Journal of Modern Applied Statistical Methods

Volume 19 | Issue 1 Article 17

8-16-2021

Performance of the Beta-Binomial Model for Clustered Binary Responses: Comparison with Generalized Estimating Equations

Seongah Im
University of Hawai'i at Mānoa, seongahi@hawaii.edu

Recommended Citation

Im, S. (2020). Performance of the beta-binomial model for clustered binary responses: Comparison with generalized estimating equations. *Journal of Modern Applied Statistical Methods, 19*(1), eP3078. doi: 10.22237/jmasm/1619482380

Journal of Modern Applied Statistical Methods
May 2020, Vol. 19, No. 1, eP3078.
doi: 10.22237/jmasm/1619482380

ISSN 1538 − 9472

Performance of the Beta-Binomial Model for Clustered Binary Responses: Comparison with Generalized Estimating Equations

Seongah Im
University of Hawai'i at Mānoa
Honolulu, HI

This study examined performance of the beta-binomial model in comparison with GEE using clustered binary responses resulting in non-normal outcomes. Monte Carlo simulations were performed under varying intracluster correlations and sample sizes. The results showed that the beta-binomial model performed better for small sample, while GEE performed well under large sample.

Keywords: Correlated binary responses, clustered data, non-normal, summed scores, intracluster correlation, overdispersion, beta-binomial, generalized estimating equations

Introduction

A series of item responses coded as 1 or 0 is often observed in various research studies in the social and behavioral sciences. Binary responses usually indicate pass/fail, agree/disagree, yes/no, and correct/incorrect. Subjects' traits measured by these items are often summarized by averaging or summing up the binary values. In educational research, the normal linear models (e.g., linear regression, ANOVA, mixed linear model) have been commonly accepted to analyze the summed scores. They can be reasonable choices when the central limit theorem holds. However, this standard approach may be limited in some situations where the normal models cannot properly approximate binomial probabilities of summary scores, often with skewed distributions (Warton & Hui, 2011). In addition, the fact that the summed test scores bound within a lower-to-upper limit is contradictory to the characteristic of a normal distribution having an infinite range, possibly leading to predicted

doi: 10.22237/jmasm/1619482380 | Accepted: March 19, 2019; Published: August 16, 2021.
Correspondence: Seongah Im, seongahi@hawaii.edu

values that are out of range. This could be an important consideration in educational research especially when predicted scores and estimates from a statistical model are utilized for making decisions on teachers and students. If normal linear modeling using summed or aggregate scores is not appropriate, models based on the binomial distribution can be an alternative that can address concerns related to non-normality and limited range (Ferrari & Comelli, 2016; Jaeger, 2008).

When item responses constitute total scores or binomial counts in binomial modeling, it should be noted that repeated responses within an individual respondent are more alike than those from different respondents, thus leading to extra-binomial variation, a feature known as overdispersion. When an analysis ignores dependence within a subject or cluster (i.e., intracluster correlation), the results can produce biased estimates and distort standard errors, leading to invalid inferences (Jaeger, 2008). Thus, a study that proposes appropriate statistical approaches to properly deal with correlated binary responses resulting in skewed outcomes would be necessary and important. Among the statistical approaches that are capable of dealing with the concerns mentioned above, this study employed the beta-binomial model (Williams, 1982) and examined its performance using Monte Carlo simulation.

Studies Using the Beta-Binomial Model

The beta-binomial model is a commonly used approach for overdispersion that is often a manifestation of intracluster correlation. The beta distribution allows a binomial parameter to be heterogeneous and enables the beta-binomial model to explain various shapes of distributions with negative to positive skewness. Empirical studies using the beta-binomial regression with clustered binary data can be found in some fields including public health, economics, and ecology. Among them, some relevant examples are studies in health sciences that used survey data to fit varying shapes of health-status survey scores from different subdomains of health (e.g., Arostegui et al., 2007; Khan & Morris, 2014; Lamu & Olsen, 2018). In these studies, the beta-binomial regression worked well for the aggregate scores that had various non-normal distributions, compared to the normal linear models or some nonparametric approaches such as those with extended estimating equations (Basu & Rathouz, 2005) and quantile regressions. Recently, Najera-Zuloaga et al. (2017) suggested the extend use of the beat-binomial model in the context of the longitudinal design to identify risk factor for chronic disease patients and found that the application leaded to clinically relevant results.

The main interest of the beta-binomial analysis is on the marginal distributions of aggregate counts or proportions, rather than modeling of initial binary observations at data collection. When the binary responses from the initial data collection are available and the marginal scores in relation to covariates are of interest, a population averaged or marginal model such as the generalized estimating equations (GEE) of Liang and Zeger (1986) would be desirable because taking the summary scores in each cluster means information regarding the trials or items is lost in the aggregation process. GEE analyzes the mean response and the within-cluster dependence, assuming the primary interest is in the former and regarding the latter as a nuisance that must be taken into account for valid inference. The parameter estimates of GEE are known to be consistent even when the working correlation matrix reflecting dependence within clusters is misspecified (Wang & Carey, 2003).

Often researchers who employ the GEE contrasted it with generalized linear mixed models (GLMM). The studies noted that GEE produced slightly overestimated parameters and GLMM reduced power. Type 1 errors were elevated with small sample sizes in both GEE and GLMM, while either of the two methods clearly outperformed (Hallgren et al., 2016; Hubbard et al., 2010). While both approaches can adeptly handle clustered outcomes, they address dependency differently. The way to interpret the regression coefficients of the two models are different, and the size of the coefficients also substantially differ depending on the extents of correlation within clusters (Hu et al., 1998; Lee & Nelder, 2004). Rather the GEE regression estimates are similar to those from the beta-binomial model and other quasi-likelihood binomial models with beta-binomial type variance (Agresti, 2013). The nuisance parameter, i.e., the within-cluster correlation, can be compared to intracluster correlation estimated from the beta-binomial model.

The beta-binomial model was also compared to GLMM with aggregate data in a few studies. Harrison (2015) found that the beta-binomial model performed well even with data sets generated under random intercepts GLMM with experimental factors in ecology. Ferrari and Comelli (2016) also found that the beta-binomial model was powerful and outperformed GLMM in a range of experimental conditions. Yet, studies that systematically compare the beta-binomial model and GEE have been rare. In an empirical study in health service, Dilba and Aerts (2004) compared the beta-binomial model and GEE for clustered data and found that Type 1 error rates of the two methods were close under an exchangeable correlation matrix, yet the generalizability of the finding was questionable as the study was performed using an empirical data set.

In education, early researchers (e.g., Carlin & Rubin, 1991; Gross & Shulman, 1980; Huynh, 1976) used the beta-binomial distribution to depict varying distributions from criterion referenced testing. Later, Fox (2008) applied the model to a multivariate randomized response data. These studies focusing on the depiction of the non-normal outcomes, however, did not aim to examine regression modeling of aggregate or clustered scores in relation to a set of covariates. Knowing that the beta-binomial model could be suitable for modeling non-normal test scores from dependent binary responses, yet the systematic examination compared with a well-known marginal model has been rare, this study investigated the performance of the beta-binomial model in comparison with GEE using correlated binary responses simulated under three experimental conditions that are relevant in real world testing situations. In the next section, GEE with logit link and the beta-binomial logistic model are described, followed by the simulation design, and results of analyses from the two models.

