and transform the log sum expression to obtain a closed form update for parameters $\boldsymbol{\phi}$. For any set of weights $\left\{w_g^{(k)}(\alpha, t)\right\}$ such that $\sum_{\alpha=t}^{R} w_g^{(k)}(\alpha, t) = 1$, we have:

$$\begin{aligned}
\log(q_t^{(g)}(\theta_g, \boldsymbol{q}^{(0)})) &= \log\left\{\sum_{\alpha=t}^{R} \binom{\alpha}{t} \theta_g^t (1-\theta_g)^{\alpha-t} q_\alpha^{(0)}\right\} \\
&= \log\left\{\sum_{\alpha=t}^{R} w_g^{(k)}(\alpha, t) \frac{\binom{\alpha}{t}\theta_g^t(1-\theta_g)^{\alpha-t} q_\alpha^{(0)}}{w_g^{(k)}(\alpha, t)}\right\} \\
&\geq \sum_{\alpha=t}^{R} w_g^{(k)}(\alpha, t) \log\left\{\frac{\binom{\alpha}{t}\theta_g^t(1-\theta_g)^{\alpha-t} q_\alpha^{(0)}}{w_g^{(k)}(\alpha, t)}\right\} \\
&= \sum_{\alpha=t}^{R} w_g^{(k)}(\alpha, t)\Big[\log\binom{\alpha}{t} + t\log(\theta_g) + \alpha\log(1-\theta_g) - \\
&\quad t\log(1-\theta_g) + \log(q_\alpha^{(0)}) - \log(w_g^{(k)}(\alpha, t))\Big] \\
&= t\log(\theta_g) - t\log(1-\theta_g) + \sum_{\alpha=t}^{R} w_g^{(k)}(\alpha, t)\Big[\log\binom{\alpha}{t} + \\
&\quad \alpha\log(1-\theta_g) + \log(q_\alpha^{(0)})\Big] - \sum_{\alpha=t}^{R} w_g^{(k)}(\alpha, t)\log w_g^{(k)}(\alpha, t).
\end{aligned} \tag{12}$$

Therefore, the lower bound of the log-likelihood function is:

$$H(\boldsymbol{\phi}_g|\boldsymbol{\phi}_g^{(k)}) = \sum_{t=0}^{R} {n_t^{(0)}}'^{(k)} \log(q_t^{(0)}) + \sum_{g=1}^{G}\sum_{t=0}^{R} {n_t^{(g)}}'^{(k)} \Big[t \log\left(\frac{\theta_g}{1-\theta_g}\right) + \sum_{\alpha=t}^{R} w_g^{(k)}(\alpha,t)\left(\log\binom{\alpha}{t} + \alpha\log(1-\theta_g) + \log(q_\alpha^{(0)})\right) - \sum_{\alpha=t}^{R} w_g^{(k)}(\alpha,t)\log(w_g^{(k)}(\alpha,t))\Big]. \tag{13}$$

The weight function $w_g^{(k)}(\alpha,t)$ is selected so that $H(\boldsymbol{\phi}_g^{(k)}|\boldsymbol{\phi}_g^{(k)}) = \log(L(\boldsymbol{\phi}_g^{(k)}))$. Notice that the weight at the k^{th} iteration is a function of α and t, shown as follows:

$$w_g^{(k)}(\alpha,t) = \frac{\binom{\alpha}{t}\left[\theta_g^{(k)}\right]^t (1-\theta_g^{(k)})^{\alpha-t} {q_\alpha^{(0)}}^{(k)}}{\sum_{\gamma=t}^{R}\binom{\gamma}{t}\left[\theta_g^{(k)}\right]^t (1-\theta_g^{(k)})^{\gamma-t} {q_\gamma^{(0)}}^{(k)}}. \tag{14}$$

The Lagrangian of the lower bound of log-likelihood function $H(\boldsymbol{\phi}_g|\boldsymbol{\phi}_g^{(k)})$ is adding a constraint that the $\boldsymbol{q}^{(0)}$ sums up to 1. By taking the derivatives with respective to each parameter, we can get the updates for $\boldsymbol{\phi}^{(k+1)}$ given the previous k^{th} iteration's parameter values, shown as follows:

$$\theta_g^{(k+1)} = \frac{\sum_{t=0}^{R} {n_t^{(g)}}'^{(k)} t}{\sum_{t=0}^{R} {n_t^{(g)}}'^{(k)} \sum_{\alpha=t}^{R} w_g^{(k)}(\alpha,t)\alpha}; \tag{15}$$

$${q_\beta^{(0)}}^{(k+1)} = \frac{{n_\beta^{(0)}}'^{(k)} + \sum_{g=1}^{G}\sum_{\rho=0}^{\beta} {n_\rho^{(g)}}'^{(k)} {w_g}^{(k)}(\beta,\rho)}{\sum_{\gamma=0}^{R}({n_\gamma^{(0)}}'^{(k)} + \sum_{g=1}^{G}\sum_{\rho=0}^{\gamma} {n_\rho^{(g)}}'^{(k)} {w_g}^{(k)}(\gamma,\rho))}. \tag{16}$$