Two Models

Generalized Estimating Equations

The Generalized Estimating Equations of Liang and Zeger (1986) was designed to develop a marginal or population-averaged model that tests influence of covariates on exponentially distributed (e.g., Poisson, Binomial, Gamma) response variables, and facilitates analysis of data collected in repeated, longitudinal, and panel designs. GEE is an extension of generalized linear model (GLM: McCullagh & Nelder, 1989), in which a specific type of correlation structure is incorporated into the variance function. The dependence among response variables across trials is specified as a working correlation matrix that accounts for the form of within-subject correlation.

For a case of binary responses, y_{ij} denotes a response of subject i at repeated trials j with either taking the value of 1 for correct or positive endorsement or 0 for incorrect or negative endorsement, where $i = 1,\ldots, N$ and $j = 1,\ldots, m$. The GEE with logit link is

$$\log\left(\frac{u_{ij}}{1-u_{ij}}\right) = \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta},$$

where $\mathbf{u}_i = (u_{i1}, u_{i2}, u_{im})$ is the vector of the means. The parameter estimates $\boldsymbol{\beta}$ are the solutions of the following equation (1), in which $\mathbf{V} = \boldsymbol{\phi}\mathbf{D}_i^{1/2}\mathbf{R}_i\mathbf{D}_i^{1/2}$ is the working covariance matrix of Y_i where $\mathbf{D}$ is the diagonal vector and $\mathbf{R}$ is a working correlation matrix.

$$\sum_{i=1}^{N} \frac{\partial \mathbf{u}'}{\partial \boldsymbol{\beta}} \mathbf{V}^{-1}\left(Y_i - \mathbf{u}_i\left(\boldsymbol{\beta}\right)\right) = 0\,. \tag{1}$$

As likelihood estimation methods require that each observation in the model is independent of the others, GEE uses a quasi-likelihood method for estimating parameters using the first two moments and produces parameter estimates that are asymptotically normal (Liang & Zeger, 1986). A consistent estimator of the asymptotic covariance matrix of $\boldsymbol{\beta}$, the solution of the equation (1), is obtained by the robust (also often called sandwich) estimator. The GEE parameter estimates are more efficient and consistent when the number of clusters increase even under misspecification of working correlation matrix. In this study, exchangeable correlation matrix was used, assuming the correlation among responses between any two observations is a constant because the correlation structure is appropriate to use with level-1 clustered data (Hilbe, 2009, p. 447).

Beta-Binomial Model

The beta-binomial model (Williams, 1982) is a well-known conjugate mixture model for clustered binary data. It applies with totals or aggregates from binary items or trials. Suppose

$$Y_i = \sum_{j=1}^{m_i} y_{ij}\,; i = 1, 2, \ldots, N;\ j = 1, 2, \ldots, m_i\,,$$

where the Bernoulli random variable y_{ij} takes value 1 with a probability p and 0 with a probability $q = 1 - p$, and $Y_i = 0, 1, \ldots, m_i$. Under conventional binomial model, $Y_i \sim \text{Binomial}(m_i, p)$ and the correlation coefficient between y_{ij} and y_{ik} is 0 for every $j \neq k$. When the independence assumption is violated, some alternative approaches to deal with the dependence should be considered. The beta-binomial model can be legitimately chosen to revamping extra-binomial variation or overdispersion due to intracluster dependence. The beta-binomial assumes that a success probability p_i for cluster i (i.e., a subject i) is beta-distributed with two shape parameters $\alpha > 0$ and $\beta > 0$. The appropriateness of the beta-binomial distribution is dependent upon how well the beta distribution can represent the population of p_is. The versatility of the beta distributions with different combinations of α and β allows one to handle various shapes of distributions. Its probability mass function of beta-binomial distribution is

$$\Pr\left(Y_i = y_i \mid m_i, \alpha, \beta\right) = \binom{m_i}{y_i} \frac{\mathrm{B}\left(\alpha + y_i, \beta + m_i - y_i\right)}{\mathrm{B}\left(\alpha, \beta\right)}, \quad (2)$$

where B(., .) is the beta function. Let $\pi_i = \alpha_i / (\alpha_i + \beta_i)$ satisfy a logistic relationship with a predictor vector $\mathbf{X}^{\mathrm{T}}$. Then Y_i follows a beta-binomial distribution with the mean and the variance as follows:

$$\mathrm{E}\left(Y_i\right) = \mu_i = m_i \pi_i, \quad (3)$$

$$\mathrm{V}\left(Y_i\right) = \sigma_i^2 = m_i \pi_i \left(1 - \pi_i\right)\left[1 + \rho_i \left(m_i - 1\right)\right], \quad (4)$$

where intracluster correlation $\rho_i = (\alpha_i + \beta_i + 1)^{-1}$. As shown in equation (3), the beta-binomial form of expectation is the summation of individual means that is the same as that for binomial distribution, but the variance (4) is different from the binomial variance, $\pi_i(1 - \pi_i)$. The multiplicative component of $[1 + \rho_i(m_i - 1)]$ in (4) specifically designates the role of statistical dependence among trials or items (Agresti, 2013; McCullagh & Nelder, 1989). The mean equation for the binomial logistic regression model with the grouped or summed outcome variable is

$$\log\left(\frac{\mu_i}{m_i - \mu_i}\right) = \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta},$$

where $\mathbf{X}^{\mathrm{T}}$ is a transposed vector of explanatory variables and $\boldsymbol{\beta}$ is a vector of regression parameter estimates. The parameter ρ_i can be formulated as a function of predictors, but more typically it is taken to be a constant and thereby restricting $\alpha_i + \beta_i$ to be constant for all i (Simonoff, 2003). In this study, the intracluster correlation was modeled as a constant, assuming correlations among responses are exchangeable, so as to compare results of beta-binomial model to those from GEE.

Method

Experimental Factors of Simulation Design

The simulation design of this study included two types of non-normal marginal distributions, negative and positive skewed distributions. Three experimental factors were manipulated for each type of distributions, totaling 36 experiments. Data generation in each experiment was performed with 1000 repetitions.

Intracluster Correlation. The impact of clustering on statistical results crucially depends on the strength of intracluster correlation. Specifically, if the intracluster correlation is relatively strong, the failure to take clustering into account is likely to have more profound effect on the outcomes of statistical analysis (Galbraith et al., 2010). This study set three different exchangeable correlations, $r = 0.2$, 0.4, and 0.6. The three correlations chosen reflect a range of magnitudes of correlations among items that are often found in surveys, tests, and questionnaires in education.

Number of Items. Three different numbers for items or trials were used, i.e., $m = 8$, 15, and 30. As this study tried to mimic real testing situations, those numbers were chosen based on hypothetical reliability coefficients calculated based on the Spearman-Brown formula (Spearman, 1910), which predicts a psychometric reliability for a lengthened test. All three conditions of 8, 15, and 30 resulted in predicted reliability around or above 0.8 (a satisfactory condition for reliability) for each of working correlation coefficients of 0.2, 0.4, and 0.6. An exception was made for the case of the working correlation 0.2 with m of 8, which resulted in 0.6 that is considered an acceptable cutoff.

Sample Size. Two different sample or cluster sizes were chosen, $N = 50$ and 200. The two conditions were chosen based on Pan's recommendation (2001b) using varying conditions of the GEE simulations with exchangeable correlations of 0.2, 0.4, and 0.6. The study showed that when $N = 50$, in most cases the power was less than or around 0.8 in most of the experimental conditions, and $N = 200$ always led to the power above 0.9.

Data Generation

The use of the GEE approach to generate correlated binary responses was advocated by Prentice (1988) because fully parametric approaches to estimating marginal response probabilities from a series of binary responses could be intractable except paired binary cases. Emrich and Piedmonte (1991) developed an algorithm reflecting Prentice's suggestion. The first step in the algorithm is to obtain correlated binary responses by applying threshold approaches to correlated continuous responses satisfying a desired marginal logit model. Each set of correlated binary observations or Bernoulli responses is treated like a block fulfilling the marginal condition. Thus, the 0 or 1 pattern within a block does not necessarily need to be the same as the patterns in other blocks as long as it meets the marginal condition, indicating the binary response patterns can vary across different blocks or individuals. Then, the dependence structure within the responses is parameterized in term of the given correlation matrix of the latent continuous responses (Emrich & Piedmonte, 1991; Touloumis, 2016).

An R package `SimCorMultRes` (Touloumis, 2017) adopting the suggested steps was used to simulate data sets with a multilevel structure (e.g., items at level 1 and subjects at level 2 in this study) given varying exchangeable correlations, sample sizes, and item numbers as mentioned before. The simulated multilevel data sets were analyzed using GEE. In the second step, all of the simulated data sets were made level-1 aggregated or clustered, which is the same as the process of summing up the binary responses to produce test total scores in practice. Two predictors included in data generation were $X1 \sim N(0, 1)$ and $X2 \sim Po(3)$. Both of the two covariates X1 and X2 were the level-2 predictors because the beta-binomial model cannot accommodate any level-1 predictor as the model utilizes the aggregate outcomes only and the two models can be only comparable under the same conditions.

For each type of distribution (i.e., positively and negatively skewed), a set of true parameters for the coefficients ($\beta 0$, $\beta 1$, $\beta 2$) were submitted. For the negatively skewed distributions, the three true parameters were 0.7, 0.3, and 0.2, respectively. In the positively skewed distributions, each parameter was fixed at –0.7, 0.3, and -0.2 correspondingly. Those true values and the distributions of the two covariates reflected results of an empirical analysis performed with an educational test. For GEE modeling, the R package `geepack` (Højsgaard et al., 2016) and another package, `aod` (Lesnoff & Lancelot, 2012) for beta-binomial modeling with maximum likelihood estimation, were used.

Evaluation Criteria

Performance of GEE and the beta-binomial model was evaluated using six evaluation criteria. They reported different ways to appreciate accuracy and consistency of estimates and inferences from the two models. All of the calculations under each experimental condition were based on 1000 replications. To evaluate the parameter estimates, the bias and mean squared error (MSE) were reported. In addition, the coverage probability (CP) of 95% confidence intervals for each model was calculated. To evaluate inference results, Type 1 error and power rates were reported. Empirical Type 1 error rates were computed as the ratio that a test statistic produced a *p*-value less than 0.05 under the null hypothesis. Power was calculated as the ratio of a *p*-value below 0.05 where each coefficient was different from 0. Additionally, agreement rates of the inferential decisions on each coefficient at the nominal significance level of 0.05 were reported.

Results

Errors in the GEE and the Beta-Binomial Estimates

This section reports the errors in the parameter estimates of the two models. The bias, mean squared error (MSE), and coverage probability (CP) of 95% confidence intervals for the coefficient estimates are reported in Table 1. As the positively and negatively skewed shapes did not seem to influence the sizes and patterns of the errors, the MSE values averaged over the two types of shapes are displayed in Figure 1, depending on the degree of intracluster correlations, the number of items, and sample size.

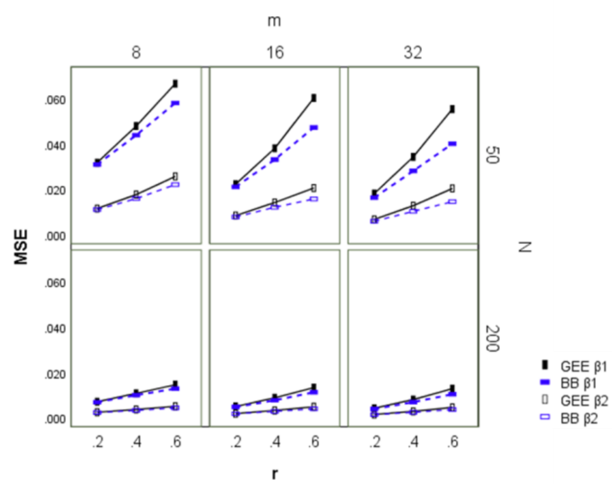

Figure 1. MSE of the GEE (solid) and Beta-Binomial (dotted) estimates

Table 1. Bias, MSE, and CP of the β1 and β2 estimates of the GEE and Beta-Binomial model

					Bias				MSE				CP			
Exp	**Type**	***r***	***m***	***N***	$\beta_G 1$	$\beta_B 1$	$\beta_G 2$	$\beta_B 2$	$\beta_G 1$	$\beta_B 1$	$\beta_G 2$	$\beta_B 2$	$\beta_G 1$	$\beta_B 1$	$\beta_G 2$	$\beta_B 2$
1	J	0.2	8	50	0.010	0.006	0.002	-0.001	0.030	0.029	0.012	0.011	93.3	94.1	92.9	93.4
2	J	0.2	8	200	-0.001	-0.006	0.002	-0.002	0.007	0.007	0.003	0.002	93.5	93.6	94.1	94.9
3	J	0.2	16	50	0.004	-0.004	0.007	0.000	0.023	0.022	0.009	0.008	92.3	94.7	92.6	94.3
4	J	0.2	16	200	0.000	-0.008	0.001	-0.006	0.005	0.005	0.002	0.002	94.6	94.8	94.8	94.7
5	J	0.2	32	50	0.003	-0.008	0.002	-0.008	0.018	0.016	0.007	0.006	93.7	95.3	92.8	94.6
6	J	0.2	32	200	0.005	-0.008	0.002	-0.008	0.004	0.004	0.001	0.001	95.5	95.3	95.1	94.7
7	J	0.4	8	50	0.016	0.005	0.007	-0.003	0.045	0.041	0.017	0.016	93.7	94.4	92.6	92.9
8	J	0.4	8	200	0.002	-0.011	0.002	-0.007	0.011	0.010	0.004	0.003	93.8	93.4	93.3	94.3
9	J	0.4	16	50	0.007	-0.010	0.011	-0.005	0.038	0.033	0.014	0.012	93.6	95.6	92.9	94.0
10	J	0.4	16	200	0.001	-0.017	0.002	-0.012	0.009	0.008	0.003	0.003	94.4	95.0	93.8	93.9
11	J	0.4	32	50	0.007	-0.017	0.004	-0.015	0.033	0.027	0.013	0.010	93.7	95.1	92.9	94.1
12	J	0.4	32	200	0.006	-0.021	0.003	-0.016	0.008	0.007	0.003	0.002	95.1	94.5	94.1	94.0
13	J	0.6	8	50	0.018	0.001	0.010	-0.001	0.065	0.056	0.025	0.022	93.6	94.9	93.0	93.6
14	J	0.6	8	200	0.004	-0.013	0.002	-0.011	0.015	0.013	0.005	0.005	93.9	93.4	93.6	94.4
15	J	0.6	16	50	0.014	-0.012	0.015	-0.006	0.055	0.046	0.022	0.017	94.8	95.6	92.0	94.6
16	J	0.6	16	200	0.001	-0.023	0.003	-0.014	0.013	0.011	0.005	0.004	93.9	93.8	95.2	94.0
17	J	0.6	32	50	0.013	-0.019	0.010	-0.017	0.054	0.040	0.020	0.014	93.4	95.0	93.3	94.1
18	J	0.6	32	200	0.007	-0.028	0.004	-0.020	0.013	0.010	0.005	0.004	95.5	94.8	94.7	93.7

Table 1 (cont.)