We use the above expressions to update the parameters in the proposed semi-parametric model. An appropriate choice of starting values for parameters is important for quick convergence and consistent estimates. We recommend to assign non-parametric estimates of the baseline marginal probabilities as the starting values of $\boldsymbol{q}^{(0)}$; and non-parametric estimates of $\frac{\lambda_1^{(g)}}{\lambda_1^{(0)}}$ to the g^{th} dose level's parameter θ_g as the starting values. This Expectation Maximization Minorize-Maximize algorithm preserves the ascent property of the EM algorithm. The iteration stops when the pre-specified convergence criteria are met.

4. Simulation

We conducted a simulation study to compare the performance of the proposed method to a relative risk GEE, which shared the same model for the marginal event probability. The simulation shared the same setting as the boron acid mouse study. We set the maximum cluster size to 10, and the true values of θ as 0.25, 0.5 and 0.75 corresponding to three dose levels. Each dose level had 250 clusters with varying cluster sizes and the baseline marginal probabilities were subject to a beta-binomial distribution with probability of success being 0.5 and overdispersion parameter being 1. We conducted 1000 simulations and applied both our proposed semi-parametric model and a relative risk GEE model to it. The simulation results are shown in table 1. Notice that the parameter estimates in table 1 are Monte-Carlo means and standard errors. We can see that the θ estimates are quite similar between two models, but the proposed semi-parametric model provides smaller standard errors.

Table 1: Simulation θ estimates for different dose levels

Dose level	True value of θ	Semi-parametric model		Relative risk GEE model	
		θ estimate	SE of θ	θ estimate	SE of θ
0	reference				
1	0.25	0.249	0.022	0.250	0.025
2	0.5	0.500	0.033	0.502	0.042
3	0.75	0.748	0.037	0.749	0.056

5. Application

We applied the proposed semi-parametric model with the EM MM algorithm to the Boron acid mouse dataset and compared the semi-parametric estimates of $\boldsymbol{\lambda}$ from different dose levels with non-parametric estimates. In figure 3, the black dotted lines represent non-parametric estimates and the colored solid lines represent semi-parametric estimates according to the proposed model. We see that the $\boldsymbol{\lambda}$ estimates are similar between both models.

As for the likelihood ratio test, the log-likelihood of the non-parametric model is -352.43, whereas that of the proposed semi-parametric model is -357.46; with p-value equal to 1 on 45 degrees of freedom, indicating the semi-parametric model fits the data well and is more parsimonious than the non-parametric model.

Estimates of θ from different dose levels are shown in table 2. The standard error of θ is calculated using the bootstrap method. Based on the marginal compatibility assumption and model fitting estimates, we can calculate the entire distribution for varying cluster sizes, which cannot be estimated using the GEE models. Table 3 lists the baseline univariate marginal probabilities for

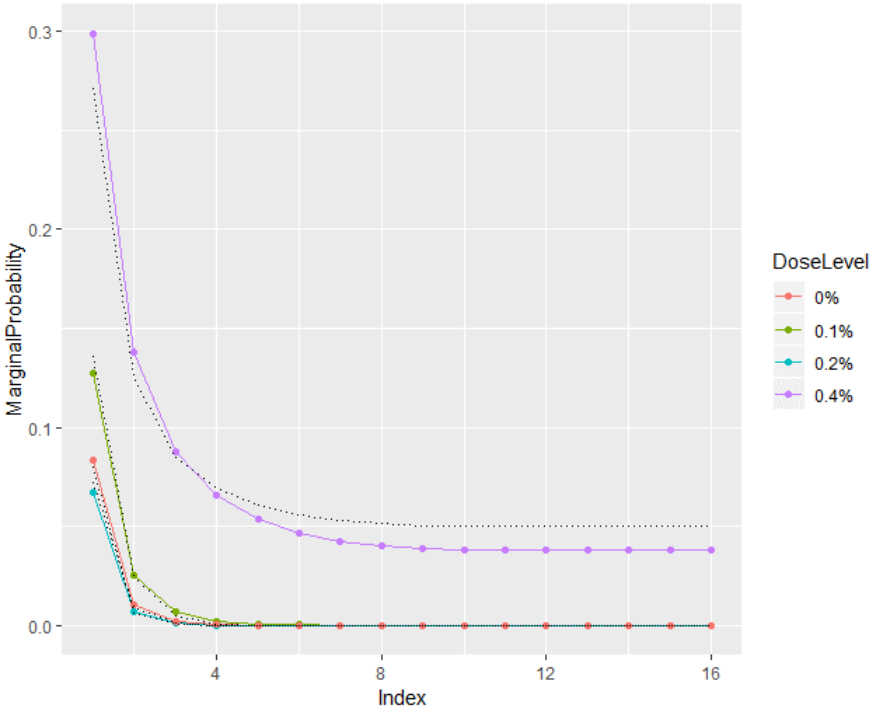

Figure 3: Compare estimates of $\boldsymbol{\lambda}$ between semi-parametric and non-parametric models

cluster sizes ranging from 1 to 7.