					Bias				MSE				CP			
Exp	**Type**	***r***	***m***	***N***	**β_G1**	**β_B1**	**β_G2**	**β_B2**	**β_G1**	**β_B1**	**β_G2**	**β_B2**	**β_G1**	**β_B1**	**β_G2**	**β_B2**
19	L	0.2	8	50	0.014	0.009	-0.002	0.001	0.033	0.033	0.012	0.011	91.7	92.5	92.6	94.7
20	L	0.2	8	200	-0.002	-0.007	-0.002	0.002	0.007	0.006	0.002	0.002	94.6	94.5	95.2	95.8
21	L	0.2	16	50	0.008	0.001	-0.003	0.005	0.022	0.020	0.008	0.008	93.7	95.4	92.4	93.8
22	L	0.2	16	200	-0.001	-0.009	0.000	0.007	0.005	0.005	0.002	0.002	94.7	94.6	93.9	94.5
23	L	0.2	32	50	-0.005	-0.005	-0.004	0.006	0.019	0.017	0.007	0.006	91.9	93.7	91.8	94.2
24	L	0.2	32	200	0.002	-0.009	0.000	0.010	0.004	0.004	0.001	0.001	94.7	94.1	94.9	94.5
25	L	0.4	8	50	0.022	0.011	-0.006	0.002	0.050	0.046	0.018	0.016	91.9	92.8	92.2	93.7
26	L	0.4	8	200	0.001	-0.011	-0.003	0.006	0.010	0.010	0.004	0.003	93.6	93.9	95.1	95.0
27	L	0.4	16	50	0.015	-0.001	-0.007	0.008	0.039	0.033	0.014	0.012	93.3	94.1	92.0	93.8
28	L	0.4	16	200	0.000	-0.019	-0.001	0.014	0.009	0.008	0.003	0.003	94.8	94.3	93.8	94.2
29	L	0.4	32	50	0.007	-0.016	-0.008	0.012	0.036	0.029	0.013	0.011	91.9	94.6	93.0	93.8
30	L	0.4	32	200	0.002	-0.022	-0.001	0.019	0.008	0.007	0.003	0.002	93.7	93.4	94.6	93.2
31	L	0.6	8	50	0.025	0.010	-0.010	0.001	0.068	0.060	0.026	0.022	92.6	93.4	91.6	93.9
32	L	0.6	8	200	0.003	-0.014	-0.004	0.008	0.014	0.012	0.005	0.004	95.3	94.5	95.8	96.5
33	L	0.6	16	50	0.022	0.001	0.009	0.014	0.066	0.049	0.019	0.015	92.0	95.0	92.2	93.2
34	L	0.6	16	200	0.002	-0.022	-0.002	0.018	0.013	0.011	0.005	0.004	94.4	94.0	94.0	93.7
35	L	0.6	32	50	0.009	-0.018	-0.012	0.012	0.057	0.040	0.021	0.015	92.6	94.8	92.6	94.1
36	L	0.6	32	200	0.002	-0.029	-0.002	0.022	0.013	0.011	0.005	0.004	93.2	92.8	94.5	93.6

The MSE values of the beta-binomial estimates were found to be smaller than those of the GEE estimates across all of the experimental conditions. Regardless of the models, the higher correlations resulted in the higher MSE values. The higher the sample size, the lower the MSE values. The added number of items had smallest effect on the MSE values.

Table 1 shows that the coverage probabilities demonstrated almost nominal 95% coverage in both models across most of the experimental conditions, although they seem to be slightly liberal. Recalling that $\beta1 = 0.3$ for both L and J shapes, and $\beta2 = 0.2$ for J-shapes and -0.2 for L-shapes, signs of biases as shown in Table 1 indicated that the GEE coefficients were overestimated, while the beta-binomial coefficients were underestimated in most of the experimental conditions (see also Table 2).

The GEE and Beta-Binomial Estimates

The parameter estimates of GEE and the beta-binomial model are presented in Table 2. Figure 2 summarizes the estimates under different conditions of sample size and intracluster correlation. When the sample size was small, the beta-binomial estimates were closer to the true parameters than the GEE estimates in most of the experimental conditions. In contrast, when the sample size got larger, it was noteworthy that the GEE estimates were consistent, staying close to the true parameters even with the highest correlation of 0.6, but the beta-binomial estimates were diverged with the added correlations. The differences between the estimates of the two models increased with the added correlations.

Figure 3 compares the density plots of the GEE and beta-binomial estimates along with histograms of the aggregate outcomes in three experimental conditions that were chosen to contrast the small, medium, and large mean differences in the parameter estimates of the two models. One should note that the ranges of the estimates at the bottom plot were much larger than those shown in the upper two plots, although they do not look much different at first glance. In the uppermost plot where the mean differences were small, the density lines of the GEE and beta-binomial estimates look very close. In the plot at the bottom with the large differences between the estimates, the lines were quite farther apart. The differences grew larger with higher intracluster correlations.

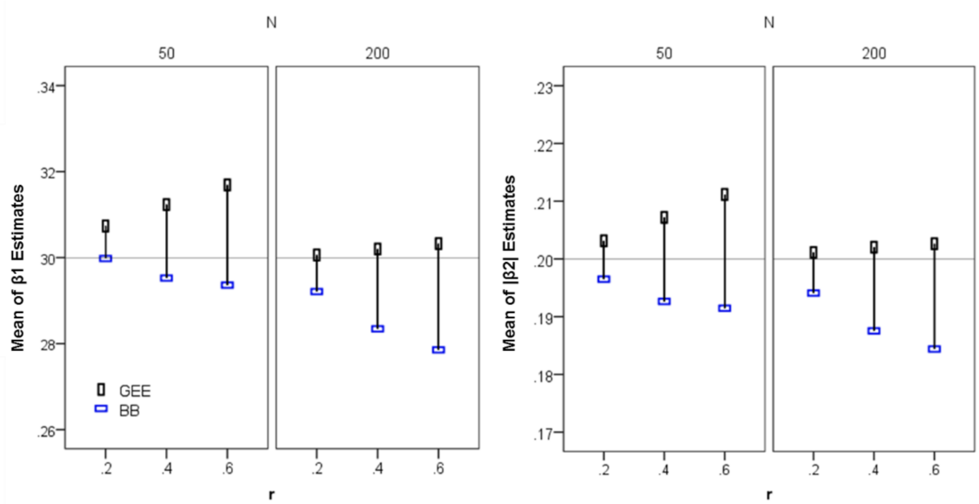

Figure 2. The β1 (left) and |β2| (right) estimates of GEE and the Beta-Binomial model

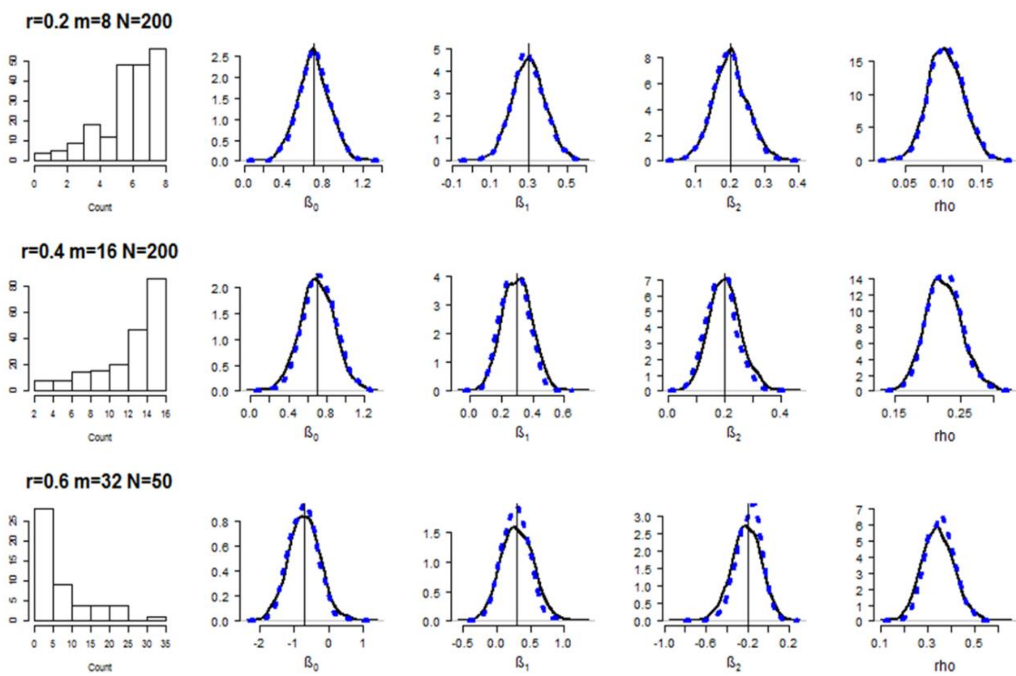

Figure 3. Example histograms of aggregate outcomes and density plots of the GEE (solid) and Beta-Binomial (dotted) estimates with the small, medium, and large mean differences

Table 2. Mean estimates and standard errors of the GEE and Beta-Binomial model

					Bias						MSE					
Exp	**Type**	***r***	***m***	***N***	**β1**	**se1**	**β2**	**se2**	**ρ**	**seρ**	**β1**	**se1**	**β2**	**se2**	**ρ**	**seρ**
1	J	0.2	8	50	0.310	0.162	0.202	0.099	0.093	0.043	0.306	0.164	0.199	0.100	0.095	0.047
2	J	0.2	8	200	0.299	0.082	0.202	0.050	0.103	0.022	0.294	0.081	0.198	0.049	0.104	0.026
3	J	0.2	16	50	0.304	0.139	0.207	0.084	0.096	0.030	0.296	0.139	0.200	0.084	0.097	0.035
4	J	0.2	16	200	0.300	0.070	0.201	0.043	0.104	0.016	0.292	0.069	0.194	0.042	0.105	0.020
5	J	0.2	32	50	0.303	0.126	0.202	0.075	0.098	0.024	0.292	0.125	0.192	0.075	0.099	0.030
6	J	0.2	32	200	0.305	0.064	0.202	0.039	0.104	0.012	0.292	0.062	0.192	0.037	0.104	0.017
7	J	0.4	8	50	0.316	0.199	0.207	0.121	0.205	0.059	0.305	0.197	0.197	0.119	0.212	0.087
8	J	0.4	8	200	0.302	0.100	0.202	0.061	0.222	0.030	0.289	0.096	0.193	0.058	0.225	0.045
9	J	0.4	16	50	0.307	0.182	0.211	0.109	0.208	0.047	0.290	0.175	0.195	0.105	0.213	0.075
10	J	0.4	16	200	0.301	0.092	0.202	0.056	0.224	0.024	0.283	0.086	0.188	0.052	0.225	0.040
11	J	0.4	32	50	0.307	0.174	0.204	0.104	0.211	0.040	0.283	0.163	0.185	0.097	0.213	0.069
12	J	0.4	32	200	0.306	0.088	0.203	0.053	0.223	0.021	0.279	0.080	0.184	0.048	0.222	0.037
13	J	0.6	8	50	0.318	0.237	0.210	0.145	0.342	0.072	0.301	0.229	0.199	0.139	0.354	0.154
14	J	0.6	8	200	0.304	0.118	0.202	0.071	0.364	0.036	0.287	0.111	0.189	0.067	0.369	0.072
15	J	0.6	16	50	0.314	0.224	0.215	0.134	0.342	0.060	0.288	0.208	0.194	0.125	0.353	0.138
16	J	0.6	16	200	0.301	0.112	0.203	0.068	0.365	0.031	0.277	0.101	0.186	0.061	0.368	0.067
17	J	0.6	32	50	0.313	0.218	0.210	0.131	0.344	0.054	0.281	0.196	0.183	0.116	0.354	0.135
18	J	0.6	32	200	0.307	0.110	0.204	0.066	0.365	0.028	0.272	0.096	0.180	0.057	0.368	0.065

Note: The standard deviations of the parameter estimates (not included in this table) were very similar out to the third decimal point of the standard error values

Table 2 (cont.)