Table 2: θ estimates for different dose levels

dose level (%)	θ estimate	SE of θ
0.4	reference	
0.2	0.224	0.164
0.1	0.426	0.206
0	0.280	0.126

6. Conclusions

In this paper, we propose a semi-parametric model for clustered binary outcomes that combines a non-parametric baseline describing within-cluster dependence structure with a parametric between-group effect, based on the exchangeability and marginal compatibility assumptions. The proposed

Table 3: Matrix of marginal probabilities for varying cluster sizes (baseline dose level)

		Number of responses (s)							
		s=0	s=1	s=2	s=3	s=4	s=5	s=6	s=7
Cluster size (r)	r=1	0.701	0.299						
	r=2	0.541	0.321	0.138					
	r=3	0.4306	0.3300	0.1520	0.0874				
	r=4	0.3497	0.3238	0.1743	0.0865	0.0657			
	r=5	0.2881	0.3081	0.1932	0.0973	0.0595	0.0538		
	r=6	0.2408	0.2839	0.2145	0.1005	0.0705	0.0432	0.0466	
	r=7	0.2045	0.2538	0.2322	0.1134	0.0625	0.0612	0.0300	0.0424

model provides estimates of the entire distribution for varying cluster sizes, so that we can infer about quantities other than the marginal means, such as the probability of at least one adverse event. It also provides smaller standard errors of estimates compared to GEE.

Some data may not be well described by a constant dependence ratio structure (but we can detect the lack of fit) and the model can only describe decrease from the baseline dose level. Our future work will focus on eliminating some of these restrictions.

7. Bibliography

Jerrold J Heindel, Catherine J Price, and Bernard A Schwetz. The developmental toxicity of boric acid in mice, rats, and rabbits. *Environmental health perspectives*, 102(suppl 7):107–112, 1994.

GM Tallis. The use of a generalized multinomial distribution in the estimation of correlation in discrete data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 24(2):530–534, 1962.

L L Kupper and J K Haseman. The use of a correlated binomial model for the analysis of certain toxicological experiments. *Biometrics*, 34(1):69–76, Mar 1978.

J N Rao and A J Scott. A simple method for the analysis of clustered binary data. *Biometrics*, 48 (2):577–85, Jun 1992.

D A Williams. The analysis of binary responses from toxicological experiments involving reproduction and teratogenicity. *Biometrics*, 31(4):949–52, Dec 1975.

R L Prentice and L P Zhao. Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses. *Biometrics*, 47(3), Sep 1991.

Stuart R Lipsitz, Nan M Laird, and David P Harrington. Generalized estimating equations for correlated binary data: using the odds ratio as a measure of association. *Biometrika*, 78(1): 153–160, 1991.

D Bowman, J J Chen, and E O George. Estimating variance functions in developmental toxicity studies. *Biometrics*, 51(4):1523–8, Dec 1995.

Dale Bowman and E Olusegun George. A saturated model for analyzing exchangeable binary data: Applications to clinical and developmental toxicity studies. *Journal of the American Statistical Association*, 90(431):871–879, 1995.

Catalina Stefanescu and Bruce W Turnbull. Likelihood inference for exchangeable binary data with varying cluster sizes. *Biometrics*, 59(1):18–24, 2003.

Zhen Pang and Anthony YC Kuk. Test of marginal compatibility and smoothing methods for exchangeable binary data with unequal cluster sizes. *Biometrics*, 63(1):218–227, 2007.

Anthony Y. C. Kuk. A litter-based approach to risk assessment in developmental toxicity studies via a power family of completely monotone functions. *Applied Statistics*, 53:369–386, 2004. ISSN 0035-9254. doi: 10.1046/j.1467-9876.2003.05369.x.

Anders Ekholm, Jukka Jokinen, John W McDonald, and Peter WF Smith. Joint regression and association modeling of longitudinal ordinal data. *Biometrics*, 59(4):795–803, 2003.

Gianfranco Lovison. A generalization of the binomial distribution based on the dependence ratio. *Statistica Neerlandica*, 69:126–149, 2015. ISSN 0039-0402. doi: 10.1111/stan.12053.

Forrest W. Crawford and Daniel Zelterman. Markov counting models for correlated binary responses. *Biostatistics*, 16:427–440, 2015. ISSN 1468-4357. doi: 10.1093/biostatistics/kxv006.