					Bias						MSE					
Exp	**Type**	***r***	***m***	***N***	**β1**	**se1**	**β2**	**se2**	**ρ**	**seρ**	**β1**	**se1**	**β2**	**se2**	**ρ**	**seρ**
19	L	0.2	8	50	0.314	0.161	-0.202	0.098	0.094	0.043	0.309	0.165	-0.199	0.100	0.096	0.047
20	L	0.2	8	200	0.298	0.082	-0.202	0.049	0.103	0.022	0.293	0.081	-0.198	0.049	0.104	0.026
21	L	0.2	16	50	0.308	0.139	-0.203	0.083	0.096	0.030	0.301	0.139	-0.195	0.084	0.098	0.036
22	L	0.2	16	200	0.299	0.070	-0.200	0.042	0.103	0.016	0.291	0.068	-0.193	0.041	0.103	0.019
23	L	0.2	32	50	0.305	0.136	-0.204	0.076	0.098	0.024	0.295	0.125	-0.194	0.075	0.098	0.024
24	L	0.2	32	200	0.302	0.064	-0.200	0.039	0.104	0.012	0.291	0.061	-0.190	0.037	0.103	0.017
25	L	0.4	8	50	0.322	0.198	-0.206	0.120	0.207	0.059	0.311	0.197	-0.198	0.119	0.214	0.086
26	L	0.4	8	200	0.301	0.100	-0.203	0.060	0.222	0.030	0.289	0.096	-0.194	0.058	0.225	0.045
27	L	0.4	16	50	0.315	0.183	-0.207	0.109	0.209	0.047	0.299	0.176	-0.192	0.105	0.213	0.078
28	L	0.4	16	200	0.300	0.092	-0.201	0.056	0.223	0.024	0.281	0.086	-0.186	0.051	0.223	0.039
29	L	0.4	32	50	0.307	0.174	-0.208	0.104	0.210	0.040	0.284	0.163	-0.188	0.097	0.212	0.072
30	L	0.4	32	200	0.302	0.088	-0.201	0.053	0.224	0.021	0.278	0.080	-0.181	0.048	0.222	0.037
31	L	0.6	8	50	0.325	0.237	-0.210	0.142	0.345	0.072	0.310	0.229	-0.199	0.138	0.357	0.147
32	L	0.6	8	200	0.303	0.118	-0.204	0.071	0.363	0.036	0.286	0.111	-0.192	0.067	0.369	0.072
33	L	0.6	16	50	0.322	0.235	-0.209	0.122	0.334	0.063	0.301	0.216	-0.186	0.112	0.347	0.181
34	L	0.6	16	200	0.302	0.112	-0.202	0.068	0.366	0.031	0.278	0.101	-0.182	0.061	0.369	0.068
35	L	0.6	32	50	0.309	0.219	-0.212	0.132	0.345	0.054	0.282	0.197	-0.188	0.117	0.352	0.138
36	L	0.6	32	200	0.302	0.110	-0.202	0.066	0.366	0.028	0.271	0.096	-0.178	0.057	0.367	0.065

Note: The standard deviations of the parameter estimates (not included in this table) were very similar out to the third decimal point of the standard error values

The last two columns under each model in Table 2 show that the estimates of intracluster correlation stayed very close across all of the experimental conditions. The sizes of the estimated intracluster correlations were slightly different, but the differences were fringe. For the simulated data sets with $r = 0.2$, the averaged correlation estimates were 0.099 for GEE and 0.100 for the beta-binomial model. When $r = 0.4$, the estimates were 0.214 and 0.217, respectively. Under $r = 0.6$, the estimates were 0.352 and 0.360 for each model correspondingly. The smaller estimates than the true correlations used for the data generation indicated that the two hypothetical predictors explained some amount of overdispersion due to dependent responses within clusters. In a follow-up analysis, the estimated correlations from the intercept only model without the two predictors were much closer to the true parameters, though they were about 0.04 to 0.1 lower than the true values on average for both.

Inference on the Beta-Binomial and GEE Estimates

The power and empirical Type 1 error rates and percent agreement of statistical decision (AD) on the estimates are reported in Table 3. The Type 1 error rates in both models ranged from 0.031 to 0.075 and were slightly larger than the nominal 0.05 level of significance on average. But they remained fairly stable across all of the conditions of correlations, sample sizes, and numbers of items.

The two models agreed upon rejection or acceptance for the coefficients at $\alpha = 0.05$. On average, the agreement rate was about 95% for the large sample and dropped to 88% for the small sample. The inference on the correlation estimates (not included in Table 3) showed higher agreement, 92% for the small sample condition and 99.98% for the large sample condition on average. The power rates of the GEE and beta-binomial estimates were averaged over the shapes and the number of items in Figure 4. Notably, the power of the estimates drastically changed from the small sample condition to the large sample condition. The higher intracluster correlations diminished the power of the estimates.

Table 3. Power, Type 1 error rate, and percent agreement on decision at $\alpha = 0.05$

					Power		Type 1			Power		Type 1		
Exp	**Type**	***r***	***m***	***N***	$\beta_G 1$	$\beta_B 1$	$\beta_G 1$	$\beta_B 1$	**AD1**	$\beta_G 2$	$\beta_B 2$	$\beta_G 2$	$\beta_B 2$	**AD2**
1	J	0.2	8	50	0.487	0.460	0.065	0.068	89.9	0.545	0.527	0.056	0.059	89.6
2	J	0.2	8	200	0.943	0.948	0.066	0.059	97.1	0.976	0.982	0.065	0.051	98.8
3	J	0.2	16	50	0.599	0.574	0.073	0.071	90.3	0.686	0.660	0.047	0.054	89.8
4	J	0.2	16	200	0.990	0.990	0.055	0.052	99.0	1.000	1.000	0.052	0.053	100.0
5	J	0.2	32	50	0.663	0.650	0.059	0.066	90.3	0.755	0.730	0.049	0.052	91.5
6	J	0.2	32	200	0.998	0.998	0.045	0.049	99.8	1.000	1.000	0.047	0.053	100.0
7	J	0.4	8	50	0.380	0.339	0.064	0.070	88.3	0.417	0.397	0.053	0.064	88.6
8	J	0.4	8	200	0.834	0.839	0.031	0.038	94.7	0.918	0.933	0.067	0.057	96.1
9	J	0.4	16	50	0.415	0.399	0.058	0.071	87.4	0.486	0.458	0.042	0.051	83.8
10	J	0.4	16	200	0.905	0.911	0.057	0.062	94.4	0.961	0.957	0.051	0.061	96.2
11	J	0.4	32	50	0.422	0.440	0.061	0.067	86.6	0.501	0.476	0.043	0.052	83.5
12	J	0.4	32	200	0.936	0.948	0.050	0.059	95.4	0.971	0.980	0.056	0.060	97.7
13	J	0.6	8	50	0.277	0.259	0.067	0.065	88.2	0.324	0.292	0.046	0.058	87.4
14	J	0.6	8	200	0.722	0.735	0.063	0.067	90.5	0.821	0.830	0.067	0.057	91.7
15	J	0.6	16	50	0.303	0.282	0.051	0.075	84.5	0.371	0.341	0.043	0.050	82.8
16	J	0.6	16	200	0.749	0.765	0.062	0.049	88.2	0.863	0.864	0.063	0.060	90.1
17	J	0.6	32	50	0.313	0.293	0.062	0.063	86.2	0.364	0.352	0.049	0.053	82.6
18	J	0.6	32	200	0.784	0.819	0.046	0.054	88.1	0.872	0.901	0.053	0.063	91.1

Table 3 (cont.)

					Power		Type 1			Power		Type 1		
Exp	**Type**	***r***	***m***	***N***	**β_G1**	**β_B1**	**β_G1**	**β_B1**	**AD1**	**β_G2**	**β_B2**	**β_G2**	**β_B2**	**AD2**
19	L	0.2	8	50	0.513	0.475	0.077	0.067	90.4	0.546	0.512	0.067	0.046	90.8
20	L	0.2	8	200	0.957	0.957	0.054	0.048	97.4	0.987	0.990	0.055	0.042	98.9
21	L	0.2	16	50	0.608	0.576	0.058	0.069	89.8	0.679	0.649	0.044	0.058	90.8
22	L	0.2	16	200	0.993	0.995	0.053	0.061	99.4	0.996	0.997	0.054	0.055	99.7
23	L	0.2	32	50	0.669	0.647	0.074	0.074	89.6	0.759	0.736	0.053	0.052	90.5
24	L	0.2	32	200	0.998	0.998	0.053	0.051	99.8	0.997	0.997	0.053	0.055	100.0
25	L	0.4	8	50	0.370	0.353	0.072	0.073	88.9	0.416	0.380	0.063	0.056	90.0
26	L	0.4	8	200	0.855	0.861	0.064	0.049	95.6	0.925	0.929	0.061	0.050	95.2
27	L	0.4	16	50	0.424	0.410	0.073	0.068	88.8	0.481	0.455	0.048	0.062	85.2
28	L	0.4	16	200	0.911	0.915	0.052	0.062	95.2	0.945	0.951	0.058	0.057	96.8
29	L	0.4	32	50	0.434	0.425	0.074	0.066	85.5	0.498	0.478	0.047	0.058	83.0
30	L	0.4	32	200	0.936	0.936	0.063	0.055	95.0	0.965	0.975	0.066	0.068	96.8
31	L	0.6	8	50	0.294	0.274	0.065	0.075	89.8	0.333	0.305	0.057	0.054	86.4
32	L	0.6	8	200	0.730	0.748	0.047	0.042	90.0	0.827	0.832	0.055	0.035	90.5
33	L	0.6	16	50	0.309	0.286	0.068	0.048	86.7	0.418	0.389	0.048	0.062	82.3
34	L	0.6	16	200	0.778	0.798	0.056	0.060	88.2	0.832	0.861	0.060	0.063	90.9
35	L	0.6	32	50	0.315	0.301	0.063	0.066	83.4	0.354	0.345	0.047	0.053	81.9
36	L	0.6	32	200	0.786	0.805	0.068	0.055	86.7	0.871	0.889	0.072	0.064	90.6

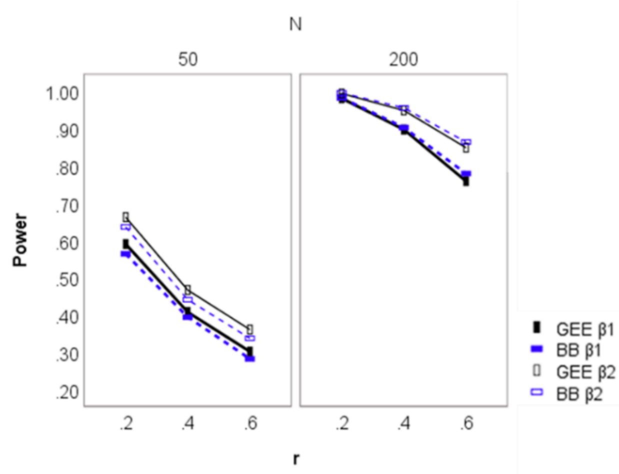

Figure 4. Power of the GEE (solid) and Beta-Binomial (dotted) estimates

---

Discussion and Conclusions

This study examined performance of the beta-binomial model for aggregate outcomes from clustered binary responses leading to non-normal marginal distributions. Its performance was compared to the results of GEE. The simulated data sets used for each model had different structures. In GEE, the binary observations signifying correct or incorrect responses to each simulated item were the units of analysis, and the binary responses were nested within subjects. In the beta-binomial model, the units of analysis were the aggregate outcomes of correlated binary responses.

This study found that GEE tended to overestimate the regression parameters, consistent with previous findings from the Hallgren et al. (2016) study. GEE under the large sample size performed consistently and was only slightly affected by the different extents of intracluster correlation. However, under small sample size, the estimates were prone to bias with higher correlations. Though it is known that GEE is less dependent on the sample size (Kenward et al., 1994; Muth et al., 2015), some researchers (e.g., Gunsolley et al., 1995) expressed concern about the performance of GEE when used for small sample. In the current study, GEE estimates under small sample or cluster size were found to be vulnerable for a particular case of binomial data with high intracluster correlation. In such a case, use of a modified variance estimator (Pan, 2001a) or different tests such as the robust score test (Guo et al., 2005) can be preferable.

It was found that the beta-binomial model can be a good alternative for GEE when sample size is small. Its estimates were closer to the true parameters than those of GEE. Indeed, the strength of the beta-binomial model would be valuable because many researchers and practitioners employ small sample or cluster size conditions when collecting data sets for their pilot studies or in experimental design settings. The beta-binomial model could be readily applied when sample size of clustered binary data is not large enough.

Under the large sample size condition, the beta-binomial model generated more biased estimates. The consistency of the beta-binomial estimates was questionable especially with high correlations, opposite to the results of GEE that showed stable performance with the large sample. Hence, GEE and the beta-binomial model can complement each other.

This study also found that in both small and large sample size conditions, the beta-binomial model was more efficient than GEE. The absolute parameter estimates of the beta-binomial model were underestimated throughout experimental conditions, which was consistent with the result of Harrison (2015). The two models agreed upon rejection or acceptance for the coefficients at the significance level of 0.05. The power of the parameter estimates was low for the small sample, yet very high at around 0.9 for the large sample size, which was similar to the result of Pan's (2001b) study. The Type 1 error probabilities of the two models were very close across the different conditions, confirming the empirical finding of Dilba and Aerts (2004). Though the values were slightly larger than the nominal significance level of 0.05, they remained fairly stable across the conditions in both models. Among the three experimental factors, the number of items or trials had the smallest effect on the parameter estimates, akin to the finding regarding the beta-binomial estimates in Harrison (2015). In the current study, the extent of intracluster correlation had profound effects on the estimation as well as on the power of the estimates for both models as found in Galbraith et al. (2010).

As this study examined cases mimicking educational tests where aggregate or clustered outcomes have non-normal and skewed distributions from correlated binary responses, generalizability of the conclusions within a broader context of aggregate outcomes might be limited. To overcome this limitation, a future study can include other aggregate or clustered designs such as a cross-sectional design where the intraclass correlation could be much smaller. Findings of this study warrant further investigation on the effect of varying degrees of intracluster correlation under more refined sample size conditions. Different data generation processes can be considered especially to further investigate the biased results of the two models under different sample size conditions. One can also ponder a systematic study including other statistical approaches that take into account overdispersion differently such as GLMM. This might answer the reason for somewhat underestimated correlation estimates even when the two predictors were absent in both models because other possible source of overdispersion such as inter-individual variability could not be captured by the two models.

References

Agresti, A. (2013). *Categorical data analysis* (3rd edition). Hoboken, NJ: Willey and Sons.

Arostegui, I., Nunez-Anton, V., & Quintana, J. M. (2007). Analysis of the Short Form-36 (SF36): The beta-binomial distribution approach. *Statistics in Medicine, 26*(6), 1318-1342. doi: 10.1002/sim.2612

Basu, A., & Rathouz, P. (2005). Estimating marginal and incremental effects on health outcomes using flexible link and variance function models. *Biostatistics*, *6*(1), 93-109. doi: 10.1093/biostatistics/kxh020

Carlin, J. B., & Rubin, D. B. (1991). Summarizing multiple-choice tests using three informative statistics. *Psychological Bulletin, 110*(2), 338-349. doi: 10.1037/0033-2909.110.2.338

Dilba, G., & Aerts, M. (2004). A comparative study of models for correlated binary data with applications to health services research. *SINET: Ethiopian Journal of Science*, *27*(2), 97-104. doi: 10.4314/sinet.v27i2.18239

Emrich, L. J., & Piedmonte, M. R. (1991). A method for generating high-dimensional multivariate binary variates. *American Statistician, 49*(4), 302-304. doi: 10.1080/00031305.1991.10475828

Ferrari, A., & Comelli, M. (2016). A comparison of methods for analysis of binomial clustered outcomes in behavioral research. *Journal of Neuroscience Methods, 274*, 131-140. doi: 10.1016/j.jneumeth.2016.10.005

Fox, J.-P. (2008). Beta-binomial ANOVA for multivariate randomized response data. *British Journal of Mathematical and Statistical Psychology*, *61*(2), 453-470. doi: 10.1348/000711007X226040

Galbraith, S., Daniel, J. A., & Vissel, B. (2010). A study of clustered data and approaches to its analysis. *The Journal of Neurosciences, 30*(32), 10601-10608. doi: 10.1523/JNEUROSCI.0362-10.2010

Gross, A. L., & Shulman, V. (1980). The applicability of the beta-binomial model for criterion referenced testing. *Journal of Educational Measurement, 17*(3), 195-201. doi: 10.1111/j.1745-3984.1980.tb00826.x

Gunsolley, J. C, Getchell, C., & Chinchilli, V. M. (1995). Small sample characteristics of generalized estimating equations. *Communications in Statistics – Simulation and Computation*, *24*(4), 869-878. doi: 10.1080/03610919508813280

Guo, X., Pan, W., Connett, J, E., Hannan, P. J., & French, S. A. (2005). Small-sample performance of the robust score test and its modifications in generalized estimating equations. *Statistics in Medicine*, *24*(22), 3479-3495. doi: 10.1002/sim.2161

Hallgren, K. A., Atkins, D. C., & Witkiewitz, K. (2016). Aggregating and analyzing daily drinking data in clinical trials: A comparison of type l errors, power, and bias. *Journal of Studies on Alcohol and Drugs*, *77*(6), 986-991. doi: 10.15288/jsad.2016.77.986

Harrison, X. A. (2015). A comparison of observation-level random effect and beta-binomial models for modeling overdispersion in binomial data in ecology and evolution. *PeerJ*, *3*, e1114. doi: 10.7717/peerj.1114

Hilbe, J. (2009). *Logistic regression models.* New York: CRC Press. doi: 10.1201/9781420075779

Højsgaard, S., Halekoh, U., & Yan, J. (2016). geepack: Generalized estimating equation package [R package, ver. 1.2-1]. Retrieved from: https://cran.r-project.org/package=geepack

Hu, F. B., Goldberg, J., Hedeker, D., Flay, B. R., & Pentz, M. A. (1998). Comparison of population-averaged and subject-specific approaches for analyzing repeated binary outcomes. *American Journal of Epidemiology*, *147*(7), 694-703. doi: 10.1093/oxfordjournals.aje.a009511

Hubbard, A. E., Ahern, J., Fleischer, N. L., van der Laan, M., Lippman, S. A., Jewell, N., Bruckner, T., & Satarianob, W. A. (2010). To GEE or not to GEE: comparing population average and mixed models for estimating the associations between neighborhood risk factors and health. *Epidemiology*, *21*(4), 467-474.

Huynh, H. (1976). Statistical consideration of mastery scores. *Psychometrika*, *41*(1), 65-78. doi: 10.1007/BF02291698

Jaeger, T. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434-446. doi: 10.1016/j.jml.2007.11.007

Kenward, M. G., Lesaffre, E., & Molenberghs, G. (1994). An application of maximum likelihood and generalized estimating equations to the analysis of ordinal data from a longitudinal study with cases missing at random. *Biometrics, 50*(4), 945-953. doi: 10.2307/2533434

Khan, I., & Morris, S. (2014). A non-linear beta-binomial regression model for mapping EORTC QLQ-C30 to the EQ-5D-3L in lung cancer patients: A comparison with existing approaches. *Health and Quality of Life Outcomes, 12* 163. doi: 10.1186/s12955-014-0163-7

Lamu, A. N., & Olsen, J. A. (2018). Testing alternative regression models to predict utilities: Mapping the QLQ-C30 onto the EQ-5D-5L and the SF-6D. *Quality of Life Research, 27*(11), 2823-2839. doi: 10.1007/s11136-018-1981-6

Lee, Y., & Nelder, J. A. (2004). Conditional and marginal models: Another view. *Statistical Science, 19*(2), 219-238. doi: 10.1214/088342304000000305

Lesnoff, M., & Lancelot, R. (2012). aod: Analysis of overdispersed data [R package, ver. 1.3]. Retrieved from: https://cran.r-project.org/package=aod

Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika, 73*(1), 13-22. doi: 10.1093/biomet/73.1.13

McCullagh, P., & Nelder, J. (1989). *Generalized linear models* (2nd edition). New York: Chapman and Hall.

Muth, C., Bales, K. L., Hinde, K., Maninger, N., Mendoza, S. P., & Ferrer, E. (2015). Alternative models for small samples in psychological research: Applying linear mixed effects models and generalized estimating equations to repeated measures data. *Educational and Psychological Measurement, 76*(1), 64-87. doi: 10.1177/0013164415580432

Najera-Zuloaga, J., Lee, D.-J., & Arostegui, I. (2018). A beta-binomial mixed-effects model approach for analyzing longitudinal discrete and bounded outcomes. *Biometrical Journal, 61*(3), 600-615. doi: 10.1002/bimj.201700251

Pan, W. (2001a). On the robust variance estimator in generalized estimating equations. *Biometrika, 88*(3), 901-906. doi: 10.1093/biomet/88.3.901

Pan, W. (2001b). Sample size and power calculations with correlated binary data. *Controlled Clinical Trials, 22*(3), 211-227. doi: 10.1016/S0197-2456(01)00131-3

Prentice, R. (1988). Correlated binary regression with covariates specific to each binary observations. *Biometrics, 44*(4), 1033-1048. doi: 10.2307/2531733

Simonoff, J. S. (2003). *Analyzing categorical data.* New York: Springer. doi: 10.1007/978-0-387-21727-7

Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology*, *3*(3), 271-295. doi: 10.1111/j.2044-8295.1910.tb00206.x

Touloumis, A. (2016). Simulating correlated binary and multinomial responses under marginal model specification: The SimCorMultRes package. *The R Journal*, *8*(2), 79-91.

Touloumis, A. (2017). SimCorMultRes: Simulates correlated multinomial responses [R package, ver. 1.5.0]. Retrieved from: https://cran.r-project.org/package=SimCorMultRes

Wang, Y. G., & Carey, V. (2003). Working correlation structure misspecification, estimation and covariate design: Implications for generalised estimating equations performance. *Biometrika*, *90*(1), 29-41. doi: 10.1093/biomet/90.1.29

Warton, D., & Hui, F. (2011). The arcsine is asinine: The analysis of proportions in ecology. *Ecology*, *92*(1), 3-10. doi: 10.1890/10-0340.1

Williams, D. (1982). Extra-binomial variation in logistic linear models. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 31*(2), 144-148. doi: 10.2307/2347977